Measuring and Modelling the Quality of Work and Employment for Employees in the United Kingdom Using Item Response Theory

Nhlanhla Ndebele

Department of Sociology and Criminology,

City, University of London

January 2024

Thesis submitted for the fulfilment of the requirements for the Degree of Doctor of Philosophy

Declaration:

I, Nhlanhla Ndebele, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature: [redacted]

Table of Contents

Acknowledgements

Firstly, I would like to thank my supervisors, Dr Matt Barnes and Dr Eric Harrison, without whose support, guidance and patience I would not have made it this far. I would also like to thank Dr Sally Stares for advice throughout the course of my research studies and Dr Vanessa Gash as the reader for my upgrade. I am also grateful to the Department of Sociology at City, University of London, and the City Q-Step Centre for sponsoring my research studies and for the opportunity to develop as a researcher. Thank you to all the City Q-Step researchers that helped me along the way, and especially Dr Sonila Dardha. Thank you also to Dr Kathrin Thomas for encouragement along the way.

Finally, I would like to thank my family for putting up with my non-stop working, hopefully I will make up for the time.

Abstract

Rationale and Aims: Standard analyses of the labour market tend to focus more on job quantity than work and employment (QWE), however, there is increasing interest at national and international levels to highlight QWE as a salient social and labour market policy issue. This, though, has been hindered by a lack of consensus on a definition, and inherent challenges in the conceptualisation and operationalisation of QWE, evidenced by the substantial number of measurement instruments in the literature. Although there is consensus about the concept being multidimensional, it cannot be directly measured, and there is no consensus on what attributes to include in the measure, how these should be aggregated and weighted, or whether to report *overall QWE* and/or different *dimensions*. Furthermore, the measurement of QWE is limited by the lack of a single source of data capturing all the relevant attributes. At the same time, there is also a lack of evaluation of the measurement equivalence of the instruments, which is a prerequisite for between-group comparisons. This study aimed to make theoretical contributions to the conceptualisation of QWE, as well as methodological contributions in the measurement of QWE. The study also sought to make substantive contributions by investigating how *overall* and/or *dimensions of QWE* varied by demographic, socio-demographic, and socio-economic characteristics in the UK employee population.

Methods: The study used data from the UK Household Longitudinal Study (UKHLS) and applied item response theory (IRT) modelling to develop a measure of QWE. Competing measurement models (*unidimensional*, *correlated-factors*, *second-order factor*, and *bifactor models*) were estimated and compared, and psychometric properties of the measurement instruments were examined. Differential item functioning (DIF) was used to evaluate the measurement equivalence of the instruments between different groups. Multiple group analysis and multiple indicators multiple causes (MIMIC) models were used to investigate the effects

of the demographic, socio-demographic, and socio-economic characteristics on *overall QWE* and other *dimensions of QWE*.

Results: The theoretical framework of QWE consisted of six dimensions: *economic compensation*, *training and progression*, *employment security*, *working conditions*, *work-life balance* (or *work-time scheduling*), and *social dialogue*. The *social dialogue* dimension was measured by a single item and in the subsequent IRT modelling, responses to this item were not sufficiently explained by the model resulting in the exclusion of the item. Results suggested that the measurement of QWE was better represented by a bifactor model. An evaluation of the psychometric properties of the measure of QWE suggested that *training and progression* and *employment security* were not a good representation of these latent traits and were excluded in subsequent analysis. DIF analyses indicated that while some items measuring *overall* and other *dimensions of QWE* exhibited differential performance between some groups, the magnitude of DIF was negligible and between-group comparison was feasible.

Substantively, results from the study suggested that demographic or socio-demographic characteristics did not explain much of the variation in *overall* or other *dimensions of QWE*, with the effect sizes either small or negligible. In contrast, socio-economic characteristics explained more of the variation in the latent traits. The bifactor model also provided a more nuanced understanding of differences in QWE between some groups that would otherwise not be feasible with other methods. The study found that females had better *working conditions* than males, while younger employees were more aware of and had better access to other forms of *work-time scheduling* than older employees. Results suggested that employees with longstanding illnesses or disability had poorer *economic compensation* and *working conditions* but were more aware of and had better access to forms of *work-time scheduling* than those without a longstanding illness or disability, while there were no differences in *overall QWE*. The study highlighted longstanding regional disparities in the labour market with advantages for those in the London and Southern England regions. However, it also highlighted better *working conditions* and more awareness of and better access to other forms of *work-time scheduling*, along with comparable *economic compensation* for employees in Scotland relative to those in London, but employees in London had better *overall QWE* than those in Scotland. Results supported evidence of better outcomes for employees in public sector organisations than those in private sector organisations. Employees in public sector organisations had better *economic compensation*, *working conditions*, and more awareness of and better access to forms of *work-time scheduling* but poorer *overall QWE* than those in private sector organisations.

Conclusions: The study contributed to the conceptualisation of QWE by developing a theoretical framework for measuring QWE and made methodological contributions by applying IRT modelling to address some of the shortcomings of existing measures. These shortcomings related to the aggregation and weighting of items measuring QWE, including whether to report *overall* and/or other *dimensions of QWE*, as well as evaluating measurement equivalence of the instrument. The study presented new knowledge that suggested the measurement of QWE was better represented by a bifactor model. The study also made some substantive contributions, which suggested that socio-economic characteristics explained more of the variation in *overall* or other *dimensions of QWE* than demographic or socio-demographic characteristics. While results from IRT modelling largely replicated those of other methods, there were some discrepancies, and the bifactor model provided a more nuanced understanding of differences in QWE between some groups.

List of Tables

List of Figures

Acronyms and Abbreviations

1-PL	One-parameter logistic
2-PL	Two-parameter logistic
3-PL	Three-parameter logistic
4-PL	Four-parameter logistic
ABIC	Adjusted Bayesian Information Criterion
AIC	Akaike Information Criterion
ANOVA	Analysis of variance
APS	Annual Population Survey
BEIS	Department for Business, Energy and Industrial Strategy
BHPS	British Household Panel Survey
BIC	Bayesian Information Criterion
CAPI	Computer-assisted personal interviewing
CATI	Computer-assisted telephone interviewing
CAWI	Computer-assisted web interviewing
CFI	Comparative fit index
CIPD	Chartered Institute of Personnel and Development
CME	Coordinated market economy
CRF	Category response function
CTT	Classical test theory
DGB	German Confederation of Trade Unions
DGBI	DGB Good Work Index
DIF	Differential item functioning
DWI	Decent Work Index
EAP	Expected a posterior
ECHP	European Community Household Panel
ECV	Explained common variance
EFA	Exploratory factor analysis
EIJQI	European Intrinsic Job Quality Index
EJQI	European Job Quality Index
ELFS	European Labour Force Survey
EM	Expectation maximisation
ETUI	European Trade Union Institute
EU	European Union
EUL	End User Licence
EVS	European Values Survey
EWCS	European Working Conditions Survey
FA	Factor analysis
FD	Factor determinacy
GRM	Graded response model
ILO	International Labour Organisation
IRF	Item response function
IRS	Item response surface
IRT	Item response theory
ISER	Institute for Social and Economic Research

Continued…

ISSP	International Social Survey Programme	NMW	National minimum wage
IV	Independent variable	NS-SEC	National Statistics Socio-Economic Classification
JQI	Job Quality Index	OECD	Organisation for Economic Co-operation and Development
LCA	Latent class analysis	ONS	Office for National Statistics
LD	Local dependence	PCA	Principal components analysis
LFS	Labour Force Survey	PISA	Programme for International Student Assessment
LL	Log likelihood	PPC	Bayesian posterior predictive check
LMD	Labour Market Dynamics	PPP	Bayesian posterior predictive p-value
LME	Liberal market economy	PRO	Patient-Reported Outcomes
LRT	Likelihood ratio test	PROMIS	Patient-Reported Outcomes Measurement Information System
LUHC	Department for Levelling Up, Housing and Communities	PRT	Power resources theory
MCMC	Markov chain Monte Carlo	PSR	Potential scale reduction
MDIFF	Multidimensional difficulty parameter or index	QoE	Quality of Employment
MDISC	Multidimensional discrimination parameter or index	QoEI	Quality of Employment Index
MIMIC	Multiple indicators multiple causes	QoWI	Quality of Work Index
MIRT	Multidimensional Item Response Theory	QuInnE	Quality of Jobs and Innovation Generated Employment Outcomes
ML	Maximum likelihood	QWE	Quality of work and employment
MLR	Maximum likelihood with robust standard errors	RMSEA	Root mean square error of approximation
MME	Mixed (or Mediterranean) market economy	SF-36	Short-form health survey
MML	Maximum marginal-likelihood	SILC	Statistics on Income and Living Conditions
NISRA	Northern Ireland Statistics and Research Agency	SOC	Standard Occupational Classification
NLW	National living wage	SRMSR	Standardised root mean square residual

Continued…

Continued…

TCF	Test characteristic function
TLI	Tucker-Lewis index
TUC	Trade Union Congress
UK	United Kingdom
UKDS	UK Data Service
UKHLS	UK Household Longitudinal Study
UKWL	UK Working Lives
UNECE	United Nations Economic Commission for Europe
VoC	Varieties of capitalism

Chapter 1 Introduction

Standard analyses of the labour market tend to focus more on job quantity, which is associated with aspects of the labour market that can be directly measured and expressed numerically, than the quality of work and employment (QWE), which cannot be directly measured and relates to job characteristics that impact workers' well-being. Job quantity indicators such as the employment and unemployment rates, redundancy rates, job vacancies, and rates of economic inactivity within the working-age population are among some of the national statistics reported in the United Kingdom's (UK) *Labour Market Overview*, while reporting of QWE is limited to indicators related to working hours and earnings (Office for National Statistics (ONS) 2022). There have been concerns about the quality of jobs created in the labour market (Leschke and Watt 2014), while evidence also suggests that QWE is associated with work productivity (Arends, Prinz, and Abma 2017), economic competitiveness, including the well-being of workers (Carrié et al. 2012).

1.1 Background of the Study

According to Marx, work is fundamental to our humanity, and he argued it is what sets humans apart from other animals, and his belief in the centrality of work is foundational to the social theory around work and employment (Warren 2016). Indeed, other than sleep, workers spend substantially more of their adult lives on work activities than any other activity (Sinclair, Morgan, and Johnson 2020), and this places great importance on QWE in the social agenda. This was the case in the 1960s and 1970s among industrial capitalist societies due to waves of strikes in the late 1960s attributed to inhumane working conditions with the rise of Fordism, characterised by assembly lines of giant car factories, but the strikes were curtailed by neoliberal policies in the 1980s (Gallie 2007b). However, issues of QWE returned to the fore of the social agenda among many post-industrial societies in the last decade of the twentieth century with interest from policymakers, governments and researchers at international and national levels (Gallie 2007b; UNECE 2010).

At the international level, the International Labour Organisation (ILO 1999) presented a *decent work* agenda due to changes in employment patterns, labour markets and relations as a result of globalisation, which had brought both prosperity and inequalities. The ILO defined *decent work* as "productive work for women and men in conditions of freedom, equity, security and human dignity" (1999:3). The aim was to promote fundamental principles and rights at work, create greater employment opportunities for women and men, provide social protection, and promote social dialogue (ILO 1999). Similarly, the Organisation for Economic Co-operation and Development (OECD) also pursued the issue of QWE as part of its well-being agenda through its *Better Life Initiative* (OECD 2013). This considered QWE as a critical dimension shaping workers' well-being at the workplace and endorsed an OECD Job Quality Framework that aimed to promote the quality of earnings, reduce labour market insecurity, and promote good working conditions (Cazes, Hijzen, and Saint-Martin 2015; OECD 2013).

Meanwhile, in Europe in 2000, the European Union presented the *Lisbon Agenda* (or *Strategy*), which proposed a strategy for *more and better jobs* in response to globalisation and challenges of a new knowledge-driven economy (European Parliament 2010). The *Lisbon Agenda* aimed to prepare for the transitioning to a new knowledge-driven economy and society by developing better policies for the information society, as well as modernising the European social model, investing in people and combating social exclusion, while sustaining healthy economic growth (European Parliament 2010, 2014). Within Europe, in 2007, the United Nations Economic Commission for Europe (UNECE) also set up a Task Force to develop a concept for the statistical measurement of quality of employment to help national statistics offices in compiling statistics on quality of employment (UNECE 2010, 2015).

At the national level, and specifically in the UK, the *making bad jobs better* initiative became a policy priority for the government with the aim of promoting job stability, job progression, skills development and improving pay levels (HM Treasury 2008). However, this lacked evidence upon which to base the policy (Kenway 2008; Lawton 2009) with the government also acknowledging constraints in achieving such improvements (HM Treasury 2008). In 2017, the government commissioned Taylor et al. (2017) to review modern working practices and develop proposals that promote fair and decent work within the UK economy in response to advances in technology and the pace of change in the modern economy. The review proposed a framework for quality of work that considered wages, employment quality, education and training, working conditions, work-life balance, and consultative participation and collective representation (Taylor et al. 2017).

Despite all these efforts, an obstacle to concrete policy actions on QWE has been the challenge of defining and measuring QWE (Cazes et al. 2015). The conceptualisation of QWE is invariably intertwined with methodological issues relating to its measurement (Burchell et al. 2014). While there is consensus in the literature about the concept being multidimensional, there is no consensus on what aspects of QWE should be included in the measure, how these are aggregated, nor indeed how much weight they contribute to the measure, whether to report *overall* and/or different *dimensions of QWE*. This is evidenced by the substantial number of measurement instruments of QWE in the literature. Some instruments include indicators unrelated to QWE, others omit important indicators, while some present unaggregated dimensions and where dimensions are aggregated, weighting is often arbitrary or equal without a theoretical explanation, and other instruments use macro-level data which make group comparison at micro-level untenable (Muñoz de Bustillo et al. 2009, 2011a, 2011b). Furthermore, a rarely considered aspect in the measurement of QWE is the testing of measurement equivalence or invariance of the instrument, despite that it is a prerequisite for between-group comparisons. Measurement equivalence provides an evaluation of whether observed indicators measure the same underlying theoretical construct, and between-group comparison is feasible if the instrument has adequate measurement equivalence (Wang and Wang 2020).

1.2 Rationale for the Study

The aim of this study is to develop a multidimensional measurement instrument of QWE at the employee level that considers *overall* and different *dimensions of QWE*, while also considering the weight of the indicators on the measurement instruments. This will allow descriptive analysis by different groups of employees but also enable modelling to predict levels of *overall* and *dimensions of QWE* for employees with different characteristics. This study also seeks to contribute new knowledge in the analysis of work and employment by applying *item response theory* (IRT) modelling to develop a multidimensional measurement instrument of QWE using secondary data. IRT modelling is part of a class of latent variable models widely used to develop measurement instruments in educational measurement and psychometrics (Desjardins and Bulut 2018) but rarely so in the research on work and employment. The study is of the UK employee population and excludes the self-employed because of a lack of comparable attributes of QWE between employees and the self-employed in social survey data, while some attributes may have different implications in measuring QWE between the two populations. The QWE for the would benefit from a separate analysis. The measurement instrument will be used to compare QWE between different groups of UK employees in the labour market based on demographic, socio-demographic, and socio-economic characteristics and the objective and research questions are outlined below.

1.3 Research Objectives

1.3.1 Objective 1

Review and evaluate existing efforts to create indices or measurement instruments of QWE.

1.3.2 Objective 2

Develop a multidimensional theoretical framework for measuring QWE.

1.3.3 Objective 3

Apply item response theory modelling to construct a measurement instrument of QWE for the UK employee population that addresses the limitations of existing measures, including considering overall and different dimensions of QWE.

1.3.4 Objective 4

Apply differential item functioning to evaluate measurement equivalence of the measurement instrument for different groups of UK employees.

1.3.5 Objective 5

Conditional on adequate measurement equivalence of the QWE measurement instrument, compare and predict levels of QWE for different groups of UK employees.

1.4 Overview of Thesis

Chapter 2 presents a review of the literature on QWE including issues related to its measurement. It considers the concept of QWE, focusing on challenges of its definition and approaches to its measurement. This also includes contributions from various traditions of the social sciences, a review of some of the existing measurement instruments, as well as developing a theoretical framework of QWE to be used in this study. The chapter also briefly considers the impact of institutional regimes (power resources theory and varieties of capitalism) on QWE, specifically focusing on the UK labour market. It also offers a review of the literature of predictors of QWE and introduces IRT as a method of creating measurement instruments, including a brief history and examples of its applications.

Chapter 3 provides an overview of the methodology of conducting the research. This outlines the exploration of secondary survey data sources with topics on work and employment and provides a detailed critique of the survey questions as well as the strengths and limitations of the data. The chapter builds on the introduction to IRT and considers the assumptions related to the models, mathematical formulations of some of the models, including the graded response and multidimensional IRT models. The chapter also introduces the extensions of IRT modelling to differential item function (DIF) for evaluating item-level measurement equivalence between respondents from different groups, multiple group modelling for comparing levels of QWE between different groups for a single predictor, and multiple indicators multiple causes models involving modelling QWE with multiple predictors. This chapter also explores estimation methods of IRT model parameters and considers frequentist and Bayesian approaches.

Chapter 4 introduces the indicators and predictors QWE used in this study. This presents the results of the univariate analysis of the indicators and predictors of QWE and the bivariate analysis examining associations between the predictors of QWE and each of the indicators of QWE in the UK employee population. The chapter discusses the results, particularly of the bivariate analysis in relation to previous literature and the UK labour market.

Chapter 5 applies IRT to construct a multidimensional measurement instrument of QWE for the UK employee population that addresses the limitations of existing measures. The chapter assesses the dimensionality of the indicators of QWE and how this informs potential measurement models. Different measurement models are compared and results of the measurement model that better fits the data are presented in detail, including the item slope-intercept parameters, model diagnostics, predicted latent trait scores, and an evaluation of the psychometric properties of the instrument.

Chapter 6 applies DIF to evaluate the measurement equivalence of the measurement instrument of QWE developed in Chapter 5 for different groups of employees. Conditional on adequate measurement equivalence, this chapter conducts multiple group analysis to compare levels of QWE for individual predictors without controlling for any other characteristics. The chapter also discusses the findings considering what characteristics predicts QWE and how the results compare with other literature.

Following on from the multiple group analysis in Chapter 6, Chapter 7 considers multiple predictors of QWE and presents results of the multiple indicators multiple causes (MIMIC) models using frequentist and Bayesian methods. MIMIC models model the effects of the demographic, socio-demographic, and socio-economic characteristics on different aspects of QWE. The chapter also includes an extensive discussion of the results and considers how these compare with those from other literature.

Lastly, Chapter 8 provides a summary of the research and highlights the contributions of the study to the topic of measuring QWE. This considers theoretical contributions to the conceptualisation of QWE and developing a framework for measuring QWE. This chapter also highlights the methodological contributions of the study in relation to limitations with the data and how it can be improved, as well as the application of IRT modelling and how it addresses limitations of existing measures of QWE. This also considers the application of frequentist and Bayesian approaches to model parameter estimation. The chapter also considers the substantive contributions and highlights, particularly, findings from this study that are inconsistent with previous literature. Finally, it considers the study's limitations, outlines how the research could be further developed, and draws conclusions from the research.

Chapter 2 Review of the Literature

This chapter aims to review the literature on the quality of work and employment including issues related to its measurement. The first section considers the concept of quality of work and employment, focusing on challenges of its definition and approaches to its measurement. This also includes contributions from various traditions of the social sciences on what dimensions are important, a review of some of the existing measurement instruments of quality of work and employment, as well as developing a framework to be used in this study. The second section introduces item response theory (IRT) as a method of creating measurement instruments, including a brief history of IRT and examples of its applications. The third section briefly considers theoretical perspectives of analysing the quality of work and employment in terms of the impact of institutional regimes (power resources theory and varieties of capitalism), although this this is limited to the UK. The last section reviews predictors of quality of work and employment, and groups these by demographic, socio-demographic, and socio-economic characteristics.

2.1 Quality of Work and Employment

2.1.1 Defining the Concept

There is a consensus in the literature about the concept of job quality being necessarily multidimensional as it is associated with different attributes or characteristics of jobs (Cazes et al. 2015; Felstead et al. 2019; Gallie 2007b; Green 2006; Kalleberg 2011; Leschke, Watt, and Finn 2008; Muñoz de Bustillo et al. 2009, 2011b). However, the concept is complex and elusive because, while it is intuitively understood within the social sciences, it cannot be directly measured; thus, it is unobservable or latent, and difficult to precisely define (Muñoz de Bustillo et al. 2009, 2011a, 2011b). Furthermore, there is no consensus in the literature on what exactly constitutes a ‘good job’ (Burchell et al. 2014; Findlay, Kalleberg, and Warhurst 2013; Kalleberg

2011). In part, this is due to the fact that workers have different motivations in relation to what constitutes good job quality, and moreover, these motivations vary over the course of their working lives (Taylor et al. 2017). Proposed definitions recognise the diverse attributes related to job quality (Kalleberg 2011), while also being worker-centred and associated with well-being;[1] thus, it can generally be defined as the degree to which a job has work and employment attributes that enhance or diminish the well-being of workers (Burchell et al. 2014; Felstead et al. 2019; Green 2006; Holman 2013b; Muñoz de Bustillo et al. 2011a).

Further to the lack of consensus on a definition, terminology used in the literature adds to the confusion, for example, expressions such as 'quality of working life', 'quality of work', 'quality of employment', 'decent work' are often used interchangeably without a clear rationale (Burchell et al. 2014). The concept can be decomposed into two broad components; thus, *employment quality* and *work quality* (Cazes et al. 2015; Muñoz de Bustillo et al. 2011a). Employment quality relates to attributes linked to employment relations that have an impact on worker well-being (Cazes et al. 2015; Muñoz de Bustillo et al. 2011a), and includes aspects such as remuneration, contractual arrangements, career development (Muñoz de Bustillo et al. 2011a). On the other hand, work quality relates to attributes associated with the work activity itself and conditions under which the work takes place that impact on worker well-being (Cazes et al. 2015; Muñoz de Bustillo et al. 2011a), and includes aspects such as work autonomy, work time scheduling, social environment (Muñoz de Bustillo et al. 2011a). In this research study, hereon in, job quality will be referred to as *quality of work and employment* (QWE).

The conceptualisation of QWE is invariably intertwined with methodological aspects relating to its measurement (Burchell et al. 2014), and it is necessary to identify what aspects ought to be included in the measure, as well as their impact (Muñoz de Bustillo et al. 2011b).

[1] The concept of well-being of workers is, in itself, elusive and multidimensional like that of job quality (Muñoz de Bustillo et al. 2011a), while also unobservable.

The following sections will consider approaches to measuring QWE and dimensions of QWE, including developing a conceptual framework to be used in this study and a critical review of some of the current QWE measurement instruments in the literature.

2.1.2 Approaches to Measuring Quality of Work and Employment

There are different approaches to measuring QWE based on whether evaluations of the attributes of QWE are subjective or objective (Brown, Charlwood, and Spencer 2012; Burchell et al. 2014; Felstead et al. 2019; Gallie 2007b; Green 2006; Muñoz de Bustillo et al. 2009, 2011b). Subjective approaches use self-reported subjective evaluations such as job satisfaction, as a proxy measure of QWE, and interpreted as positively associated with QWE (Brown et al. 2012; Burchell et al. 2014; Felstead et al. 2019; Muñoz de Bustillo et al. 2009, 2011b). While it might be advantageous that job satisfaction as a measure of QWE reduces a multidimensional concept to a unidimensional measure that is easier to understand and incorporates workers' preferences (Muñoz de Bustillo et al. 2009, 2011a, 2011b), it oversimplifies a complex concept and the influence of workers' preferences has implications for comparative analysis (Burchell et al. 2014; Felstead et al. 2019; Green 2006). Indeed, empirical evidence suggests that workers often report being satisfied with jobs of objectively poor quality (Brown et al. 2012). However, it still offers useful insights into workers' expectations and experiences about a job (Brown et al. 2012; Felstead et al. 2019). An important disadvantage is that job satisfaction does not focus on the attributes of the work and employment but is rather an outcome based on individual preferences, which may be influenced by other attributes unrelated to QWE (Wright et al. 2018). As such it cannot be decomposed into different indicators or dimensions of QWE to adequately understand the attributes influencing the overall score in order to inform policy, and is therefore not a suitable measure of QWE (Brown et al. 2012; Burchell et al. 2014; Felstead et al. 2019; Muñoz de Bustillo et al. 2009, 2011b).

An alternative approach to measuring QWE is to illicit workers' opinions in evaluating job attributes they consider important for QWE (Burchell et al. 2014; Muñoz de Bustillo et al. 2009, 2011b; Wright et al. 2018). Similarly to job satisfaction, this approach is based on a subjective evaluation of QWE and has the advantage of incorporating workers' preferences (Muñoz de Bustillo et al. 2009, 2011b). However, more importantly and in contrast to job satisfaction, is that workers are presented with specific job attributes to evaluate, which may include subjective evaluation of objective attributes associated with the job (Burchell et al. 2014; Muñoz de Bustillo et al. 2009, 2011b; Mussmann 2009; Wright et al. 2018). This means that measures of QWE based on this approach can be decomposed by different indicators or dimensions of QWE and therefore, potentially more informative for policy development. Apart from the disadvantages associated with subjective evaluations being based on individual preferences, a consequential challenge is determining what job attributes to present to the workers for evaluation (Muñoz de Bustillo et al. 2009, 2011b; Wright et al. 2018). This fundamentally relates to the conceptualisation of QWE (Burchell et al. 2014), therefore, this approach cannot be the sole basis for measuring QWE (Muñoz de Bustillo et al. 2009, 2011b).

In contrast to the previous two approaches, this last approach aims to use; as far as is possible; objective evaluations of job attributes that have an impact on workers' well-being (Brown et al. 2012; Burchell et al. 2014; Cazes et al. 2015; Felstead et al. 2019; Muñoz de Bustillo et al. 2011a, 2011b). Similarly to the previous approach, it requires that workers be presented with specific job attributes to evaluate; however, these attributes are drawn upon solid theoretical foundations and a body of empirical literature on work and employment from different traditions of the social sciences (Burchell et al. 2014; Muñoz de Bustillo et al. 2009, 2011a, 2011b), and is the approach applied in this study. In this regard, theoretical justifications and empirical evidence for the job attributes to consider in a measure of QWE are grounded in literature. However, evaluations of the job attributes are often collected using social survey instruments by asking workers about their experiences of the job and as such, some evaluations will be subjective evaluations of objective attributes (Brown et al. 2012; Cazes et al. 2015; Felstead et al. 2019). A criticism of the objective approach is that it assumes work and employment is of central importance to workers' lives and does not take in account variations between individuals, for instance, individuals might opt for jobs with a seemingly poor attribute while this better serves their non-work lives and vice versa (Green 2006).

2.1.3 Dimensions of Quality of Work and Employment

Muñoz de Bustillo et al. (2009) drew upon theoretically founded empirical evidence from different traditions of the social sciences to identify a potential list of dimensions that have an impact on workers' well-being to consider in developing a measurement instrument of QWE. While these dimensions were mainly drawn from economic and sociological traditions, they included contributions from the literature on labour market segmentation, occupational medicine along with health and safety literature, and work-life balance studies (Gallie 2007b; Muñoz de Bustillo et al. 2009, 2011b) and are displayed in Table 2.1.

Table 2.1: Dimensions of Quality of Work and Employment from Different Traditions of the Social Sciences

Orthodox Economic Approach: Compensating Differentials	Radical Economic Approach: Exploitation	Traditional Sociological Approach: Alienation and Intrinsic Quality of Work	Institutional Approach: Segmentation and Employment Quality	Occupational Medicine, and Health and Safety Literature: Risks and Impact of Work on Health	Work-life Balance Studies
Labour compensation	*Power relations*	*Objective strand*		*Conditions*	*Working time*
1. Wages	2. Industrial democracy as a compensating power	3. Skills	9. Contractual status and stability of employment	11. Physical risks	15. Duration
		4. Autonomy	10. Skills development and career progression	12. Psychosocial risks	16. Scheduling
		Subjective strand		*Outcomes*	17. Flexibility
		5. Powerlessness		13. Perceived impact of work on health	18. Regularity
		6. Meaninglessness		14. Absenteeism	19. Clear boundaries
		7. Social isolation			*Intensity*
		8. Self-estrangement			20. Pace of work and workload
					21. Stress and exhaustion

Adapted from Muñoz de Bustillo et al. (2009:51).

Economic Traditions

The economic tradition has provided different perspectives on the topic of QWE, although the dominant school of thought has focused on pecuniary indicators, with QWE equated to levels and stability of economic compensation, particularly wages (Cascales Mira 2021; Kalleberg 2011; Muñoz de Bustillo et al. 2011b). However, the implications of wages are still debatable (Gallie 2007b). Orthodox economic approaches are dominated by the theory of compensating wage differentials (Muñoz de Bustillo et al. 2011b). This suggests that for similarly skilled workers and controlling for other characteristics, workers in jobs with poorer conditions are more likely to be paid more than those in jobs with better conditions (Green 2006; Muñoz de Bustillo et al. 2011b).[2] Critics of this theory, within and outside the economics domain, have argued that the assumptions are stringent, and despite the wage compensations, wage adjustments may not sufficiently compensate for the poorer conditions (Felstead et al. 2019). Furthermore, if workers were fully compensated for the differences in their conditions, then the issue of QWE is not an economic argument (Muñoz de Bustillo et al. 2011b).

On the other hand, radical economic approaches championed by neo-Marxists argued that in capitalist economies unequal power relations existed between the working and capitalist classes (Gallie 2007b). According to neo-Marxists workers were always exploited regardless of competition in the labour market, though less so in more competitive markets (Grint and Nixon 2015; Kalleberg 2011; Muñoz de Bustillo et al. 2009, 2011b). As a result, in such economies, poor working conditions were associated with low wages, unless workers enhanced their power resources and gained enough economic and political force, typically through organised labour movement (Muñoz de Bustillo et al. 2011b). From this perspective, the fight for the betterment of working conditions is more a political issue reflecting power relations

[2] This theory assumes perfect competition, with workers having complete information about the associated working conditions, and there is full employment in the labour market (Muñoz de Bustillo et al. 2011b).

between working and capitalist classes (Esping-Andersen 1990; Grint and Nixon 2015; Korpi 1985, 2006; Muñoz de Bustillo et al. 2009, 2011b; Olsen and O'Connor 2018). Thus, the exchange of labour for wages was really an exchange of labour power and defining 'good' working conditions would become a political issue reflecting dynamics in society (Grint and Nixon 2015; Muñoz de Bustillo et al. 2009, 2011b). However, the prevailing school of thought in the literature is that higher QWE is characterised by higher wages, including other fringe benefits (Felstead et al. 2019; Green 2006; Holman 2013b; Kalleberg 2011).

Sociological Traditions

While economic traditions focused on pecuniary aspects of QWE, sociological traditions highlighted intrinsic aspects (Table 2.1) associated with work and employment, such as skills and autonomy (Cascales Mira 2021; Felstead et al. 2019; Gallie 2007b; Green 2006; Kalleberg 2011; Muñoz de Bustillo et al. 2009). From the sociological perspective, these aspects are rooted in Marx's theory of alienation and are integral to the measurement of QWE (Edgell and Granter 2020; Gallie 2007b; Gallie and Zhou 2013; Kalleberg 2011; Muñoz de Bustillo et al. 2009, 2011b). According to Marx's ([1959] 2000) theory of alienation critiqued work in industrial capitalist societies and suggested that this stifled innate human creativity by alienating workers from the product of their labour, activity of the work, their essential nature and ultimately from each other. In his *Labour and Monopoly Capital* published in 1974, Harry Braverman, a sociologist, argued that the sense of alienation was exacerbated by Taylorism[3] and the rise of Fordism[4] in the first half of the twentieth century with the development of the

---

[3] Taylorism is a system that took a scientific approach to the management of the organisation of work by centralising the labour process, including every aspect of the production process. This involved gathering knowledge of the labour processes previously held by the workforce, monopolising this knowledge among managers, and managers using this monopoly to control all aspects of the labour process (Edgell and Granter 2020).

[4] The concept of Fordism has various interrelated meanings, firstly, this can refer to the production system or labour process characterised by mass production, which is the context applied in this study. Secondly, this can relate to an economic system

industrial organisation which introduced machinery and assembly-line processes (Edgell and Granter 2020; Gallie 2007b). According to Braverman, the approach to the organisation of the labour process based on Taylorism principles of management and Fordist production techniques had the effect of monopolising knowledge and authority among the management, resulting in the deskilling and degradation of labour and a workforce with limited autonomy (Edgell and Granter 2020; Gallie 2007b; Muñoz de Bustillo et al. 2011b).

In contrast to Braverman, another sociologist, Daniel Bell, published work around the same time arguing that the impact of modern technology had rather had the effect of 'upskilling' for workers (Edgell and Granter 2020; Muñoz de Bustillo et al. 2011b). According to Bell, modern technology had resulted in industrial societies transforming into post-industrial societies in which the economic structure was dominated by service work, with knowledge the key factor of production and a professionally skilled workforce with increased autonomy (Edgell and Granter 2020). The debate of the effect of changes in the economic structure of industrial societies in terms of deskilling and upskilling continues in the literature, with empirical evidence suggesting a complicated picture of deskilling but also upskilling and skill polarisation (Martinaitis, Christenko, and Antanavičius 2021). This debate has important implications for workers' experience of alienation and analysis of QWE (Muñoz de Bustillo et al. 2011b), and according to neo-Marxists, in order for workers to establish de-alienation, they needed control of the conception and execution of their work activities (Edgell and Granter 2020).

While Marx's ([1959] 2000) theory of alienation, traditionally, referred to the alienation or state of separation of workers from the means and product of their labour in an objective sense, other sociologists and social psychologists introduced a subjective component to the

---

characterised by mass consumption, and thirdly, a regulation system that supports mass production (e.g. physically healthy workforce) and mass consumption (e.g. financially healthy consumers) (Edgell and Granter 2020).

theory. For example, consideration of aspects such as job satisfaction (Muñoz de Bustillo et al. 2011b), or Seeman's (1959) five different dimensions or meanings of alienation;[5] which were limited by the sociologist, Robert Blauner, to powerlessness (lacking freedom and control at work), meaninglessness (lack of understanding and purpose of the work activity), social isolation (lacking a sense of belonging and identity), and self-estrangement (lack of involvement and fulfilment at work) (Table 2.1) (Edgell and Granter 2020; Muñoz de Bustillo et al. 2011b). However, the subjective interpretation of alienation was insensitive to Marx's traditional concept and introduces important limitations of the use of subjective dimensions in the measurement of QWE. Thus, workers can report being satisfied despite being alienated from the means and product of their labour (Muñoz de Bustillo et al. 2011b).

Other Social Sciences Traditions

The segmentation theory arose as a reaction to the neoclassical economic approach to the labour market, which viewed the labour market as a single market governed by competitive rules of supply and demand of labour (Muñoz de Bustillo et al. 2011b). The segmentation theory was initially proposed by Peter Doeringer and Michael Piore in 1970, although this was in terms of *dualism* and depicted the labour market as, rather, segmented into primary and secondary sector workforces (Gallie 2007b; Muñoz de Bustillo et al. 2011b). With the rise of large-scale unemployment in the UK in the 1980s, literature on labour market segmentation highlighted some aspects central to the notion of QWE (Gallie 2007b); thus, contractual status and employment stability on the one hand, and opportunities for skills development and career progression on the other (Table 2.1) (Gallie 2007b; Muñoz de Bustillo et al. 2011b). The primary sector workforce was characterised by high levels of job security with progression opportunities, while the secondary sector workforce was in dead-end jobs with chronic job

[5] Thus, powerlessness, meaninglessness, normlessness, isolation, and self-estrangement (Seeman 1959).

insecurity and non-standard (temporary, part-time) contracts (Edgell and Granter 2020; Gallie 2007b).

The health and safety of workers is a fundamental aspect in the measuring of QWE and is associated with the structure and organisation of work (Sinclair et al. 2020). The literature on health and safety is rooted in the traditions of occupational medicine, which pre-dated the Industrial Revolution, and health and safety studies date back to the early years of the Industrial Revolution (Muñoz de Bustillo et al. 2011b). Occupational medicine systematically investigated diseases associated with certain occupations, while the Industrial Revolution ushered in a transformation from small scale production of goods in pre-industrial society to large scale production but resulted in appalling working conditions for workers during the early years (Muñoz de Bustillo et al. 2011b; Sinclair et al. 2020; Vidal 2016). However, the development of modern labour regulation in the first half of the twentieth century improved health and safety standards at the workplace. For example, the Health and Safety at Work etc. Act 1974 was the first to cover all workplaces and place greater accountability of maintaining health and safety at work on both employers and workers in the UK. These standards considered whether workplace environments exposed workers to physical and psychological risks, although the types of risks and hazards have evolved with changes in the economic structures of industrial societies (Muñoz de Bustillo et al. 2011b; Sinclair et al. 2020). Thus, while in the 1970s the risks and hazards focused more on physical aspects associated with industrial societies, these shifted towards psychosocial risks associated with the service sector after the 1980s (Muñoz de Bustillo et al. 2011b). Notably, Muñoz de Bustillo et al. (2011b) considered some aspects that are outcomes, such as absenteeism and perceived impact of work on health (Table 2.1), rather than actual attributes of the work or employment, which should not be included in a measure of QWE.

Another important dimension in the measurement of QWE identified in the social sciences literature is the notion of work-life balance, which relates to the balance between work and non-working life (Gallie 2007b; Muñoz de Bustillo et al. 2011b). The conflicting demands between work and non-working life have always been a present dynamic in the labour market; however, this became a salient issue in the course of the 1990s due to an increase of women into the labour market from the 1970s (Gallie 2007b; Muñoz de Bustillo et al. 2011b; Scherer and Steiber 2007). This is not an indication that prior to the 1970s work-life balance was not issue, but rather a reflection of the division of economic roles which designated women reproductive roles at home and men productive roles in the labour market (Muñoz de Bustillo et al. 2011b; Scherer and Steiber 2007). As a dimension in the measurement of QWE, indicators can be subjective or objective, but the subjective approach has important limitations that have already been highlighted in measuring QWE. From an objective approach, indicators for the work-life balance dimension consider working time; such as duration of working hours, working times, flexible working arrangements, regular work pattern, and clear boundaries separating work and non-working time; as well as intensity, including the workload and pace of work, and stress and exhaustion (Table 2.1) (Felstead et al. 2019; Gallie 2007b; Kalleberg 2011; Muñoz de Bustillo et al. 2009, 2011b; Sinclair et al. 2020). However, indicators of stress and exhaustion are outcomes of rather than components of QWE, while workload and pace of work are more indicative of the working conditions.

2.1.4 Review of Existing Measurement Instruments of Quality of Work and Employment

This section reviews some of the existing (or proposed) measurement instruments of QWE in the literature based on the work by Muñoz de Bustillo et al. (2009, 2011a, 2011b), but also considers more recent measurement instruments. The measures of QWE are listed in Table 2.2, along with some of the features of the measures and the review considers the shortcomings of the existing measures, while highlighting some of the challenges of developing a measure of QWE.

Table 2.2: Existing (or Proposed) Measurement Instruments of Quality of Work and Employment

Index (Acronym)	**Coverage**	**Author** (Year)	**Aggregation / Weighting** Dimensions	**Aggregation / Weighting** Overall	**Data Source** (Periodicity) / **Level**	**Measurement Equivalence**
The Measuring Job Quality Working Group proposal	UK	Irvine et at.(2018)	No / N/a	No / N/a	LFS (Quarterly) / Micro	N/a
Chartered Institute of Personnel and Development (CIPD) Job Quality Index	UK	Gifford (2018)	Yes / Equal	No / N/a	UKWL (Annually) / Micro	No
Job Quality Indicators	UK	ONS (2019, 2022)	No / N/a	Yes / Equal (2019)	LFS, APS (Annually) / Micro	No
Job Quality	UK	Zwysen and Demireva (2020)	Yes / Model based	No / N/a	UKHLS (Annually) / Micro	No
DGB Good Work Index (DGBI)	Germany	Mussmann (2009)	Yes / Equal	Yes / Equal	Ad hoc survey (Annually) / Micro	No
Quality of Work Index (QoWI) and Quality of Employment Index (QoEI)	Luxembourg	Steffgen et al. (2020)	Yes / Equal	No / N/a	Ad hoc survey (single exercise) / Micro	No
Job Quality Index	South Africa	Monnakgotla and Oosthuizen (2021)	Yes / Equal	Yes / Equal	LFS (Quarterly), LMD (Annually) / Micro	No
The European Job Quality Index (EJQI)	European Union	Leschke et al. (2008; 2012), Piasna (2017)	Yes / unequal	Yes / Equal	Multiple sources / Macro	No
Job Quality Index (JQI)	European Union	Muñoz de Bustillo et al. (2011b)	Yes / Mostly equal	Yes / Equal	EWCS (5 years) / Micro	No
European Working Conditions Survey (EWCS)	European Union	Eurofound (2012, 2017b)	Yes / Equal	No / N/a	EWCS (5 years) / Micro	No

Continued…

Continued…

Index (Acronym)	**Coverage**	**Author** (Year)	**Aggregation / Weighting**		**Data Source** (Periodicity) / **Level**	**Measurement Equivalence**
			Dimensions	Overall		
European Intrinsic Job Quality Index (EIJQI)	European Union	Cascales Mira (2021)	Yes / Model based	Yes / Equal	EWCS (5 years) / Micro	Yes
Laeken indicators of job quality (Laeken)	European Union	European Commission (2008)	Yes / Data based	No / N/a	ECHP, ELFS, SILC (Annually) / Macro	No
Tangian's proposal (Tangian)	European Union	Tangian (2007)	Yes / Equal	Yes / Unequal	EWCS (5 years) / Micro	No
OECD Job Quality Framework	OECD and non-OECD countries	Cazes et al. (2015)	Yes / Equal	No / N/a	OECD Statistics database / Micro	No
The Quality of Employment (QoE)	Latin American developing countries	Sehnbruch et al. (2020)	Yes / Equal	No / N/a	Multiple sources (single exercise) / Macro	No
Decent Work Index-1 (DWI-1)	Developed and developing countries	Ghai (2003)	Yes / Equal	Yes / Equal	ILO databases (single exercise) / Macro	No
Decent Work Index-2 (DWI-2)		Bonnet et al. (2003)	Yes / Equal	Yes / Equal		No
Decent Work Index-3 (DWI-3) (proposal)		Anker et al. (2003)	Yes / N/a	No / N/a		No
Decent Work Index-4 (DWI-4)		Bescond et al. (2003)	No / N/a	Yes / Equal		No

* *Notes*: N/a – Not applicable; APS – Annual Population Survey; ECHP – European Community Household Panel; ELFS – European Labour Force Survey; ILO – International Labour Organisation; LMD – Labour Market Dynamics; SILC – Statistics on Income and Living Conditions; EWCS – European Working Conditions Survey; UKWL: UK Working Lives. The DWI-3 is a proposal rather than an operationalised measure of QWE. Adapted from Muñoz de Bustillo et al. (2011b:141–42).

Composition of the Measures

Considering the measures of QWE in Table 2.2, one of the limitations rooted in the lack of consensus on a definition of QWE and inherent challenges in its conceptualisation is the composition of measures of QWE, evidenced by the inclusion of dimensions related to the labour market but unrelated to QWE in some indices or systems of indices (Muñoz de Bustillo et al. 2009). This is particularly the case with the Laeken Indicators of Job Quality (*Laeken*), which include gender equality, access to the labour market as well as overall economic performance and productivity dimensions (European Commission 2008), and the ILO's Decent Work Indices which include the Gini index, unemployment rate, inflation or absolute poverty as indicators of QWE (Anker et al. 2003; Bescond et al. 2003; Bonnet et al. 2003; Ghai 2003). The framework proposed by the Measuring Job Quality Working Group[6], and also employed by the Chartered Institute of Personnel and Development (CIPD) Job Quality Index (Gifford 2018), included a health, safety and psychosocial well-being dimension as a measure of QWE (Irvine et al. 2018). However, considering the definition of QWE relates to attributes that enhance or diminish the well-being of workers (Burchell et al. 2014; Felstead et al. 2019; Green 2006; Holman 2013b; Muñoz de Bustillo et al. 2011a), well-being is arguably an outcome rather than a component of QWE. On the other hand, attributes relating to health and safety are more indicative of working conditions associated with the job, but these attributes need to focus on the work activity, such as exposure to hazardous chemicals.

Conversely, some measures of QWE omitted some dimensions considered important within social sciences literature in the measurement of QWE (Muñoz de Bustillo et al. 2009).

[6] This Working Group was co-chaired by Martyn Evans (Carnegie UK Trust) and Matthew Taylor (Royal Society of Arts (RSA) Future Work Centre) and included members from the Office for National Statistics (ONS), Chartered Institute of Personnel and Development (CIPD), Resolution Foundation, Trades Union Congress (TUC), as well as expert contributions of members from other influential organisations in the topic of work and employment, including the Department of BEIS (Irvine, White, and Diffley 2018).

A majority of the indices listed in Table 2.2 do not include indicators on work intensity or health and safety, except in general, those that use the EWCS, and this is largely attributed to data availability (Muñoz de Bustillo et al. 2009). However, some indices have important omissions beyond availability of data. For example, the ILO's Decent Work Index-1 (Ghai 2003), the *Laeken* (European Commission 2008), as well as the European Intrinsic Job Quality Index (EIJQI) (Cascales Mira 2021) omit a dimension related to wages, which is an important omission in a measure of QWE.

There are also considerable variations in the number of indicators or dimensions among measures of QWE (Muñoz de Bustillo et al. 2009). For example, for the list in Table 2.2, the number of indicators included in different indices or systems of indices ranged from five for the Quality of Employment (QoE) (Sehnbruch et al. 2020) to over 100 indicators for Tangian's proposal, which uses the EWCS (Tangian 2007). On the other hand, in terms of dimensions, the OECD's Job Quality Framework (Cazes et al. 2015) and the QoE (Sehnbruch et al. 2020) had three dimensions, while the ILO's Decent Work Index-3 had 11 dimensions (Anker et al. 2003). Furthermore, some measures, such as the German Confederation of Trade Unions' (DGB) Good Work Index (DGBI) (Mussmann 2009), are based on purely subjective indicators even for objective attributes of work and employment such as pay, with respondents asked, for example, whether their pay was appropriate considering the work they did. While this provides workers' evaluation of good work as argued by the DGB (Holler 2013; Mussmann 2009), this is based on individual preferences which depend on individual circumstances with implications for comparative analysis (Muñoz de Bustillo et al. 2009). Additionally, empirical evidence suggests that workers often report being satisfied with jobs of objectively poor quality (Brown et al. 2012).

Aggregation

One of the methodological challenges in creating a measure of QWE is the aggregation of indicators (Muñoz de Bustillo et al. 2009). The measure proposed by the Measuring Job Quality Working Group (Irvine et al. 2018); and employed in the ONS's Job Quality Indicators (ONS 2019, 2022);[7] recommended presenting unaggregated results for individual indicators of QWE, grouped by dimensions outlined in their framework and suggested that this provided a multi-faceted picture of QWE. Furthermore, the Working Group were also against reporting composite indices of their dimensions and an overall measure. They argued that a single measure was less transparent and may misrepresent important trends in individual indicators, while also oversimplifying a complex concept.[8] However, they proposed that if reporting composite indices, these should be accompanied by the unaggregated results of the individual indicators (Irvine et al. 2018). Presenting individual indicators grouped by dimension considers individual indicators in isolation without accounting for the complexity of the relationships that exist between the indicators in evaluating the quality of a job and implies that all indicators are of equal importance in measuring QWE.

For most of the measures of QWE in Table 2.2, aggregation was carried out at one or two levels. Firstly, individual indicators were aggregated within a dimension of QWE and this was the case for all the measures, except for the measure proposed by the Measuring Job Quality Working Group (Irvine et al. 2018), the ONS's Job Quality Indicators (ONS 2019, 2022), the ILO's Decent Work Index-4 (Bescond et al. 2003), and instances where a dimension was measured by one indicator, such as the QoE (Sehnbruch et al. 2020). Secondly, dimensions

[7] Although the initial exploratory analysis of Job Quality Indicators reported a binary composite measure with equal weighting of the individual indicators (ONS 2019).

[8] These arguments are, in part, similar to those levelled against the use of job satisfaction as a measure of QWE, although important differences are the subjectivity of job satisfaction as well as the inability to decompose job satisfaction into different dimensions to understand what attributes are influencing the overall score to adequately inform policy (Muñoz de Bustillo et al. 2011a, 2011b).

were aggregated to create an overall measure of QWE for most of the measures, including the European Job Quality Index (EJQI) by the European Trade Union Institute (ETUI) (Leschke et al. 2008, 2012; Piasna 2017), the Job Quality Index (JQI) by Muñoz de Bustillo et al. (2011b) as well as the EIJQI (Cascales Mira 2021).

On the other hand, other measures such as the CIPD's Job Quality Index (Gifford 2018), Eurofound's (2012, 2017b) EWCS measures, including the European Commission's (2008) *Laeken* did not aggregate dimensions to an overall measure, while Steffgen et al. (2020) aggregated dimensions to measures of the Quality of Work Index (QoWI) and Quality of Employment Index (QoEI) but not into an overall measure. Notably, the reason for not aggregating dimensions for the CIPD index was partly to avoid making decisions about weighting, although they acknowledged that some dimensions were more consequential than others (Sarkar and Gifford 2018). Differences over whether to aggregate indicators or not, highlight the lack of consensus in the operationalisation of measuring QWE. However, aggregated measures synthesise information capturing different aspects of work and employment and can be impactful in summarising multidimensional data (Muñoz de Bustillo et al. 2009).

Weighting

The aggregation of indicators, inevitably, requires decisions to be made about the weighting of the indicators on the aggregate measure but there is also no consensus in the literature on how weights should be assigned. Often these are assigned in an arbitrary manner or assumed to be equal without a theoretical explanation (Muñoz de Bustillo et al. 2009, 2011a, 2011b), but are also influenced by the aggregation method. Most of the measures in Table 2.2 used equal weighting, including Sehnbruch et al. (2020), who argued that this was to highlight the equal importance among the dimensions in their QoE measure (labour income, employment stability, and employment conditions), but also for the ease with which the measure can be interpreted for policymaking purposes.

For their JQI, Muñoz de Bustillo et al. (2011b) also applied equal weighting of indicators within most of their dimensions but assigned a higher weight for physical risks compared to psychosocial risks in their health and safety dimension, based on arguments in the health and safety literature. Muñoz de Bustillo et al. (2011b) then estimated the arithmetic means of the indicators within dimensions; however, for the overall measure, all dimensions were equally weighted and a geometric mean[9] was estimated. While analytically, this approach may be simple, estimating means for indicators with ordinal levels of measurement; as is the case for indicators used in the JQI and for most of the measures in Table 2.2, is conceptually inaccurate. This is because intervals or distances between response categories of indicators with ordinal levels of measurement are not necessarily equal (Agresti and Finlay 2014; Field, Miles, and Field 2012), and values assigned to the categories merely distinguish between the categories and reflect their natural relative ordering or ranking.

The weighting strategy used by Muñoz de Bustillo et al. (2011b) was similar to the one employed by the ETUI for their EJQI, although this was arbitrary without a clear rationale, but they conducted a sensitivity analysis of their weighting strategy within dimensions (Leschke et al. 2008, 2012; Piasna 2017). In contrast, Tangian's (2007) proposal applied equal weighting of indicators within dimensions and unequal weighting of dimensions on the overall measure based on the premise that the number of indicators within a dimension reflected the dimension's importance; thus, dimensions with more indicators were assigned higher weights. However, conceptually, the number of indicators in a dimension has no bearing on the relative importance of the dimension.

[9] Muñoz de Bustillo et al. (2011b) estimated geometric means because they do not assume a linear relationship between the dimensions and are not influenced by extreme scores compared to arithmetic means.

On the other hand, Cascales Mira (2021) and Zwysen and Demireva (2020) applied alternative approaches to weighting in developing their measures. Cascales Mira (2021) used exploratory factor analysis (EFA) to determine the number of latent variables (dimensions) that account for the relationship among indicators of their EIJQI, while Zwysen and Demireva (2020) applied latent class analysis (LCA) to determine the number of latent classes (dimensions) based on responses to indicators in their measure.[10] Both these approaches do not assume a prior latent structure among indicators, and weights are determined by the relationship between the indicator and the latent variable or class membership it is measuring, based on a statistical model that can be tested by evaluating how well it fits the data (Bartholomew et al. 2008; Bartholomew, Knott, and Moustaki 2011). However, the latent variables or class membership need to be conceptually interpretable. In developing the *Laeken*, the European Commission (2008) applied principal components analysis (PCA)[11] to determine the weighting of indicators on the dimensions. In this instance, the weights of indicators on the dimensions are based on the relationship between the indicator and the principal components (dimensions), (Bartholomew et al. 2008), but this is a descriptive technique that cannot be tested.

These model or data-based approaches reduce the arbitrariness of assigning the weights of indicators on the dimensions. Criticisms often levelled at these approaches are that the results may be counterintuitive and not necessarily align with the purposes of public policy or

[10] EFA and LCA are applied in the analysis of multivariate data and part of latent variable models, which theorise that the relationships among a set of observed variables (alternatively, responses to a set of observed variables) (indicators) are dependent on some underlying unobserved (or latent) variable(s) / class membership (dimensions). EFA and LCA are mainly distinguished by the level of measurement of the set of observed and the underlying unobserved variables, which are both continuous in EFA, while they are both categorical in LCA (Bartholomew et al. 2008; Bartholomew, Knott, and Moustaki 2011).

[11] While PCA is also applied in the analysis of multivariate data, it is a descriptive technique that reduces the dimensionality of a set of observed variables (or summarises patterns of correlations among a set of observed variables) (indicators) into a smaller number of principal components, which are uncorrelated variables (dimensions) while retaining as much information from the data as possible (Bartholomew et al. 2008).

consensus view on the importance of indicators (Sehnbruch et al. 2020). However, with survey data often collected from the workers, perhaps patterns in the data may be more informative in estimating weights, while it can also be argued that public policy should be driven by data and consensus built on empirical evidence. Both EFA and PCA assume that indicators have interval or ratio levels of measurement (Bartholomew et al. 2008), but notably, Cascales Mira (2021) and the European Commission (2008) created their measures using indicators with ordinal levels of measurement. Indeed, it is the case that observed indicators of different aspects of QWE in social survey data are often categorical. Furthermore, to estimate the overall score of the EIJQI, Cascales Mira (2021) calculated the arithmetic mean of the dimensions, which assumes the dimensions had equal weight on the overall index and the rationale was that based on their model, the dimensions explained a similar proportion of the variance among the indicators.

Data Sources

Another challenge with creating a measure of QWE is related to availability of data (Muñoz de Bustillo et al. 2009), with no single source of data capturing all the relevant indicators of QWE (Cazes et al. 2015; Wright et al. 2018). To address this limitation multiple data sources, often derived from surveys, are often used but this means that the measures are developed at macro-level and cannot be disaggregated at individual or micro-level (Muñoz de Bustillo et al. 2009) as is the case for some measures in Table 2.2. On the other hand, some measures use single sources of data, and while they may be hampered by missing indicators, the measures can be disaggregated by groups of interest, including analysis at country-level for cross-national surveys, such as the EWCS. However, measures of QWE based on survey data are also affected by the periodicity of the survey in terms of how regularly they can be reviewed, which can have implications for enacting and reviewing policies (Muñoz de Bustillo et al. 2009). Measures of QWE based on the EWCS, for example, can only be reviewed every five years due to its periodicity; however, empirical evidence found relatively marginal changes in most aspects of QWE over a 15-year period (Eurofound 2007). Notably, the CIPD Job Quality Index uses data from the UK Working Lives (UKWL) survey, an annual survey of UK adults in work, with participants selected using a quota sampling design (Sarkar and Gifford 2018). While the periodicity of the survey allows for a regular review of the index, the sample is based on a non-probability sampling design and this has implications for inferences that can be drawn about the population of UK adults in work based on the sample; thus, results based on this sample may be biased (Heeringa, West, and Berglund 2017).

Measurement Equivalence

A seldom considered aspect in the literature on the measurement of QWE is the testing of measurement equivalence of the measurement instrument, despite that it is a prerequisite for between-group comparisons. Thus, measurement equivalence provides an evaluation of whether the observed indicators are measuring the same underlying theoretical construct and if it holds, then between-group comparison is feasible (Wang and Wang 2020). Except for the EIJQI by Cascales Mira (2021) who evaluated equivalence of the factor loadings by country, all measures of QWE in Table 2.2 did not evaluate measurement equivalence of their measurement instruments prior to between-group comparisons and implicitly assumed adequate measurement equivalence. This may, perhaps, be due to limited application of robust methodology in the measurement of QWE.

2.1.5 Frameworks of Quality of Work and Employment

In terms of frameworks of QWE, different studies have included different dimensions in their measures. Table 2.3 displays frameworks from more recent studies with a focus on measures from the UK, Europe, and some international frameworks, although this does not include ILO frameworks. ILO frameworks tend to include dimensions related to the labour market but are unrelated to QWE.

Table 2.3 Frameworks of Quality of Work and Employment

Framework	Department for BEIS	QuInnE Project	Measuring Job Quality Working Group	Chartered Institute of Personnel Development	European Trades Union Institute	European Working Conditions Survey	United Nations Economic Commission for Europe	OECD Job Quality Framework	Job Quality Index
Author	Taylor et al. (2017)	European Commission (2018)	Irvine et at. (2018)	Gifford (2018)	Leschke et al. (2008, 2012), Piasna (2017)	Eurofound (2012, 2017b)	UNECE (2010)	Cazes et al. (2015)	Muñoz de Bustillo et al. (2011b)
Dimensions	1. Wages 2. Employment quality 3. Education and training 4. Working conditions 5. Work-life balance 6. Consultative participation and collective representation	1. Wages 2. Employment quality 3. Education and training 4. Working conditions 5. Work-life balance 6. Participation and representation	1. Pay and benefits 2. Terms of employment 3. Job design and nature of work (including skills use, control, progression opportunities) 4. Social support and cohesion 5. Health, safety and psychosocial well-being 6. Work-life balance 7. Voice and representation	1. Pay and benefits 2. Terms of employment (including development opportunities) 3. Job design and nature of work (including use of skills) 4. Social support and cohesion 5. Health and well-being 6. Work-life balance 7. Voice and representation	1. Wages 2. Non-standard forms of employment 3. Skills and career development 4. Working conditions and job security 5. Working time and work-life balance 6. Collective interest representation	1. Earnings 2. Prospects 3. Skills and discretion 4. Work intensity 5. Physical environment 6. Working time quality 7. Social environment	1. Income and benefits 2. Security of employment and social protection 3. Skills development and training 4. Employment-related relationships and work motivation 5. Working time and work-life balance 6. Social dialogue 7. Safety and ethics	1. Earnings 2. Labour market security 3. Quality of the working environment (including learning opportunities, work autonomy, working long hours)	1. Pay 2. Employment quality 3. Intrinsic quality of work (including skills and autonomy) 4. Health and safety 5. Work-life balance

Overall, the frameworks consisted of between five and seven similar dimensions, although the OECD Job Quality Framework (Cazes et al. 2015) consisted of three dimensions; however, this included learning opportunities, work autonomy, and working long hours indicators under one dimension of *quality of the working environment*. The assigning of indicators to dimensions was not consistent across frameworks. For example, the framework by Muñoz de Bustillo et al. (2011b) had an *intrinsic quality of work* dimension that included skills and autonomy indicators, whereas in other frameworks there were specific *skills and development* dimensions, with work autonomy included under the *working conditions* dimension. On the other hand, the *working conditions* dimension for the ETUI framework (Leschke et al. 2008, 2012; Piasna 2017) included job security indicators, which are included *employment quality* or *terms of employment* dimensions in other frameworks. As mentioned in the previous section, frameworks by the Measuring Job Quality Working Group (Irvine et al. 2018) and the CIPD (Gifford 2018) include dimensions with well-being measures, which are arguably outcomes rather than inputs of QWE, while the framework for the United Nations Economic Commission for Europe (UNECE 2010) includes a safety and ethics dimension which is unrelated to QWE.

The EWCS framework by Eurofound (2012, 2017b) presented a measure of QWE with seven dimensions, consisting of *earnings*, *prospects* for career progression and job security, *skills and discretion* measuring learning and training opportunities, *work intensity* measuring work demands, *physical environment* relating to physical risks at the workplace, *working time quality* measuring aspects such as atypical working time, and *social environment* relating to social relationships at the workplace. Taylor et al. (2017) proposed a framework of Quality of Work as part of their review of modern working practices on behalf of the Department for BEIS. This consisted of six dimensions based on the framework developed for the Quality of Jobs and Innovation Generated Employment Outcomes (QuInnE) project, which was a

European Commission interdisciplinary project examining the mutual impact between job quality and innovation, including the effects on job creation and quality (Erhel and Guergoat-Larivière 2016; European Commission 2018). The six dimensions were *wages*, *employment quality*, *education and training*, *working conditions*, *work-life balance*, and *participation and representation* (Erhel and Guergoat-Larivière 2016; European Commission 2018; Taylor et al. 2017).

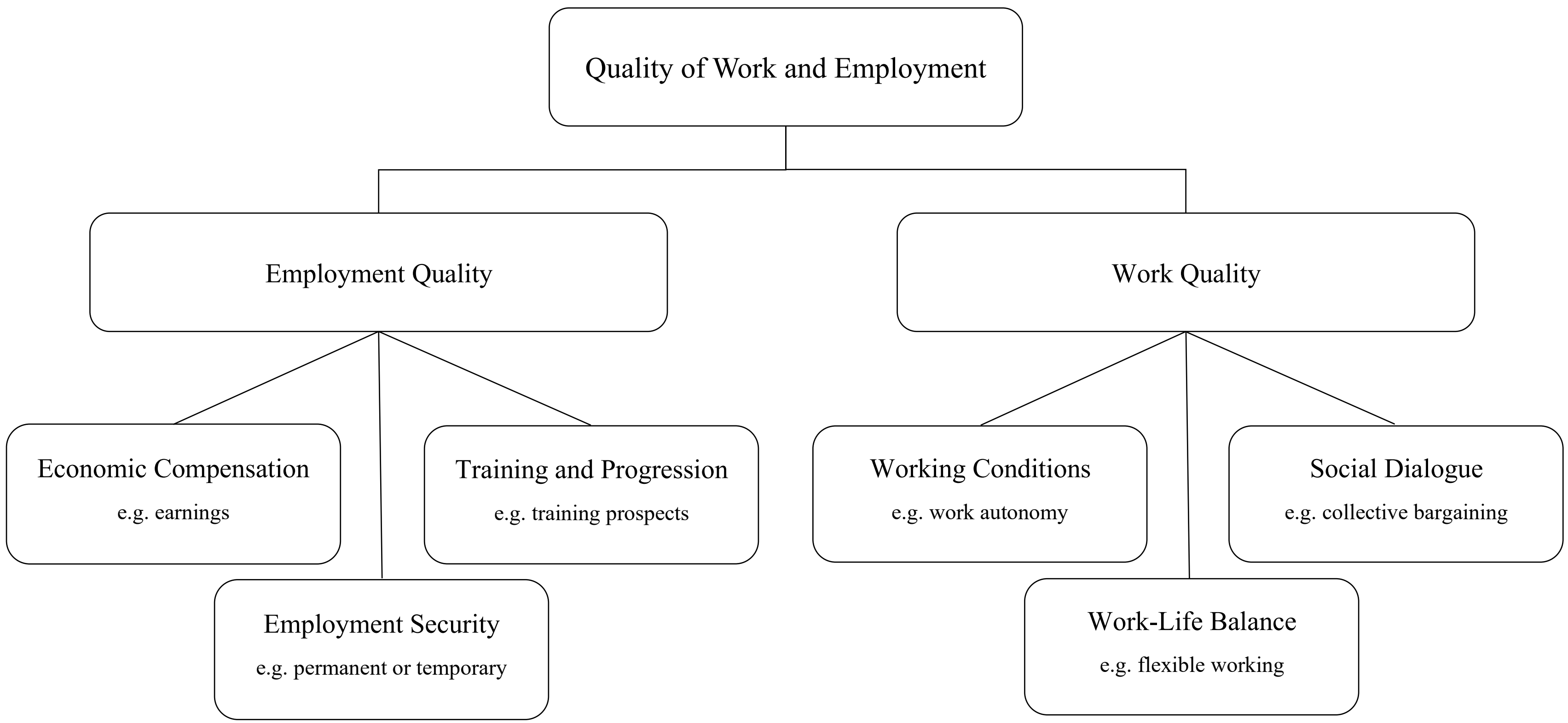
Framework of Quality of Work and Employment for this Study

The framework of QWE for this study will be based on the frameworks from the QuInnE project (Erhel and Guergoat-Larivière 2016; European Commission 2018) and Taylor et al. (2017). This was because these frameworks presented more comprehensive dimensions of a measure of QWE compared to other frameworks and excluded dimensions that are not appropriate for QWE. For example, they excluded dimensions that are outcomes of or are unrelated to QWE such as well-being in the case of frameworks by the CIPD (Gifford 2018) and the Measuring Job Quality Working Group (Irvine et al. 2018), or safety and ethics in the case of UNECE (2010). On the other hand, some frameworks excluded important dimensions, for example, Muñoz de Bustillo et al.'s (2011b) framework excluded collective representation.

Other frameworks aggregated dimensions that should be disaggregated, for example, the OECD's framework aggregated working conditions, skills development, and work-life balance into one dimension (Cazes et al. 2015). Conversely, other frameworks disaggregated dimensions that should, arguably, be aggregated, such as work intensity and physical environment for the Eurofound (2012, 2017b) framework, which are indicators of working conditions. Lastly, the ETUI framework included job security as part of the working conditions dimension, when perhaps this fitted better with the non-standard forms of employment dimension (Leschke et al. 2008, 2012; Piasna 2017).

The QuInnE project (Erhel and Guergoat-Larivière 2016; European Commission 2018) and the Department of BEIS (Taylor et al. 2017) frameworks consisted of the *wages*, *employment quality*, *education and training*, *working conditions*, *work-life balance*, and *participation and representation* dimensions. However, notwithstanding the comprehensive nature of these frameworks, the conceptualisation of their dimensions could be improved. For example, the 'wages' dimension is framed as being measured by pecuniary rewards, whereas this should capture non-wage pecuniary rewards as well and was framed as the *economic compensation* dimension in the framework for this study. The 'education and training' dimension was framed as the *training and progression* dimension to associate skills development and progression at the workplace. Additionally, 'employment quality' is used to describe a broad component of QWE linked to employment relations that have an impact on workers' well-being (Cazes et al. 2015; Muñoz de Bustillo et al. 2011a), therefore this was framed as the *employment security* dimension. The QWE framework to be used in this study will consist of six dimensions, namely *economic compensation*, *training and progression*, *employment security*, *working conditions*, *work-life balance*, and *social dialogue* dimensions. The framework is displayed in Figure 2.1.

Figure 2.1: Theoretical Framework of Quality of Work and Employment



Considering the employment quality aspects in Figure 2.1, the economic compensation dimension relates to the remuneration of workers for their labour. This can be captured by pecuniary rewards such as pay, and non-wage pecuniary rewards such as access to pension schemes, paid annual leave or sickness absence, and any other remuneration-related benefits provided by the employer. The training and progression dimension is concerned with opportunities for training and progression provided by the employer. This may be work-related training, on or away from the job, provided by the employer, as well as opportunities for career advancement. On the other hand, employment security is the extent to which the employment offers protection against job loss. This can be captured by the type of employment, whether it is permanent or temporary, fixed or predictable working hours, or information on perceived job security.

In terms of work quality in Figure 2.1, the working conditions dimension considers the conditions under which workers carry out their work. This can relate to the health and safety aspects such as exposure to hazardous work activities, work intensity, job variety, as well as control over work activities. The work-life balance dimension relates to the aspects of work that may conflict with non-work time or the extent to which work can be reconciled with private life. This may include indicators such as flexible working arrangements, working sociable or unsociable times, including working long hours. Lastly, the social dialogue dimension considers the degree to which workers have the freedom of association, with the right to organise and bargain collectively. This can be captured by the availability of recognised trade unions or staff associations to represent workers at the workplace.

2.2 The United Kingdom Labour Market

Theoretical debates suggest institutional structures have important implications for the economic performance and social processes of advanced capitalist economies, including the functioning of labour markets and quality of work and employment (Gallie 2007b; Green et al.

2013; Howell 2003). Empirical evidence from comparative analysis studies has indicated cross-national variations in levels of QWE based on institutional regimes (Gallie 2007a, 2009; Green et al. 2013; Holman 2013a). There are diverse competing approaches to these institutional regimes, and two influential approaches are the *power resources theory* (or *employment regimes*) and *varieties of capitalism* (or *production regimes*) (Gallie 2007b; Korpi 2006).[12]

The UK labour market is an archetypal example of a *market regime* in the PRT approach (Gallie 2007b; Holman 2013a), while it is an archetypal example of an *liberal market economy* (LME) in the context of the VoC approach (Gallie 2007b; Hall and Soskice 2001; Soskice 1999). Labour markets in *market regimes* are characterised by minimal regulation and assume market adjustments will, in the long-run, result in relatively high levels of employment (Gallie 2007b; Holman 2013a), while remunerations for employees are associated with their marginal productivity (Gallie 2007b). In *market regimes*, organised labour has limited involvement in the decision-making within organisations or influence on the government (Gallie 2007b; Holman 2013a), with employment levels not particularly regarded as an appropriate political aim, while employment conditions are considered a concern for individual organisations (Gallie 2007b). Consequently, labour markets in these regimes are relatively fluid, with few restrictions on employers hiring and firing employees regardless of contract type, less willingness by employers to train their labour force due to likely poor returns on such investments. They are also characterised by highly restricted work-family support, strong stratification by class, and differentials; such as wages, the complexity of job design, hence autonomy; that primarily reflect skill differences, as well as minimal public sector involvement (Gallie 2007b; Holman 2013a).

[12] Refer to Esping-Andersen (1990), Gallie (2007b), Korpi (1985, 2006), and Olsen and O'Connor (2018) for background literature on *power resources theory*. For *varieties of capitalism* see Gallie (2007b), Hall and Soskice (2001), Korpi (2006), and Soskice (1999).

Labour markets in *LME*s are deregulated, and companies resolve their coordination problems primarily through market arrangements with minimal state involvement, resulting in highly fluid markets (Gallie 2007b; Hall and Soskice 2001; Soskice 1999). These labour markets are characterised by strong inter-company competition and anti-collusion policies; corporate governance structures with financial systems that provide short-term investment for companies, limiting their financial capital, hence investment in human capital; and education and training systems emphasising general education (Hall and Soskice 2001; Soskice 1999, 2005). The education and training systems lack industry-specific skills post-compulsory secondary education, except for employees with sufficient general education, leading to a highly polarised labour force with weakly developed low-skilled and strongly developed high-skilled workers (Gallie 2007b; Soskice 1999). In terms of industrial relations systems, employers have no legal obligation to establish workplace employee representation for collective bargaining over wages and working conditions, while trade unions have limited roles; though in some sectors, they have a strong influence (Hall and Soskice 2001; Soskice 1999). This, in turn, facilitates unilateral decision-making by top management, including the authority to hire and fire employees with implications for job security, and flexibility in wage-setting to attract employees with appropriate skills (Hall and Soskice 2001; Soskice 1999).

2.3 Predictors of Quality of Work and Employment

2.3.1 Demographic Characteristics

Sex

Inequalities by sex in the labour market are well documented in literature (Korpi 2018), and females tend to be in precarious work characterised by part-time and/or temporary employment contracts than males (Edgell and Granter 2020; Fredman 2004; Pollert and Charlwood 2009). The precarious nature of employment for females means that, compared to males, they were more likely to be in employment with lower levels of overall QWE marked by poor employment security, greater impediment in accessing training, little or non-linear career progression pathways (Fredman 2004; Piasna and Plagnol 2018) due to career breaks as a consequence of childrearing (Lindley 2015; Piasna and Plagnol 2018), and more likely to be unrepresented or excluded from collective bargaining agreements (Pollert and Charlwood 2009). Furthermore, females were more likely to earn less than males (Fredman 2004; Lindley 2015; Piasna and Plagnol 2018), although this gap has narrowed in the UK, partly as a result of an increase in rates of higher educational attainment by females (Lindley and Machin 2012). However, variations in subjects of degrees and choice of occupation by sex mean the gap persists (Lindley 2015). In terms of job control, studies have found no differences between females and males in the UK employee population (Gallie and Zhou 2013; Lindley 2015; Wu, Xu, and He 2021). On the other hand, evidence suggests that female employees in the UK were more likely to have higher levels of work-life balance than male employees; however, this was attributed to job design with employers seeking low-cost and flexible labour rather than female employees' preferences or the need to accommodate family responsibilities (Piasna and Plagnol 2018; Tomlinson 2007).

Ethnic Group

Disparities in labour market experiences by ethnic group can, in part, be attributed to historical roots which defined race and ethnicity as marks of inferiority (Dillon 2020; Korpi 2018). In their study, Zwysen and Demireva (2020) found that UK-born employees from ethnic minority backgrounds were less likely to be in jobs with high levels of economic compensation, work-life balance, job security, and intrinsic satisfaction compared to UK-born employees from a White ethnic background. There were also differences within ethnic minority groups, with UK-born Pakistani and Bangladeshi, South-Asia and Black migrant employees consistently in jobs of poor quality (Zwysen and Demireva 2020). These findings were supported by Clark and Ochmann (2022) in their study, who found that male UK employees from Black Caribbean, Black African, Pakistani and Bangladeshi ethnic backgrounds were more likely to be in low-paid, temporary or involuntary part-time jobs compared to males from a White British ethnic background.

Age Group

The experience of different forms of precarious work in the labour market varies by age (Kim and Kurz 2001), and empirical evidence has shown that, generally, younger employees fare worse off than older employees (Arranz, García-Serrano, and Hernanz 2019), although this is not a linear relationship. In their study of the UK employee population, Kim and Kurz (2001) found that marginal part-time work (< 15 hours per week) was more common among older than younger or middle-aged employees, while fixed-term contracts were more common among younger employees. These contracts offer limited employment security and lower levels of economic compensation, particularly in low-skilled occupations (Kim and Kurz 2001) and often not covered by collective bargaining agreements (Bosch 2009). Other empirical evidence has indicated that compared to older employees, younger employees were more likely to participate in work-related training (Canduela et al. 2012; Dieckhoff, Jungblut, and O'Connell 2007). This could be in part, due to employers being reluctant to train older workers because of perceived limited returns on their investment (Canduela et al. 2012), but also a willingness to learn for younger employees when they enter the labour market. Studies have also suggested generational differences in relation to the centrality of work in employees' lives, with younger employees less likely to feel work should be an important aspect of their lives than older employees (Smola and Sutton 2002) and an emphasis of 'working to live, not living to work' approach to work-life balance (Sturges and Guest 2004). However empirical evidence has indicated that younger employees had poor work-life balance than older employees, which can be partly attributed to them prioritising establishing their careers by working long hours (Sturges and Guest 2004).

2.3.2 Socio-demographic Characteristics

Relationship Status

Research in social sciences has shown that relationship status is associated with a number of outcomes including productivity and economic outcomes, and in terms of QWE, evidence suggests marriage is positively associated with earnings (Bardasi and Taylor 2008; Ribar 2004). Much of the empirical research has focused on the marriage premium, specifically for males, and in the UK, evidence has indicated that married males were more likely to have higher wages than single males (Bardasi and Taylor 2008; Ribar 2004; Schoeni 1995). However, evidence pertaining to females is more ambiguous with research often framed in terms of marriage penalties (Ribar 2004). Furthermore, there are debates about causality, thus whether marriage increases productivity among males hence the higher earnings, or it is a case of self-selection into marriage among more productive, high earning males or indeed a combination of both (Bardasi and Taylor 2008; Ribar 2004). The observed marriage premium can be attributed to spousal support, which allows spouses to concentrate on their activity of specialisation (market or non-market), stabilising influence of marriage which can lead to accumulation of human capital resulting in increased productivity and/or opportunities for progression, and employer bias towards married employees due to societal norms and latent high valued characteristics associated with marriage (Bardasi and Taylor 2008; Ribar 2004) although this may be unsustainable in a competitive market economy (Ribar 2004).

Parental Status

Parenthood has an impact on the QWE of employees and evidence suggests that working parents, especially those with young children (Bryan and Sevilla 2017), make trade-offs and place increased priority on flexible working arrangements, often to the detriment of other aspects of QWE, thus they undertake less secure, low-paying jobs below their skill level and forgo progression opportunities (King’s College London and Working Families 2021). On the other hand, for employees without children, especially in professional occupations and regardless of sex, control over work schedules results in poor work-life balance with increased unpaid overtime hours than for working parents (Chung and van der Horst 2020). Empirical evidence has also indicated that lone parents were particularly disadvantaged and more likely to be in precarious employment compared to coupled parents (Nieuwenhuis and Maldonado 2018), due to having fewer options to use as bargaining power to access jobs of higher quality (Esser and Olsen 2018; King's College London and Working Families 2021). Furthermore, while there is heterogeneity among lone parents, they tend to be predominantly female (Esser and Olsen 2018; Klett-Davies 2016; Nieuwenhuis and Maldonado 2018), and empirical evidence from the UK suggests the precarity of their employment can in part be explained by the nature of work lone mothers do, thus they are more likely to be in low or lower-middle skilled occupations compared to coupled mothers, who are twice as likely to be in high skilled professions than lone mothers (Klett-Davies 2016).

Illness or Disability

Much of the emphasis on disability and the labour market in the UK has focused on government initiatives on improving participation of the disabled in paid employment and supporting them at the workplace (Grover and Piggott 2015; Lewis, Dobbs, and Biddle 2013), such as the Equality Act 2010 which prohibits discrimination in employment and recruitment on the basis of disability among other characteristics, direct or indirectly (Powell 2021). Evidence indicates that people with disabilities are increasingly joining the workforce (Department for BEIS 2018). However, in terms of the QWE, disabled employees experience prejudice in the labour market compared to non-disabled employees (Grover and Piggott 2015; TUC 2021b). Empirical evidence based on UK data suggests that, compared to the non-disabled, disabled employees tend to be in non-standard employment (Davidson and Kemp 2008) characterised by job insecurity (Meager and Hill 2005), low pay, often ineligible for sick pay or occupational pensions, limited pathway to promotion or career progression, lower levels of job autonomy, and less likely to have union representation (McGovern, Smeaton, and Hill 2004). Disabled employees are also more likely to be in part-time employment, which is associated with poor QWE (Grover and Piggott 2015) but affords better work-life balance (Lyonette 2015). This can be attributed to a variety of reasons, for example, socially embedded barriers; such as the physical environment and discriminatory attitudes of employers; while it can also be framed as an individual supply-side issue, such as skills differentials (Grover and Piggott 2015).

Region

There are longstanding disparities in the labour market, including QWE, within and across regions and nations of the UK (Jones and Green 2009), partly driven by a shift from heavy industry to a knowledge economy as a result of globalisation and technological progress (Department for LUHC 2022; Hepworth, Binks, and Ziemann 2005). Jones and Green (2009) found that based on monetary rewards, the London region had the highest proportion of high-quality jobs and the lowest proportion of low-quality jobs compared to other regions. This was followed by the South East region on both measures, while there was little variation between other regions, although the North East, Wales and Northern Ireland had the lowest proportion of high-quality jobs (Jones and Green 2009). High levels of pay in London and the South East were supported by evidence in other literature (Department for LUHC 2022; Low Pay Commission 2021), and this can be attributed to the high-skilled workforce working in the knowledge economy that is highly centralised in these regions (Hepworth et al. 2005; Jones and Green 2009). Notably, London is an extreme case where the knowledge economy is at its most competitive, whilst also least inclusive (Hepworth et al. 2005; TUC 2021b), thus according to the TUC (2021b) eight of ten constituencies with the highest proportion of low paid workers were located in London.

2.3.3 Socio-economic Characteristics

Educational Attainment

Education is one of the most important investments in human capital; thus, knowledge, skills and qualifications (Okay-Somerville and Scholarios 2013; Solomon, Nikolaev, and Shepherd 2022). Empirical evidence suggests that individuals with higher levels of education have greater job resources, such as pay, job variety and autonomy, while they also experience greater job demands, such as work intensity (Solomon et al. 2022), which have implications on QWE. In their study of UK employees, Okay-Somerville and Scholarios (2013) supported the view of higher levels of QWE in terms of economic compensation, job security, skills utilisation and development, and job autonomy for graduates however there were variations among graduates. This variation was attributed, in part, to the expansion of access to higher education, which resulted in the over-supply and underemployment of university graduates in the labour market, leading to graduates being employed in non-graduate occupations (Green and Zhu 2010; Okay-Somerville and Scholarios 2013; Warhurst 2008), as well as variations in subjects of degrees (Lindley 2015).

Occupational Classification

Empirical evidence suggests QWE varies by occupational classification, attributed, in part, to skills differentials in the occupational hierarchy (Gallie 2015; Wheatley 2022). Based on the UK employee population and using major groups of the 2020 Standard Occupational Classification, Wheatley (2022) found relatively higher levels of subjective pay, employee voice, and job design; such as autonomy, skills and development prospects; among employees in high-skilled than low-skilled occupations. However, work-life balance and types of contract did not necessarily vary in an occupational hierarchical manner, with employees in administrative and secretarial occupations having higher levels compared to other occupations (Wheatley 2022). These findings were partly supported by Gallie (2015), who used the

National Statistics Socio-Economic Classification (NS-SEC) and found higher levels of pay, job skills, task discretion, and organisational participation among UK employees in higher than those in lower occupational groups. On the other hand, employees in higher occupational groups experienced greater levels of work intensity compared to those in lower occupational groups, while levels of job insecurity were higher and comparable between employees in routine (low-skilled) and higher managerial and professional (high-skilled) occupations than other occupational groups (Gallie 2015).

Full or Part-time

Another predictor of QWE is whether employees are in part-time or full-time employment and evidence from literature of the UK labour market suggested that part-time jobs, by design, required few skills and low levels of training and educational attainment compared to full-time jobs (Lyonette, Baldauf, and Behle 2010; Warren and Lyonette 2015). Part-time jobs are dominated by females, often to accommodate family responsibilities while also contributing to household income; however, the proportion of males in part-time employment has been increasing (Lyonette 2015; Warren and Lyonette 2015); and prevalent in low-skilled occupations, leading to the occupational downgrading of highly skilled employees who opt to work reduced hours, particularly females, due to limited jobs in high-skilled occupations (Lyonette et al. 2010).

Compared to part-time jobs, full-times jobs are characterised by higher levels of economic compensation, better prospects for promotion, job security, as well as unionised organisations (Warren and Lyonette 2015), but poor work-life balance (Lyonette 2015). This was supported by McGovern, Smeaton, and Hill (2004), who found that UK employees in non-standard forms of employment were less likely to have union representation, as well as job autonomy than those in standard forms of employment, while Hoque and Kirkpatrick (2003) found that UK employees in non-standard employment were more likely to be marginalised in terms of training and development, and consultation at work regardless of occupational level than those on standard forms of employment.

Organisational Sector

Empirical evidence also suggests there are variations in QWE by type of organisational sector (Leschke and Keune 2008). Evidence of the public sector pay premium for UK employees is well established, and this is partly attributable to a more skilled workforce with higher levels of education in the public sector relative to the private sector, although there are variations to this pattern (Cribb, Emmerson, and Sibieta 2014; Murphy et al. 2020; Rubery 2013). While the pay distribution is more uniform within the public sector, it is less so in the private sector and at the top of the distribution, pay in the private sector is higher than that in the public sector (Cribb et al. 2014; Lucifora and Meurs 2006). Additional to the pay differential, public sector employment offers, on average, more valuable pensions and greater coverage than private sector employment (Cribb and Emmerson 2014). Regarding other dimensions of QWE, empirical evidence from the UK employee population suggested better outcomes in the public sector than in the private sector; thus, employment security (Fontaine et al. 2020), training prospects (Leschke and Keune 2008), provisions for work-life balance (Rubery 2013), and access to and level of unionisation (Charlwood and Terry 2007).

Organisation Size

Theoretical literature suggests there is some ambiguity about the effect of firm size on QWE; however, empirical evidence using UK population data indicated that employees in small-sized firms were likely to report greater QWE for non-pecuniary indicators than those in large-sized firms (Bryson, Erhel, and Salibekyan 2021). This was supported by Storey et al. (2010), who found a negative association between employee-reported job quality and firm size in the UK population. In their study, Bryson et al. (2021) also found increased job demands, lower levels of autonomy, poor work-life balance, and employee-employer relations for employees in large-sized firms compared to those in small-sized firms, while large-sized firms were more likely to offer training than small-sized firms and firm size had limited impact on job security and skill development. However, a study by Forth et al. (2006) of a UK employee population found higher levels of pay in medium and large-sized firms compared to small-sized firms, less likelihood of formal practices that support work-life balance; although they were more likely to report having access to various flexible working arrangements if needed; or union involvement in negotiating pay and conditions in small and medium-sized firms than in large-sized. Forth et al. (2006) also found that employees in small-sized firms reported greater job security and autonomy, experienced less work intensity than those in large-sized firms.

2.4 Introduction to Item Response Theory and Applications

Similarly to (exploratory) FA and LCA, item response theory (IRT) is also part of a class of latent variable models but consists of a family of mathematical models that model the relationship between a set of observed items or variables and the underlying latent trait(s) influencing responses to those items. IRT modelling is widely used in educational measurement and psychometrics (Desjardins and Bulut 2018) but seldom so in research on work and employment. This is despite that observed items measuring different attributes of work and employment in social survey data are often categorical, for which IRT is appropriate. Central to IRT is how respondents with different levels of the underlying latent trait(s) respond to each of the items measuring the underlying latent trait(s) and places characteristics of the items, for example, the (relative) item difficulty, on the same scale as the underlying latent trait(s) (Raykov and Marcoulides 2011, 2018). The models describe the probability of a respondent's response to an item given their latent trait level(s) and the item parameters, which are characteristics of the item such as its difficulty and how well it distinguishes between respondents (van der Linden 2016; van der Linden and Hambleton [1997] 2010). IRT modelling is appropriate for categorical observed variables, with the latent traits measured on a continuous scale (Bartholomew et al. 2008, 2011).

2.4.1 Brief History of Item Response Theory

IRT has its foundations in the work by Alfred Binet published in 1905, with his colleague Théodore Simon, developing a measure to differentiate between students with different learning abilities in Parisian schools for the purpose of tailoring their educational needs (van der Linden 2016; van der Linden and Hambleton [1997] 2010). This was, however, overshadowed by Charles Spearman's work published in 1904, which had introduced the idea that an observed score to a test can be decomposed into a true score and some random error. Spearman's work generated follow-up interest and became the basic assumption of what was to be known as the classical test theory (CTT) model (van der Linden 2016; van der Linden and Hambleton [1997] 2010; Lord and Novick [1968] 2008). While Binet's work did not immediately generate much follow-up interest, he realised that he had to measure a complex variable that could not be directly observed; unlike in other scientific disciplines where simple physical quantities were measured and manipulated; and these unobservable variables would later be referred to as *latent variables* (van der Linden 2016; van der Linden and Hambleton [1997] 2010).

To create a measure differentiating students with different learning abilities, Binet designed a broad spectrum of tasks thought to measure major mental functioning, such as working memory, reasoning capacity, judgement, and abstraction to capture the "richness of intelligence". The tasks were then used in a fully standardised test, where the testing materials, the administration process, and scoring rules were carefully protocolled. However, as there was no natural scale for the measurement of intelligence, Binet used the students' chronological ages as given quantities in pretesting and estimated curves (with unknown shapes) of the proportion of correct answers as a function of age for each item to identify their scale values.

These scale values were then used to estimate the mental age of the students' actual performance (van der Linden 2016a).

While the importance of Binet's work was not immediately recognised, it is his introduction of the idea of scaling that would contribute to the practice of IRT (van der Linden 2016a). This was recognised by Louis Thurstone who, in 1925, introduced a scaling method that, in contrast to Binet, disassociated age from the measurement of intelligence (van der Linden 2016; van der Linden and Hambleton [1997] 2010). Thurstone assumed an unknown latent scale for the set of items but imposed a known shape on the curves of the proportion of correct answers; thus, a cumulative normal distribution or normal-ogive function, and used their estimated location parameters as scale values for the items. In 1928, Thurstone published worked that expanded the application of this scaling method to measure other vague constructs such as attitude, based on a set of items with agree or disagree response options to attitudinal questions (van der Linden 2016a). However, this method was plagued by the confusion between the use of the normal-ogive functions as response functions and distribution functions for estimating scores in the populations of interest (van der Linden 2016; van der Linden and Hambleton [1997] 2010).[13]

Frederic Lord in 1952 and George Rasch in 1960 were among the first to overcome the confusion experienced by Thurstone and other authors relating to the distribution functions and their use as response functions (van der Linden 2016; van der Linden and Hambleton [1997] 2010). Lord formulated a two-parameter normal-ogive model, a mathematical model based on a normal-ogive function modelling the probability of a correct response to an item given an unknown level of ability (van der Linden 2016; Lord and Novick [1968] 2008). The parameters related to the characteristics of the items; thus, the *difficulty parameter*, representing the

[13] The normal-ogive model was also used by other authors as response functions for test items around this period, such as Richardson in the mid-1930s, Ferguson, Lawley, Mosier, and Tucker in the early to mid-1940s (van der Linden 2016; van der Linden and Hambleton [1997] 2010).

location on the ability scale where a respondent had a 0.5 probability of answering an item correctly, and the *discrimination parameter*, representing the degree to which the item discriminated between the probabilities of respondents with abilities below and above the difficulty parameter answering the item correctly (van der Linden and Hambleton [1997] 2010). On the other hand, Rasch formulated a model, the Rasch model, that modelled the probability of a correct response to an item given an unknown level of ability as the ratio of a respondents' ability over the sum of their ability and the difficulty parameters of the items (van der Linden 2016; van der Linden and Hambleton [1997] 2010).[14]

While much of the developments in IRT modelling, thus far, were limited to dichotomously scored items, in 1961 Rasch put forward a general unidimensional model that modelled responses for polytomous items, of which models for dichotomous items were a special case (van der Linden 2016a). On the other hand, there were no further developments of the normal-ogive model by Lord, partly due to the model not accounting for the probability of respondents guessing a correct answer for multiple-choice questions (van der Linden 2016; van der Linden and Hambleton [1997] 2010). However, a statistician, Allan Birnbaum, who was working in relative isolation in the late 1950s to make the normal-ogive model statistically feasible, proposed replacing the normal-ogive model with a logistic model, but his work only became known through his contributions to Lord and Novick ([1968] 2008). The formulation with two parameters was called the two-parameter logistic (2-PL) model, with the difficulty and discrimination parameters retaining their interpretation from the two-parameter normal-ogive model. Furthermore, Birnbaum proposed a third parameter to account for the probability of respondents with low ability answering a multiple-choice item correctly by guessing, and

[14] Rasch was aware that this formulation could be transformed into a form approximating the normal-ogive function using a logistic transformation, however, he did not use this transformation (van der Linden 2016a).

this was called the three-parameter logistic (3-PL) model (van der Linden 2016; van der Linden and Hambleton [1997] 2010; Lord and Novick [1968] 2008).[15]

Later landmark contributions in the development of IRT modelling can be attributed to work published in 1969 and 1972 by Fumiko Samejima ([1997] 2010, 2016) which proposed the *graded response model* for polytomous items with ordered categories. In 1972, R. Darrell Bock ([1997] 2010) also proposed an IRT model for polytomous items with unordered categories. Other IRT models modelling respondents' performance on ordered and unordered polytomous items are variations of the graded response model by Samejima (Masters and Wright [1997] 2010; Muraki [1997] 2010) and nominal categories model by Bock (Thissen and Steinberg [1997] 2010). Since the 1970s, the field of IRT modelling has matured, with alternative models such as multidimensional IRT models (Reckase 2009, 2016), explanatory response models (De Boeck and Wilson 2016) as well as modelling longitudinal item response data (Cai 2016) developed. However, some of the most effective developments have been the rigor in the application of statistical techniques including tests of model fit, and computational power that has improved the efficiency of parameter estimation (van der Linden 2016a).

2.4.2 Applications of Item Response Theory Modelling

Since the 1980s, IRT modelling has been extensively studied and applied in educational measurement, with the research including the measurement of achievement, aptitude, and ability constructs (Bock and Gibbons 2021; Reise and Revicki 2015). However, it is only from the early 2000s that the application of IRT evolved beyond the confines of educational measurement, into other typical performance domains such as psychopathology, personality, patient-reported outcomes, and health-related quality-of-life measurement, as well as in market research (Bock and Gibbons 2021; van der Linden 2016a; Reise and Revicki 2015). The

[15] Although this model no longer defined a logistic function (van der Linden and Hambleton [1997] 2010).

application of IRT modelling in these domains was motivated, much like in educational assessment, by the need to address practical and technical issues associated with measurement (Reise and Revicki 2015), similarly to some of the current challenges with the measurement of QWE (Muñoz de Bustillo et al. 2009, 2011a, 2011b).

Educational Measurement

The Programme for International Student Assessment (PISA) was launched by the OECD in 1997 in response to the general public and governments' need for comparable evidence of educational outcomes at national and international levels and applied IRT modelling (OECD 2019).[16] This aimed to develop and conduct a large-scale international assessment for monitoring educational system outcomes related to student achievement, within a common international framework and provide information for informing policy grounded in empirical data (OECD 1999). The PISA survey has been administered every three years since 2000 to students aged 15 years from OECD and other participating countries, within a framework focusing on reading, mathematical, and scientific literacy domains. The PISA does not primarily evaluate the mastery of specific curriculum content on these domains, but also places emphasis on the mastering of processes, understanding of concepts and assesses the extent of students' application of knowledge and skills within these domains to new situations (OECD 1999, 2019).

Health Measurement

In health measurement, the Patient-Reported Outcomes Measurement Information System (PROMIS) project, launched in 2004, was one of the first major projects to emphasise

[16] The PISA is not the first international comparative survey to employ IRT modelling for monitoring educational outcomes based on student achievement (OECD 1999). Other international comparative surveys include the International Association for the Evaluation of Educational Achievement, established in 1958 and the International Assessment of Educational Progress, created by the Education Testing Service in 1988.

the application of IRT modelling in developing item banks for patient-reported health outcomes (Cella et al. 2010; Gershon, Hays, and Kallen 2016; Revicki, Chen, and Tucker 2015). This was driven by the desire from healthcare providers, insurance companies, and the government to incorporate input from patients in clinical decision-making through the assessment of patient-reported outcomes (PROs) and assessing outcomes as part of healthcare evaluation (Cella et al. 2010; Gershon et al. 2016). However, the assessment of health outcomes was not systematic or consistent, with multiple assessments of varying quality and for different purposes developed; for example, instruments developed for a specific clinical study or the simultaneous development of multiple instruments measuring the same health domain by different researchers (Revicki et al. 2015). The PROMIS project aimed to develop item banks of different health domains providing efficient, flexible, and publicly available measurements of PROs, including health-related quality of life for the clinical research community (Cella et al. 2010; Gershon et al. 2016).

Psychological Tests

There are different types of psychological tests employed in psychometrics and some important tests are in the areas of intelligence, personality, and clinical testing (De Boeck 2016). Much of the traditional application of IRT modelling has been limited to unidimensional models, however, responses to some test items reflect more than a single latent construct and require more complex modelling to reflect the multidimensionality (Bonifay 2020; Reckase 2009). In clinical testing, alterations of cortical thickness in psychosis syndromes are well established (Stan et al. 2020). Thus, cortical thinning is observed, particularly, in the temporal and frontal lobes of patients with schizophrenia (van Haren et al. 2011), those experiencing first episodes of psychosis (Buchy et al. 2011; Gutiérrez-Galve et al. 2010), chronic stages of the illness (Knöchel et al. 2016), and those at ultra-high risk for psychosis (Jung et al. 2011). However, the relationship between cortical thinning in specific regions of the brain and specific psychotic symptoms had not been established (Bock and Gibbons 2021; Stan et al. 2020). Stan et al. (2020) applied multidimensional IRT modelling to investigate whether there were specific regions in which cortical thinning was associated with a particular profile of psychotic symptom ratings. They used data from the National Institute of Mental Health, consisting of respondents with a psychotic disorder, a sample of their first-degree relatives with or without a psychiatric diagnosis as well as a sample of healthy respondents.

2.5 Summary

This chapter has discussed the concept of QWE and considered challenges of defining it and the implications for its operationalisation. Different approaches to measuring QWE, such as job satisfaction, exclusively subjective approaches, and objective approaches, were explored and this study will use the objective approaches. Dimensions from different traditions of social sciences considered important in the measurement of QWE were discussed, with economic traditions focusing on wages and power relations, sociological traditions focusing on skills and autonomy, and other social sciences traditions considering contractual arrangements, skills development and career progression, as well as health and safety at work, and work-life balance.

Existing measurement instruments of QWE were reviewed and evaluated, with a particular focus on the limitations of these measures. These limitations related to the composition of the measures, with some measures including components unrelated to QWE, while others omitted important components of QWE. Issues related to the aggregation of indicators were highlighted, with no consensus on whether to report unaggregated results of individual indicators, aggregate individual indicators within dimensions, and/or aggregate the dimensions into an overall measure. Intertwined with the aggregation issues is the weighting of indicators on the aggregate measure, and there is no consensus on how these should be assigned. Often these were assigned in an arbitrary manner or assumed to be equal without a theoretical explanation. Another limitation was related to data sources, with no single source capturing all the relevant indicators of QWE. Lastly, much of the literature on measuring QWE did not consider the measurement equivalence of QWE instruments, and this was implicitly assumed, yet it is a prerequisite for between-group comparisons.

The review of the literature also considered different frameworks for measuring QWE. Some frameworks included dimensions that were outcomes rather than components of QWE, such as well-being, while others omitted important components, such as collective representation. Furthermore, other frameworks put together dimensions that should be separated, and conversely, some separated dimensions that should be put together. A theoretical framework for measuring QWE with six dimensions was proposed for use in this study, and this consisted of *economic compensation*, *employment security*, *training and progression*, *working conditions*, *work-life balance* and *social dialogue*.

The literature also considered characteristics of the UK labour market from the perspectives of PRT and VoC. The UK labour market in an archetypal example of a *market regime* from the perspective of PRT and an *LME* in the context of VoC. This has implications on some aspects of QWE for the UK employees. Thus, overall, the UK labour market is characterised by minimal state involvement, with employers having no legal obligations to establish workplace employee representation for collective bargaining over wages and working conditions. Trade unions also have limited roles, though they have strong influence in some sectors, while employers have few restrictions on hiring and firing employees, are less willing to train their labour force, and their top management have flexibility on wage-setting to attract employees with appropriate skills. Predictors of the QWE were considered and categorised in terms of demographic, socio-demographic, and socio-economic characteristics.

Lastly, the chapter introduced IRT modelling as well as its applications and proposed it as a method for developing a measurement instrument of QWE that addresses some of the shortcomings of existing measures of QWE. IRT modelling allows not only for observed items measuring QWE to be aggregated into overall and/or other dimensions of QWE depending on the specified model, but also aggregation of items with nominal or ordinal levels of measurement. It can also serve to evaluate the hypothesised latent structure of the items and how well the model fits the data. Furthermore, IRT modelling also addresses issues related to weighting of observed items on the overall and/or other dimensions of QWE. Thus, similarly to FA, LCA or PCA, which use model or data-based approaches to assign weights, IRT modelling uses item parameter estimates to determine weights of items on the measure of QWE. Finally, IRT modelling can be extended to evaluate the measurement equivalence of the QWE measurement instrument, which is a prerequisite for between-group comparisons and holds when estimated item parameters based on an IRT model are the same between respondents from different groups.

The next chapter will consider the overarching methodology for conducting the research. This will include exploring sources of data and ethical considerations of the research. It will provide an in-depth focus of IRT modelling, including its extension to evaluating measurement equivalence, and methods of comparing and predicting QWE for different groups of employees.

Chapter 3 Methodology

This chapter aims to introduce and explain the methods applied in the study. The first section explores secondary survey data sources that include topics on work and employment. This will focus, specifically, on social surveys with appropriate survey items for measuring the quality of work and employment (QWE) based on the framework proposed in Chapter 2. Different sources of data will be compared, including a critique of the survey questions as well as the strengths and limitations of the data. The second section highlights ethical considerations for the study in the context of secondary data analysis. The third section builds on the introduction to item response theory (IRT) in Chapter 2. This considers the assumptions related to IRT models, mathematical formulations of some of the models, including the graded response and multidimensional IRT models, as well as model diagnostics and comparison. The fourth section introduces the extension of IRT modelling to differential item function for evaluating item-level measurement equivalence between respondents from different groups. The fifth section introduces multiple group modelling for comparing levels of QWE between different groups for a single predictor. The sixth section introduces multiple indicators multiple causes models involving modelling levels of QWE with multiple predictors. Lastly, the seventh section explores estimation methods of IRT model parameters and considers frequentist and Bayesian approaches.

3.1 Sources of Data

This study is an analysis of secondary survey data primarily focusing on the UK employee population, therefore, UK-specific data sources and cross-national surveys with a UK sample were considered. The primary criterion for data selection was whether the data included appropriate survey items for measuring QWE, that is work and employment aspects that have an impact on employees' well-being (Muñoz de Bustillo et al. 2009). This was based on the QWE framework presented in Figure 2.1, with items capturing objective aspects preferred. However, some aspects of QWE, such as work autonomy, cannot be objectively captured by social survey instruments (Wright et al. 2018), therefore subjective items were also considered (Brown et al. 2012; Cazes et al. 2015; Felstead et al. 2019). A limitation of subjective items is that they introduce individual preferences in the evaluation of work and employment characteristics.

Other considerations included the level or unit of analysis for the survey. Survey studies that collected data at the individual level were preferred so that the measure of QWE can be developed for individual employees and allow between-group comparisons (Muñoz de Bustillo et al. 2011b; Wright et al. 2018). Related to the unit of analysis was the sample size and sampling design of the survey studies, with large sample sizes drawn using a probability sampling design preferred. Large sample sizes would ensure disaggregation of the analysis by demographic, socio-demographic, and socio-economic characteristics was feasible, while the probability sampling design would enable generalisability of the results to a wider population from where the sample was drawn. Furthermore, the periodicity of the survey was also considered so that the measure of QWE can be regularly updated and reviewed over time (Muñoz de Bustillo et al. 2011b), be it trends over time based on cross-sectional study designs or changes over time employing longitudinal study designs.

Various social surveys with items relevant to QWE were considered and an initial comparison presented in Appendix 3.1. These survey studies included the Annual Population Survey (APS), the Labour Force Survey (LFS), Understanding Society: The United Kingdom Household Longitudinal Study (UKHLS), the European Values Survey (EVS), the European Working Conditions Survey (EWCS), and the International Social Survey Programme (ISSP). Based on the availability of items for different dimensions of QWE, the EWCS, the LFS, and the UKHLS were explored further, and survey items examined in more detail.

3.1.1 European Working Conditions Survey

The EWCS is a cross-national survey whose target population includes all individuals aged 15 years (16 years in some countries) or over who are in employment and live in private households across some European countries. The survey has been conducted by Eurofound every five years since 1990 – 1991, with the sixth edition conducted in 2015 and this covered the 28 European Union (EU) Member States, as well as Norway, Switzerland, Turkey, North Macedonia, Serbia, Montenegro, and Albania (Eurofound 2017a). The EWCS aims to evaluate and quantify working conditions of employees and the self-employed, with the survey covering different aspects of the respondents' working lives such as earnings, employment status, learning and training, work organisation, working time, work-life balance, voice and participation, physical and psychosocial risk factors among other aspects (Eurofound 2016b).

The study has a complex sample design, with sampling plans designed for each country and involving multistage stratified, cluster sampling with a known non-zero probability of selection for respondents. In terms of sample size, the study has a reference sample size of 1,000 per country, while some countries have larger reference sample sizes, for example that for the UK was 1,600 for the sixth edition (Eurofound 2016a) and approximately 44,000 respondents were interviewed. For the sixth edition, computer-assisted personal interviewing (CAPI) was used across all countries and all interviews for data collection (Eurofound 2016b).

3.1.2 Labour Force Survey

The LFS is a survey of households living at private addresses in the UK and is the main source for statistics on employment, unemployment and economic inactivity. It provides information to help develop, manage, evaluate and offer insights on labour market policies and the questionnaire includes questions on earnings, health and safety, employment status, working patterns and hours of work, education and training, along with demographic, socio-demographic, and socio-economic characteristics of individuals within households. The survey is conducted by the Office for National Statistics (ONS) in Great Britain (and the Central Survey Unit of the Northern Ireland Statistics and Research Agency (NISRA) in Northern Ireland) and was first carried out in the UK in 1973 but has been conducted in its current quarterly periodicity since 1992. The target population of the LFS is based on the UK resident population and includes all individuals residing in private households, as well as in National Health Service accommodation and young people living away from their parental home in student halls of residence (ONS 2016).

For the most part, the LFS has a single-stage sampling design, with addresses randomly selected from a postcode address file that is implicitly geographically stratified and all adults within the household sampled. However, for addresses with multiple household-occupancy, only one household is randomly selected, meaning that households, strictly, have an unequal probability of selection. For the year 2016, the survey consisted of approximately 39,500 households with around 90,000 individuals every quarter. The LFS also has a rotational sampling design, where some selected addresses are retained in the sample for five consecutive quarters (ONS 2016) and this provides a longitudinal sample, with data published as two-quarter and five-quarter longitudinal data sets. In addition to the quarterly LFS data for the UK, there is also a cross-national version of the LFS, the EU LFS, covering the 28 EU Member States, Iceland, Norway, Switzerland, Turkey, and North Macedonia. The questionnaires are administered mainly through CAPI when households are first included in the sample and through computer-assisted telephone interviewing (CATI) in subsequent quarters.

3.1.3 United Kingdom Household Longitudinal Study

The UKHLS is an annual panel survey of households and individuals in the UK which started in 1991 as the British Household Panel Survey (BHPS) and subsequently replaced by the UKHLS in 2009. The UKHLS is conducted by the Institute for Social and Economic Research (ISER), University of Essex and consisted of approximately 40,000 UK households at Wave 1 (2009 – 2010) (Fumagalli, Knies, and Buck 2017) and had a sample of approximately 39,200 respondents aged 16 years and over at Wave 8 (2016 – 2017). It aims to provide high-quality longitudinal data on multi-topics to help understand UK life in the twenty-first century and how this is changing at the household and individual level. The individual questionnaire includes modules with questions on current employment for employees and the self-employed, a two-year rotating module on working conditions, job satisfaction, as well as questions on demographic, socio-demographic, and socio-economic characteristics of the respondents (University of Essex, Institute for Social and Economic Research 2018).

The study has a complex sample design involving multistage stratified, cluster sampling with a known unequal probability of selection for respondents (University of Essex, Institute for Social and Economic Research 2018). The sample design has five components, including a general population sample which is representative of the UK general population. The sampling frame is based on residential addresses or delivery points (Fumagalli et al. 2017; Lynn 2009) and excludes all those not included in the postcode address file. In terms of data collection, the UKHLS employs a mixed mode design, with CAPI and computer-assisted web interviewing (CAWI) mainly used at Wave 8 (2016 – 2017). However, CATI was also used to administer the questionnaire for individuals and households that had not responded through CAPI and CAWI (Carpenter 2018).

3.1.4 Data Selection

A summary of the review of the data sets is presented in Table 3.1. The table summarised different features of the data sets, including potential indicators of QWE based on the theoretical framework in Figure 2.1. The survey items for indicators of QWE from the EWCS (Sixth Edition, 2015) and the LFS (2016) are outlined in Appendix 3.2 and Appendix 3.3 respectively, while those for the UKHLS (Wave 8, 2016 – 2017) are outlined in Table 3.2.

Table 3.1: Review of Data Sources

Data Feature			EWCS (6th Edition, 2015)	LFS (2016)	UKHLS (Wave 8, 2016 – 2017)
Level or unit of analysis			Individuals and country level	Individuals and households	Individuals and households
Coverage			35 European countries, incl. UK	UK (EU LFS available)	UK
Sampling design			Complex sampling design	Single-stage sampling design	Complex sampling design
Periodicity			Every five years	Quarterly	Annually (2-year rotational module)
Study design			Cross-sectional design	Cross-sectional design with panel sample	Panel design with cross-sectional sample
Approximate sample size (UK individuals aged 16 years and over)			1,600	70,000 at every quarter	39,200
Dimensions of QWE and Indicators	Economic compensation	Gross pay	×	✓	✓
		Adequate pay	✓	×	×
		Pecuniary and non-pecuniary rewards	✓	×	✓
		Paid holiday	×	✓	×
		Pension provision	×	×	✓
		Pay progression	×	×	✓
	Training and progression	Training participation / prospects	✓	×	✓
		Training days	✓	×	×
		Progression prospects	✓	×	✓
	Employment security	Employment type	✓	✓	✓
		Job security	✓	×	✓
		Predictable hours	✓	✓	×
	Working conditions	Health and safety	✓	×	×
		Work intensity	✓	×	×
		Job variety	✓	×	×
		Work autonomy	✓	×	✓
	Work-life balance	Working hours	✓	✓	✓
		Flexible working arrangements	✓	×	✓
		Working times	✓	✓	✓
		Work and non-work time	✓	×	×
	Social dialogue	Direct participation and support	✓	×	×
		Collective bargaining	✓	✓	✓

Notes: (✓): Appropriate indicator available. (×): Indicator not available or available but not appropriate.

Table 3.2: Survey Items for Indicators of QWE from the UKHLS

Dimension	Indicator	Survey Question	Response Options (excluding options for missing responses)
Economic Compensation	Gross pay	The last time you were paid, what was your total (gross) pay before any deductions? This is before any deductions for tax, National Insurance or pension contributions, student loan repayments, union dues and so on. Please include any overtime, bonuses, commission, tips or tax refunds.	1. [Value >= 0]
	Pension provision	Does your present employer run a pension scheme or superannuation scheme for which you are eligible?	1. Yes 2. No
	Pay bonuses	In the last 12 months have you received any bonuses such as Christmas or quarterly bonus, profit-related pay or profit sharing bonus, or an occasional commission?	
	Pay progression	Some people can normally expect their pay to rise every year by moving to the next point on the scale, as well as receiving negotiated pay rises. Are you paid on this type of incremental scale?	
Training and Progression	Progression prospects	Even though you would not like this to happen, do {JBLKCHA = 2} / Do you think this actually will happen in the coming twelve months? (Get a better job with your current employer)	1. Yes 2. No 3. Doesn't apply
	Training prospects	Even though you would not like this to happen, do {JBLKCHB = 2} / Do you think this actually will happen in the coming twelve months? (Take up work related training)	1. Yes 2. No
Employment Security	Employment type	Leaving aside your own personal intentions and circumstances, is your job…	1. A permanent job 2. Or is there some way it is not permanent?
	Job security	I would like you to think about your employment prospects over the next 12 months. Thinking about losing your job by being sacked, laid-off, made redundant or not having your contract renewed, how likely do you think it is that you will lose your job during the next 12 months? Is it...	1. Very likely 2. Likely 3. Unlikely 4. Very unlikely
Working Conditions	Work autonomy	In your current job, how much influence do you have over... 1. What tasks you do in your job? (Job tasks) 2. The pace at which you work? (Work pace) 3. How you do your work? (Work manner) 4. The order in which you carry out tasks? (Task order) 5. The time you start or finish your working day? (Work hours)	1. A lot 2. Some 3. A little 4. None

Continued…

Continued…

Work-time Scheduling	Working hours	Thinking about your (main) job, how many hours, excluding overtime and meal breaks, are you expected to work in a normal week?	1. [Value >= 0]
	Formal flexibility	I would like to ask about working arrangements at the place where you work. If you personally needed any, which of the following arrangements are available at your workplace? Code all that apply: 1. Part-time working 2. Working term-time only 3. Job sharing 4. Flexi-time 5. Working compressed hours 6. To work annualised hours 7. To work from home on a regular basis 8. Other flexible working arrangements	0. Not mentioned 1. Mentioned
	Informal flexibility	Aside from any formal arrangements for flexible working you have, are {if JBFlex less than 9} / Are {JBFlex = none} you able to vary your working hours on an informal basis, for example by re-arranging your start or finish times if you need to?	1. Yes 2. No 3. Sometimes
	Working times	Which times of day do you usually work?	1. Mornings only 2. Afternoons only 3. During the day 4. Evenings only 5. At night 6. Both lunchtimes and evenings 7. Other times of day 8. Rotating shifts 9. Varies/no usual pattern 10. Daytime and evenings 97. Other
	Weekend working	Do you ever work at weekends?	1. Yes – most / every weekend 2. Yes – some weekends 3. No weekend working
Social Dialogue	Collective bargaining	Is there a trade union, or a similar body such as a staff association, recognised by your management for negotiating pay or conditions for the people doing your sort of job in your workplace?	1. Yes 2. No

Source: UK Household Longitudinal Study, Wave 8 (2016 – 2017).

For the economic compensation dimension, while the EWCS had more indicators than the LFS or the UKHLS, these focused on what made up the pecuniary and non-wage pecuniary rewards. Other measures considered the respondents' satisfaction with their pay, which is subjective, and earnings were based on net rather than gross earnings (Appendix 3.2). A limitation of net earnings is that it measures disposable income, and different respondents will have different deductions unrelated to QWE and is not an appropriate measure. While the LFS and UKHLS had measures of gross earnings (Table 3.1), the LFS had limited measures of economic compensation, with its other measure related to the number of paid holidays (Appendix 3.3). On the other hand, the UKHLS had a broad set of indicators which also included pension provision, bonus payments, as well as pay progression (Table 3.2).

In terms of the training and progression dimension, the LFS has some indicators related to training opportunities, but the survey items referred to previous or current employers, and it is therefore not clear which job the respondents are referring to (Appendix 3.3). For this dimension, indicators from the UKHLS were subjective and measured the respondents' perceived prospects of training or progression, and the dimension had only two indicators (Table 3.2). On the other hand, the EWCS had a broader set of indicators, with objective indicators for training participation and training days but subjective indicators for progression prospects (Appendix 3.2).

Considering the employment security dimension, both the LFS and the UKHLS had two indicators and considered employment type, and while the LFS also measured predictable hours, the UKHLS measured job security (Table 3.1). However, the job security indicator was subjective as it measured the respondents' perception of how likely they were to lose their job in the next 12 months (Table 3.2). The EWCS measured employment type, predictable hours as well as job security, but indicators of job security were also subjective (Appendix 3.2).

The EWCS had a wider range of indicators for the working conditions dimension than the LFS or the UKHLS. These included health and safety focusing on exposure to hazards at work, work intensity, job variety, and autonomy indicators and consisted of both subjective and objective indicators (Table 3.1 and Appendix 3.2). The working conditions indicators from the UKHLS focused exclusively on autonomy and were a subjective evaluation of the respondents' degree of influence over various aspects of their work (Table 3.2). Indicators for the LFS were related to health and safety; however, these focused on either the outcomes of work rather than the work characteristic or referred to the previous or current employer and were therefore not appropriate measures (Appendix 3.3).

Regarding the work-life balance dimension, all the three data sets considered working hours and working times (Table 3.1). While they all also considered indicators on flexible working arrangements, those from the LFS were based on the respondents' actual agreed formal flexible working arrangements (Appendix 3.3). A limitation of such measures is that they only measure a respondent's preferred working arrangements rather than an evaluation of the job and do not consider the range of flexible arrangements offered by the employer. These survey items are therefore not appropriate indicators of QWE. In contrast, indicators from the UKHLS considered formal flexible working arrangements available at a respondents' workplace. However, these indicators measured respondents' awareness of formal flexible working arrangements available at their organisation. The UKHLS also includes a measure of informal flexibility (Table 3.2). On the other hand, the flexible working arrangements indicator from the EWCS provided a general measure of working arrangements with no specific arrangements offered. Additionally, the EWCS also considered whether respondents' working time arrangements were regularly changed, and importantly, how working hours fitted with non-work time (Appendix 3.2). The implication of LFS and UKHLS not including a measure of work and non-work time means framing this dimension as work-life balance would not be appropriate with these data sets. Therefore, for the LFS and UKHLS, the dimension will be framed as *work-time scheduling* to broadly capture the scheduling of respondents working arrangements.

Lastly, in relation to the social dialogue dimension, the LFS (Appendix 3.3) and the UKHLS (Table 3.2) only had one appropriate indicator measuring whether a workplace had a trade union or staff association to represent respondents in collective bargaining. In contrast, the EWCS had a broader set of indicators that included not only a collective bargaining indicator but also indicators related to direct participation and support at the workplace (Appendix 3.2).

For the other features of the data sets, the LFS and the UKHLS had large sample sizes conducive for disaggregating the analysis by demographic, socio-demographic, and socio-economic characteristics. In contrast, the EWCS had a small sample size for individual countries which might have implications for the feasibility of disaggregating the analysis by different groups within countries, particularly for characteristics with a high number of groups or categories. However, the EWCS would be suitable for comparing QWE between countries. A limitation of the LFS is that not all indicators are available at every quarter, and while it has the added advantage of a panel sample for a potential longitudinal analysis, the sample is retained for five quarters which is not long enough to capture change for a concept such as QWE. Conversely, the quarterly periodicity of the survey is also too frequent to capture change in QWE over time. In comparison, the UKHLS has a panel design with a rotational module on working conditions every two years. This, along with the duration of the study means the UKHLS is ideal for cross-sectional and potential longitudinal analysis.

Overall, the EWCS had a broader set of indicators across different dimensions based on the framework for QWE in Figure 2.1, compared to other data sets, except for the economic compensation dimension. However, an important limitation of the EWCS is the small sample size for individual countries. The LFS had limited indicators across dimensions of QWE, with indicators for training and progression, and working conditions dimensions not appropriate for measuring QWE, while the social dialogue dimension had one indicator. On the other hand, the UKHLS had appropriate indicators across different dimensions of QWE, but the training and progression, and employment security dimensions each had two indicators, while the social dialogue dimension also had one indicator. The working conditions dimension was also measured exclusively by different aspects of work autonomy. However, as the focus of this research study is on the UK employee population, the UKHLS will be used due to its large sample size compared to the EWCS, as well as the potential to conduct a longitudinal analysis.

3.2 Ethical Considerations

Ethical approval for this research study was sought and obtained from City, University of London. The study had a low risk as it used secondary data and ethical issues related to accessing and the use of the data. Data were accessed from the UK Data Service (UKDS), and these were safeguarded data, which are anonymised with a remote risk of respondents being identifiable. The UKDS End User Licence (EUL) agreement sets the terms and conditions under which researchers can use the data and a summary of these is outlined in Appendix 3.4.

By accepting the terms and conditions of the EUL, researchers agree to use the data as stipulated in the EUL, including preserving the confidentiality of individuals, households or organisations in the data, and not to use the data for commercial purposes without seeking permission. Other considerations also include the correct citation and acknowledgement in publications, notifying the UKDS of any published work based on their data collection, as well as destroying all copies of the data at the end of the access period.

3.3 Item Response Theory

Item response theory (IRT) is part of a class of latent variable models and consists of a framework of mathematical models modelling the relationship between a set of observed items or variables and the underlying latent trait(s) influencing responses to those items (van der Linden 2016; van der Linden and Hambleton [1997] 2010). Fundamentally, IRT relates to the conditional probability of a random respondent selecting a particular response to an item given their latent trait level(s) and the item parameters (Desjardins and Bulut 2018; Hambleton, Robin, and Xing 2000; Reckase 2009). This conditional probability is called an *item response function* (IRF) (Bartolucci, Bacci, and Gnaldi 2016; van der Linden and Hambleton [1997] 2010), and item parameters are characteristics of an item which can include the *difficulty*, *discrimination*, *lower asymptote*, and *upper asymptote* parameters (described in the following sections) (Hambleton et al. 2000). Table 3.3 shows the classification of IRT relative to other latent variable models, although the list is not exhaustive (Bartholomew et al. 2008, 2011; Collins and Lanza 2010; Heinen 1996). IRT is appropriate if the observed items are categorical and the latent variables are assumed to be continuous (Bartholomew et al. 2008).

Table 3.3: Classification of Latent Variable Models

Latent variables	**Observed variables**	
	Continuous (interval / ratio)	**Categorical** (nominal / ordinal)
Continuous (interval / ratio)	Factor analysis	Item response theory or Latent trait analysis
Categorical (nominal / ordinal)	Latent profile analysis	Latent class analysis and Discrete latent trait analysis

Adapted from Bartholomew et al.(2008, 2011), Collins and Lanza (2010), and Heinen (1996).

3.3.1 Assumptions of Item Response Theory Models

The main assumptions of IRT models relate to *dimensionality*, *local independence*, and *monotonicity* (Bartolucci et al. 2016). *Dimensionality* relates to the number of underlying latent traits theorised to explain the dependence between a given set of items in a population of interest (Raykov and Marcoulides 2018; Reckase 2009), alternatively, the number of underlying latent traits required to achieve local independence (Raykov and Marcoulides 2011). While most IRT models assume unidimensional latent structures, this is a strong assumption and needs to be tested (Bartolucci et al. 2016; Reckase 2009).

The *local independence* (or conditional independence) assumption is related to dimensionality and assumes that given the underlying latent trait(s), a respondent's responses to a set of items are conditionally independent (Bartolucci et al. 2016; Raykov and Marcoulides 2018). This means that a respondent's response to an item solely depends on their latent trait(s) level and item parameters, and is independent of responses to other items and responses by other respondents (Reckase 2009).

Monotonicity is concerned with the function describing the conditional probability of a response to an item as a function of the latent trait(s) (Raykov and Marcoulides 2018). For dichotomously scored items, this is assumed to be a monotonic non-decreasing function of the latent trait(s); that is, the conditional probability of selecting a positive response is constant or increases with increasing level of the latent trait(s) (Bartolucci et al. 2016; Reckase 2009). However, for polytomous ordinal items, only the conditional probabilities of selecting the lowest or highest response categories are assumed to be monotonically decreasing or increasing functions, respectively, with increasing latent trait levels (Raykov and Marcoulides 2018). This suggests that probability of selecting the lowest response category decreases with increasing levels of the latent trait(s), while the probability of selecting the highest response category increases with increasing levels of the latent trait(s).

3.3.2 Item Response Theory Models

IRT models can be considered as extensions of standard linear factor models (Bartholomew et al. 2011). Consider a k set of continuous items, x_i ($i = 1, 2, \ldots, k$), whose responses are thought to be dependent on a single continuous latent variable or trait, θ, and a sample of n independent respondents, for the jth respondent, the standard linear factor model can, mathematically, be represented as:

$$x_{ij} = \mu_i + a_i\theta_j + \varepsilon_{ij} \tag{3.1}$$

where θ_j is the latent trait level for respondent j, a_i is the *factor loading* (slope) for x_i on θ, ε_{ij} represents the *unique factor* or *residual* for respondent j to x_i, and μ_i is a constant term or intercept for x_i. The model assumes the latent trait follows a standard normal distribution, that is, a mean of zero and a unit variance ($\theta \sim N(0, 1)$), while the residuals for each item follow a normal distribution with a mean of zero but variances (σ_i^2) may differ ($\varepsilon_i \sim N(0, \sigma_i^2)$). The residuals for each item are orthogonal to each other and the latent trait (Bartholomew et al. 2008; Skrondal and Rabe-Hesketh 2004).

The expectation from both sides of Equation 3.1 can be recast statistically as:

$$E\,(x_{ij} \mid \theta_j) = \mu_i + a_i\theta_j \tag{3.2}$$

where $E\,(x_{ij} \mid \theta_j)$ is respondent j's expected score on x_i given θ_j. From Equation 3.2, the expected scores of the items are a linear function of the latent trait (Raykov and Marcoulides 2018) and the form of the expression $\mu_i + a_i\theta_j$ is called the slope-intercept parameterisation (Reckase 2009).

3.3.3 Item Response Theory Models for Dichotomous Items

Consider a case where the items are categorical, with x_i having u response options or categories. Let x_i be dichotomously scored, then u is either 0 or 1, where zero represents a negative response and one a positive response. In this case, the expected score in Equation 3.2

for the jth respondent is the conditional probability of a positive response to x_i given θ_j (Bartholomew et al. 2008; Raykov and Marcoulides 2011, 2018). A link function mapping the range of the conditional probabilities between [0, 1] onto the unrestricted range of the right-hand side of Equation 3.2 needs to be specified and be a monotonic function of the latent trait, typically a logit (or probit) link function is used (Bartholomew et al. 2008; Skrondal and Rabe-Hesketh 2004). Equation 3.2 can then be expressed as:

$$logit\,\{P\,(x_{ij} \;=\; u \mid \theta_j)\} = \mu_i \;+\; a_i\theta_j = \; a_i\,(\theta_j \;-\; b_i) \tag{3.3}$$

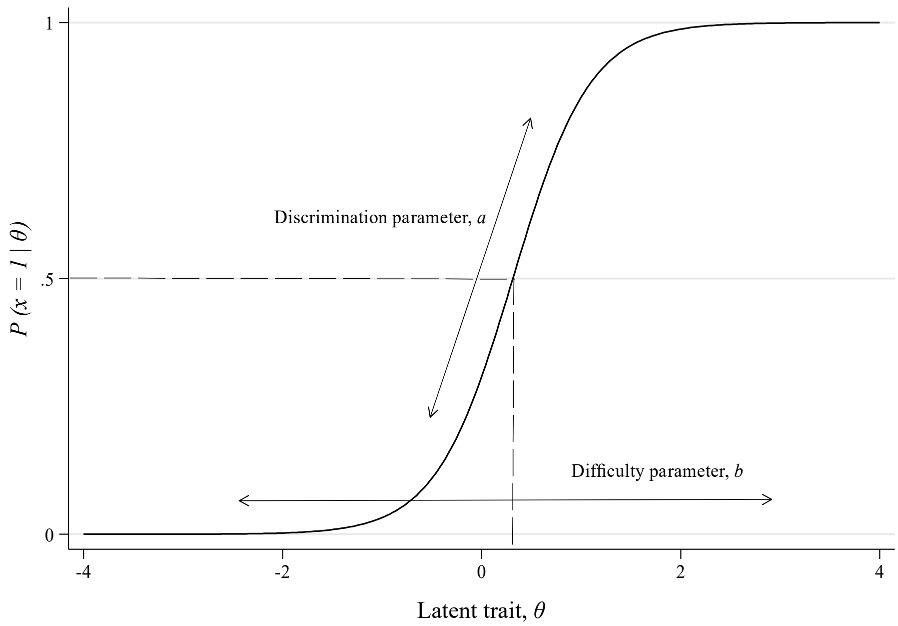
where u is respondent j's response to x_i, $P\ (x_{ij} = u \mid \theta_j)$ is the conditional probability of that response given θ_j, and the slope-intercept parameterisation transformed to a slope-threshold parameterisation using $\mu_i = - a_i b_i$ (Mair 2018) with a_i and b_i being *discrimination* and *difficulty parameters* for x_i, respectively. The *difficulty parameter* (b_i) is the latent trait level (or threshold) required for a respondent to have a 0.5 probability of selecting a positive response to an item, while the *discrimination parameter* (a_i) is the degree to which the item differentiates between the probabilities of a positive response of respondents with latent trait levels below and above the difficulty parameter (Bartolucci et al. 2016; Hambleton et al. 2000; van der Linden and Hambleton [1997] 2010; Skrondal and Rabe-Hesketh 2004). In principle, items with higher discrimination parameters provide more information, hence more precision, in the measurement of the latent trait, particularly around the difficulty parameter, than those with lower discrimination parameters (Raykov and Marcoulides 2018).

The logit of the conditional probability in Equation 3.3 can be transformed back to conditional probabilities (Raykov and Marcoulides 2018) into Equation 3.4:

$$P\ (x_{ij}\ =\ u\ |\ \theta_j, a_i, b_i) = \frac{e^{\ (logit)}}{1 +\ e^{\ (logit)}} = \frac{e^{a_i\ (\theta_j\ -\ b_i)}}{1 +\ e^{a_i\ (\theta_j\ -\ b_i)}} \tag{3.4}$$

where e is the base of the natural logarithms, and other parameters are as previously defined and represent the IRF. The IRF for a dichotomous item is illustrated graphically in Figure 3.1, with the horizontal axis representing the latent trait continuum and the vertical axis depicting the probability of selecting a positive response (x = 1) to the item given the latent trait.

Figure 3.1: Item Response Function for a Dichotomous Item



Notes: The figure illustrates an IRF for a dichotomous item and considers a 2-PL model, i.e., models the difficulty and discrimination parameters for the item. The item has a difficulty parameter of approximately b = 0.2 and the steeper the slope around the difficulty parameter, the higher the discrimination parameter and the more the item distinguishes between respondents with different latent trait levels around θ = 0.2 and vice versa. Items with a difficulty parameter b > 0.2 will require higher latent trait levels to have a 0.5 probability of selecting x = 1 and vice versa.

From the illustration, the difficulty parameter for the item is approximately 0.2; thus, a respondent requires a latent trait level of approximately θ = 0.2 to have a 0.5 probability of selecting a positive response, while the slope of the curve around the difficulty parameter indicates the discrimination parameter. As mentioned, the discrimination parameter is the degree to which an item differentiates between respondents with different latent trait levels around the difficulty parameter (Bartolucci et al. 2016; Hambleton et al. 2000; Skrondal and Rabe-Hesketh 2004), and the steeper the slope around the difficulty parameter, the more the item differentiates between respondents. The probability of selecting a positive response increases with increasing latent trait level; thus, respondents with higher levels of θ have a greater chance of selecting $x = 1$. While Figure 3.1 illustrates the IRF for a single item, in practice this will include IRFs for all the items in the measurement model, thus, placing each of the items and their parameters on the same scale as the underlying latent trait.

The formulation in Equation 3.4 was proposed by Birnbaum ([1968] 2008) and is a *two-parameter logistic* (2-PL) *model* as each item is associated with two parameters; thus, the difficulty and discrimination parameters. However, there are other formulations of IRT models that are beyond the scope of this study. For instance, if the discrimination parameters in the formulation in Equation 3.4 are constrained to be equal across items, the 2-PL model is reduced to a *one-parameter logistic* (1-PL) *model* (van der Linden 2016a; Reckase 2009). On the other hand, a specific form of a 1-PL model where the discrimination parameters across items are constrained to be equal to one is called the *Rasch model* (von Davier 2016; van der Linden and Hambleton [1997] 2010). There is also a *three-parameter logistic* (3-PL) *model* incorporating a lower asymptote parameter representing the probability of respondents with infinitely low latent trait levels selecting a positive response by chance or guessing, and a *four-parameter logistic* (4-PL) *model* modelling an upper asymptote parameter representing the probability that respondents with infinitely high latent trait levels can select a negative response (Desjardins and Bulut 2018; Mair 2018).

3.3.4 Item Response Theory Models for Polytomous Items

IRT models for dichotomously scored items can be generalised to polytomous items (Desjardins and Bulut 2018). There are a range of polytomous IRT models for ordered and unordered responses and the models are distinguished by how the link function describes the relationship between response categories (Bartolucci et al. 2016).[17] For these models, the conditional probabilities relate to selecting a particular response category to an item given the underlying latent trait and item parameters (Raykov and Marcoulides 2018). This study will apply the *graded response model* (GRM).

Graded Response Model

The GRM is a generalisation of the 2-PL model for dichotomous items to polytomous ordinal items proposed by Samejima ([1997] 2010, 2016). The IRF for a GRM is a cumulative logit function modelling the conditional probability of a respondent selecting a particular response category or higher given their latent trait level (Samejima 2016). Consider the 2-PL model in Equation 3.4 and let x_i have u ($u = 0, 1, \ldots, u - 1$) finite ordered response categories that are mutually exclusive and collectively exhaustive. The cumulative conditional probability for the jth respondent selecting a response category u or higher to x_i is:

$$P\left(x_{ij} \geq u \mid \theta_j, a_i, b_{iu}\right) = \frac{e^{a_i\,(\theta_j - b_{iu})}}{1 + e^{a_i\,(\theta_j - b_{iu})}} \tag{3.5}$$

where b_{iu} is the difficulty parameter of selecting the uth response category or higher to x_i and other parameters are as previously defined. In this formulation Equation 3.5, the difficulty parameter is the latent trait level at which the respondent has a 0.5 probability of selecting a particular category or higher (Hambleton et al. 2000; Samejima 2016). As the GRM is a cumulative logit function and dichotomises successive response categories into $P(x \geq u \mid \theta)$ and $P(x < u \mid \theta)$, the conditional probability of the jth respondent selecting response category

[17] These include the graded response, partial credit, rating scale, and nominal response models (Raykov and Marcoulides 2018).

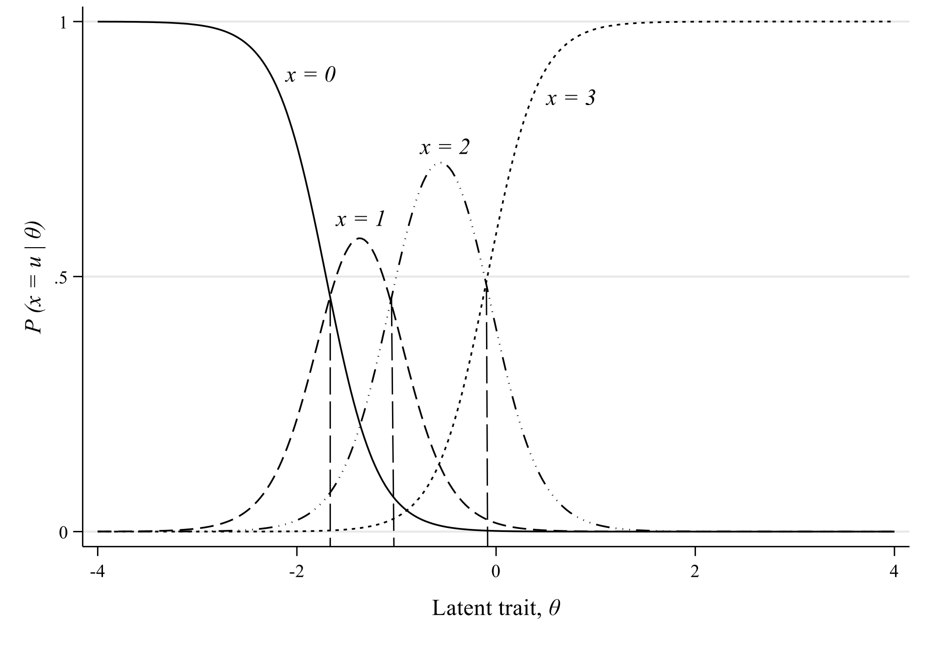
u given θ is the difference between adjacent cumulative logit functions and represents a *category response function* (CRF) (Equation 3.6), that is:

$$P\ (x_{ij}\ =\ u\mid\theta_j, a_i, b_{iu}) = P\ (x_{ij} \geq\ u\mid\theta_j, a_i, b_{iu}) - P\ (x_{ij} \geq\ u+1\mid\ \theta_j, a_i, b_{iu+1}) \qquad (3.6)$$

where $P\ (x_{ij} = u \mid \theta_j, a_i, b_{iu})$ is the conditional probability of selecting response category u, $P\ (x_{ij} \geq u \mid \theta_j, a_i, b_{iu})$ is the conditional probability of selecting response category u or higher, and $P\ (x_{ij} \geq u + 1 \mid \theta_j, a_i, b_{iu+1})$ is the conditional probability of selecting the response category u + 1 or higher (Samejima 2016).

While IRFs for a polytomous item can be depicted based on the difficulty parameter for each response category and discrimination parameter for the item, Figure 3.2 is an illustration of the CRFs for an ordinal item with four categories. The sum of conditional probabilities across all response categories at any point along the latent trait continuum is one (Raykov and Marcoulides 2018), and the intersection points between adjacent categories represent locations or thresholds along the continuum as which respondents select one category or another. Note that the CRFs of the lowest and highest response categories are decreasing and increasing monotonic functions of the latent trait, respectively, indicating a decreasing probability of selecting higher categories for respondents with low latent trait levels, and an increasing probability of selecting higher categories for respondents with high latent trait levels.

Figure 3.2: Category Response Functions for a Polytomous Item



Notes: The figure illustrates CRFs for a polytomous item with four ordered categories. The intersection points between adjacent categories represent locations along the latent trait continuum at which respondents select one response category or another. Thus, respondents with a latent trait level approximately: i) below –1.7 have the highest probability of selecting response category $x = 0$, ii) between –1.7 and –1 are most likely to select $x = 1$, iii) between –1 and –0.1 are most likely to select $x = 2$, and iv) above –0.1 have the highest probability of selecting $x = 3$ compared to any other categories.

3.3.5 Multidimensional Item Response Theory Models

Thus far, the assumption has been that responses to the set of items were dependent on a single latent trait or had a unidimensional latent structure. However, some real-world constructs in social sciences and other scientific disciplines are inherently multifaceted and require models that more accurately reflect the dimensionality of the latent structures (Reckase 2009), as is the case with the concept of QWE (Cazes et al. 2015; Felstead et al. 2019; Gallie 2007b; Green 2006; Kalleberg 2011; Leschke et al. 2008; Muñoz de Bustillo et al. 2011a, 2011b). IRT models for unidimensional latent structures can be extended to multidimensional IRT (MIRT) models. MIRT models can be classified into two major types, thus, compensatory and partially compensatory models. This is based on how the latent traits interact in specifying the probability of responding positively or in a particular category to an item (Bonifay 2020; Desjardins and Bulut 2018; Reckase 2009).

In *compensatory MIRT models*, the probability of a response to an item is based on a linear combination or sum of the latent traits weighted by the item slope (discrimination) parameter (Desjardins and Bulut 2018; Reckase 2009). The linear combination means that, in estimating the probability of responding to an item, latent traits can compensate for each other (Bonifay 2020; Desjardins and Bulut 2018; Reckase 2009). In contrast to compensatory models, in *partially compensatory MIRT models*, the probability of a response to an item is based on the product of the probabilities for each latent trait (Desjardins and Bulut 2018; Reckase 2009). The multiplication of the probabilities results in a non-linear combination of the latent traits and means the probability of a response to an item cannot exceed the highest of these probabilities, thus, reducing the degree to which the latent traits can compensate for each other (Bonifay 2020; Desjardins and Bulut 2018; Reckase 2009).[18] In this study it is assumed that the probability of a response to an item is based on a linear combination of the latent traits where they compensate for each other, hence compensatory MIRT models will be estimated.

Additional to MIRT models being classified in terms of compensatory or partially compensatory models, they can be further differentiated by whether they model between-item or within-item multidimensional structures (Adams, Wilson, and Wang 1997). In *between-item MIRT models*, each item in a set of test items is associated with only one latent trait and this structure is also referred to as a simple structure. On the other hand, *within-item MIRT models* are characterised by at least one item being associated with more than one latent trait and the structures are also referred to as complex structures (Adams et al. 1997; Desjardins and Bulut 2018; Paek and Cole 2020).

[18] In compensatory and partially compensatory MIRT models, different combinations of the latent traits can result in the same probability of a positive response or responding in a particular category to an item (Bonifay 2020).

Consider the GRM in Equation 3.5, but rather than responses to the k set of items, x_i (i = 1, . . ., k) being dependent on a single underlying latent trait, θ, let them be dependent on m underlying latent traits, θ_m (m = 1, 2, . . ., M). The unidimensional model in Equation 3.5 can be extended to a multidimensional model (Bonifay 2020; Reckase 2009). However, the slope-threshold parameterisation of this equation does not generalise to a multidimensional model well (Cai, Yang, and Hansen 2011) as only a single intercept for an item can be estimated (Paek and Cole 2020) and slopes for an item between latent traits will not necessarily be equal. The exponent in the equation can be transformed to a slope-intercept parameterisation; thus, if $d_i = -a_i b_i$, then $a_i (\theta_j - b_i) = a_i \theta_j + d_i$, where d_i is the intercept for x_i (Bonifay 2020; Cai et al. 2011; Reckase 2009). Let the multidimensional model be a compensatory model and have a complex latent structure, the cumulative conditional probability for the jth respondent selecting a response category u or higher to x_i is:

$$P\left(x_{ij} \geq u \mid \boldsymbol{\theta_j}, \boldsymbol{a_i}, d_{iu}\right) = \frac{e^{(a_{i1}\theta_{j1} + a_{i2}\theta_{j2} + \cdots + a_{im}\theta_{jm} + d_{iu})}}{1 + e^{(a_{i1}\theta_{j1} + a_{i2}\theta_{j2} + \cdots + a_{im}\theta_{jm} + d_{iu})}} \qquad (3.7)\ [19]$$

where $\boldsymbol{\theta_j}$ is the vector of respondent j's latent trait levels ($\boldsymbol{\theta_j} = \theta_{j1}, \theta_{j2}, \ldots, \theta_{jm}$), $\boldsymbol{a_i}$ is the vector of slope (discrimination) parameters for item x_i associated with each of the latent traits ($\boldsymbol{a_i} = a_{i1}, a_{i2}, \ldots, a_{im}$), and d_{iu} is the multidimensional intercept related to selecting the uth response category or higher to x_i. Similarly to the unidimensional GRM, the conditional probability of the jth respondent selecting response category u given the vector of latent trait levels is the difference between adjacent cumulative logit functions as in Equation 3.6.

[19] Bold terms represent vectors but are expanded on the right-hand side of the equation to demonstrate how elements of the vectors interact. For a model with a simple latent structure, the respondent j's response to x_i will be influenced by one latent trait, while slopes for other latent traits will be constrained to be equal to zero (Cai et al. 2011).

The multidimensional intercept is not the same as the difficulty parameter in the sense of a unidimensional model, as it is not a unique indicator of the item difficulty (Reckase 2009). This rather indicates the easiness of responding positively or in a particular response category to an item given the latent traits (Cai et al. 2011; Paek and Cole 2020; Reckase 2009). This, however, can be transformed into a *multidimensional difficulty parameter or index* (MDIFF) analogous to the difficulty parameter in a unidimensional model using the formulation adopted from Reckase (2009) in Equation 3.8. Thus, for the model in Equation 3.7, the MDIFF required for the respondent to have the probability > 0.5 of selecting the *u*th response category to item x_i, B_{iu}, is:

$$B_{iu} = \frac{-d_{iu}}{\sqrt{(a_{i1}^2 + a_{i2}^2 + \cdots + a_{im}^2)}} \tag{3.8}$$

The vectors of slope parameters in MIRT models can also be transformed to a *multidimensional discrimination parameter or index* (MDISC) analogous to the discrimination parameter in a unidimensional model using the formulation in Equation 3.9 (Reckase 2009). For the model in Equation 3.7, the MDISC for item x_i, A_i, is:

$$A_i = \sqrt{(a_{i1}^2 + a_{i2}^2 + \cdots + a_{im}^2)} \tag{3.9}$$ [20]

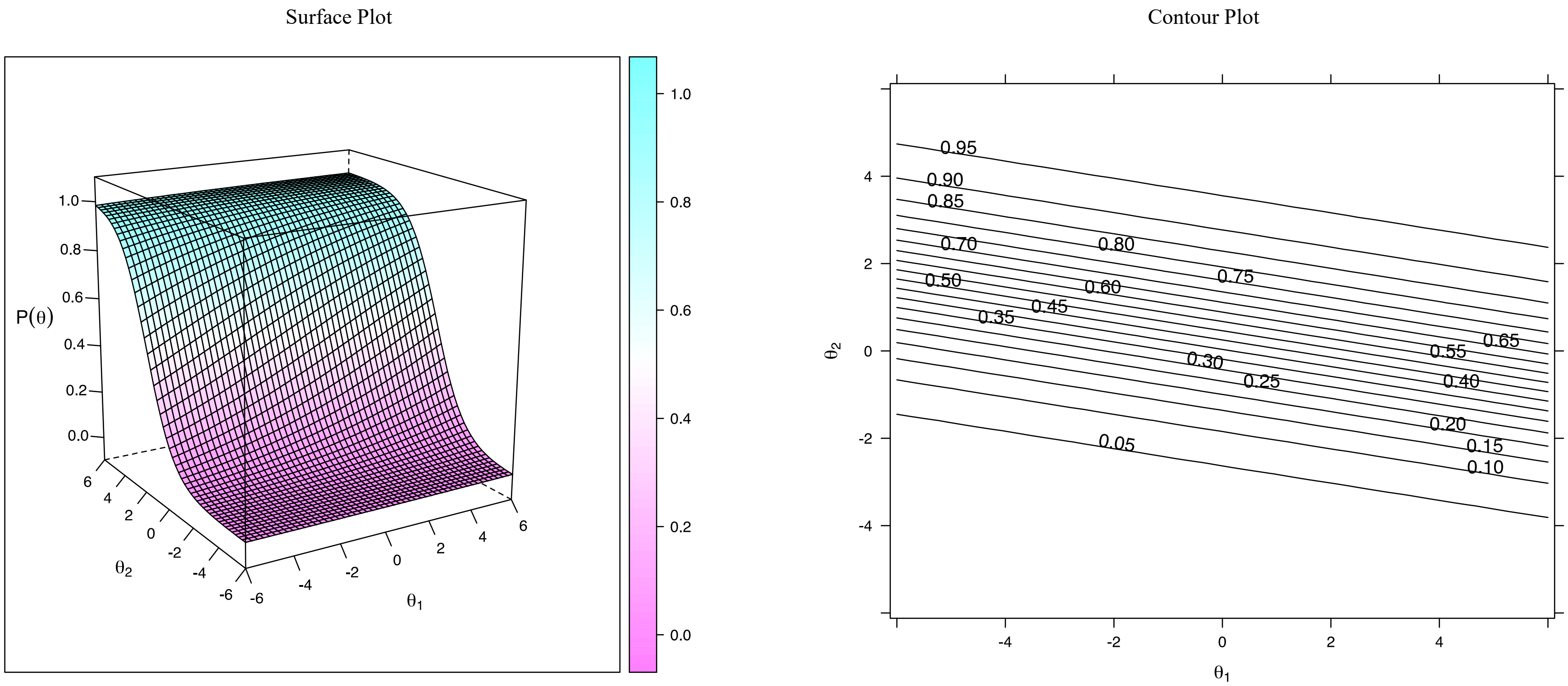
An obvious distinction between unidimensional and multidimensional IRT models is the dimensionality of the underlying latent traits, with the IRF for the former being 2-dimensional (Equations 3.4 – 3.5 and Figures 3.1 – 3.2), while that for the latter models the

[20] The MDISC has the same mathematical form as the denominator of the MDIFF in Equation 3.8, and the MDIFF can, alternatively, be expressed as $B_{iu} = -d_{iu} / A_i$ (Reckase 2009).

probability in an m-dimensional space and is referred to as an *item response surface* (IRS) (Equation 3.7) (Bonifay 2020; Reckase 2009). This has implications for the interpretation of the MDISC and MDIFF, but also results in an important feature of MIRT models. Thus, in the case of a compensatory MIRT model in Equation 3.7, if the probability of a response to an item is set to a constant value, p, then all combinations of the latent traits that satisfy the expression $p = a_{i1}\,\theta_{j1} + a_{i2}\,\theta_{j2} + \ldots + a_{im}\,\theta_{jm} + d_{iu}$ will fall on a straight line (Reckase 2009).

For simplicity, let the probability of a positive response to x_i in Equation 3.7 be dependent on two latent traits, θ_1 and θ_2 and this can be visualised in an IRS and contour plot of the surface (Bonifay 2020; Reckase 2009) displayed in Figure 3.3. The plot on the left illustrates the IRS with the x-axis and y-axis representing the latent traits, θ_1 and θ_2, respectively, and the z-axis representing the probability of a positive response. The plot on the right also displays θ_1 and θ_2 on the x-axis and y-axis, respectively, but the probabilities are depicted in contour lines. The straight equiprobable contours demonstrate the compensatory nature of the model, while the probability of a positive response is a monotonically increasing function of the latent traits (Reckase 2009).

Figure 3.3: Item Response Surface and Contour Plot



Notes: The plots represent the latent traits θ_1 and θ_2 on the *x*-axis and *y*-axis, respectively, with the probability of a positive response shown on the *z*-axis in the surface plot (left) and as contour lines in the contour plot (right). Both plots illustrate a monotonically increasing probability of a positive response as a function of the latent traits; however, it is difficult to discern the probability associated with various combinations of (θ_1, θ_2) from the surface plot. This is clearer from the contour plot, with points along a straight line indicating combinations of (θ_1, θ_2) with an equal probability of selecting the positive response. The straight equiprobable contours demonstrate the compensatory nature of the model, with high levels of θ_2 compensating for low levels of θ_1 and vice versa, resulting in high probabilities of a positive response to the item. Furthermore, the directional impact of the slope associated with the IRS for MIRT models is apparent with the rate of change of the probability of a positive response greater along the direction of θ_2 than θ_1 from the point of origin (0, 0), indicating that θ_2 was more influential in responding to this item than θ_1.

An additional consideration for MIRT models is that the slope of the IRS has a direction within the latent trait space (Bonifay 2020; Reckase 2009). The origin of the (θ_1, θ_2)-plane, (0, 0), in Figure 3.3 represents the average difficulty of the item relative to both latent traits and the interpretation of the MDISC and MDIFF only applies in a specific direction, determined by the direction of the steepest slope from the origin. Thus, the MDISC indicates the extent to which an item differentiates between respondents with low and high levels of the latent traits around the point of the steepest slope in a particular direction from the origin (Bonifay 2020; Reckase 2009). From Figure 3.3, the direction of the steepest slope is along θ_2 than θ_1, as the rate of change of the probability of a positive response is greater along the direction of θ_2 than θ_1 (Reckase 2009). On the other hand, the MDIFF relates to the levels of the latent traits required for a respondent to have a probability > 0.5 of selecting a positive or particular response category along the direction of the steepest slope from the origin (Bonifay 2020; Reckase 2009); thus, along the direction of θ_2 in Figure 3.3.

3.3.6 Item Response Theory Model Diagnostics and Comparison

Model Test Statistics

For IRT models, model test statistics are conducted at the *item*, *person*, and *model* level (Desjardins and Bulut 2018). The signed χ^2 statistics ($S - \chi^2$) will be used to assess *item fit* (Orlando and Thissen 2000, 2003; Toland 2014), and this tests the null hypothesis of no difference between expected and observed response proportions by item response category (Morizot, Ainsworth, and Reise 2007). Ideally, p-values > 0.05 for the $S - \chi^2$ statistics are desired, so that differences between observed and model-predicted response proportions are not statistically significant. However, the test is sensitive to sample size and likely to yield statistically significant results for trivial non-zero differences in population parameters for large samples (Morizot et al. 2007).

Person fit measures how statistically likely a respondent's response pattern to a set of items is, given the estimated model (Morizot et al. 2007), and there are various person fit indices for IRT models. The standardised fit index proposed by Drasgow et al. (1985), the Z_h statistic will be used in this study. The Z_h statistic is a standardised statistic, $Z_h \sim N(0, 1)$, with the expected values of zero suggesting that response patterns are aligned with the item parameters estimated by the model. On the other hand, large negative expected values ($Z_h < -3$) indicate aberrant response patterns, while large positive expected values are also indicative of a higher likelihood of the response pattern than predicted by the model (Desjardins and Bulut 2018; Paek and Cole 2020).

Model fit for IRT models evaluates how well the overall IRT model fits the data based on univariate and bivariate marginal tables, and the M_2 limited information goodness-of-fit statistic will be used (Cai et al. 2006; Maydeu-Olivares and Joe 2005, 2006, 2014). The M_2 statistic tests the null hypothesis of no difference between the expected and observed marginal tables (Maydeu-Olivares and Joe 2005, 2006). Similarly to the $S - \chi^2$ statistic, p-values > 0.05 are desired, so that differences between expected and observed marginal tables are not statistically significant. However, for large samples, the test is likely to yield statistically significant results for trivial non-zero differences in population parameters (Morizot et al. 2007).

In addition to the model test statistics, the local dependence (LD) pairwise residuals between items will be evaluated to determine whether the local independence assumption is tenable given the model (Chen and Thissen 1997, Paek and Cole 2020). This will be based on the signed G^2 statistics (G^2 LD), with values close to zero suggesting that the assumption is tenable (Chen and Thissen 1997); that is, a respondent's response to an item solely depends on their latent trait(s) level and associated item parameters (Reckase 2009). Standardised G^2 LD statistics (signed Cramer's V coefficients) will be estimated to aid interpretation for polytomous items (Chalmers 2012). Signed Cramer's V coefficients range from [–1, 1] (Paek and Cole 2020) and Morizot et al. (2007) suggested values | > 0.20 | may indicate a violation of the assumption and possible poor local fit.

Approximate Fit Indices

Approximate fit indices are not tests of statistical significance, but rather are continuous measures of model-data correspondence (Kline 2016); thus, they measure how well the model fits the data. The approximate fit indices considered in this study are the *root mean square error of approximation* (RMSEA), the *standardised root mean square residual* (SRMSR), the *comparative fit index* (CFI), and the *Tucker-Lewis index* (TLI). Table 3.4 displays the cut-off criteria for the approximate fit indices, with those for the RMSEA and SRMSR based on suggestions by Maydeu-Olivares and Joe (2014) for categorical data used in IRT models. On the other hand, the cut-off criteria for the CFI and TFI were suggested by Hu and Bentler (1999).

Table 3.4: Cut-off Criteria for Approximate Fit Indices

Criterion	***RMSEA**	***SRMSR**	****CFI**	****TLI**
Adequate fit	≤ 0.089	≤ 0.05	≥ 0.95	≥ 0.95
Close fit	≤ 0.05	≤ 0.027		
Excellent fit	≤ 0.05 / (u – 1)	≤ 0.027 / (u – 1)		

Notes: The cut-off criteria for RMSEA and SRMSR were suggested by Maydeu-Olivares and Joe (2014) for categorical data and u is the number of categories. *Maydeu-Olivares and Joe (2014). **Hu and Bentler (1999).

The RMSEA is a standardised measure of the lack of fit of a specified model to the population (Baldwin 2019; Wang and Wang 2020); that is, it measures how well an IRT model reproduces the bivariate tables (Maydeu-Olivares and Joe 2014). It is a scaled as a badness-of-fit statistic, with zero indicating no model misfit, while values greater than zero indicate some degree of misfit and is reported along with its 90% confidence interval (Baldwin 2019; Finch and French 2015; Kline 2016; Maydeu-Olivares and Joe 2014; Wang and Wang 2020). The RMSEA is also a parsimony corrected measure and imposes a penalty for model complexity (Baldwin 2019; Maydeu-Olivares and Joe 2014).

The SRMSR is a residual-based index and measures the average difference between observed and model estimated correlation matrices (Baldwin 2019; Maydeu-Olivares and Joe 2014; Wang and Wang 2020). The SRMSR ranges between [0, 1] and is also scaled as a badness-of-fit measure, with higher values indicating a worse fit (Maydeu-Olivares and Joe 2014; Wang and Wang 2020).

The CFI and TLI both quantify how much better a specified model fits than the baseline or null model; thus, one that assumes no covariance between the observed items (Baldwin 2019; Finch and French 2015; Wang and Wang 2020). While the CFI ranges between [0, 1], the TLI is not guaranteed to vary between this range (Hu and Bentler 1999; Wang and Wang 2020) and the TLI tends to be lower than the CFI, although the estimates are often close (Baldwin 2019; Wang and Wang 2020). However, higher values of the indices indicate a better model fit to the data, while the TLI is also corrected for parsimony (Baldwin 2019; Finch and French 2015; Hu and Bentler 1999; Wang and Wang 2020).

Information Criteria Indices

Information criteria indices are relative model-fit statistics for comparing models (Wang and Wang 2020) and measure the variance that is not explained by the specified model (Finch and French 2015). Commonly used indices include the *Akaike Information Criterion* (AIC) and *Bayesian Information Criterion* (BIC), and their corrected AIC (AICc) and sample-size adjusted BIC (ABIC) indices (Finch and French 2015; Kline 2016; Wang and Wang 2020). The indices are based on the model chi-square (χ^2) and in comparing between models, a model with the lower value is considered to exhibit a better fit to the data (Finch and French 2015; Kline 2016). The indices impose a penalty for model complexity, although to varying degrees and do not require models to be nested like the *chi-square difference* ($\Delta\chi^2$) *test* (Finch and French 2015; Kline 2016; Wang and Wang 2020).

3.4 Introduction to Differential Item Functioning

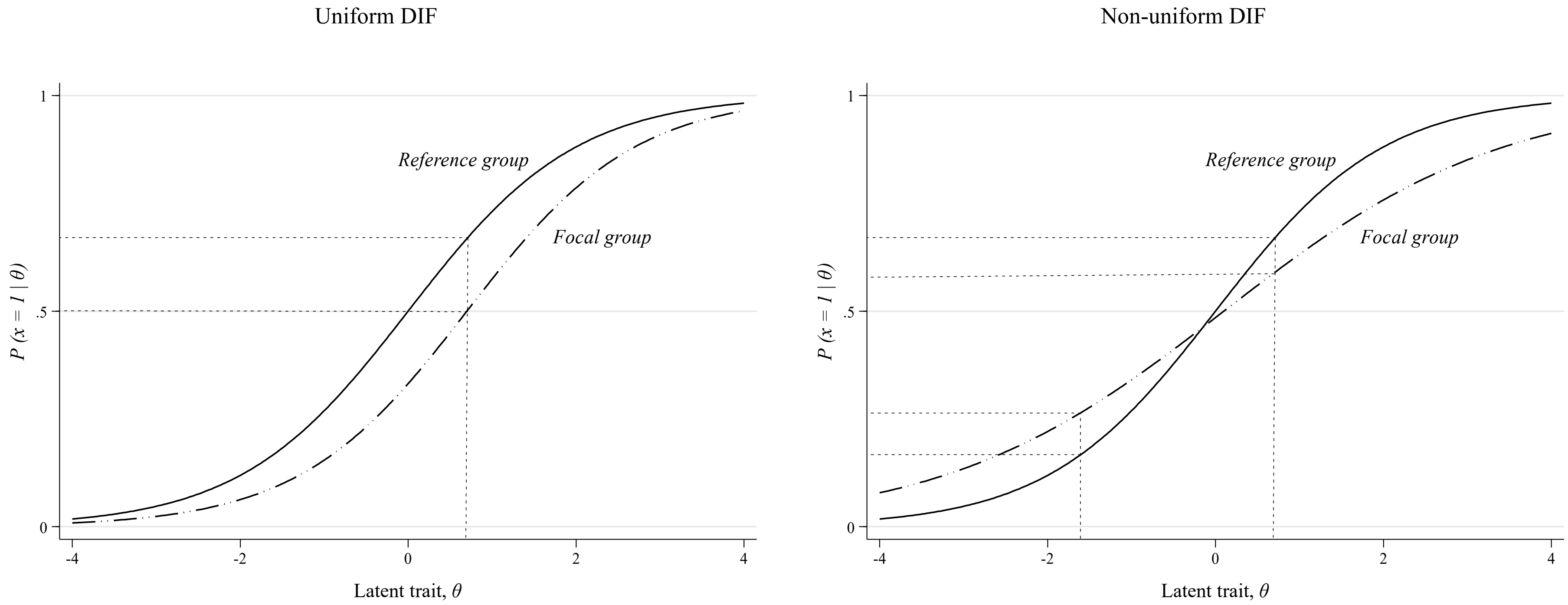
Differential item functioning (DIF) evaluates item-level measurement equivalence to identify differential item performance between respondents from different groups (Angoff 2009; Choi, Gibbons, and Crane 2011; Kim et al. 2007). The group of interest is the *focal group*, while the group to which the item performance is compared to is the *reference group* (Holland and Wainer [1993] 2009). Formally, an item exhibits DIF if respondents from different groups with equal latent trait levels have different conditional probabilities of selecting the same response category, that is, different IRFs (Angoff 2009; Raykov and Marcoulides 2018). Mathematically, an item exhibits DIF if Equation 3.10 holds for at least one value along the latent trait continuum:

$$P_f\ (x = u \mid \theta)\ \neq\ P_r\ (x = u \mid \theta) \tag{3.10}$$

where u is a response category to item x given θ, and the subscripts f and r correspond to focal and reference groups, respectively (Raykov and Marcoulides 2018). Measurement instruments with items that exhibit DIF are likely to have reduced validity for between-group comparisons as this may indicate the instrument is measuring different underlying latent trait(s) between groups (Angoff 2009).

DIF can either be uniform or non-uniform. *Uniform DIF* occurs if an item has a different difficulty parameter between groups given the latent trait, and the effect is a constant higher or lower differential item performance for one group. *Non-uniform DIF* occurs if an item has a different discrimination parameter between groups given the latent trait, resulting in one group having a higher differential item performance for part of the continuum and a lower performance in other parts (Bartolucci et al. 2016; Mair 2018). Figure 3.4 displays IRFs illustrating uniform and non-uniform DIF for a dichotomous item, x and latent trait, θ. The plot on the left illustrates uniform DIF, with the reference group having a higher differential performance compared to the focal group on the item across the range of the latent trait continuum. On the other hand, the plot on the right illustrates non-uniform DIF, with the reference group having a lower differential performance at lower levels of the latent trait and a higher differential performance at higher levels of the latent trait than the focal group.

Figure 3.4: Item Response Functions Illustrating Differential Item Function



Notes: Comparing respondents with the same latent trait level (e.g., $\theta \approx 0.6$), respondents in the reference group have a higher probability of selecting $x = 1$ compared to those in the focal group, and this differential item performance occurs along the whole range of the latent trait continuum.

Notes: Comparing respondents with the same latent trait level, at lower levels of the latent trait (e.g., $\theta \approx -1.7$) respondents in the reference group have a lower probability of selecting $x = 1$ compared to those in the focal group, while at higher levels of the latent trait (e.g., $\theta \approx 0.6$) respondents in the reference group have a higher probability of selecting $x = 1$ than those in the focal group.

There are various methods of detecting DIF, this study applied an iterative hybrid ordinal logistic regression/IRT approach by Choi et al. (2011). Consider a set of ordinal items, x_i (i = 1, . . ., k) with u_q response categories (q = 0, . . ., q-1). For this method, a set of ordinal logistic regression models, specifically proportional-odds models, are specified estimating the cumulative probabilities that a respondent's response to an item falls in a particular category or higher, $P\ (x_i \geq u_q)$, as a function of the latent trait estimated by the IRT model and group membership (Bartolucci et al. 2016; Choi et al. 2011; Mair 2018). The set of nested proportional-odds models are formulated in Equations 3.11 as:

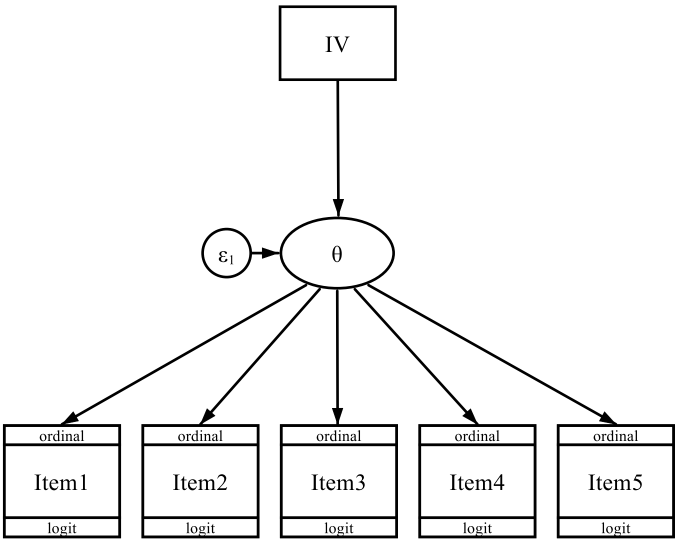
$$
\begin{aligned}
&Model\ 1:\ logit\left\{P\left(x_i \geq u_q\right)\right\} = \mu_i + \beta_1\theta \\
&Model\ 2:\ logit\left\{P\left(x_i \geq u_q\right)\right\} = \mu_i + \beta_1\theta + \beta_2\tau \\
&Model\ 3:\ logit\left\{P\left(x_i \geq u_q\right)\right\} = \mu_i + \beta_1\theta + \beta_2\tau + \beta_3\theta\tau
\end{aligned}
\tag{3.11}
$$

where β represents regression coefficients for all cumulative logits, μ are varying intercepts, θ is the latent trait estimated by the IRT model, and τ is the group membership variable (Choi et al. 2011; Mair 2018). *Model* 1 estimates the cumulative probability of responding in a particular response category as a function of the latent trait only, while *Model* 2 estimates this probability as a function of the latent trait and group membership, and *Model* 3 adds an interaction term between the latent trait and group membership. Comparison between *Model* 2 v *Model* 1 evaluates uniform DIF, while *Model* 3 v *Model* 2 evaluates non-uniform DIF, and *Model* 3 v *Model* 1 evaluates total DIF effect and a statistically significant difference between the models would indicate presence of DIF (Bartolucci et al. 2016; Choi et al. 2011; Mair 2018).

3.5 Multiple Group Analysis

Multiple group analysis is applied where respondents from different populations or groups, thought to have different levels of an underlying latent trait(s), respond to the same set of items (or a subset of the same items) measuring the latent trait(s) (Bock and Zimowski 2010; Finch and French 2015; Paek and Cole 2020; Wang and Wang 2020). The aim of multiple group IRT analysis is to model these group differences by simultaneously estimating the item parameters and distributions of the latent trait(s) for respondents in mutually exclusive groups (Bock and Zimowski 2010). Subject to adequate measurement equivalence, multiple group analysis can be used to compare latent trait means between groups (Finch and French 2015). Consider the path diagram in Figure 3.5 consisting of a measurement model depicting a set of observed ordinal items, Item1 – Item5, measuring a single underlying latent trait, θ. The path diagram also includes a structural model in which an observed independent variable, IV, predicts θ and an error term, ε_i, capturing the variance in θ not explained by the IV.[21]

Figure 3.5: Path Diagram for a Multiple Group Model



Notes: Path diagram of a multiple group model depicting a measurement model with observed ordinal items (Item1 – Item5) whose responses are dependent on a single latent trait (θ), and a structural model in which θ is dependent on an observed independent grouping variable (IV) and ε_i is an error term capturing the variance in θ not explained by the IV.

[21] The observed ordinal items and the latent trait are also referred to as *endogenous variables* as they are influenced by another variable in the model, while the observed predictor variable is an *exogenous variable* as it is not influenced by any other variable in the model (Kline 2016; Wang and Wang 2020).

Let the observed predictor variable, IV, have u mutually exclusive groups ($u = 1, 2, . . ., u$), and to compare levels of θ between the groups, all items and their parameters; for example, the difficulty and discrimination parameters, and the error variances; are specified to be equal between all the groups. However, the means and variances of θ for the groups are allowed to vary, with one group, a reference group, constrained to have a mean of zero and a unit variance, while the means and variances for other groups are freely estimated (Finch and French 2015; Paek and Cole 2020). The estimated population means, and their variances can then be compared between groups based on the IRT modelled mean structures, and as the mean and variance for the reference group are constrained to zero and one, respectively, the estimated means and variances indicate how means between groups differ and variation in θ among respondents in different groups (Bock and Zimowski 2010; Finch and French 2015; Paek and Cole 2020).

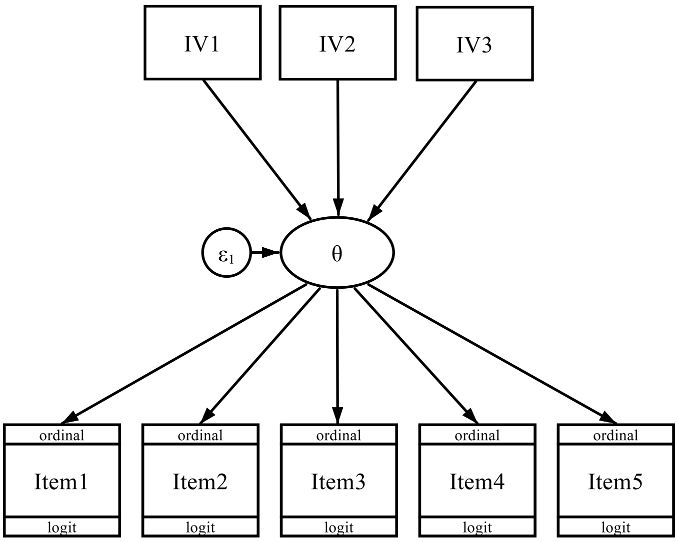
3.6 Multiple Indicators Multiple Causes Models

Multiple indicators multiple causes (MIMIC) models are a special case of structural equation models where there are multiple indicators measuring an underlying latent trait(s), and the latent trait(s) is/are influenced by multiple causes (Finch and French 2015; Wang and Wang 2020). In the context of IRT modelling, the models simultaneously estimate the item parameters of observed categorical indicators measuring a latent trait(s) and the population regression coefficients based on the regression of the latent trait(s) on the observed predictor variables (Bartolucci et al. 2016; Desjardins and Bulut 2018; Paek and Cole 2020).[22] Figure 3.6 illustrates a path diagram of a MIMIC model consisting of a measurement model in which a set of observed ordinal items, Item1 – Item5, measure a single underlying latent trait, θ, and

[22] The models are also referred to as *latent regression* or *explanatory IRT* models (Bartolucci, Bacci, and Gnaldi 2016; Desjardins and Bulut 2018; Paek and Cole 2020).

a structural model with θ regressing on observed predictor variables, IVs, and an error term, ε_i, capturing the variance in θ not explained by the IVs in the model.

Figure 3.6: Path Diagram for a MIMIC Model



Notes: Path diagram of a MIMIC model depicting a measurement model with observed ordinal items (Item1 – Item5) whose responses are dependent on a single latent trait (θ), and a structural model in which θ is dependent on multiple observed independent variables (IVs) and ε_i is an error term capturing the variance in θ not explained by the IVs.

In contrast to multiple group modelling, MIMIC models allow multiple observed predictor variables, which enables controlling for other person or respondent characteristics, and similarly to regression modelling, the predictors can be categorical or continuous (Finch and French 2015; Paek and Cole 2020; Wang and Wang 2020). Since the latent trait(s), $\theta \sim N$ (0, 1), for model identification, the intercept(s) in a MIMIC model is constrained to be zero (Paek and Cole 2020; Wang and Wang 2020).

3.7 Model Parameter Estimation

IRT models can be estimated using frequentist or Bayesian approaches, however, these approaches have some fundamentally different philosophies to statistical modelling and inference (Fox 2010; Levy and Mislevy 2016; van der Linden 2016b). Frequentist approaches, such as maximum likelihood (ML) estimation, consider observed sample data as repeatable random samples with a conditional distribution, and model parameters as unknown but fixed and constant across the repeated random samples (Levy and Mislevy 2016; Wang and Wang 2020). In contrast, the Bayesian approach treats the observed sample data as fixed, while model parameters are unknown and random (uncertain) with a distribution to capture uncertainty in the parameters (Fox 2010; Levy and Mislevy 2016; Wang and Wang 2020). Let $\boldsymbol{x}$ and $\boldsymbol{\Theta}$ be vectors representing collections of observed sample data and unknown model parameters, respectively.

3.7.1 Frequentist Approach

Frequentist approaches construct a model by specifying the conditional distribution of the observed sample data given the model parameters; thus, $p\ (\boldsymbol{x} \mid \boldsymbol{\Theta})$, and in this paradigm, observed sample data are considered to be random (uncertain), while model parameters are fixed (constant) and unknown (Levy and Mislevy 2016; Wang and Wang 2020). The aim of model fitting and parameter estimation is then to find point estimates of model parameters that yield the highest probability of reproducing the observed data (Levy and Mislevy 2016). Good point estimates maximise this probability (or likelihood), and is the conditional probability $p\ (\boldsymbol{x} \mid \boldsymbol{\Theta})$, expressed as a function of model parameters, $\boldsymbol{\Theta}$. Since values of $\boldsymbol{x}$ are observed, when substituted into the conditional probability expression, the likelihood function can be expressed as $L\ (\boldsymbol{\Theta} \mid \boldsymbol{x})$, reflecting that values of the sample data are observed, and the likelihood function varies over different values of $\boldsymbol{\Theta}$.[23] Estimation then relates to finding values of $\boldsymbol{\Theta}$ that maximise $L\ (\boldsymbol{\Theta} \mid \boldsymbol{x})$ (Levy and Mislevy 2016).

In IRT modelling, ML is usually used for model estimation, and ML estimators are full information estimators that use all available information or raw response patterns in model

[23] The likelihood function, $L\ (\boldsymbol{\Theta} \mid \boldsymbol{x})$, is the same as the expression for conditional probability, $p\ (\boldsymbol{x} \mid \boldsymbol{\Theta})$ (Levy and Mislevy 2016).

estimation (Bartholomew et al. 2008; Paek and Cole 2020; Wang and Wang 2020).[24] Estimation of model parameters relate to person parameters (latent traits for individual respondents), and item parameters (Glas 2016). However, estimation of a large number of parameters associated with IRT models can lead to inconsistent parameter estimates (Glas 2016; Haberman 2016). To address this, person parameters are assumed to have common distributions and integrated out of the likelihood function to obtain a marginal likelihood, and the maximisation of this likelihood is referred to as maximum marginal-likelihood (MML) estimation (Glas 2016). The MML estimator is the commonly used estimator in IRT modelling (Edwards et al. 2015; Glas 2016), with the expectation maximisation (EM) algorithm for obtaining parameter estimates for models containing latent variables (Rijmen, Jeon, and Rabe-Hesketh 2016).

ML estimation is carried out under a normality assumption of the data, and where there is severe non-normality, robust estimators that allow for data non-normality can be used. For example, M*plus* (Muthén and Muthén 2017) provides a ML estimator, the MLR, that estimates standard errors robust to non-normality and non-independence of observations associated with data obtained from complex sampling designs (Wang and Wang 2020).

3.7.2 Bayesian Approach

Bayesian analysis aims to estimate a posterior model, which models the probability distribution of unknown model parameters ($\boldsymbol{\Theta}$) given observed sample data ($\boldsymbol{x}$), whilst also incorporating prior knowledge or beliefs about the parameters. This probability distribution is the posterior distribution (Fox 2010; Levy and Mislevy 2016; Wang and Wang 2020). Model construction involves assigning a *prior distribution* to model parameters, $p\ (\boldsymbol{\Theta})$, based on the prior knowledge representing uncertainty associated with the parameters and combining this with (new) evidence from the observed sample data given the model parameters, which is the

[24] This is as opposed to limited information estimation, where covariance or correlation matrices are used as input data for model estimation (Paek and Cole 2020).

likelihood (p ($\boldsymbol{x}$ | $\boldsymbol{\Theta}$)), to yield a *posterior distribution*, p ($\boldsymbol{\Theta}$ | $\boldsymbol{x}$) (Fox 2010; Johnson and Sinharay 2016; Levy and Mislevy 2016; Wang and Wang 2020). The formal process of combining this information is provided by the Bayes' theorem[25] such that the posterior distribution of model parameters, $\boldsymbol{\Theta}$, given the observed sample data, $\boldsymbol{x}$, is:

$$p\ (\boldsymbol{\Theta} \mid \boldsymbol{x}) = \frac{p\ (\boldsymbol{x} \mid \boldsymbol{\Theta}) * p\ (\boldsymbol{\Theta})}{p\ (\boldsymbol{x})} \propto p\ (\boldsymbol{x} \mid \boldsymbol{\Theta}) * p\ (\boldsymbol{\Theta}) \tag{3.12}$$

where $\propto$ denotes proportionality as p ($\boldsymbol{x}$) in Equation 3.12 does not vary with the values of the model parameters, and dropping the term indicates that the posterior distribution is proportional to the product of the likelihood and prior distribution (Fox 2010; Johnson and Sinharay 2016; Levy and Mislevy 2016). Alternatively, the posterior distribution is yielded by weighting prior information about model parameters by new information from the observed sample data (Wang and Wang 2020).

As the model parameters in Bayesian analysis are unknown and random (uncertain), their posterior distributions are estimated via simulations and methods, such as Markov chain Monte Carlo (MCMC), used for Bayesian inference. MCMC methods are iterative algorithms that use a Markov chain process to draw random samples of parameter values from a posterior distribution with numerous iterations and Monte Carlo integration used to estimate posterior point estimates of the parameters, including their standard deviations and credible intervals (Bartholomew et al. 2011; Fox 2010; Junker, Patz, and VanHoudnos 2016; Levy and Mislevy 2016; Wang and Wang 2020). The Markov chain process involves sequential draws of parameter values, with values dependent on the previous one, resulting in a chain

[25] Considering an interaction between two random variables, A and B, with p () indicating the probability mass (for discrete variables) or density (for continuous variables) function, the rule of conditional probability is p (A | B) = p (A, B) / p (B). This can be used to derive Bayes' theorem, which states that: p (B | A) = p (A | B) * p (B) / p (A) (Fox 2010; Johnson and Sinharay 2016; Levy and Mislevy 2016; Wang and Wang 2020).

(Bartholomew et al. 2011; Levy and Mislevy 2016; Wang and Wang 2020). More than one chain can be generated for a parameter by specifying different starting values and setting different seeds for random draws (Wang and Wang 2020). Popular sampling methods for estimating parameters of latent variable models include the Metropolis-Hastings algorithm and the Gibbs sampler as they simplify estimation of complex models (Bartholomew et al. 2011; Johnson and Sinharay 2016; Junker et al. 2016).

A source of controversy in Bayesian analysis is the selection of prior distributions for model parameters, which is regarded as arbitrary and subjective (Fox 2010), yet they are influential in the statistical modelling and inference (Johnson and Sinharay 2016). The specification of the prior distribution is subjective as it is based on the researcher's prior knowledge (Fox 2010; Johnson and Sinharay 2016; Levy and Mislevy 2016). However, it is not arbitrary as it reflects the researcher's thoughts based on prior information which may be from observed data, relevant new information, or expert opinion (Fox 2010). Furthermore, objective priors, also known as diffuse or non-informative priors, which reflect complete ignorance about the model parameters, can be specified (Fox 2010; Johnson and Sinharay 2016). Bayesian estimates obtained with non-informative priors will be close to estimates obtained from frequentist methods as the posterior distribution would strongly resemble the likelihood (Johnson and Sinharay 2016; Levy and Mislevy 2016; Muthén and Asparouhov 2012).

3.7.3 Bayesian versus Frequentist Approaches

For a comparison between Bayesian and frequentist approaches, model estimation with frequentist approaches is entirely based on observed sample data, while Bayesian approaches use prior information and (new) information from observed sample data (Fox 2010; Wang and Wang 2020). Updating prior information with new information from observed sample data may result in improved reliability of the statistical inferences (Fox 2010). Another difference between the approaches is that frequentist analysis approximates point estimates of the unknown model parameters, whereas Bayesian analysis estimates the entire distribution of the model parameters, capturing uncertainty in parameter estimates (Fox 2010; Levy and Mislevy 2016; Wang and Wang 2020). There are also differences in the interpretation of frequentist and Bayesian confidence intervals (CIs). For example, the Bayesian 95% CI (also referred to as the credible interval) is interpreted as a 95% probability that the true (unknown) estimate of a population parameter is between the lower and upper limits of the interval, given (new) evidence from the observed sample data. On the other hand, the frequentist 95% CI is interpreted as that based on hypothesised repeated experiments and the CIs computed for all the experiments, then we can be 95% confident that the true (unknown) estimate of a population parameter would lie between the lower and upper limits of the interval. The frequentist CI is often interpreted, mistakenly, as the Bayesian CI due to the simplicity of the latter. However, as the population parameter in the frequentist approach is unknown but fixed, the estimate would either be inside or outside the interval. Furthermore, frequentist estimators make distributional assumptions about the data, and in contrast, Bayes estimators make no such assumptions and are robust to data non-normality, while incorporating prior information means they have superior performance when working with small samples of observed data (Muthén and Asparouhov 2012; Wang and Wang 2020). Similarly to the frequentist approach's ML estimator, Bayes estimators are full information estimators (Fox 2010; Levy and Mislevy 2016; Wang and Wang 2020). However, the estimation of complex models that are computationally cumbersome with the ML estimator, such as IRT models with a high number of latent traits, are computationally less demanding with Bayes estimators (Muthén and Asparouhov 2012).

3.8 Summary

This chapter has considered the overarching methodology of conducting this research. Sources of secondary survey data with a UK employee population were explored and the primary criterion for data selection was the availability of appropriate survey items for measuring QWE. This was based on the theoretical framework of QWE presented in Chapter 2 which consisted of six dimensions; thus, *economic compensation*, *employment security*, *training and progression*, *working conditions*, *work-life balance* (*or work-time scheduling*) and *social dialogue*. Survey items capturing objective attributes of QWE were preferred, although some aspects of QWE, such as work autonomy, cannot be objectively captured by social survey instruments. The EWCS, the LFS, and the UKHLS were explored in more detail, and their advantages and disadvantages were highlighted. While the EWCS had a broader set of items across different dimensions of QWE, except for the *economic compensation* dimension, particularly compared to the UKHLS, an important limitation was the small UK sample size compared to the LFS or the UKHLS. On the other hand, the LFS had limited items across dimensions of QWE, with indicators for *training and progression*, and *working conditions* dimensions not appropriate for measuring QWE, while the *social dialogue* dimension had one indicator. The UKHLS had appropriate items across different dimensions of QWE, but the *training and progression*, and *employment security* dimensions each had two indicators, while the *social dialogue* dimension also had one indicator. The *working conditions* dimension was also exclusively measured by different aspects of work autonomy. However, as the focus of this research study is on the UK employee population, the UKHLS was selected due to its large sample size compared to the EWCS and broader set of items across different dimensions of QWE compared to the LFS. Furthermore, the UKHLS has the potential to be used to investigate changes in QWE over time by using the panel design of the study.

The chapter highlighted ethical considerations for the research study, with ethical approval sought and obtained from City, University of London. As this study used secondary data, it had low risk and ethical issues related to accessing and using the data according to the terms and conditions set out in the UKDS's End User Licence agreement.

This chapter also built on the introduction to IRT modelling in Chapter 2 as a method for developing a measurement instrument of QWE. It highlighted the classification of IRT relative to other latent variable models and demonstrated how standard linear factor models can be extended to IRT models by using a link function to model the conditional probability of selecting a positive response (for dichotomous observed items) or responding in a particular category (for polytomous observed items). Assumptions of IRT models in terms of dimensionality, that is, the number of latent traits theorised to explain the dependence between observed items, and the local independence assumption, which assumes that given the latent trait(s), the observed items are conditionally independent. The monotonicity assumption was also considered, and this relates to the function describing the conditional probability of a response to an item as a function of the latent trait(s). For dichotomous items, the conditional probability of selecting a positive response is described by a function that is constant or increases with increasing levels of the latent trait(s). This means that the higher the respondent's latent trait level, the more likely they are to select a positive response to an item. On the other hand, for polytomous ordinal items, only the conditional probability of selecting the lowest response category decreases with increasing levels of the latent trait(s), and that of selecting the highest response category increases with increasing levels of the latent trait(s). This means that the higher the respondent's latent trait level, the less likely they are to select the lowest category but the more likely they are to select the highest category. While most of the IRT models assume unidimensional latent structures, the chapter highlighted that these models can be extended to MIRT models to reflect the multidimensionality of some real-world constructs, such as QWE, more accurately. The chapter also presented model diagnostics for IRT models to help evaluate the latent structures of the observed data and test whether the hypothesised latent structures fit the data in developing the measure of QWE.

DIF was also introduced as an extension of IRT modelling and a method to evaluate item-level measurement equivalence of the QWE measurement instrument. Measurement equivalence is a prerequisite for between-group comparisons but is seldom considered in the literature on the measurement of QWE. Multiple group analysis was also introduced as a method to compare levels of QWE between groups of a single observed predictor of QWE, conditional on adequate measurement equivalence, while MIMIC models were also introduced to compare levels of QWE between groups for multiple observed predictors of QWE. Lastly, estimation methods of IRT models were considered, and this was from the perspective of frequentist and Bayesian approaches. Frequentist approaches consider observed sample data as repeatable random samples with a conditional distribution, and model parameters as unknown but fixed and constant across the repeated random samples. In contrast, the Bayesian approach treats the observed sample data as fixed, while model parameters are unknown and random (uncertain) with a distribution to capture uncertainty in the parameters.

The next chapter will introduce the variables that will be used in this research study, including how they were processed. This will focus on the frequency distributions of individual indicators and predictors of QWE, but also how each indicator is associated with each predictor of QWE. The chapter will also consider how the associations between the indicators and predictors of QWE compare with published literature.

3.9 Appendices

3.9.1 Appendix 3.1: Comparison of Survey Data Sets

Dimension	Indicator	Potential Measure for Indicator	Data Sets					
			UKHLS	LFS	APS	EWCS	ISSP	EVS
	Earnings	Pay from main job	✓	✓	✓	✓		
	Adequate pay	Paid adequately for their efforts in their job				✓	✓	✓
	Pension provision	Employer offers pension scheme	✓					
Economic Compensation		Paid sickness absence		✓				✓
	Non-wage pecuniary benefits	Paid annual leave		✓				
		Paid special leave		✓				
	Other benefits	Other benefits provided by employer	✓			✓		
	Job tenure	Years with current employer	✓	✓	✓	✓		✓
Employment Security	Job security	Likelihood to lose job in near future	✓			✓		
	Employment contract	Permanent or temporary employment	✓	✓	✓	✓	✓	
	Predictable working hours	Working hours vary from week to week		✓	✓	✓	✓	
	Progression opportunities	Prospects for progression with employer	✓			✓	✓	✓
	Training participation	Participated in work-related training funded or provided by employer	✓	✓		✓	✓	✓
Training and Progression	Training opportunities	Offered work-related training funded or provided by employer	✓	✓	✓			
	Training duration	Duration of training funded or provided by employer	✓			✓		
	Transferability	Training funded or provided by employer has accredited qualification	✓			✓	✓	

Continued…

Continued…

Dimension	**Indicator**	**Potential Measure of Indicator**	**Data Sets**					
			UKHLS	LFS	APS	EWCS	ISSP	EVS
Working Conditions	Work intensity	Workload in main job				✓		
		Use of own time to meet demands of work				✓		
		Working to tight deadlines				✓		✓
	Autonomy	Degree of control over execution of work responsibilities	✓			✓	✓	✓
	Health and safety	Duties hazardous to health		✓		✓	✓	✓
	Job variety	Job involves monotonous tasks				✓	✓	✓
Work-Life Balance	Weekly working hours	Hours usually worked per week in main job	✓	✓	✓	✓		✓
	Working time	Times of day and days usually works	✓	✓	✓	✓	✓	✓
	Flexible working arrangements	Negotiable working time arrangements with employer	✓	✓		✓	✓	
	Work and non-work time	Balance between work and non-work time				✓		
Social Dialogue	Trade union	Recognised trade union or staff association at workplace	✓	✓		✓	✓	
	Direct participation	Involved in improving work organisation or work processes				✓		✓
	Collective bargaining	Pay and conditions affected by collective bargaining agreements	✓	✓				

Notes: UKHLS: UK Household Longitudinal Study; LFS: Labour Force Survey; APS: Annual Population Survey; EWCS: European Working Conditions Survey; ISSP: International Social Survey Programme; EVS: European Values Survey.

3.9.2 Appendix 3.2: Survey Questions for Indicators of QWE from the EWCS

Dimension	Indicator	Survey Question	Response Options (excluding options for missing responses)
Economic Compensation	Adequate pay	To what extent do you agree or disagree with the following statements about your job? (Q89) A. Considering all my efforts and achievements in my job, I feel I get paid appropriately.	1. Strongly agree 2. Tend to agree 3. Neither agree nor disagree 4. Tend to disagree 5. Strongly disagree
	Pecuniary and non-wage pecuniary rewards	Thinking about your earnings from your main job, what do they include? (Q101) A. Basic fixed salary / wage B. Piece rate or productivity payments C. Extra payments for additional hours of work / overtime D. Extra payments compensating for bad or dangerous working conditions E. Extra payments compensating for Sunday work F. Payments based on your individual performance G. Payments based on the performance of your team / working group / department H. Payments based on the overall performance of the company (profit sharing scheme) where you work I. Income from shares in the company you work for J. Advantages of other nature (for instance medical services, access to shops, etc.)	1. Yes 2. No
	Earnings	Please can you tell us how much are your NET monthly earnings from your main paid job? (Q104)	Net monthly earnings from the main job in national currency: [Amount]
		Perhaps you can provide the approximate range instead. What letter best matches your total net earnings from your main job (SHOW CARD Q105)? Use the part of the show card that you know best: weekly, monthly or annual earnings. (Q105)	1. D 2. B 3. I 4. O 5. T 6. G 7. P 8. A 9. F 10. E 11. Q 12. H
Training and Progression	Training participation	Over the past 12 months, have you undergone any of the following types of training to improve your skills? [IF Q17=00: Since you started your main paid job…] (Q65) A. Training paid for or provided by your employer C. On-the-job training (co-workers, supervisors)	1. Yes 2. No
	Training days	Over the past 12 months, how many days in total did you spend in training paid for or provided by your employer? [IF Q17=00: Since you started your main paid job…] (Q66)	1. 1 day or less 2. 2-3 days 3. 4-5 days 4. 6-9 days 5. 10-19 days 6. 20 days or more
	Progression prospects	Do you agree or disagree with the following statements on the training received over the last 12 months paid for and provided by your employer [IF Q65A=1 AND Q65B≠1] (Q67) exclude self-funded training. A. The training has helped me improve the way I work. B. I feel my prospects for future employment are better	1. Strongly agree 2. Tend to agree 3. Neither agree nor disagree 4. Tend to disagree 5. Strongly disagree
		To what extent do you agree or disagree with the following statements about your job? (Q89) B. My job offers good prospects for career advancement	

Continued…

Continued…

Employment Security	Employment type	What kind of employment contract do you have in your main job? (Q11)	1. Contract of unlimited duration (UK: permanent) 2. Contract of limited duration (UK: fixed-term) 3. A temporary employment agency contract 4. An apprenticeship or other training scheme 5. No contract
	Job security	Do you agree or disagree with the following statements on the training received over the last 12 months paid for and provided by your employer [IF Q65A=1 AND Q65B≠1] (Q67) exclude self-funded training. B. I feel that my job is more secure because of my training To what extent do you agree or disagree with the following statements about your job? (Q89) G. I might lose my job in the next 6 months	1. Strongly agree 2. Tend to agree 3. Neither agree nor disagree 4. Tend to disagree 5. Strongly disagree
	Predictable hours	Do you work...? (Q39) A. The same number of hours every day B. The same number of days every week C. The same number of hours every week	1. Yes 2. No
Working Conditions	Health and safety	Please tell me, using the following scale, are you exposed at work to...? (Q29) (hazards at work) A. Vibrations from hand tools, machinery, etc. B. Noise so loud that you would have to raise your voice to talk to people C. High temperatures which make you perspire even when not working D. Low temperatures whether indoors or outdoors E. Breathing in smoke, fumes (such as welding or exhaust fumes), powder or dust (such as wood dust or mineral dust) etc. F. Breathing in vapours such as solvents and thinners G. Handling or being in skin contact with chemical products or substances H. Tobacco smoke from other people I. Handling or being in direct contact with materials which can be infectious, such as waste, bodily fluids, laboratory materials, etc. Please tell me, using the same scale, does your main paid job involve...? (Q30) (specific tasks) A. Tiring or painful positions B. Lifting or moving people C. Carrying or moving heavy loads D. Sitting E. Repetitive hand or arm movements F. Dealing directly with people who are not employees at your workplace such as customers, passengers, pupils, patients, etc G. Handling angry clients, customers, patients, pupils etc. H. Being in situations that are emotionally disturbing for you I. Working with computers, laptops, smartphones etc.	1. All of the time 2. Almost all of the time 3. Around ¾ of the time 4. Around ½ of the time 5. Around ¼ of the time 6. Almost never 7. Never

Continued…

Continued…

Working Conditions	Health and safety	Does your job ever require that you wear personal protective equipment? (Q31)	1. Yes 2. No
		Regarding the health and safety risks related to the performance of your job, how well informed would you say you are? (Q33)	1. Very well informed 2. Well informed 3. Not very well informed 4. Not at all well informed
		Does the following exist at your company or organisation…? (Q71) B. Health and safety delegate or committee?	1. Yes 2. No
		Over the last month, during the course of your work have you been subjected to any of the following? (Q80) A. Verbal abuse? B. Unwanted sexual attention? C. Threats? D. Humiliating behaviours?	
		And over the past 12 months, during the course of your work have you been subjected to any of the following? [IF Q17=00: And since you started your main paid job…] (Q81) A. Physical violence B. Sexual harassment C. Bullying / harassment	
	Work intensity	Over the last 12 months, how often have you worked in your free time to meet work demands? [IF Q17=00: Since you started your main paid job…] (Q46)	1. Daily 2. Several times a week 3. Several times a month 4. Less often 5. Never
		And does your job involve… (Q49) A. Working at very high speed B. Working to tight deadlines	1. All of the time 2. Almost all of the time 3. Around ¾ of the time 4. Around ½ of the time 5. Around ¼ of the time 6. Almost never 7. Never
		For each of the following statements, please select the response which best describes your work situation (Q61) G. You have enough time to get the job done	1. Always 2. Most of the time 3. Sometimes 4. Rarely 5. Never
	Job variety	Please tell me, does your job involve short repetitive tasks of less than… (Q48) A. 1 minute B. 10 minutes	1. Yes 2. No
		Generally, does your main paid job involve… (Q53) C. Solving unforeseen problems on your own D. Monotonous tasks E. Complex tasks F. Learning new things	

Continued…

Continued…

Working Conditions	Autonomy	Are you able to choose or change… (Q54) A. Your order of tasks B. Your methods of work C. Your speed or rate of work	1. Yes 2. No
Work-life Balance	Working hours	How many hours do you usually work per week in your main paid job? (Q24)	[Number of hours per week]
	Working times	Normally, how many times a month do you work…? (Q37) A. at night, for at least 2 hours between 10.00pm and 05.00am? B. on Sundays? C. on Saturdays? D. More than 10 hours a day?	[2-digit response] (01 – 31) Never (00)
	Flexible working arrangements	How are your working time arrangements set? (Q42)	1. They are set by the company/organisation with no possibility for changes 2. You can choose between several fixed working schedules determined by the company/organisation 3. You can adapt your working hours within certain limits (e.g. flextime) 4. Your working hours are entirely determined by yourself
		Do changes to your working time arrangements occur regularly? (IF YES) How long before are you informed about these changes? (Q43)	1. No 2. Yes, the same day 3. Yes, the day before 4. Yes, several days in advance 5. Yes, several weeks in advance
	Work and non-work time	In general, how do your working hours fit in with your family or social commitments outside work? (Q44)	1. Very well 2. Well 3. Not very well 4. Not at all well
Social Dialogue	Direct participation and support	For each of the following statements, please select the response which best describes your work situation (Q61) B. Your manager helps and supports you. C. You are consulted before objectives are set for your work D. You are involved in improving the work organisation or work processes of your department or organisation N. You can influence decisions that are important for your work	1. Always 2. Most of the time 3. Sometimes 4. Rarely 5. Never
		Does the following exist at your company or organisation…? (Q71) C. A regular meeting in which employees can express their views about what is happening in the organisation?	1. Yes 2. No
	Collective bargaining	Does the following exist at your company or organisation…? (Q71) A. Trade union, works council or a similar committee representing employees?	1. Yes 2. No

Source: European Working Conditions Survey, Sixth Edition (2015).

3.9.3 Appendix 3.3: Survey Questions for Indicators of QWE from the LFS

Dimension	**Indicator**	**Survey Question**	**Response Options** (excludes options for missing responses)
Economic Compensation	Gross pay*	What was your gross pay, that is your pay before any deductions, the last time you were paid? Do not include expenses (if possible)	[Amount = pounds] 99995 = 99995 or more
	Paid holidays‡	How many days of paid holiday are you entitled to per year - please exclude public holidays?	Enter number of days 97 = 97 days or more
Training and Progression	Training opportunities*	May I just check, in the last 3 months, beginning [date], has your (previous or current) employer offered you any training or education either on, or away from, your job?	1. Yes, education or training offered 2. No, not offered
Employment Security	Employment type*	Leaving aside your own personal intentions and circumstances, was your job…	1. A permanent job 2. Or is there some way it is not permanent?
	Predictable hours*	Does the total number of hours you work tend to vary from week to week?	1. Yes 2. No
Working Conditions	Accident†	Thinking of the twelve months since [full date], have you had any accident resulting in injury at work or in the course of your work?	1. Yes 2. No
	Ill health†	(Apart from the accident you have told me about,) within the last twelve months have you suffered from any illness, disability or other physical or mental problem that was caused or made worse by your job or by work you have done in the past?	

Continued…

Continued…

Work-time Scheduling	Working hours*	How many hours per week do you usually work in your (main) job / business – please exclude meal breaks?	Enter number of hours 97 = 97 or more
	Formal flexibility⸸⸸	Some people have special working hours arrangements that vary daily or weekly. In your (main) job is your agreed working arrangement any of the following… Code up to 3: 1. Flexitime (flexible working hours) 2. An annualised hours contract 3. Term-time working 4. Job sharing 5. A nine-day fortnight 6. A four-and-a-half day week 7. Zero hours contract 8. On-call working, or 9. None of these?	1. Yes 2. No
	Working times⸷	Within your regular pattern of work is it usual for you to work... Code all that apply: 1. During the day? 2. During the evening? 3. At night?	1. Usual to work… 2. Not usual to work…
		Do you do shift work in your (main) job…	1. Most of the time 2. Occasionally 3. Or never?
	Working days⸷	May I just check, on which days do you usually work?	1. Monday 2. Tuesday 3. Wednesday 4. Thursday 5. Friday 6. Saturday 7. Sunday
Social Dialogue	Collective bargaining⸸	Are your pay and conditions of employment directly affected by agreements between your employer and any trade union(s) or staff association(s)?	1. Yes 2. No

Source: Labour Force Survey (2016).

Notes: Availability of questions by quarter of data set: * All quarters (Q1 – Q4), ⸶ Q1 only, ⸷ Q2 only, ⸸ Q4 only, and ⸸⸸ Q2 and Q4 only.

3.9.4 Appendix 3.4: UK Data Service End User Licence

7. End User Licence summary text

Below are seventeen points to help you understand the EUL. These pointers are for general guidance and you must read and understand the full EUL before agreeing to it. By accepting the EUL, you agree:

1. To use the data in accordance with the EUL and to notify the UK Data Service of any non-compliance you are aware of.
2. Not to use the data for commercial purposes without obtaining permission and, where relevant, an appropriate licence if commercial use of the data is required.
3. That the EUL does not transfer any interest in intellectual property to you.
4. That the EUL and Data Collection(s) are provided without warranty or liability of any kind.
5. To abide by any further conditions notified to you.
6. To give access to the Data Collections only to Registered Users with a registered use (who have accepted the terms and conditions, including any relevant further conditions). There are some exceptions regarding the use of Data Collection(s) for teaching and the use of Data Collection(s) for Commercial purposes set out in an additional Commercial Licence.
7. To ensure that the means of access to the data (such as passwords) are kept secure and not disclosed to anyone else.

8. To comply with all obligations to preserve the confidentiality of, and not attempt to identify, individuals, households or organisations in the data.
9. To use the correct methods of citation and acknowledgement in publications.
10. To send the UK Data Service bibliographic details of any published work based on our Data Collections.
11. That Personal Data about you may be held for validation and statistical purposes and to manage the service, and that these will only be passed on to the following, in specific circumstances: the Data Collection Depositor or their nominee, your own institute or your research funder.
12. To notify the UK Data Service of any errors discovered in the Data Collections.
13. That Personal Data submitted by you are accurate to the best of your knowledge and kept up to date by you.
14. To meet any charges that may apply.
15. To offer for deposit any new Data Collection(s) which have been derived from the materials supplied, where permission to do so has been granted.
16. To, destroy **all** copies of the data to the standards specified in point 1.16 to the standards specified in the Research Data Handling and Security: Guide for Users.
17. That any non-compliance with the EUL will lead to immediate termination of your access to the services and could result in legal action against you.

Chapter 4 Univariate and Bivariate Analysis

The aim of this chapter is to introduce the indicators and predictors of quality of work and employment (QWE) used in this study and understand their association in the UK employee population. The first section considers the methodology, focusing on the data and sample selection, introduces the variables used in the study, and describes the methods applied to conduct the analyses. The second section presents the results of the univariate analysis of the indicators and predictors of QWE and the bivariate analysis examining associations between the predictors of QWE and each of the indicators of QWE in the UK employee population. The last section will discuss the results, particularly of the bivariate analysis in relation to previous literature and the UK labour market.

4.1 Methodology

4.1.1 Data and Sample

Data from Wave 8 (2016 – 2017) of *Understanding Society: The United Kingdom Household Longitudinal Study* (UKHLS) were used for this study. The UKHLS is a representative study of the UK general population and has a complex sample design involving multistage stratified, cluster sampling with a known unequal probability of selection for respondents (University of Essex, Institute for Social and Economic Research 2018). The sample was limited to employees aged 16 years old or over, who were in a paid job, and participated in full interviews and the base sample was 16,981. This excluded self-employed respondents as well as responses from proxy interviews, as work and employment characteristics of the self-employed differ compared to those of employees and would benefit from a separate analysis, while the subjective nature of some indicators of QWE used in this study renders proxy responses unsuitable.

4.1.2 Variables

Indicators of Quality of Work and Employment

The descriptions of the indicators of QWE are shown in Table 4.1 (refer to Table 3.2 for the survey items) and grouped according to dimensions of QWE. The assigned item numbers are used in subsequent chapters to identify different indicators in path diagrams. The indicators consisted of dichotomous and polytomous items, all with ordinal levels of measurement and were (re)coded such that higher values represented higher levels of QWE.

Table 4.1: Indicators of Quality of Work and Employment

Dimension	Indicator	Item	Description
Economic Compensation	Effective gross pay	Item1	Effective gross hourly pay relative to NMW / NLW.
	Pension provision	Item2	Employer runs a pension or superannuation scheme for which you are eligible.
	Pay bonuses	Item3	Receives any bonus such as quarterly bonus, profit-sharing bonus, or commission.
	Pay progression	Item4	Pay includes annual increments.
Training and Progression	Progression prospects	Item5	Expect better job with current employer in next 12 months.
	Training prospects	Item6	Expect work-related training with current employer in next 12 months.
Employment Security	Employment type	Item7	Current job permanent or temporary.
	Job security	Item8	Perceived likelihood to lose job in the next 12 months.
Working Conditions	Job tasks	Item9	Autonomy over what tasks you do in your job.
	Work pace	Item10	Autonomy over the pace at which you work.
	Work manner	Item11	Autonomy over how you do your work.
	Task order	Item12	Autonomy over the order in which you carry out tasks.
	Work hours	Item13	Autonomy over the time you start or finish your working day.
Work-time Scheduling	Part-time	Item14	Availability of part-time working.
	Term-time	Item15	Availability of working term-time only.
	Job sharing	Item16	Availability of job sharing.
	Flexi-time	Item17	Availability of flexi-time working.
	Compressed hours	Item18	Availability of working compressed hours.
	Annualised hours	Item19	Availability of working annualised hours.
	Home working	Item20	Availability of working from home on a regular basis.
	Other flexibility	Item21	Availability of other flexible working arrangements.
	Informal flexibility	Item22	Informal flexible working arrangements.
	Working times	Item23	Times of day usually worked.
	Weekend working	Item24	Ever work at weekends.
Social Dialogue	Collective bargaining	Item25	Union or staff association recognised for negotiating pay or conditions at workplace.

Source: UKHLS, Wave 8 (2016 – 2017).

Considering the economic compensation dimension, *pension provision*, *pay bonuses*, and *pay progression* were dichotomously scored and recoded with 0 representing '*no*' and 1 '*yes*'. *Effective gross pay* measured respondents' hourly pay relative to the national minimum wage (NMW) or national living wage (NLW). This was computed from respondents' gross pay the last time they were paid, or their usual pay and total hours usually worked a week, including paid and unpaid overtime. Those who worked less than five hours (124 (0.73%) cases) or more than 100 hours (12 (0.07%) cases) a week were excluded**.** Unpaid hours are not usually considered in estimating hourly pay, in part because they are often not captured in social survey data (Bell et al. 2000). However, these are captured in UKHLS and in labour market research, particularly in relation to QWE, accounting for unpaid hours is arguably a more accurate reflection of pay differentials (Bell et al. 2000). The effective gross hourly pay was recoded into a polytomous item with five categories relative to the NMW or NLW at the time of interview and considering the respondent's age, with 0 representing '*below NMW or NLW*', and other categories based on a quartile distribution above the NMW or NLW. The NMW applied to respondents aged less than 25 years old, while those aged 25 years old or more were eligible for the NLW (Department for BEIS 2021).

In terms of the training and progression dimension, *progression prospects* and *training prospects* were (re)coded as dichotomous items such that 0 represented '*no*', and 1 '*yes*'. The *progression prospects* item had a '*doesn't apply*' valid response category[26] and this was recoded as '*no*'. Survey questions for these indicators were subjective in nature and measured respondents' expectations of training and progression prospects.

Regarding the employment security dimension, *employment type* was dichotomously scored and recoded such that 0 represented '*temporary*' employment and 1 '*permanent*' employment. *Job security* was a polytomous item with four response categories, 0 representing

[26] The UKHLS Support Team suggested these were respondents who may have reached the top tier in their roles.

‘*very likely*’, 1 ‘*likely*’, 2 ‘*unlikely*’ and 3 ‘*very unlikely*’. The survey question for the *job security* item was also subjective in nature and measured respondents’ perceived likelihood of losing their job.

The working conditions dimension measured different aspects of work autonomy (*job tasks*, *work pace*, *work manner*, *task order*, and *work hours*) and these were polytomous items with four response categories. They were recoded such that 0 represented ‘*none*’, 1 ‘*a little*’, 2 ‘*some*’ and 3 ‘*a lot*’ in terms of the respondents’ level of work autonomy.

The work-time scheduling dimension measured the awareness of different aspects of formal flexible working arrangements (*part-time*, *term-time*, *job sharing*, *flexi-time*, *compressed hours*, *annualised hours*, *home working*, and *other flexibility*) available at a workplace, as well as *informal flexibility*, *working times*, and *weekend working*. Formal flexible working arrangement items were dichotomously scored with 0 representing ‘*not mentioned*’ and 1 ‘*mentioned*’ as being available. *Informal flexibility* was a polytomous item with three categories and was recoded with 0 representing ‘*no*’, 1 ‘*sometimes*’ and 2 ‘*yes*’. The *working times* item was recoded into a dichotomous item with usual working times that included working during the evening, at night, rotating shifts or varying patterns considered ‘*unsociable times*’ and coded as 0, while 1 represented ‘*sociable times*’. The *weekend working* item measured whether respondents ever worked during the weekend and was a polytomous item with three categories with 0 representing ‘*most or every weekend*’, 1 ‘*some weekends*’, and 2 ‘*no weekend working*’.

Lastly, the *social dialogue* dimension had one indicator, *collective bargaining*, and this considered whether respondents had a trade union or staff association recognised by management for negotiating pay or conditions at their workplace. This was a dichotomous item and recoded such that 0 represented ‘*no*’ and 1 ‘*yes*’.

Predictors of Quality of Work and Employment

The descriptions of predictors of QWE are outlined in Table 4.2 (refer to Appendix 4.1 for the survey items) and are classified into *demographic*, *socio-demographic*, and *socio-economic characteristics*.

Table 4.2: Predictors of Quality of Work and Employment

	Predictor	**Description**
Demographic Characteristics	Sex	Sex of respondent.
	Ethnic group	Ethnic background of respondent.
	Age group	Age group of respondent in years.
Socio-demographic Characteristics	Relationship status	Relationship status of respondent.
	Parental status	Parental status relative to relationship status and primary school age children (5 – 11 years old) in household.
	Illness or disability	Long-standing (over a period of at least 12 months) physical or mental impairment, illness or disability.
	Region	Government office region of the respondent's household postcode.
Socio-economic Characteristics	Education	Highest educational qualification attained ever reported by respondent.
	Occupational classification	Standard Occupational Classification based on the 2000 code frame of respondent's current job.
	Full or part-time	Full-time or part-time employment (full-time > 30 hours per week).
	Organisational sector	Type of organisational sector of respondent's current employer.
	Organisation size	Organisation size based on number of employees at a respondent's workplace.

Source: UKHLS, Wave 8 (2016 – 2017).

In terms of demographic characteristics, s*ex* was dichotomous ('*females*' v '*males*'), while *ethnic group* was recoded into '*White*', '*Mixed*', '*Asian or Asian British*', and '*Black or Black British*'. *Age group* was recoded into five age groups, '*16-24*', '*25-34*', '*35-49*', '*50-64*', and '*65+*' years old.

For socio-demographic characteristics, *relationship status* considered whether employees were 's*ingle*', '*married or cohabiting*', '*divorced or separated*' or '*widowed*', while *parental status* was based on an employee's relationship status and whether they had primary school age children (5 – 11 years old) in the household and had three categories, thus '*lone parents with primary school age children*', '*coupled parents with primary school age children*' and '*employees without primary school age children*'. *Illness or disability* was dichotomous and considered whether employees had a long-standing illness or not ('*yes*' v '*no*'), and the *region* was based on the government office region of an employee's household postcode and was recoded into eight regions, '*London*', '*Southern England*', '*East of England*', '*the Midlands*', '*Northern England*', '*Wales*', '*Scotland*', and '*Northern Ireland*'.

For socio-economic characteristics, *education* measured the highest educational qualification attained ever reported by an employee and was recoded into '*no qualifications*', '*GCSE / O-level or lower*', '*up to A-level*', '*up to a diploma in higher education*', '*university or higher degree*', and due to a high number of missing cases (1,927 (11.35%) cases), a category for employees with no data recorded about their educational qualifications ('*no recorded data*') was included. *Occupational classification* was based on the 2000 code frame of the Standard Occupational Classification (SOC2000) and limited to major groups, thus '*managers and senior officials*', '*professional*', '*associate professional and technical*', '*administrative and secretarial*', '*skilled trades*', '*personal service*', '*sales and customer service*', '*process, plant and machine operatives*' and '*elementary*' occupations. In terms of *full or part-time* employment,[27] employees were classified as being in '*full-time*' employment if they worked more than 30 hours a week (OECD 2022), otherwise, they were classified as being in '*part-time*' employment, while *organisational sector* was also dichotomous ('*private*' v '*public*') based on the type of organisational sector of an employee's employer. *Organisation size* was measured based on the number of employees at a respondent's workplace and recoded

[27] There are no specific cut-off criteria for the number of hours distinguishing between full or part-time workers; however, full-time workers tend to work 35 hours or more per week (Department for BEIS n.d.).

into four categories using the criteria for filing annual financial accounts for all companies, thus '*micro*' (< 10 employees), '*small*' (10 – 49 employees), '*medium*' (50 – 199[28] employees), and '*large*' (≥ 200 employees) size companies (Department for BEIS 2022).

4.1.3 Methods

Analyses were conducted in *Stata* 17 (StataCorp 2021), while the '*ggplot2*' package (Wickham 2016) in *R* (R Core Team 2020) was used for graphical visualisation. Univariate analyses of the indicators and predictors of QWE, limited to estimates of the sample proportions, were conducted, including an examination of the missing data. Contingency tables were estimated for bivariate analyses to examine the association between the predictors of QWE and each of the indicators of QWE in the UK employee population and pairwise deletion was used to handle missing data.

As the UKHLS has a complex sample design, complex survey weighting was applied to the analyses. The statistic used to test the null hypothesis of independence between indicators and predictors of QWE was based on the design-adjusted Rao-Scott *F*-test statistic, which is a design-corrected standard χ^2 test statistic (Heeringa et al. 2017). While the standard χ^2 test compares unweighted observed frequencies in a contingency table to unweighted expected frequencies, the design-corrected test uses weighted frequencies (Agresti and Finlay 2014; Heeringa et al. 2017). Alternatively, the *Wald* χ^2 *test statistic* can be used (Heeringa et al. 2017; Sribney 1998). To examine whether there was an association between indicators and predictors of QWE, the null hypotheses (H_0) of independence between a pair of variables in the UK employee population tested were, thus:

[28] This is defined as companies with no more than 250 employees (Department for BEIS 2022) but was limited to 200 due to response options in the UKHLS.

H_0: An indicator and a predictor of QWE are statistically independent.

Hypothesis (4.1)

H_1: An indicator and a predictor of QWE are not statistically independent.

The analyses were conducted at the 5% significance level, with *p*-values < 0.05 suggesting a statistically significant association between the variables in the UK employee population and rejection of the H_0. On the other hand, *p*-values > 0.05 indicated variables were statistically independent in the UK employee population and failure to reject the H_0. The larger the value of the design-adjusted Rao-Scott *F*-test statistic, the greater the evidence against the H_0 (Agresti and Finlay 2014).

4.2 Results

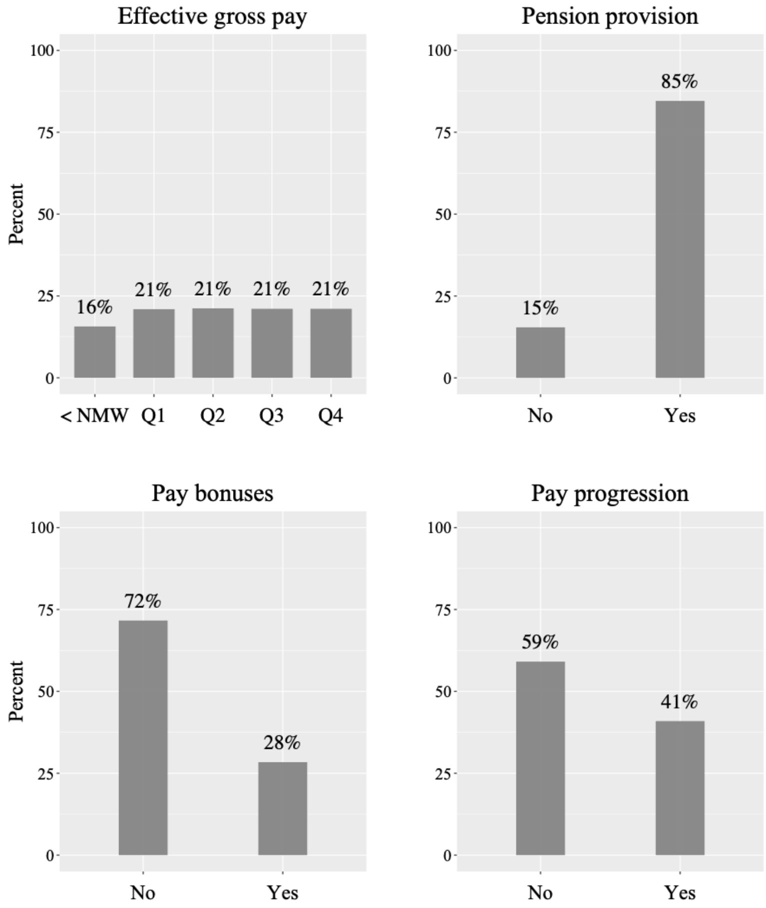
4.2.1 Univariate Analysis

Indicators of Quality of Work and Employment

Economic Compensation

For economic compensation indicators (Figure 4.1), 16% of employees in this sample had an *effective gross pay* below the NMW or NLW, while the categories of pay at or above the NMW or NLW were based on a quartile distribution. *Pension provision* was widely available to employees, with 85% of respondents in the sample eligible for a pension or superannuation scheme run by their employer. On the other hand, most employees were in employment where economic compensation did not include *pay bonuses* (72%) or *pay progression* in the form of annual increments (59%).

Figure 4.1: Distributions of Economic Compensation Indicators

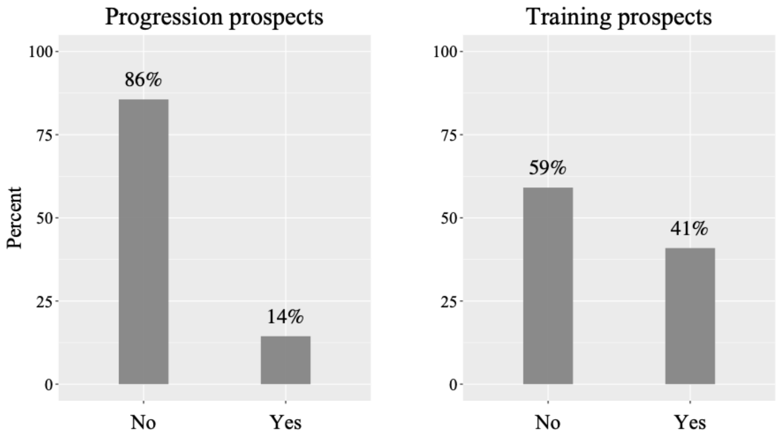


Source: UKHLS, Wave 8 (2016 – 2017). *Notes*: Weighted percentages.

Training and Progression

Employees in this sample were more likely to express not having training or progression prospects with their employers (Figure 4.2). Thus, 86% of employees expressed they did not expect *progression prospects*, while 59% did not expect work-related *training prospects* with their employer in the following 12 months.

Figure 4.2: Distributions of Training and Progression Indicators

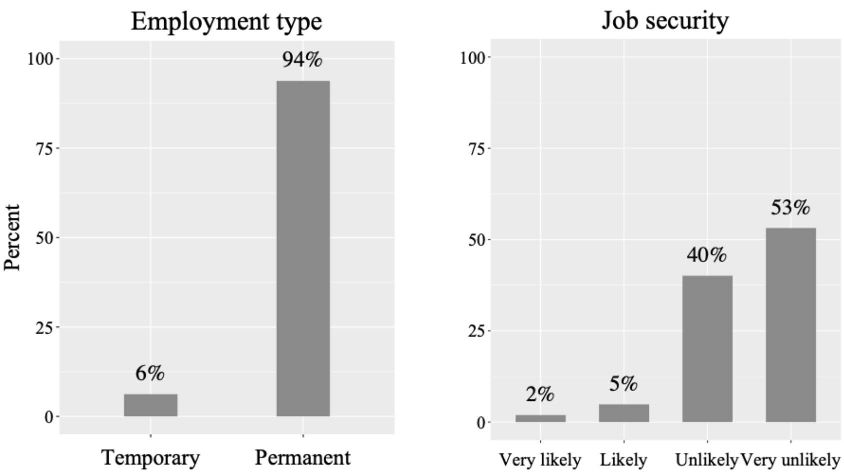


Source: UKHLS, Wave 8 (2016 – 2017). Weighted percentages.

Employment Security

Regarding employment security indicators, employees in this sample were more likely to be or feel secure in their employment (Figure 4.3), with 94% of the employees in *employment type* that was permanent, while in terms of *job security*, 93% perceived they were unlikely or very unlikely to lose their job in the following 12 months.

Figure 4.3: Distributions of Employment Security Indicators

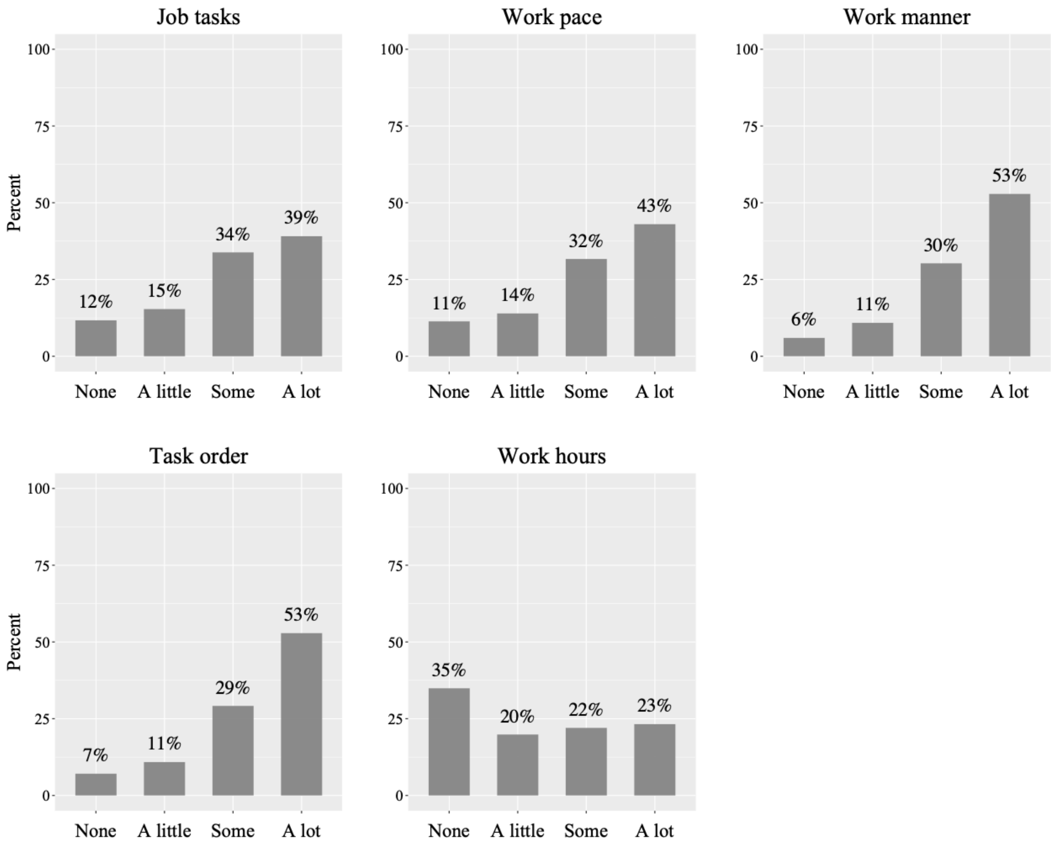


Source: UKHLS, Wave 8 (2016 – 2017). Weighted percentages.

Working Conditions

For working conditions indicators measuring work autonomy (Figure 4.4), either a plurality or majority of employees in the sample reported having a lot of work autonomy in terms of *job tasks* (39%), *work pace* (43%), *work manner* (53%), or *task order* (53%). However, for *work hours*, a plurality of employees (35%) reported having no autonomy over the time they started or finished their working day.

Figure 4.4: Distributions of Working Conditions Indicators

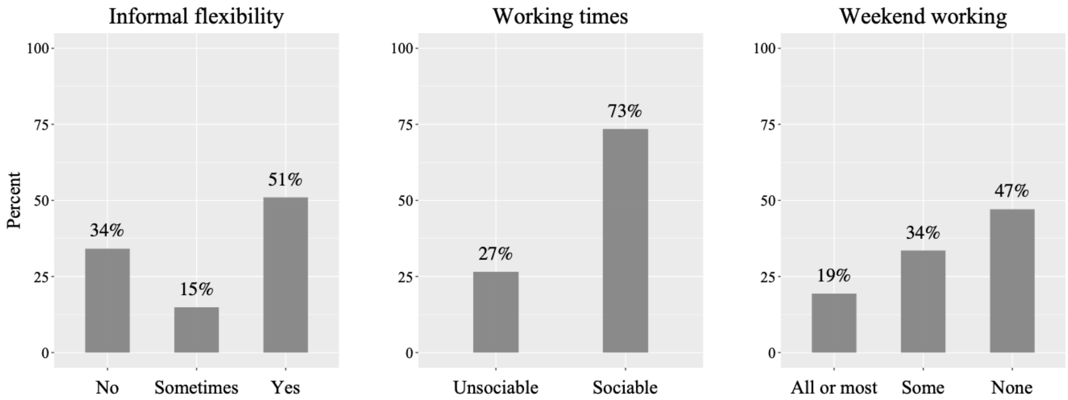


Source: UKHLS, Wave 8 (2016 – 2017). Weighted percentages.

Work-time Scheduling

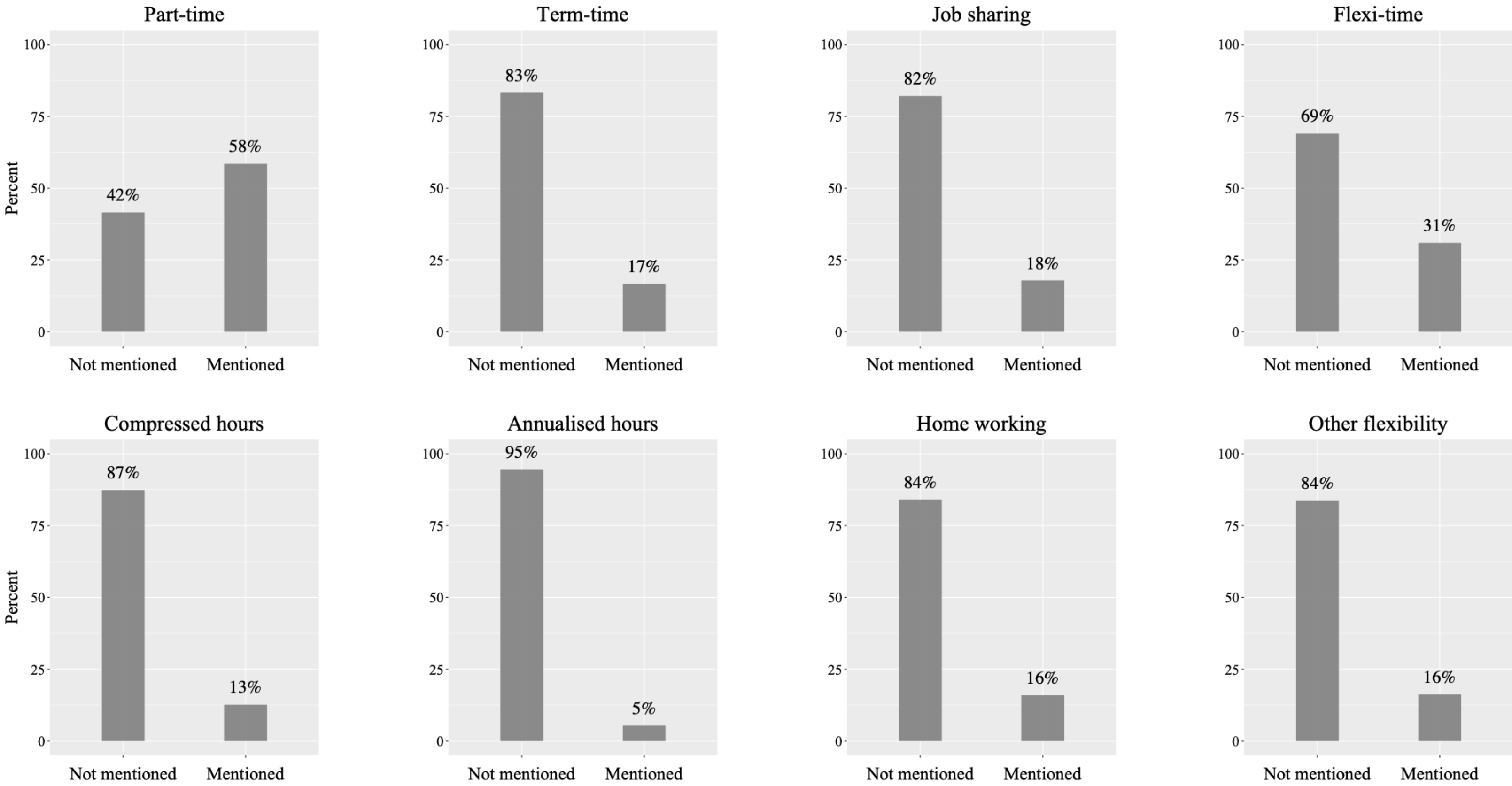
In terms of work-time scheduling indicators (Figure 4.5) and considering awareness of formal flexible working arrangements items, employees in this sample were more likely to mention *term-time* (83%), *job sharing* (82%), *flexi-time* (69%), *compressed hours* (87%), *annualised hours* (95%), *home working* (84%), or *other flexibility* (84%) not being available at their workplace, while a minority mentioned *part-time* (42%) working not being available. For *informal flexibility*, 34% of employees reported not being able to vary their working hours on an informal basis, while for *working times* 27% reported that they usually worked during unsociable times, and in terms of *weekend working* 53% reported that they worked all / most or some weekends.

Figure 4.5: Distributions of Work-time Scheduling Indicators



Continued...

Continued…

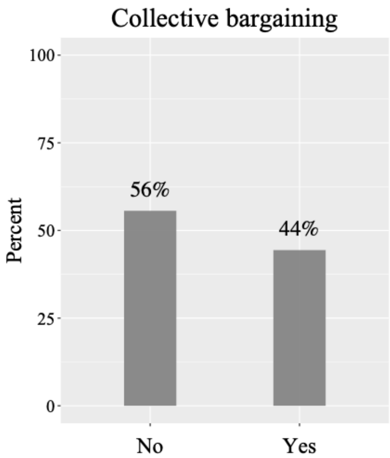


Source: UKHLS, Wave 8 (2016 – 2017). *Notes*: Weighted percentages.

Social Dialogue

For the social dialogue indicator focusing on *collective bargaining*, a plurality of employees in this sample reported that there was no trade union or staff association recognised by management at their workplace for negotiating pay or condition (56%), compared to 44% who did (Figure 4.6).

Figure 4.6: Distribution of Social Dialogue Indicator



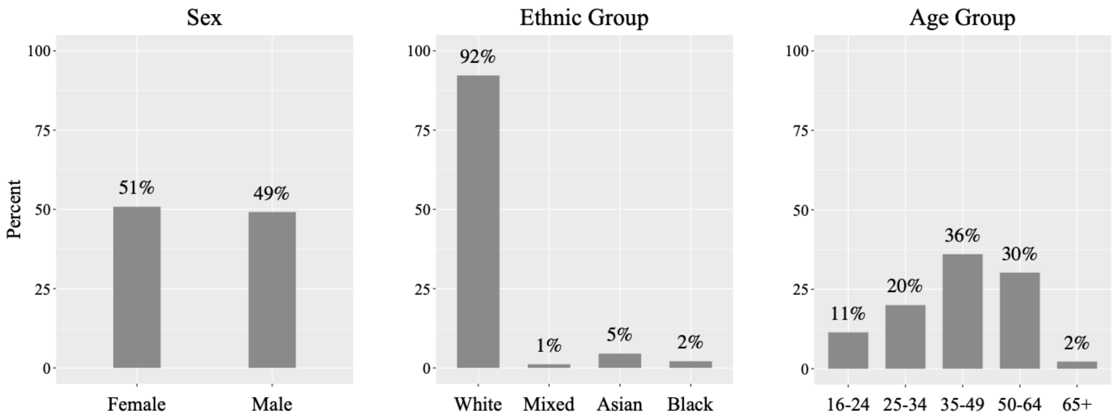
Source: UKHLS, Wave 8 (2016 – 2017). Weighted percentages.

Predictors of Quality of Work and Employment

Demographic Characteristics

In terms of demographic characteristics (Figure 4.7), the sample was approximately evenly distributed by sex (51% '*females*' v 49% '*males*'), while 92% of the employees were from a '*White*' ethnic background, compared to 1% from '*Mixed*', 4% from '*Asian or Asian British*', and 2% from '*Black or Black British*' ethnic backgrounds. The modal age group for the sample was employees aged '*35-49*' years old (36%), with those aged '*65+*' years old (2%) having the least proportion.

Figure 4.7: Distributions of Demographic Characteristics

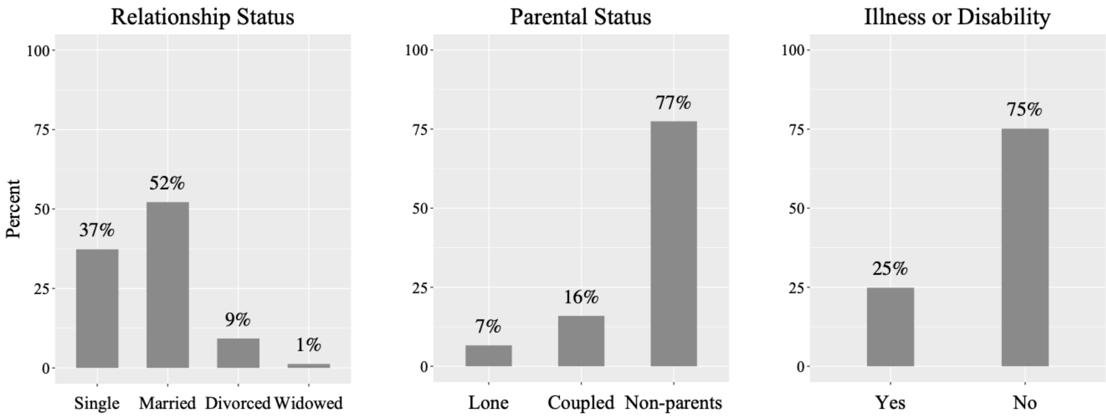


Notes: Weighted percentages and data from UKHLS, Wave 8 (2016 – 2017)

Socio-demographic Characteristics

For socio-demographic characteristics (Figure 4.8), and considering relationship status, 52% of the employees in the sample were '*married or cohabiting*', compared to 37% who were '*single*', 9% who were '*divorced or separated*', and 1% who were '*widowed*', while for parental status in relation to having primary school age children (5 – 11 years old) in the household, a majority of the sample were '*employees without school age children*' (77%), compared to 16% who were '*coupled parents with school age children*', and 7% who were '*lone parents with school age children*'. Regarding long-standing illness or disability, 75% of employees in the sample reported having '*no*' long-standing illness or disability, and in terms of region, a plurality of the employees resided in '*Northern England*' (24%) and those who resided in '*Northern Ireland*' (3%) made up the smallest proportion of the sample (Appendix 4.3).

Figure 4.8: Distributions of selected Socio-demographic Characteristics

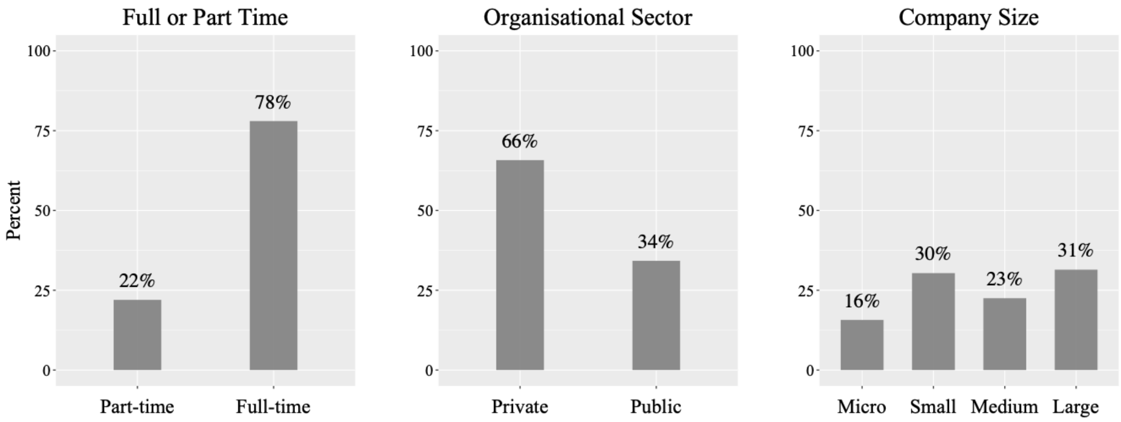


Notes: Weighted percentages and data from UKHLS, Wave 8 (2016 – 2017)

Socio-economic Characteristics

Considering socio-economic characteristics of the sample, a plurality of the employees had attained a '*university or higher degree*' (30%), while 27% had '*GCSE / O-level or lower*' qualification, 13% '*no qualifications*', 11% '*up to A-level*' qualification, 10% '*up to diploma in higher education*', and 8% had '*no recorded data*' (Appendix 4.3). For occupational classification, a slight plurality of employees in this sample were in '*associate professional and technical*' (17%) occupations, followed by '*managers and senior officials*' (16%) and '*professional*' (14%) occupations, while those in '*sales and customer service*' (8%), '*process, plant and machine operatives*' and '*skilled trades*' (both 6%) occupations made up some of the lowest proportions (Appendix 4.3). Employees in this sample were more likely to be in '*full-time*' than '*part-time*' (78% v 22%) employment, and work in a '*private sector*' than '*public sector*' (66% v 34%) organisation. In terms of organisation size, a slight plurality worked for a '*large*' size organisation (31%), compared to 30% who worked for a '*small*', 22% for a '*medium*', and 16% for a '*micro*' size organisation (Figure 4.9).

Figure 4.9: Distributions of selected Socio-economic Characteristics



Notes: Weighted percentages and data from UKHLS, Wave 8 (2016 – 2017)

4.2.2 Bivariate Analysis

For the bivariate analyses, the tests of statistical significance relate to the UK employee population and the *F*-test values refer to the design-adjusted Rao-Scott *F*-test statistic for overall contingency tables for a pair of variables. The degrees of freedom for the *F*-test statistic were estimated and but are not reported.

Table 4.3: Contingency Tables of Indicators of QWE and Demographic Characteristics

Dimension	Indicator	Response category	Sex		Ethnic group				Age group				
			Female*	Male	White*	Mixed	Asian	Black	16-24*	25-34	35-49	50-64	65+
Economic Compensation	Effective gross pay	Below NMW/NLW (%)	19	12**	15	20	17	18	29	17**	12**	13**	31
	Pension provision	No (%)	16	15	15	21	22**	21**	32	14**	13**	12**	40
	Pay bonuses	No (%)	78	65**	71	74	76	80**	72	70**	70	74**	77**
	Pay progression	No (%)	57	61**	59	60	60	65	66	54**	57**	62	76**
Training and Progression	Progression prospects	No (%)	87	85**	87	86	76**	73**	69	79**	87**	94**	99**
	Training prospects	No (%)	58	60	60	48**	47**	45**	50	50	57**	69**	85**
Employment Security	Employment type	Temporary (%)	7	5**	6	16**	6	8	15	7**	4**	5**	13
	Job (in)security	Very unlikely (%)	54	52	54	39**	49	47	55	56	52	52	60
Working Conditions	Control: Job tasks	None (%)	12	11	12	13	9	17**	14	11	10**	13	19
	Control: Work pace	None (%)	12	11	12	12	9	10	13	11	10	12	13
	Control: Work manner	None (%)	6	6	6	8	6	8	8	5	5**	6	11
	Control: Task order	None (%)	6	8	7	7	7	11**	8	6	6	8	13
	Control: Work hours	None (%)	39	31**	35	35	26**	39	43	36**	31**	36**	36

Continued…

Dimension	Indicator	Response category	Demographic Characteristics: Sex		Ethnic group				Age group				
			Female*	Male	White*	Mixed	Asian	Black	16-24*	25-34	35-49	50-64	65+
Work-time Scheduling	Availability: Part-time	Not mentioned (%)	27	56**	41	34	48**	42	38	42	42	42	27**
	Availability: Term-time	Not mentioned (%)	78	89**	83	81	87	84	86	84	82**	83	89
	Availability: Job sharing	Not mentioned (%)	78	86**	82	88**	90**	84	88	84**	80**	81**	88
	Availability: Flexi-time	Not mentioned (%)	69	69	69	67	70	70	73	66**	67**	71	79
	Availability: Compressed hours	Not mentioned (%)	87	88	87	89	93**	91	92	86**	85**	89	96
	Availability: Annualised hours	Not mentioned (%)	95	95	95	96	96	96	95	95	94	94	98
	Availability: Home working	Not mentioned (%)	86	82**	84	87	87	90**	92	83**	81**	85**	91
	Availability: Other flexibility	Not mentioned (%)	84	83	84	84	84	88	86	82**	82**	86	87
	Informal flexibility	No (%)	36	33**	33	38	43**	47**	35	33	32	37	32
	Working times	Unsociable (%)	25	28**	26	38**	31**	37**	39	27**	25**	24**	23**
	Weekend working	Most or all (%)	19	20	19	25	21	22	38	21**	16**	15**	17**
Social Dialogue	Collective bargaining	No (%)	51	61**	55	54	61**	58	75	58**	53**	50**	64**

Notes: Percentages for response categories indicating the poorest QWE, except for the '*job security*' item. *: reference category. **: statistically significant difference compared to the reference category. Absence of (*): no statistically significant difference compared to the reference category. Shading indicates percentage difference compared to reference category.

Shading key:	Statistically insignificant	| < 25 % |	| 25 - 49 % |	| 50 - 99 % |	| 100 - 199 % |	| ≥ 200 % |

Demographic Characteristics

Table 4.3 shows a contingency table of the indicators of QWE and demographic characteristics. This displays the cell percentages of the categories indicating the poorest QWE, except for the 'job security' item. The tests of significance compare differences to a reference group, with the shading of cells depicting the magnitude of the percentage points difference relative to the reference group and darker shades indicating greater differences (see shading key in Table 4.3).

Sex

In relation to economic compensation indicators, females were more likely to earn below the NMW or NLW (19% v 12%) ($F = 73.92$, $p < 0.001$), as well as not to receive pay bonuses (78% v 65%) ($F = 237.03$, $p < 0.001$) compared to males. However, females were slightly less likely to be in employment where pay did not include annual increments than males (57% v 61%) ($F = 15.79$, $p = 0.001$). The association between sex and pension provision was statistically insignificant ($F = 0.27$, $p = 0.605$).

Considering training and progression indicators, females were slightly more likely not to expect a better job with their employer in the following 12 months than males (87% v 85%) ($F = 7.38$, $p = 0.007$). However, the association between sex and training prospects was statistically insignificant ($F = 3.26$, $p = 0.071$).

In terms of employment security indicators, females were slightly more likely to be in temporary employment than males (7% v 5%) ($F = 15.19$, $p = 0.001$). The association between sex and job (in)security was not statistically significant ($F = 1.91$, $p = 0.126$).

For working conditions indicators, males were less likely to report not having any control over their work hours compared to females (31% v 39%) ($F = 26.28$, $p < 0.001$). While there were significant differences in the percentages of employees who reported not having any control over their job tasks, work pace, work manner, or task order by sex, overall associations were statistically significant. Males were slightly more likely to report having a lot of control over their job tasks (42% v 37%) ($F = 9.83$, $p < 0.001$), work pace (45% v 41%) ($F = 6.55$, $p = 0.002$), or work manner (55% v 51%) ($F = 6.86$, $p = 0.001$) compared to females. However, females were slightly more likely to report having a lot of control over their task order than males (53% v 51%) ($F = 2.64$, $p = 0.048$).

Regarding work-time scheduling indicators and awareness of formal flexible working arrangements, males were more likely to mention not being aware of part-time (56% v 27%) ($F = 937.78$, $p < 0.001$), term-time (89% v 78%) ($F = 221.98$, $p < 0.001$), and job sharing (86% v 78%) ($F = 103.36$, $p < 0.001$) being available at their workplace than females. On the other hand, females were slightly more likely to mention not being aware of home working being available than males (86% v 82%) ($F = 35.97$, $p < 0.001$). However, differences in awareness of availability of flexi-time ($F = 0.17$, $p = 0.682$), compressed hours ($F = 1.95$, $p = 0.163$), annualised hours ($F = 0.06$, $p = 0.807$), and other flexibility ($F = 2.73$, $p = 0.099$) by sex were statistically insignificant. For informal flexibility, females were slightly more likely to report not being able to vary their working hours on an informal basis than males (36% v 33%) ($F = 10.55$, $p < 0.001$). However, females were less likely to work unsociable times (25% v 28%) ($F = 15.99$, $p = 0.001$), or not to work at weekends (53% v 41%) ($F = 74.21$, $p < 0.001$) compared to males.

In terms of collective bargaining, males were more likely to report not having a recognised trade union or staff association at their workplace than females (61% v 51%) ($F = 114.10$, $p < 0.001$).

Ethnic Group

Employees from a White ethnic background were less likely to be in employment where they were not eligible for a pension scheme (15%) than employees from Asian or Asian British (22%), or Black or Black British backgrounds (21%) ($F = 7.39$, $p = 0.001$). Employees from a

White ethnic background were also less likely not to receive a pay bonus (71%) compared to Black or Black British employees (80%) ($F = 6.09$, $p = 0.004$). The associations between ethnic group and effective gross pay ($F = 1.55$, $p = 0.106$), or pay progression ($F = 1.47$, $p = 0.224$) were statistically insignificant.

For progression prospects, employees from a White ethnic background were more likely not to expect a better job with their employer in the following 12 months (87%) than Asian or Asian British (76%), or Black or Black British (73%) employees ($F = 27.02$, $p < 0.001$). Employees from White ethnic backgrounds were also the most likely not to expect work-related training with their employer in the following 12 months (60%) compared to employees from any other ethnic background; thus, 48% for Mixed, 47% for Asian or Asian British, and 45% for Black or Black British employees ($F = 22.07$, $p < 0.001$).

Employees from Mixed ethnic backgrounds were more likely to be in temporary employment (16% v 6%) ($F = 6.61$, $p = 0.006$) than those from a White ethnic background. In terms of job (in)security, employees from a Mixed ethnic background were also less likely to feel that their job was secure than those from a White ethnic background, that is, very unlikely to be of the perception that they will lose their job in the following 12 months (39% v 54%) ($F = 2.70$, $p = 0.007$).

Relating to working conditions indicators, Black or Black British employees were more likely to report not having any control over their job tasks (17% v 12%) ($F = 2.27$, $p = 0.018$) or task order (11% v 7%) ($F = 3.44$, $p = 0.005$) compared to employees from a White ethnic background. On the other hand, White employees were more likely to report not having any control over their work hours than Asian or Asian British employees (35% v 26%) ($F = 4.51$, $p < 0.001$). While there were no significant differences in the percentages of employees who reported having no control over their work pace or work manner by ethnic group, overall associations were statistically significant. Employees from White ethnic backgrounds were more likely to report having a lot of control over their work pace (44% v 30%) ($F = 4.24$, $p < 0.001$) or work manner (53% v 42%) ($F = 2.24$, $p = 0.022$) than employees from a Mixed ethnic background.

For work-time scheduling indicators, considering awareness of formal flexible working arrangements, employees from a White ethnic background were less likely to mention not being aware of part-time (41% v 48%) ($F = 4.37$, $p = 0.005$), or compressed hours (87% v 93%) ($F = 9.84$, $p < 0.001$) being available at a workplace than Asian or Asian British employees. Employees of Mixed (88%), or Asian or Asian British (90%) ethnicity were more likely to mention not being aware of job sharing being available compared to White employees (82%) ($F = 16.17$, $p < 0.001$), while White employees were also less likely to mention not being aware of home working being available than Black or Black British employees (84% v 90%) ($F = 4.82$, $p = 0.003$). Associations between ethnic group and awareness of availability of term-time ($F = 2.57$, $p = 0.056$), flexi-time ($F = 0.15$, $p = 0.922$), annualised hours ($F = 2.05$, $p = 0.106$), or other flexible working ($F = 1.35$, $p = 0.257$) were statistically insignificant. For other work-time scheduling indicators, White employees (33%) were less likely to report not being able to informally vary their working hours than Asian or Asian British (43%), or Black British employees (47%) ($F = 6.84$, $p < 0.001$), or to work unsociable hours than employees from any other ethnic group ($F = 10.87$, $p < 0.001$). On the other hand, there were no significant differences in the percentages of employees who worked most or all weekends by ethnic group, however, White employees were more likely not to work at weekends than employees of Mixed ethnicity (48% v 35%) ($F = 2.48$, $p = 0.024$).

For collective bargaining, Asian or Asian British employees were more likely to report not having a recognised trade union or staff association at their workplace compared to White employees (61% v 55%) ($F = 2.63$, $p = 0.049$).

Age Group

Employees aged 16-24 years old (29%) were more likely to earn below the NMW or NLW than those aged 25-34 (17%), 35-49 (12%), or 50-64 (13%) years old (F = 72.77, $p <$ 0.001). Employees aged 16-24 years old (32%) were also more likely to be not eligible for a pension scheme than those aged 25-34 (14%), 35-49 (13%), or 50-64 (12%) years old (F = 81.42, $p < 0.001$). While employees aged 16-24 years old (72%) were slightly more likely not to receive pay bonuses than those aged 25-34 years old (70%), they were less likely not to compared to those aged 50-64 (74%) or 65+ (77%) years old ($F = 4.52$, $p = 0.002$). Employees aged 25-34 (54%) or 35-49 (57%) years old were less likely not to receive annual increments compared to those aged 16-24 years old (66%), while those aged 65+ years old (76%) were more likely not to than those aged 16-24 years old ($F = 23.93$, $p < 0.001$).

Considering training and progression indicators, employees aged 16-24 years old (69%) were less likely not to expect a better job with their employer in the following 12 months than employees in any other age group ($F = 143.27$, $p < 0.001$). Furthermore, employees aged 16-24 years old (50%) were also less likely not to expect work-related training than those aged 35-49 (57%), 50-64 (69%), or 65+ (85%) years old ($F = 76.07$, $p < 0.001$).

Regarding employment type, employees aged 16-24 (15%) years old were more likely to be in temporary employment than those aged 25-34 (7%), 35-49 (4%), or 50-64 (5%) years old ($F = 41.92$, $p < 0.001$). Differences in the percentages of employees aged 16-24 years old compared to those in any other age group who felt they were very unlikely to lose their job were not significant, however, the overall association was statistically significant ($F = 3.00$, $p = 0.005$), with employees aged 65+ years old (60%) more likely to feel secure in their job.

Regarding working conditions indicators, employees aged 35-49 years old were slightly less likely to report not having any control over their job tasks (10% v 14%) ($F = 12.59$, $p <$ 0.001) or work manner (5% v 8%) ($F = 12.73$, $p < 0.001$) compared to those aged 16-24 years old. Those aged 16-24 years old (43%) were also slightly more likely to report not having any control over their work hours than those aged 25-34 (36%), 35-49 (31%), or 50-64 (36%) years old ($F = 23.06$, $p < 0.001$). While differences in the percentages by age group for employees who reported having no control over their work pace or task order were not significant, overall associations were statistically significant. Employees aged 16-24 years old (32%) were less likely to report having a lot of control over their work pace than those aged 25-34 (41%), 35-49 (46%), 50-64 (44%), or 65+ (51%) years old ($F = 8.17$, $p < 0.001$). Those aged 16-24 years old (41%) were also less likely to report having a lot of control over their task order than those aged 25-34 (52%), 35-49 (56%), 50-64 (54%), or 65+ (55%) years old ($F = 10.78$, $p < 0.001$).

Considering awareness of formal flexible working arrangements, employees aged 16-24 years old were more likely to mention not being aware of part-time being available at a workplace than those aged 65+ years old (38% v 27%) ($F = 6.49$, $p < 0.001$), or term-time being available (slightly) than those aged 35-49 years old (86% v 82%) ($F = 5.11$, $p = 0.005$). Those aged 16-24 years old were also slightly more likely to mention not being aware of job sharing ($F = 14.55$, $p < 0.001$), or home working ($F = 24.64$, $p < 0.001$) being available than those aged 25-34, 35-49, or 50-64 years old. Furthermore, compared to those aged 25-34 or 35-49 years old, employees aged 16-24 years old were slightly more likely to mention not being aware of flexi-time ($F = 10.41$, $p < 0.001$), compressed hours ($F = 18.32$, $p < 0.001$) or other formal flexibility ($F = 9.16$, $p < 0.001$) being available. The association between age group and awareness of availability of annualised hours was statistically insignificant ($F = 2.23$, $p = 0.066$). While there were no significant differences in the percentages of employees aged 16-24 years old who reported not being able to informally vary their working hours compared to those in any other age group, those aged 35-49 years old were slightly more likely to be able to informally vary their working hours than those aged 16-24 years old (54% v 48%) ($F = 3.89$, $p = 0.002$). Employees aged 16-24 years old were more likely to work unsociable times (39%)

($F = 26.93$, $p < 0.001$), or at most or all weekends (38%) ($F = 35.61$, $p < 0.001$) than employees in any other age group.

In terms of collective bargaining, employees aged 16-24 years old (75%) were more likely to report not having a recognised trade union or staff association at their workplace than employees in any other age group ($F = 65.28$, $p < 0.001$).

Table 4.4: Contingency Tables of Indicators of QWE and Socio-demographic Characteristics

Dimension	Indicator	Response category	Relationship status				Parental status			Disability		Region				
			Sgl.*	Mar.	Div.	Wid.	Lone*	Couple	Non-parent	Yes*	No	S. Eng.*	N. Eng.	Wal.	Sco.	N. Ire.
Economic Compensation	Effective gross pay	Below NMW / NLW (%)	21	12**	**17****	23	29	12**	15**	18	15**	12	17**	22**	14	20**
	Pension provision	No (%)	20	13**	13**	19	23	14**	14**	14	16	15	15	17	14	27**
	Pay bonuses	No (%)	72	70	77**	79	76	69**	72	73	71	69	71	77**	80**	84**
	Pay progression	No (%)	60	58	61	62	64	58	59	62	58**	60	59	60	57	60
Training and Progression	Progression prospects	No (%)	79	89**	91**	93**	82	85	86**	88	85**	83	86**	88**	89**	91**
	Training prospects	No (%)	54	61**	64**	75**	54	53	61**	62	58**	58	59	59	62	68**
Employment Security	Employment type	Temporary (%)	9	4**	6**	9	7	5	6	7	6	7	6	7	6	7
	Job (in)security	Very unlikely (%)	52	54	50	55	53	54	53	51	54	52	52	51	59**	66**
Working Conditions	Control: Job tasks	None (%)	13	10**	14	14	13	9	12	14	11**	10	12	14	15**	15
	Control: Work pace	None (%)	12	10	14	9	11	11	11	13	11**	10	12	13	12	12
	Control: Work manner	None (%)	7	5**	9	12	6	5	6	7	5**	5	6	6	7	10**
	Control: Task order	None (%)	8	6**	9	12	8	5	7	9	6**	6	7	8	8	13**
	Control: Work hours	None (%)	39	31**	41	41	40	29**	36	38	34**	30	36**	41**	43**	52**

Continued…

Dimension	Indicator	Response category	Socio-demographic Characteristics													
			Relationship status				Parental status			Disability		Region				
			Sgl.*	Mar.	Div.	Wid.	Lone*	Couple	Non-parent	Yes*	No	S. Eng.*	N. Eng.	Wal.	Sco.	N. Ire.
Work-time Scheduling	Availability: Part-time	Not mentioned (%)	41	43	38	36	33	42**	42**	39	43**	42	41	43	40	51**
	Availability: Term-time	Not mentioned (%)	86	82**	82**	82	84	81	84	82	83	83	83	84	86**	86
	Availability: Job sharing	Not mentioned (%)	85	80**	82	78	85	80**	82	81	82	83	82	83	79	83
	Availability: Flexi-time	Not mentioned (%)	70	68	72	69	71	67	69	68	69	66	70**	70	72**	78**
	Availability: Compressed hours	Not mentioned (%)	88	86	89	89	90	84**	88	87	88	86	88	90	86	89
	Availability: Annualised hours	Not mentioned (%)	95	94	94	94	94	95	95	95	94	95	94	94	95	98**
	Availability: Home working	Not mentioned (%)	87	81**	87	88	90	79**	85**	84	84	79	86**	89**	88**	94**
	Availability: Other flexibility	Not mentioned (%)	85	83**	85	86	84	80**	84	84	84	82	84	83	89**	89**
	Informal flexibility	No (%)	35	33	37	37	33	31	35	35	34	30	34**	33	44**	58**
	Working times	Unsociable (%)	31	24**	25**	27	31	24**	27	27	26	24	27**	28	30**	33**
	Weekend working	Most or all (%)	26	15**	18**	17	24	16**	20	21	19	17	20**	24**	19	19
Social Dialogue	Collective bargaining	No (%)	62	52**	51**	50**	61	54**	55**	53	57**	61	54**	50**	46**	52**

Notes: Percentages for response categories indicating the poorest QWE, except for the '*job security*' item. *: reference category. **: statistically significant difference compared to the reference category. Absence of (*): no statistically significant difference compared to the reference category. Sgl: Single. Mar: Married or cohabiting. Div: Divorced or separated. Wid: Widowed. S. Eng: South of England. N. Eng: North of England. Wal: Wales. Sco: Scotland. N. Ire: Northern Ireland. Shading indicates percentage difference compared to reference category.

Shading key:	Statistically insignificant	| < 25 % |	| 25 - 49 % |	| 50 - 99 % |	| 100 - 199 % |	| ≥ 200 % |

Socio-demographic Characteristics

Table 4.4 shows a contingency table of the indicators of QWE and socio-demographic characteristics. This displays the cell percentages of the categories indicating the poorest QWE, except for the 'job security' item. The tests of significance compare differences to a reference group, with the shading of cells depicting the magnitude of the percentage points difference relative to the reference group and darker shades indicating greater differences (see shading key in Table 4.4).

Relationship Status

Single employees were slightly more likely to earn below the NMW or NLW (21%) than married / cohabiting (12%) or divorced / separated (17%) employees (F = 70.57, $p <$ 0.001), as well as not to be eligible for a pension scheme (20%) than married / cohabiting (13%) or divorced / separated (13%) employees (F = 34.53, $p <$ 0.001). On the other hand, divorced / separated employees were slightly more likely not to receive pay bonuses compared to single employees (77% v 72%) (F = 7.32, p = 0.001). While there were no significant differences in the percentages by relationship status for employees who did not receive annual pay increments, overall association was statistically significant. Married / cohabiting employees were slightly more likely to receive pay increments (42%) than employees in any other relationship status (F = 3.25, p = 0.021).

For training and progression indicators, single employees were less likely not to expect a better job (79%) (F = 84.94, $p <$ 0.001), or work-related training (54%) (F = 23.68, $p <$ 0.001) with their employer in the following 12 months than employees in any other relationship status.

In relation to employment type, married / cohabiting (4%) or divorced / separated (6%) employees were slightly less likely to be in temporary employment compared to single employees (9%) (F = 23.93, $p <$ 0.001). The association between relationship status and job (in)security was statistically insignificant (F = 1.69, p = 0.091).

Considering working conditions indicators, single employees were slightly more likely to report not having any control over their job tasks (13% v 10%) ($F = 13.14$, $p < 0.001$), work manner (7% v 5%) ($F = 10.91$, $p < 0.001$), task order (8% v 6%) ($F = 8.37$, $p < 0.001$), or work hours (39% v 31%) ($F = 26.41$, $p < 0.001$) than married / cohabiting employees. The differences in the percentages by relationship status for employees who reported having no control over their work pace were insignificant, however, overall association was statistically significant. Married / cohabiting employees were more likely to report having a lot of control over their work pace than single employees ($F = 6.19$, $p < 0.001$).

Regarding awareness of formal flexible working arrangements, married / cohabiting employees were slightly less likely to mention not being aware of job sharing (80% v 85%) ($F = 12.15$, $p < 0.001$), home working (81%v 87%) ($F = 24.17$, $p < 0.001$), or other formal flexibility (83% v 85%) ($F = 4.72$, $p = 0.003$) being available at their workplace compared to single employees. Single employees (86%) were also slightly more likely to mention not being aware of term-time being available than married / cohabiting (82%) or divorced / separated (82%) employees ($F = 8.03$, $p < 0.001$). Differences in the percentages of employees by relationship status who mentioned not being aware of part-time or compressed hours being available were not significant, however, overall associations were statistically significant. Married / cohabiting were slightly the least likely to mention being aware of part-time working being available (57%) ($F = 3.88$, $p = 0.009$), while slightly the most likely to mention being aware of working compressed hours being available (14%) ($F = 3.04$, $p = 0.030$). The associations between relationship status and awareness of availability of flexi-time ($F = 2.58$, $p = 0.053$), or annualised hours ($F = 0.86$, $p = 0.460$) were statistically insignificant. For the other work-time scheduling indicators, single employees were more likely to work unsociable times (31%) ($F = 19.57$, $p < 0.001$), or most or all weekends (26%) ($F = 28.32$, $p < 0.001$) than married / cohabiting or divorced / separated employees. While differences in percentages of employees by relationship status who reported not having informal flexibility were not insignificant, overall association was statistically significant. Widowed employees were more likely to report having informal flexibility (55%) than employees in any other relationship status ($F = 4.25$, $p = 0.003$).

In relation to collective bargaining, single employees were more likely to report not having a recognised trade union or staff association at their workplace (62%) than employees in any other relationship status ($F = 36.86$, $p < 0.001$).

Parental Status

Lone parents with primary school age children were more likely to earn below the NMW or NLW (29%) ($F = 35.67$, $p < 0.001$), or not to be eligible for a pension scheme (23%) ($F = 12.91$, $p < 0.001$) than coupled parents with or employees without primary school age children. Lone parents with primary school age children were also more likely not received pay bonuses than couple parents (76% v 69%) ($F = 6.13$, $p = 0.002$). While differences in the percentages by parental status for employees who had no pay progression were insignificant, overall association was statistically significant. Lone parents with primary school age children were less likely to receive annual increments (36%) than employees of other parental status ($F = 3.16$, $p = 0.043$).

For training and progression indicators, lone parents with primary school age children were less likely not to expect a better job (slightly) (82% v 86%) ($F = 3.46$, $p = 0.034$) or work-related training (54% v 61%) ($F = 17.29$, $p < 0.001$) with their employer in the following 12 months than employees without primary school age children.

Differences in percentages of employees by parental status who were in temporary employment were insignificant, however, overall association was statistically significant. Lone parents with primary school age children were less likely to be in permanent employment by a small margin (93%) than employees of any other parental status ($F = 3.28$, $p = 0.038$). The association between parental status and job (in)security was statistically insignificant ($F = 0.53$, $p = 0.784$).

In terms of working conditions indicators, coupled parents with primary school age children were less likely to report not having any control over their work hours compared to lone parents with primary school age children (29% v 40%) ($F = 8.06$, $p < 0.001$). While differences in percentages of employees by parental status with no control over their job tasks, work manner or task order were insignificant, overall associations were statistically significant. Coupled parents with primary school age children were more likely to report having a lot of control over job tasks (44%) ($F = 4.04$, $p = 0.005$), work manner (57%) ($F = 2.71$, $p = 0.013$), or task order (56%) ($F = 3.91$, $p = 0.007$) compared to lone parents with or employees without primary school age children. The association between parental status and work pace was statistically insignificant ($F = 1.83$, $p = 0.090$).

Considering awareness of formal flexible working arrangements, lone parents with primary school age children were less likely to mention not being aware of part-time being available (33%) than coupled parents with (42%) or employees without (42%) primary school age children ($F = 9.59$, $p = 0.001$). However, lone parents were slightly more likely to mention not being aware of job sharing (85% v 80%) ($F = 3.56$, $p = 0.029$), compressed hours (90% v 84%) ($F = 10.15$, $p < 0.001$) or other flexibility (84% v 80%) ($F = 8.88$, $p = 0.001$) being available than coupled parents. Furthermore, lone parents were also more likely to mention not being aware of home working being available (90%) than coupled parents with (79%) or employees without (85%) primary school age children ($F = 23.84$, $p < 0.001$). Associations between parental status and awareness of availability of term-time ($F = 2.49$, $p = 0.083$), flexi-time ($F = 2.36$, $p = 0.094$), or annualised hours ($F = 0.16$, $p = 0.851$) were statistically insignificant. For other work-time scheduling indicators, while differences in percentages of employees by parental status who reported not having informal flexibility were insignificant, overall association was statistically significant. Coupled parents with primary school age children were more likely to report having informal flexibility (55%) than employees of any other parental status ($F = 3.91$, $p = 0.004$). Lone parents with primary school age children were also more likely to work unsociable hours (31% v 24%) ($F = 3.99$, $p = 0.019$), or most or all weekends (24% v 16%) ($F = 4.4$, $p = 0.001$) than coupled parents with primary school age children.

For collective bargaining, lone parents with primary school age children were slightly more likely to report not having a recognised trade union or staff association at their workplace (61%) than coupled parents with (54%) or employees without (55%) primary school age children ($F = 4.60$, $p = 0.010$).

Long-standing Illness or Disability

Employees with a long-standing illness or disability were slightly more likely to earn below the NMW or NLW (18% v 15%) ($F = 6.58$, $p < 0.001$), or not to receive annual increments (62% v 58%) ($F = 11.70$, $p = 0.006$) than those without a long-standing illness or disability. The difference in percentages of employees by long-standing illness or disability who had no access to a pension scheme was insignificant, however, overall association was statistically significant. Employees with a long-standing illness or disability were slightly more likely to be in employment where they were eligible for a pension scheme (86%) than those without (84%) ($F = 4.08$, $p = 0.044$). The association between long-standing illness or disability and pay bonuses was statistically insignificant ($F = 1.91$, $p = 0.167$).

For training and progression indicators, employees with a long-standing illness or disability were slightly more likely not to expect a better job (88% v 85%) ($F = 15.03$, $p = 0.001$), or work-related training (62% v 58%) ($F = 12.33$, $p = 0.005$) with their employer in the following 12 months than employees with no long-standing illness or disability.

Differences in percentages of employees by long-standing illness or disability who felt secure in their jobs were insignificant, however, overall association was statistically significant. Employees without a long-standing illness or disability were more likely to feel secure in their job than employees with a long-standing illness or disability by a small margin (54% v 51%) (F = 4.32, p = 0.048). The association between long-standing illness or disability and employment type was statistically insignificant (F = 2.61, p = 0.106).

Regarding working conditions indicators, employees with a long-standing illness or disability were slightly more likely to report not having any control on all the indicators than those without a long-standing illness or disability; thus, job tasks (14% v 11%) (F = 10.40, p < 0.001), work pace (13% v 11%) (F = 5.19, p = 0.014), work manner (7% v 5%) (F = 7.08, p = 0.010), task order (9% v 6%) (F = 6.94, p = 0.010), and work hours (38% v 34%) (F = 4.54, p = 0.035).

Employees with a long-standing illness or disability were slightly less likely to mention not being aware of part-time working being available at their workplace than employees with no long-standing illness or disability (39% v 43) (F = 11.53, p = 0.007). However, associations between long-standing illness or disability and awareness of availability of other formal flexible working arrangements were statistically insignificant; thus term-time (F = 3.67, p = 0.056), job sharing (F = 1.73, p = 0.189), flexi-time (F = 0.84, p = 0.360), compressed hours (F = 0.88, p = 0.348), annualised hours (F = 3.39, p = 0.066), home working (F = 0.17, p = 0.681), and other flexibility (F = 0.002, p = 0.963). Associations between long-standing illness or disability and informal flexibility (F = 0.42, p = 0.658), or working times (F = 0.10, p = 0.758) were also statistically insignificant. While the difference in percentages of employees by long-standing illness or disability who reported working most or all weekends were insignificant, overall association was statistically significant. Employees with a long-standing illness or disability were slightly more likely to work most or all weekends than those without (21% v19%) ($F = 3.04$, $p = 0.048$).

Considering collective bargaining, employees with a long-standing illness or disability were slightly less likely to report not having a recognised trade union or staff association at their workplace than those without (53% v 57%) ($F = 10.23$, $p = 0.001$).

Region

Employees in the south of England were less likely to earn below the NMW or NLW (12%) than those in the north of England (17%), Wales (22%), or Northern Ireland (20%) ($F = 11.28$, $p < 0.001$). Employees in Northern Ireland were also more likely not to be eligible for a pension scheme compared to those in the south of England (27% v 15%) ($F = 12.63$, $p < 0.001$). Furthermore, employees in the south of England were less likely not to receive pay bonuses (69%) than those in Wales (77%), Scotland (80%), or Northern Ireland (84%) ($F = 21.77$, $p < 0.001$). The association between region and pay progression was statistically insignificant ($F = 0.65$, $p = 0.623$).

For training and progression indicators, employees in Northern Ireland were more likely not to expect work-related training compared to those in the south of England (68% v 58%) ($F = 3.65$, $p = 0.007$). On the other hand, employees in the south of England were slightly less likely not to expect a better job with their employer in the following 12 months (83%) than those in any other region ($F = 8.60$, $p < 0.001$).

In terms of job (in)security, employees in the south of England were less likely to feel that their job was secure (52%) than those in Scotland (59%) or Northern Ireland (66%), that is, very unlikely to be of the perception that they will lose their job in the following 12 months ($F = 3.45$, $p = 0.001$). The association between region and employment type was statistically insignificant ($F = 0.82$, $p = 0.512$).

Considering working conditions indicators, employees in the south of England were less likely to report not having any control over their job tasks than those in Scotland (10% v 15%) ($F = 3.19$, $p = 0.002$), as well as work manner (5% v 10%) ($F = 4.20$, $p < 0.001$) or task order (6% v 13%) ($F = 2.87$, $p = 0.008$) than those in Northern Ireland. Furthermore, employees in the south of England were also less likely to report not having any control over their work hours (30%) than those in any other region ($F = 12.99$, $p < 0.001$). While differences in percentages of employees by region who reported having no control over their work pace were insignificant, overall association was statistically significant. Employees in Northern Ireland less likely to report having a lot of control over their work pace (34%) than those in any other region ($F = 2.71$, $p = 0.016$).

Employees in the south of England were less likely to mention not being aware of part-time (42% v 51%) ($F = 3.40$, $p = 0.011$) or annualised hours (slightly) (95% v 98%) ($F = 2.70$, $p = 0.035$) being available at their workplace than those in Northern Ireland, as well as term-time (slightly) being available compared to those in Scotland (83% v 86%) ($F = 2.52$, $p = 0.041$). Employees in the south of England were also less likely to mention not being aware of other formal flexibility being available (82%) than those in Scotland (89%) or Northern Ireland (89%) ($F = 11.36$, $p < 0.001$). Furthermore, employees in the south of England were less likely to mention not being aware of flexi-time being available (66%) than those in the north of England (slightly) (70%), Scotland (72%) or Northern Ireland (78%) ($F = 8.73$, $p < 0.001$), but also home working (79%) compared to employees in any other region ($F = 38.29$, $p < 0.001$). While there were no significant differences in the percentages by region for employees who mentioned not being aware of compressed hours being available, overall associations were statistically significant and those in Wales were the most likely not to mention awareness of this being available (90%) ($F = 3.41$, $p = 0.010$). The association between region and awareness of availability of job sharing was statistically insignificant ($F = 2.19$, $p = 0.075$). For other work-time scheduling indicators, employees in the south of England were less likely to report not being able to informally vary their working hours (30%) ($F = 17.96$, $p < 0.001$) or to work unsociable times (24%) ($F = 6.48$, $p < 0.001$) than those in the north of England, Scotland or Northern Ireland. Employees in the south of England were also less likely to work most or all weekends (17%) compared to those in the north of England (slightly) (20%) or Wales (24%) ($F = 3.27$, $p = 0.001$).

In terms of collective bargaining, employees in the south of England were more likely to report not having a recognised trade union or staff association at their workplace (61%) than those in any other region ($F = 24.37$, $p < 0.001$).

Table 4.5: Contingency Tables of Indicators of QWE and Socio-economic Characteristics

Dimension	Indicator	Response category	Socio-economic Characteristics													
			Full/Part-time		Org. sector		Organisation size				Education			Occ. classification		
			PT*	FT	Pri.*	Pub.	Mic.*	Sma.	Med.	Lar.	None*	Sch.	Uni.	Mngr.*	Inter.	Rout.
Economic Compensation	Effective gross pay	Below NMW / NLW (%)	28	12**	18	10**	27	20**	13**	7**	23	20	9**	6	23**	28**
	Pension provision	No (%)	29	11**	21	5**	44	18**	8**	4**	19	19	10**	10	18**	24**
	Pay bonuses	No (%)	82	69**	61	91**	73	75	73	66**	75	69**	73	70	74**	73
	Pay progression	No (%)	71	56**	68	41**	77	65**	57**	46**	64	63	52**	52	62**	72**
Training and Progression	Progression prospects	No (%)	91	84**	85	86	90	86**	86**	82**	91	85**	83**	85	86	88
	Training prospects	No (%)	69	56**	63	52**	69	59**	57**	55**	66	62**	51**	54	60**	73**
Employment Security	Employment type	Temporary (%)	11	5**	5	9**	7	6	6	6	4	6	8**	6	6	7
	Job (in)security	Very unlikely (%)	52	54	52	55**	55	55	53	51	53	53	53	55	53	48**
Working Conditions	Control: Job tasks	None (%)	18	10**	12	11	11	12	12	12	15	14	8**	6	14**	23**
	Control: Work pace	None (%)	15	10**	11	11	8	12**	12**	11**	13	14	9**	8	13**	19**
	Control: Work manner	None (%)	9	5**	6	5	4	6	7	6	9	8	4**	2	7**	13**
	Control: Task order	None (%)	11	6**	8	5**	5	8**	8**	7	12	9**	4**	2	8**	17**
	Control: Work hours	None (%)	43	33**	35	35	31	39**	39**	30	43	41	27**	22	43**	52**

Continued…

Dimension	Indicator	Response category	Socio-economic Characteristics													
			Full/Part-time		Org. sector		Organisation size				Education			Occ. classification		
			PT*	FT	Pri.*	Pub.	Mic.*	Sma.	Med.	Lar.	None*	Sch.	Uni.	Mngr.*	Inter.	Rout.
Work-time Scheduling	Availability: Part-time	Not mentioned (%)	13	50**	48	29**	48	41**	43	38**	45	43	39**	43	36**	48**
	Availability: Term-time	Not mentioned (%)	80	84**	91	68**	92	84**	79**	82**	88	86	79**	81	84**	89**
	Availability: Job sharing	Not mentioned (%)	83	82	90	68**	91	85**	83**	74**	87	85	77**	77	84**	92**
	Availability: Flexi-time	Not mentioned (%)	76	67**	72	62**	76	75	72	57**	75	72	64**	61	72**	84**
	Availability: Compressed hours	Not mentioned (%)	93	86**	91	80**	94	92	89**	78**	92	90	83**	82	91**	95**
	Availability: Annualised hours	Not mentioned (%)	96	94**	96	92**	97	96	96	91**	96	95	93**	93	96**	97**
	Availability: Home working	Not mentioned (%)	92	82**	85	82**	88	91	87	74**	93	89**	76**	73	93**	98**
	Availability: Other flexibility	Not mentioned (%)	88	83**	84	83	87	86	84	80**	88	85**	81**	79	87**	90**
	Informal flexibility	No (%)	37	33**	32	38**	30	36**	39**	31	41	35**	30**	27	36**	48**
	Working times	Unsociable (%)	29	25**	28	23**	24	27	27	27	30	31	22**	20	27**	41**
	Weekend working	Most or all (%)	23	18**	23	12**	25	23	18**	14**	23	24	14**	13	23**	31**
Social Dialogue	Collective bargaining	No (%)	60	54**	73	21**	84	66**	51**	35**	57	61	50**	52	57**	63**

Notes: Percentages for response categories indicating the poorest QWE, except for the '*job security*' item. *: reference category. **: statistically significant difference compared to the reference category. Absence of (*): no statistically significant difference compared to the reference category. PT: Part-time. FT: Full-time. Pri: Private. Pub: Public. Mic: Micro. Sma: Small. Med: Medium. Lar: Large. Sch: School qualifications. Uni: University. Mngr: Managerial and professional. Inter: Intermediate. Rout: Routine and manual. Shading indicates percentage difference compared to reference category.

Shading key:	Statistically insignificant	| < 25 % |	| 25 - 49 % |	| 50 - 99 % |	| 100 - 199 % |	| ≥ 200 % |

Socio-economic Characteristics

Table 4.5 shows a contingency table of the indicators of QWE and socio-economic characteristics. This displays the cell percentages of the categories indicating the poorest QWE, except for the 'job security' item. The tests of significance compare differences to a reference group, with the shading of cells depicting the magnitude of the percentage points difference relative to the reference group and darker shades indicating greater differences (see shading key in Table 4.5).

Education

In relation to economic compensation indicators, employees with no educational qualifications were more likely to earn below the NMW or NLW (23% v 9%) ($F = 149.94$, $p < 0.001$), not to be eligible for a pension scheme (19% v 10%) ($F = 66.26$, $p < 0.001$), or not to receive annual increments (64% v 52%) ($F = 59.50$, $p < 0.001$) than employees with university qualifications. Furthermore, employees with no educational qualifications were more likely not to receive pay bonuses than those with school qualifications (75% v 69%) ($F = 15.60$, $p < 0.001$).

Employees with no educational qualifications were more likely not to expect a better job (91%) ($F = 23.54$, $p < 0.001$), or work-related training (66%) ($F = 62.85$, $p < 0.001$) with their employer in the following 12 months than those with school or university qualifications.

Considering employment type, employees with university qualifications were slightly more likely to be in temporary employment than those with no educational qualifications (8% v 4%) ($F = 11.24$, $p < 0.001$). The association between educational qualifications and job (in)security was statistically insignificant ($F = 1.08$, $p = 0.371$).

For working conditions indicators, employees with university qualifications were less likely to report not having any control over their job tasks (8% v 15%) ($F = 20.30$, $p < 0.001$), work pace (9% v 13%) ($F = 9.27$, $p < 0.001$), work manner (4% v 9%) ($F = 19.13$, $p < 0.001$)

or work hours (27% v 43%) ($F = 41.20$, $p < 0.001$) compared to employees with no educational qualifications. Employees with no educational qualifications were also more likely to report not having any control over their task order (12%) than those with school (slightly) (9%) or university (4%) qualifications ($F = 29.74$, $p < 0.001$).

In terms of awareness of formal flexible working arrangements, employees with no educational qualifications were more likely to mention not being aware of part-time (45% v 39%) ($F = 7.08$, $p = 0.009$), term-time (86% v79%) ($F = 43.82$, $p < 0.001$), job sharing (87% v 77%) ($F = 50.24$, $p < 0.001$), flexi-time (75% v 64%) ($F = 41.05$, $p < 0.001$), compressed hours (92% v 83%) ($F = 59.35$, $p < 0.001$) or annualised hours (slightly) (96% v 93%) ($F = 12.49$, $p < 0.001$) being available than those with university qualifications. Furthermore, those with no educational qualifications were also more likely to mention not being aware of home working (93%) ($F = 179.86$, $p < 0.001$), or other formal flexibility (88%) ($F = 21.13$, $p < 0.001$) being available than those with school or university qualifications. Employees with no educational qualifications were also more likely to report not being able to informally vary their working hours (41%) than employees with school (35%) or university (30%) qualifications ($F = 15.47$, $p < 0.001$). On the other hand, employees with university qualifications were less likely to work unsociable times (22% v 30%) ($F = 37.77$, $p < 0.001$), or to work most or all weekends (14% v 23%) ($F = 36.35$, $p < 0.001$) compared to employees with no educational qualifications.

Regarding collective bargaining, employees with no educational qualifications were more likely to report not having a recognised trade union or staff association at their workplace than those with university qualifications (57% v 50%) ($F = 47.53$, $p < 0.001$).

Occupational Classification

Employees in managerial and professional occupations were less likely to earn below the NMW or NLW (6%) ($F = 366.77$, $p < 0.001$), not to be eligible for a pension scheme (10%)

($F = 113.41$, $p < 0.001$) or not to receive annual increments (52%) ($F = 110.60$, $p < 0.001$) than employees in intermediate or routine and manual occupations. Employees in managerial and professional occupations were also slightly less likely not to receive pay bonuses than those in intermediate occupations (70% v 74%) ($F = 9.44$, $p = 0.001$).

While the differences in percentages by occupational classification for employees who had no progression prospects were insignificant, overall associations were statistically significant. Employees in managerial and professional occupations were slightly more likely to expect a better job with their employer in the following 12 months (15%) than those in intermediate (14%) or routine and manual (12%) occupations ($F = 3.84$, $p = 0.022$). Employees in managerial and professional occupations were also less likely not to expect work-related training (54%) than those in intermediate (60%) or routine and manual (73%) occupations ($F = 94.88$, $p < 0.001$).

Considering job (in)security, employees in managerial and professional occupations were more likely to feel that their job was secure (55%) than those in routine and manual occupations (48%), that is, very unlikely to be of the perception that they will lose their job in the following 12 months ($F = 3.62$, $p = 0.014$). The association between occupational classification and employment type was statistically insignificant ($F = 1.98$, $p = 0.138$).

Employees in managerial and professional occupations were less likely to report having no control over any of the working conditions indicators than employees in intermediate or routine and manual occupations; thus, job tasks (6%) ($F = 104.80$, $p < 0.001$), work pace (8%) ($F = 40.57$, $p < 0.001$), work manner (2%) ($F = 73.04$, $p < 0.001$), task order (2%) ($F = 108.11$, $p < 0.001$) and work hours (22%) ($F = 135.82$, $p < 0.001$).

Regarding awareness of formal flexible working arrangements, employees in managerial and professional occupations were less likely to mention not being aware of term-time (81%) ($F = 36.65$, $p < 0.001$), job sharing (77%) ($F = 113.01$, $p < 0.001$), flexi-time (61%)

($F = 176.49$, $p < 0.001$), compressed hours (82%) ($F = 148.51$, $p < 0.001$), annualised hours (slightly) (93%) ($F = 30.25$, $p < 0.001$), home working (73%) ($F = 460.49$, $p < 0.001$) or other formal flexibility (79%) ($F = 77.67$, $p < 0.001$) being available at their workplace than employees in intermediate or routine and manual occupations. However, for part-time working, employees in managerial and professional occupations (43%) were more likely to mention not being aware of this being available than employees in intermediate occupations (36%), but less likely than employees in routine and manual occupations (48%) ($F = 42.03$, $p < 0.001$). For other work-time scheduling indicators, employees in managerial and professional occupations were less likely to report not being able to informally vary their working hours (27%) ($F = 66.74$, $p < 0.001$), to work unsociable times (20%) ($F = 132.20$, $p < 0.001$), or to work most or all weekends (13%) ($F = 75.85$, $p < 0.001$) than employees in intermediate or routine and manual occupations.

For collective bargaining, employees in managerial and professional occupations (52%) were less likely to report not having a recognised trade union or staff association at their workplace than those in intermediate (57%) or routine and manual (63%) occupations ($F = 34.99$, $p < 0.001$).

Full or Part Time Employment

Considering economic compensation indicators, part-time employees were more likely to earn below the NMW or NLW (28% v 12%) ($F = 133.50$, $p < 0.001$), have no access to pension scheme (29% v 11%) ($F = 349.49$, $p < 0.001$), not to receive pay bonuses (82% v 69%) ($F = 156.70$, $p < 0.001$), or not to receive annual increments (71% v 56%) ($F = 176.46$, $p < 0.001$) than full-time employees.

In terms of training and progression indicators, full-time employees were less likely not to expect a better job (84% v 91%) ($F = 65.94$, $p < 0.001$), or work-related training (56% v

69%) (F = 122.47, p < 0.001) with their employer in the following 12 months compared to part-time employees.

For employment type, part-time employees were more likely to be in temporary employment than full-time employees (11% v 5%) (F = 99.51, p < 0.001). The association between full or part-time employment and job (in)security was statistically insignificant (F = 2.30, p = 0.075).

Part-time employees were more likely to report having no control over any of the working conditions indicators than full-time employees; thus, job tasks (18% v 10%) (F = 53.49, p < 0.001), work pace (15% v 10%) (F = 23.25, p < 0.001), work manner (9% v 5%) (F = 36.87, p < 0.001), task order (11% v 6%) (F = 32.86, p < 0.001), and work hours (43% v 33%) (F = 35.75, p < 0.001).

Regarding awareness of formal flexible working arrangements, part-time employees were less likely to mention not being aware of part-time (13% v 50%) (F = 1136.63, p < 0.001), or term-time (80% v 84%) (F = 28.85, p < 0.001) working being available at their workplace than full-time employees. On the other hand, full-time employees were less likely to mention not being aware of flexi-time (67% v 76%) (F = 62.11, p < 0.001), compressed hours (86% v 93%) (F = 113.53, p < 0.001), annualised hours (slightly) (94% v 96%) (F = 19.64, p < 0.001), home working (82% v 92%) (F = 182.52, p < 0.001), or other flexibility (83% v 88%) (F = 45.10, p < 0.001) being available compared to part-time employees. The association between full or part-time employment and awareness of availability of job sharing was statistically insignificant (F = 2.17, p = 0.140). For other work-time scheduling indicators, part-time employees were slightly more likely to report not being able to informally vary their working hours (37% v 33%) (F = 3.74, p = 0.024), to work unsociable times (29% v 25%) (F = 7.39, p = 0.007), or to work most or all weekends (23% v 18%) (F = 82.79, p < 0.001) than full-time employees.

Considering collective bargaining, part-time employees were more likely to report not having a recognised trade union or staff association at their workplace than full-time employees (60% v 54%) ($F = 30.11$, $p < 0.001$).

Organisational Sector

Employees working in private sector organisations were more likely to earn below the NMW or NLW (18% v 10%) ($F = 69.39$, $p < 0.001$), have no access to pension scheme (21% v 5%) ($F = 487.41$, $p < 0.001$), or not to receive annual increments (68% v 41%) ($F = 677.03$, $p < 0.001$) than public sector employees. However, public sector employees were more likely not to receive pay bonuses compared to those in the private sector (91% v 61%) ($F = 1151.98$, $p < 0.001$).

In terms of training prospects, private sector employees were more likely not to expect work-related training in the following 12 months than public sector employees (63% v 52%) ($F = 99.18$, $p < 0.001$). However, the association between organisational sector and progression prospects was statistically insignificant ($F = 0.59$, $p = 0.444$).

Regarding employment security indicators, public sector employees were slightly more likely to be in temporary employment (9% v 5%) ($F = 50.13$, $p < 0.001$) or to be of the perception that they were very unlikely to lose their job in the following 12 months (55% v 52%) ($F = 4.84$, $p = 0.002$) compared to public sector employees.

Considering working conditions indicators, public sector employees were slightly less likely to report having no control over their task order compared to private sector employees (5% v 8%) ($F = 10.25$, $p < 0.001$). While the differences in percentages by organisational sector for employees with no control over their job tasks, work pace, work manner or work hours were insignificant, overall associations were statistically significant. Private sector employees were slightly more likely to report having a lot of control over their job tasks (40% v 37%) ($F = 8.02$, $p < 0.001$), work pace (45% v 39%) ($F = 12.71$, $p < 0.001$), or work manner (54% v

52%) ($F = 4.10$, $p = 0.007$) than public sector employees. However, public sector employees were slightly more likely to report having some or a lot of control over their work hours compared to private sector employees (84% v 81%) ($F = 10.25$, $p < 0.001$).

Private sector employees were more likely to mention not being aware of part-time (48% v 29%) ($F = 343.93$, $p < 0.001$), term-time (91% v 68%) ($F = 793.66$, $p < 0.001$), job sharing (90% v 68%) ($F = 704.92$, $p < 0.001$), flexi-time (72% v 62%) ($F = 112.53$, $p < 0.001$), compressed hours (91% v 80%) ($F = 244.95$, $p < 0.001$), annualised hours (slightly) (96% v 92%) ($F = 80.87$, $p < 0.001$), or home working (slightly) (85% v 82%) ($F = 17.74$, $p < 0.001$) being available at their workplace than public sector employees. The association between organisational sector and awareness of availability of other formal flexibility was statistically insignificant ($F = 2.49$, $p = 0.115$). For informal flexibility, public sector employees were more likely to report not being able to vary their working hours than private sector employees (38% v 32%) ($F = 22.34$, $p < 0.001$). However, public sector employees were less likely to work unsociable times (23% v 28%) ($F = 28.09$, $p < 0.001$), or to work most or all weekends (12% v 23%) ($F = 116.07$, $p < 0.001$) compared to private sector employees.

For collective bargaining, private sector employees were more likely to report not having a recognised trade union or staff association at their workplace than public sector employees (73% v 21%) ($F = 2534.55$, $p < 0.001$).

Organisation Size

Considering economic compensation indicators, employees in micro size organisations were more likely to earn below the NMW or NLW (27%) ($F = 74.95$, $p < 0.001$), not to be eligible for a pension scheme (44%) ($F = 393.29$, $p < 0.001$) or not to receive annual increments (77%) ($F = 160.64$, $p < 0.001$) than employees in small, medium or large size organisations. Employees in micro size organisations were also more likely not to receive pay bonuses than those in large size organisations (73% v 66%) ($F = 22.58$, $p < 0.001$).

For training and progression indicators, employees in micro size organisations were more likely not to expect a better job (90%) ($F = 15.87$, $p < 0.001$), or work-related training (69%) ($F = 29.32$, $p < 0.001$) with their employer in the following 12 months than employees in small, medium or large size organisations.

While the differences in percentages by organisation size for employees who felt they were very unlikely to lose their job were insignificant, overall associations were statistically significant. Employees in large size organisations were slightly less likely to feel that their job was secure (51%) than those in micro (55%), small (55%) or medium (53%) size organisations, that is, very unlikely to be of the perception that they will lose their job in the following 12 months ($F = 3.47$, $p = 0.003$). The association between organisation size and employment type was statistically insignificant ($F = 0.63$, $p = 0.597$).

For working conditions indicators, employees in micro size organisations were slightly less to report having no control over their work pace (8%) than those in small (12%), medium (12%) or large (11%) size organisations ($F = 8.57$, $p < 0.001$). Employees in micro size organisations were also slightly less likely to report having no control over their task order (5%) ($F = 7.82$, $p < 0.001$) or work hours (31%) ($F = 17.69$, $p < 0.001$) than those in small or medium size organisations. Differences in the percentages by organisation size for employees with no control over their job tasks or work manner were insignificant, however, overall associations were statistically significant. Employees in micro size organisations were more likely to report having a lot of control over their job tasks (50%) ($F = 9.87$, $p < 0.001$) or work manner (59%) ($F = 4.32$, $p < 0.001$) than those in small, medium or large size organisations.

In terms of awareness of formal flexible working arrangements, employees in micro size organisations were more likely to mention not being aware of part-time (48%) being available at their workplace than those in small (41%) or large (38%) size organisations ($F = 17.98$, $p < 0.001$). Employees in large size organisations were less likely to mention not being aware of flexi-time (57% v 76%) ($F = 97.83$, $p < 0.001$), annualised hours (91% v 97%) ($F = 27.33$, $p < 0.001$), home working (74% v 88%) ($F = 119.20$, $p < 0.001$) or other formal flexibility (80% v 87%) ($F = 16.07$, $p < 0.001$) being available at their workplace compared to employees in micro size organisations. Furthermore, employees in micro size organisations were more likely to mention not being aware of term-time (92%) ($F = 47.68$, $p < 0.001$) or job sharing (91%) ($F = 85.45$, $p < 0.001$) being available than employees in any other organisation size, while also more likely to mention not being aware of compressed (92%) being available than those in medium (89%) or large (78%) size organisations ($F = 120.72$, $p < 0.001$). For other work-time scheduling indicators, employees in micro size organisations were less likely to report not being able to informally vary their working hours (30%) than those in small (36%) or medium (39%) size organisations ($F = 13.53$, $p < 0.001$), but more likely to work most or all weekends (25%) than those in medium (18%) or large (14%) size organisations ($F = 24.29$, $p < 0.001$). The association between organisation size and working times was statistically insignificant ($F = 1.99$, $p = 0.113$).

Considering collective bargaining, employees in micro size organisations were more likely to report not having a recognised trade union or staff association at their workplace (84%) than those in small (66%), medium (51%) or large (35%) size organisations ($F = 419.88$, $p < 0.001$).

4.3 Discussion

This chapter has introduced the variables that are going to be used in this study, presented their univariate analyses and sought to understand how the variables are associated in the UK employee population through bivariate analyses. Results from the analysis supported previous literature highlighting inequalities in the labour market by sex (Korpi 2018). Thus, on average, males in the UK employee population were more likely to earn more and receive bonuses compared to females and Lindley (2015) attributed this, in part, to occupational differences. However, pay for females was more likely to include annual increments, which are common in public than private sector organisations and females were more likely to work in the public sector. This might also explain why females were more likely to report having recognised trade unions or staff associations at their workplace than males as evidence suggests there is greater access to unionisation in the public sector (Charlwood and Terry 2007). Results from the analysis were consistent with literature indicating females were more likely to be in temporary employment with less progression prospects, and generally lower levels of autonomy (Fredman 2004; Piasna and Plagnol 2018; Pollert and Charlwood 2009) but had better awareness of availability of flexible working arrangements compared to males (Piasna and Plagnol 2018; Tomlinson 2007).

The analysis supported evidence from literature about disparities in the labour market by ethnic group, indicating that employees from a White ethnic background experienced better outcomes in the labour market compared to those from other ethnic backgrounds (Dillon 2020; Korpi 2018). The analysis found that employees from a White ethnic background had higher levels of economic compensation; less likely to be in low paid employment, whilst also more likely to be eligible for pension schemes and receive bonuses; employment security, greater scope for autonomy, and better awareness and access to working arrangements that support work-time scheduling than employees from other ethnic backgrounds. This was consistent with findings from Clark et al. (2022) and Zwysen and Demireva (2020). However, the analysis found that employees from a Black or Black British background were more likely to expect a better job, as well as work-related than employees from other ethnic backgrounds.

Evidence has suggested variations in the different forms of precarious work by age (Kim and Kurz 2001), particularly for young workers (Arranz et al. 2019) but also for older workers. Indeed, results from this analysis indicated that employees in the 16-24 and 65+ age groups were more likely to be in insecure employment, with less scope for work autonomy, no recognised trade unions or staff association, and among the lowest earners, while those aged 65+ years old were also least likely to be eligible for a pension scheme, bonuses or annual increments. This might be due to that employees in these age groups tend to be in low-skilled occupations (Kim and Kurz 2001), with limited coverage by collective bargaining agreements (Bosch 2009). The results, however, also indicated that young employees were the most likely to expect training and progression as they develop their careers, but also a reflection of the reluctance by employers to invest in older workers (Canduela et al. 2012). In terms of work-time scheduling indicators, results suggested that middle aged employees had better awareness of availability of flexible working arrangements, but those in the 16-24 and 65+ age groups were more likely to mention being aware of part-time working being available, while those aged 65+ years old also reported a higher likelihood of being able to vary their working hours on an informal basis, and those 16-24 years old were the most likely to work unsociable times, and this might be an indication of the precarious nature of employment for these groups (Kim and Kurz 2001).

Evidence from literature has indicated a marriage premium in the labour market in terms of economic compensation, particularly for males (Bardasi and Taylor 2008; Ribar 2004) and this is supported by results from this analysis, although the analysis did not distinguish by sex. Employees in married / cohabiting relationships were the most likely to be high earners compared to employees in other relationship types, while also being among the most likely to be eligible for pension schemes, receive bonuses and annual increments. Married / cohabiting employees were also more likely to be in permanent employment with higher levels of working conditions in all aspects of work autonomy. This can be partly attributed to increased productivity associated with married / cohabiting employees as a result of human capital accumulated with spousal support (Bardasi and Taylor 2008; Ribar 2004). The analysis found mixed results in terms of awareness of availability of formal flexible working arrangements by relationship status, although married / cohabiting employees had better awareness and access to working arrangements that supported better work-time scheduling. On the other hand, single employees were more likely to expect training and progression, work unsociable times and on most / all weekends, while also the most likely not have a recognised trade union or staff association at their workplace compared to employees in other relationships. This might be associated with age as single people tend to be predominantly young and more likely to be in precarious work (Kim and Kurz 2001).

According to Nieuwenhuis and Maldonado (2018) lone parents are particularly disadvantaged in the labour market and this was supported by results from this analysis. Lone parents with primary school age children were most likely to be the lowest earners, not to be eligible for pension schemes, bonuses or annual increments, and have lower levels of work autonomy in terms of task order or work hours, with no recognised trade union or staff association at their workplace compared to coupled parents with primary school age children or employees without primary school age children. In terms of work-time scheduling indicators, lone parents with school age children were also more likely to mention not being aware of formal flexible working arrangements available at their workplace (except for part-time), work unsociable times, and most / all weekends compared to employees of other parental status. As lone parents tend to be predominantly female (Esser and Olsen 2018; Klett-Davies 2016; Nieuwenhuis and Maldonado 2018), this can be attributed to the precarious nature of the work lone mothers do which tends to be in low or lower middle skilled occupations (Klett-Davies 2016).

Evidence from literature has suggested that employees with a disability experience prejudice in the labour market compared to those without a disability (Grover and Piggott 2015; TUC 2021a) and tend to be in non-standard employment (Davidson and Kemp 2008). This was generally supported by results from this analysis which found that employees with a disability were slightly more likely to be in employment with low pay which did not include annual increments, however they were slightly more like to be in employment where they were eligible for a pension compared to those without a disability. Furthermore, the analysis found that compared to those without a disability, employees with a disability were less likely to expect training and progression, perceived their job to be less secure, had lower levels of working conditions in all aspects of work autonomy, however they were more likely to report being aware of part-time working being available and having a recognised trade union or staff association at their workplace. These findings were generally consistent with evidence from other literature (Grover and Piggott 2015; McGovern et al. 2004; Meager and Hill 2005) and can be attributed to socially embedded barriers in the labour market for employees with a disability as well as skills differentials (Grover and Piggott 2015).

In terms of region, evidence from literature has suggested disparities in the labour within and across regions and nations of the UK (Jones and Green 2009), partly driven by a shift from heavy industry to a knowledge-based economy (Department for LUHC 2022; Hepworth et al. 2005). Results from this analysis, in part, supported evidence in other literature, which highlighted the North East, Wales and Northern Ireland as having the lowest proportion of high-quality jobs, while London and the South East regions had the highest proportions of high-quality jobs (Jones and Green 2009) due a high-skilled occupational base working in the knowledge-based economy highly centralised in these regions (Hepworth et al. 2005; Jones and Green 2009). Employees in North Ireland and Wales were among the most likely to be the lowest earners, with those in London the least likely. While employees from Northern Ireland were the most likely to feel secure in their jobs, they were also the most likely not to expect work-related training and progression, have lower levels of work autonomy and have less aware of the scope for flexible working arrangements available at their workplace. On the other hand, employees in London were the most likely to expect both work-related training and progression, while also among the most likely to feel insecure about their job along with employees from Northern England, Wales, or East of England. Employees in London or Southern England were the most likely not to have recognised trade unions or staff associations at their workplace.

The results in this analysis supported evidence from previous studies that higher levels of education resulted in human capital (Okay-Somerville and Scholarios 2013; Solomon et al. 2022) that afforded individuals greater job resources (Solomon et al. 2022). The analysis found that, compared to those with lower levels of educational qualifications, employees with a university or higher degree were more likely to have higher pay, greater eligibility for a pension provision as well as more likely to receive annual increments, and greater work autonomy. Furthermore, employees with a university or higher degree had better awareness and access to working arrangements that supported better work-time scheduling, were more likely to have recognised trade union or staff associations at their workplace, and among the most likely to expect work-related training and progression. However, the analysis found that employees with a university or higher degree were among the least likely to be in permanent employment. The findings might be attributed to skills differentials between employees with different levels of education and the occupations in which graduates are employed (Green and Zhu 2010; Okay-Somerville and Scholarios 2013; Warhurst 2008).

Findings from this analysis, generally, supported evidence from previous literature in terms of variations in QWE by occupational classification with better outcomes for high-skilled employees, partly attributed to skills differentials in the occupational hierarchy (Gallie 2015; Wheatley 2022). For example, Wheatley (2022) found higher levels of pay, work autonomy, employee voice, as well as skills and development prospects among high-skilled employees compared to low-skilled employees. This analysis found lower levels on all indicators of economic compensation and work autonomy among employees in low-skilled occupations compared to those in high-skilled occupations. While the analysis found some heterogeneity among high-skilled employees in terms of collective bargaining, with employees in managerial and professional occupations the most likely not to have recognised trade union or staff associations at their workplace, other high-skilled employees were more likely to have recognised trade union or staff associations at their workplace than low-skilled employees. For work-time scheduling indicators, employees in managerial and professional occupations were less likely to mention not being aware of formal flexible working arrangements available than those in intermediate or routine and manual occupation, except for part-time working when compared to those in intermediate occupations. Employees in managerial and professional occupations were also less likely to not to be able to informally vary their working hours, work unsociable hours, or most or all weekends.

Results from this analysis supported previous literature relating to the precarious nature of part-time compared to full-time employment. By design, part-time jobs require fewer skills and lower levels of training and educational attainment compared to full-time jobs (Lyonette et al. 2010; Warren and Lyonette 2015). This analysis found that employees in part-time employment had lower levels on all indicators of economic compensation, training and progression, and working conditions dimensions, while they were also more likely to be in temporary employment, and not to have a recognised trade union or staff association at their workplace compared to employees in full-time employment. This was consistent with evidence in other literature (Hoque and Kirkpatrick 2003; McGovern et al. 2004; Warren and Lyonette 2015). Furthermore, while part-time employees were more likely to be aware of part-time or term-time working being available at their workplace than full-time employees, this analysis found that, overall, full-time employees reported better awareness of availability of flexible working arrangements compared to part-time employees, contrary to evidence from Lyonette (2015) who suggested that full-time employees had poor work-life balance than part-time employees, perhaps due to the precarious nature of part-time employment with employees not being aware of or having access to other forms of better work-time scheduling.

Previous evidence of a public sector pay premium among UK employees is well established, attributed partly to a more skilled workforce with higher levels of education in the public sector relative to the private sector (Cribb et al. 2014; Murphy et al. 2020). This was supported by findings from this analysis, thus, employees working in the private sector were more likely to be the lowest earners, not eligible for a pension scheme, or for their pay not to include annual increments; however, they were more likely to receive bonuses compared to public sector employees. On the other hand, public sector employees were more likely to expect work-related training, report better awareness of availability of flexible working arrangements, or have recognised trade unions or staff associations at their workplace compared to private sector employees, consistent with findings from other studies (Charlwood and Terry 2007; Leschke and Keune 2008; Rubery 2013). However, private sector employees were slightly more likely to be in permanent employment and report being secure in the job than public sector employees contrary to findings by Fontaine et al. (2020). This might be due to decreases in public sector employment and increases in private sector employment in the period after the 2008 financial crisis (ONS 2020). Results of working conditions in terms of different aspects of work autonomy by organisational sector were mixed.

Results from this analysis suggested that employees in large size companies had higher levels of economic compensation than those in companies of other sizes and this was consistent with findings by Forth et al. (2006). Thus, employees in large companies were the least likely to be the lowest earners, not to be eligible for a pension scheme, or for pay not include annual increments, however they were also the least likely to receive bonuses than employees in companies of other sizes. Compared to other organisation sizes, large companies were also more likely to offer better prospects for training and progression, better awareness of availability of flexible working arrangements, and less likely not to have recognised trade unions or staff associations at their workplace. This supported findings by Forth et al. (2006) and can be attributed to the resources at the disposal of large companies. On the other hand, employees in medium size companies were slightly the most likely to feel secure, while employees in micro size companies had highest levels of working conditions in a broad scope of work autonomy indicators and this was partly supported by other studies (Bryson et al. 2021; Forth et al. 2006).

Univariate and bivariate analyses are important in understanding the frequency distributions of the variables used in the study and how the indicators and predictors of QWE are related. However, limitations, specifically in relation to bivariate analysis, are that it only considers the association between two variables at a time and assumes that all predictors of QWE have an equal influence on the indicators of QWE. Subsequent chapters of this study will apply latent variable modelling to aggregate the indicators of QWE and conduct latent regression analysis to model the effect of the predictors on measures of QWE.

4.4 Appendices

4.4.1 Appendix 4.1: Survey Questions for Predictors of QWE from the UKHLS

	Predictor	Survey Question	Response Options (excluding options for missing responses)
Demographic Characteristics	Sex	What is your sex?	2. Male 3. Female
	Ethnic group	What is your ethnic group?	3. British / English / Scottish / Welsh / Northern Irish 4. Irish 5. Any other White background 6. White and Black Caribbean 7. White and Black African 8. White and Asian 9. Any other mixed background 10. Indian 11. Pakistani 12. Bangladeshi 13. Chinese 14. Any other Asian background 15. Caribbean 16. African 17. Any other Black background 18. Arab 97. Any other ethnic group
	Age group	What is [your] / [NAME's] / [ff_forname's] date of birth? 1. Enter the day of the month. 2. Enter month. 3. Enter the year.	1. [Value >= 0] [01 – 31] (day of birth) 2. [1: Jan, 2: Feb, 3: Mar, 4: Apr, 5: May, 6: Jun, 7: Jul, 8: Aug, 9: Sep, 10: Oct, 11: Nov, 12: Dec] (month of birth) 3. [Value >= 0] [yyyy] (year of birth)
Socio-demographic Characteristics	Relationship status	Since personal circumstances can change over time, we would just like to check some important information. What is [NAME's / CNAME's / ff_forname's / your] legal marital status?	3. Single, never married or never in a legally recognised Civil Partnership 4. Married 5. Civil partnership in a legally recognised civil partnership 6. Separated but legally married 7. Divorced 8. Widowed 9. Separated from Civil Partner 10. Former Civil Partner, Civil Partnership legally dissolved 11. Surviving Civil Partner (partner having died)
	Parental status	Derived variable based on the legal marital status and the number of children aged 5 - 11 years old in the household.	4. [Value >= 0] (0: None)
	Illness or disability	Do you have any long-standing physical or mental impairment, illness or disability? By 'long-standing', I mean anything that has troubled you over a period of at least 12 months, or that is likely to trouble you over a period of at least 12 months.	3. Yes 4. No

Continued…

Continued…

Socio-demographic Characteristics	Region	Government office region derived from the household's postcode.	1. North East 2. North West 3. Yorkshire and the Humber 4. East Midlands 5. West Midlands 6. East of England	7. London 8. South East 9. South West 10. Wales 11. Scotland 12. Northern Ireland
Socio-economic Characteristics	Education	Can you tell me the highest educational or school qualification you have obtained?	2. University Higher degree 3. First-degree or equivalent 4. Diploma in Higher Education 5. Teaching qualification 6. Nursing or other medical qualification (not yet mentioned) 7. Other higher degree 8. A level 9. Welsh Baccalaureate	10. International Baccalaureate 11. AS level 12. Highers (Scotland) 13. Certificate of sixth-year studies 14. GCSE / O level 15. CSE 16. Standard/ordinary / lower grade 17. Other school certificate 96. None of the above
	Occupational classification	Standard Occupational Classification 2000 (SOC2000) of current job. Condensed 3-digit version.	111 – 123: Managers and senior officials 211 – 245: Professional 311 – 356: Associate professional and technical 411 – 421: Administrative and secretarial	511 – 549: Skilled trades 611 – 629: Personal service 711 – 721: Sales and customer service 811 – 822: Process, plant and machine operatives 911 – 925: Elementary occupations
	Full or part-time	Derived variable based on total hours, i.e. including both normal and overtime hours. Employed full-time (i.e., greater than 30 hours per week).	4. Full-time employee 5. Part-time employee	
	Organisational sector	Do you work for a private firm or business or other limited company, or do you work for some other type of organisation?	7. Private firm or business, a limited company 8. Other type of organisation	
	Organisation size	How many people are employed at the place where you work?	4. 1 – 2 5. 3 – 9 6. 10 – 24 7. 25 - 49 8. 50 – 99 9. 100 – 199 10. 200 – 499 11. 500 – 999	12. 1000 or more 13. Don't but < 25 14. Don't know but ≥ 25

Source: UK Household Longitudinal Study, Wave 8 (2016 – 2017).

4.4.2 Appendix 4.2: Frequency Distributions of Indicators of QWE

Economic Compensation Indicators

Indicator		Value	Frequency (Unweighted)	Percentage (Weighted)
Effective Gross Pay				
	Below NMW / NLW	0	2609	16
	Q1 above NMW / NLW	1	3258	21
	Q2 above NMW / NLW	2	3664	21
	Q3 above NMW / NLW	3	3645	21
	Q4 above NMW / NLW	4	3585	21
Total Valid			**16761**	**100**
	Missing		220	
Total			**16981**	

Indicators		Pension Provision		Pay Bonuses		Pay Progression	
	Value	Frequency (Unweighted)	Percentage (Weighted)	Frequency (Unweighted)	Percentage (Weighted)	Frequency (Unweighted)	Percentage (Weighted)
No	0	2430	15	12360	72	9538	59
Yes	1	14218	85	4527	28	6963	41
Total Valid		**16648**	**100**	**16887**	**100**	**16501**	**100**
Don't know		305		57		441	
Refusal		18		27		29	
Missing		10		10		10	
Total		**16981**		**16981**		**16981**	

Training and Progression Indicators

Indicators		Progression Prospects		Training Prospects	
	Value	Frequency (Unweighted)	Percentage (Weighted)	Frequency (Unweighted)	Percentage (Weighted)
No	0	13853	86	9612	59
Yes	1	2380	14	6654	41
Total Valid		**16233**	**100**	**16266**	**100**
Don't know		740		707	
Refusal		7		7	
Missing		1		1	
Total		**16981**		**16981**	

Employment Security Indicators

Indicator		Value	Frequency (Unweighted)	Percentage (Weighted)
Employment Type				
	Temporary	0	1035	6
	Permanent	1	15924	94
Total Valid			**16959**	**100**
	Don't know		20	
	Refusal		2	
Total			**16981**	
Job Security				
	Very likely	0	336	2
	Likely	1	783	5
	Unlikely	2	6661	40
	Very unlikely	3	8750	53
Total Valid			**16530**	**100**
	Don't know		426	
	Refusal		14	
	Missing		11	
Total			**16981**	

Working Conditions Indicators

Indicators		Job Tasks		Work Pace		Work Manner		Task Order		Work Hours	
	Value	**Frequency** (Unweighted)	**Percentage** (Weighted)	**Frequency** (Unweighted)	**Percentage** (Weighted)	**Frequency** (Unweighted)	**Percentage** (Weighted)	**Frequency** (Unweighted)	**Percentage** (Weighted)	**Frequency** (Unweighted)	**Percentage** (Weighted)
None	0	2031	12	1919	11	1042	6	1214	7	5854	35
A little	1	2545	15	2383	14	1903	11	1858	11	3291	20
Some	2	5725	34	5419	32	5122	30	4932	29	3761	22
A lot	3	6640	39	7212	43	8866	53	8929	53	4027	23
Total Valid		**16941**	**100**	**16933**	**100**	**16933**	**100**	**16933**	**100**	**16933**	**100**
Don't know		25		33		32		33		33	
Refusal		14		14		15		14		14	
Missing		1		1		1		1		1	
Total		**16981**		**16981**		**16981**		**16981**		**16981**	

Work-time Scheduling Indicators

Indicators	Value	Part-time Frequency (Unweighted)	Part-time Percentage (Weighted)	Term-time Frequency (Unweighted)	Term-time Percentage (Weighted)	Job Sharing Frequency (Unweighted)	Job Sharing Percentage (Weighted)	Flexi-time Frequency (Unweighted)	Flexi-time Percentage (Weighted)
Not mentioned	0	6639	42	13710	83	13572	82	11422	69
Mentioned	1	9933	58	2862	17	3000	18	5150	31
Total Valid		**16572**	**100**	**16572**	**100**	**16572**	**100**	**16572**	**100**
		367		367		367		367	
Refusal		31		31		31		31	
Missing		11		11		11		11	
Total		**16981**		**16981**		**16981**		**16981**	

Indicators	Value	Compressed Hours Frequency (Unweighted)	Compressed Hours Percentage (Weighted)	Annualised Hours Frequency (Unweighted)	Annualised Hours Percentage (Weighted)	Home Working Frequency (Unweighted)	Home Working Percentage (Weighted)	Other Flexibility Frequency (Unweighted)	Other Flexibility Percentage (Weighted)
Not mentioned	0	14462	87	15689	95	13956	84	13927	84
Mentioned	1	2110	13	883	5	2616	16	2645	16
Total Valid		**16572**	**100**	**16572**	**100**	**16572**	**100**	**16572**	**100**
		367		367		367		367	
Refusal		31		31		31		31	
Missing		11		11		11		11	
Total		**16981**		**16981**		**16981**		**16981**	

Continued…

Indicator		Value	Frequency (Unweighted)	Percentage (Weighted)
Informal Flexibility				
	No	0	6054	34
	Sometimes	1	2439	15
	Yes	2	8397	51
Total Valid			**16890**	**100**
	Don't know		71	
	Refusal		9	
	Missing		11	
Total			**16981**	
Working Times				
	Unsociable times	0	4412	27
	Sociable times	1	12493	73
Total Valid			**16905**	**100**
	Other		62	
	Don't know		2	
	Refusal		12	
Total			**16981**	
Weekend Working				
	Most or every weekend	0	3047	19
	Some weekends	1	5712	34
	No weekend working	2	8209	47
Total Valid			**16968**	**100**
	Don't' know		6	
	Refusal		7	
Total			**16981**	

Social Dialogue Indicator

Indicator		Value	Frequency (Unweighted)	Percentage (Weighted)
Collective Bargaining				
	No	0	8416	56
	Yes	1	7829	44
Total Valid			**16245**	**100**
	Don't know		709	
	Refusal		17	
	Missing		10	
Total			**16981**	

Source: UKHLS, Wave 8 (2016 – 2017).

4.4.3 Appendix 4.3: Frequency Distributions of Predictors of QWE

Demographic Characteristics

Predictor		Value	Frequency (Unweighted)	Percentage (Weighted)
Sex				
	Female	0	9246	51
	Male	1	7735	49
Total			**16981**	
Ethnic Group				
	White	0	13716	92
	Mixed	1	327	1
	Asian or Asian British	2	1791	5
	Black or Black British	3	844	2
Total Valid			**16678**	**100**
	Missing		303	
Total			**16981**	
Age Group				
	16 – 24	0	1649	11
	25 – 34	1	3223	20
	35 – 49	2	6568	36
	50 – 64	3	5156	30
	65+	4	385	2
Total			**16981**	

Socio-demographic Characteristics

Predictor		Value	Frequency (Unweighted)	Percentage (Weighted)
Relationship Status				
	Single	0	5588	37
	Married / cohabiting	1	9345	52
	Divorced / separated	2	1786	9
	Widowed	3	214	1
Total Valid			**16933**	**100**
	Missing		48	
Total			**16981**	
Parental Status				
	Lone parent w sch age chd	0	1066	7
	Coupled parents w sch age chd	1	2946	16
	Employees w/o sch age chd	2	12906	77
Total Valid			**16918**	**100**
	Missing		63	
Total			**16981**	
Illness or Disability				
	Yes	0	4016	25
	No	1	12941	75
Total Valid			**16957**	**100**
	Missing		24	
Total			**16981**	
Region				
	London	0	2285	12
	Southern England	1	3293	22
	East of England	2	1444	10
	The Midlands	3	2665	16
	Northern England	4	3777	24
	Wales	5	1068	5
	Scotland	6	1434	8
	Northern Ireland	7	1006	3
Total Valid			**16972**	**100**
	Missing		9	
Total			**16981**	

Socio-economic Characteristics

Predictor		Value	Frequency (Unweighted)	Percentage (Weighted)
Education				
	No qualifications	0	2279	13
	GCSE / O-level or lower	1	3975	27
	Up to A-level	2	1758	11
	Up to Diploma in HE	3	1786	10
	University or higher degree	4	5256	30
	No recorded data	5	1927	8
Total			**16981**	**100**
Occupational Classification				
	Managers & senior officials	0	2561	16
	Professional occupations	1	2442	14
	Assoc prof. & tech occupations	2	2913	17
	Administrative & secretarial	3	1963	12
	Skilled trades occupations	4	939	6
	Personal service occupations	5	1864	10
	Sales & customer service	6	1267	8
	Process, plant & machine oper.	7	1018	6
	Elementary occupations	8	1763	11
Total Valid			**16730**	**100**
	Missing		251	
Total			**16981**	
Full or Part-time				
	Part-time	0	3875	22
	Full-time	1	12964	78
Total Valid			**16839**	**100**
	Missing		142	
Total			**16981**	
Organisational Sector				
	Private sector	0	10762	66
	Public sector	1	6096	34
Total Valid			**16858**	**100**
	Missing		123	
Total			**16981**	

Continued…

Predictor		Value	Frequency (Unweighted)	Percentage (Weighted)
Organisation size				
	Micro	0	2597	16
	Small	1	5049	30
	Medium	2	3839	23
	Large	3	5420	31
Total Valid			**16905**	**100**
	Missing		76	
Total			**16981**	

Chapter 5 Measuring Quality of Work and Employment

The objective of this chapter is to apply item response theory (IRT) to construct a multidimensional measurement instrument of quality of work and employment (QWE) for the UK employee population that addresses the limitations of existing measures, including considering overall and different dimensions of QWE. The measurement instrument will be based on the conceptual framework presented in Chapter 2 (Section 2.1.5 and Figure 2.1). The first section describes the methodology of developing the measurement instrument and considers the data and sample, the observed items or indicators of QWE, and the methods applied. The second section presents the results, starting with an assessment of the dimensionality of the indicators of QWE and how this informs potential measurement models, followed by a comparison of different measurement models. Results of the measurement model that better fits the data are presented in detail, including the item slope-intercept parameters, model diagnostics, predicted latent trait scores, and evaluation of the properties of the instrument. The final section discusses the advantages and disadvantages of the measurement instrument and sets out how it will be used in the subsequent chapters to investigate QWE for different groups of the UK employee population.

5.1 Methodology

5.1.1 Data and Sample

This study uses data from Wave 8 (2016 – 2017) of *Understanding Society: The United Kingdom Household Longitudinal Study* (UKHLS) (University of Essex, Institute for Social and Economic Research 2018). The sample was limited to employees aged 16 years old and over, who were in a paid job, and participated in full interviews and the base sample was 16,981. Refer to Chapter 3 (Section 3.1.3) for a more detailed description of the data.

5.1.2 Variables

Table 4.1 outlined the indicators used to develop the QWE measurement instrument, their descriptions, item numbers used in subsequent path diagrams. Initially, 25 indicators were considered in developing the measure of QWE, however, the '*weekend working*' and '*collective bargaining*' indicators were excluded due to a violation of the local independence assumption in the subsequent IRT modelling. That is, responses to these items were not sufficiently explained by the model. A consequence of excluding these indicators, particularly the '*collective bargaining*' indicator, is that the resulting measure of QWE does not capture any aspect of the social dialogue dimension from the conceptual framework of QWE presented in Chapter 2 (Figure 2.1). This is because the UKHLS only had one appropriate indicator for this dimension. The measurement instrument consisted of 23 items, which were dichotomously or polytomously scored, and had ordinal levels of measurement. These were (re)coded such that higher values represented higher levels of QWE as described in Chapter 4 (Section 4.1.2).

5.1.3 Methods

Analyses were conducted in *Stata* 17 (StataCorp 2021), '*mirt*' package (Chalmers 2012) in *R* (R Core Team 2020), and M*plus* 8.8 (Muthén and Muthén 2017). To cross-validate the measurement model, the base sample was split into two approximately equal-sized random samples; thus, the first random sample (n = 8,491) and second random sample (n = 8,490).

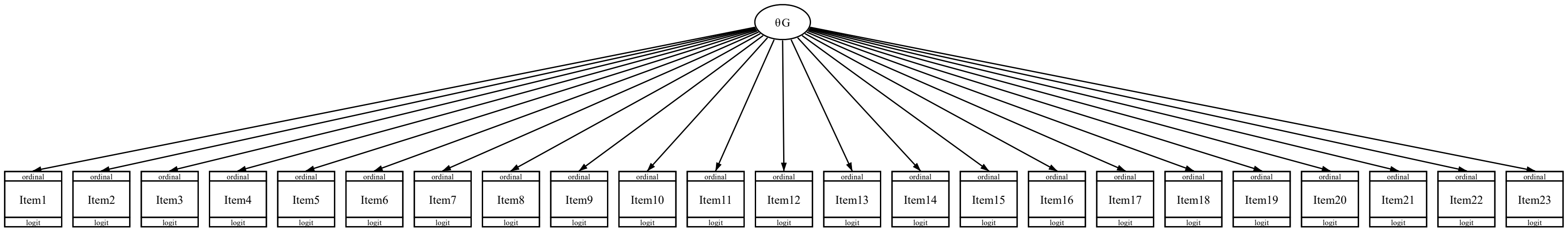
Assessing Data Structure and Measurement Models Comparison

The first random sample was used to assess the data structure and compare competing measurement models. Since the observed indicators of QWE were ordinal and dichotomously or polytomously scored, the polychoric correlation matrix was estimated to investigate the strength of the linear relationship between the indicators. The correlation coefficients were analysed based on Cohen's (1988) rule of thumb[29] to better understand the data structure.

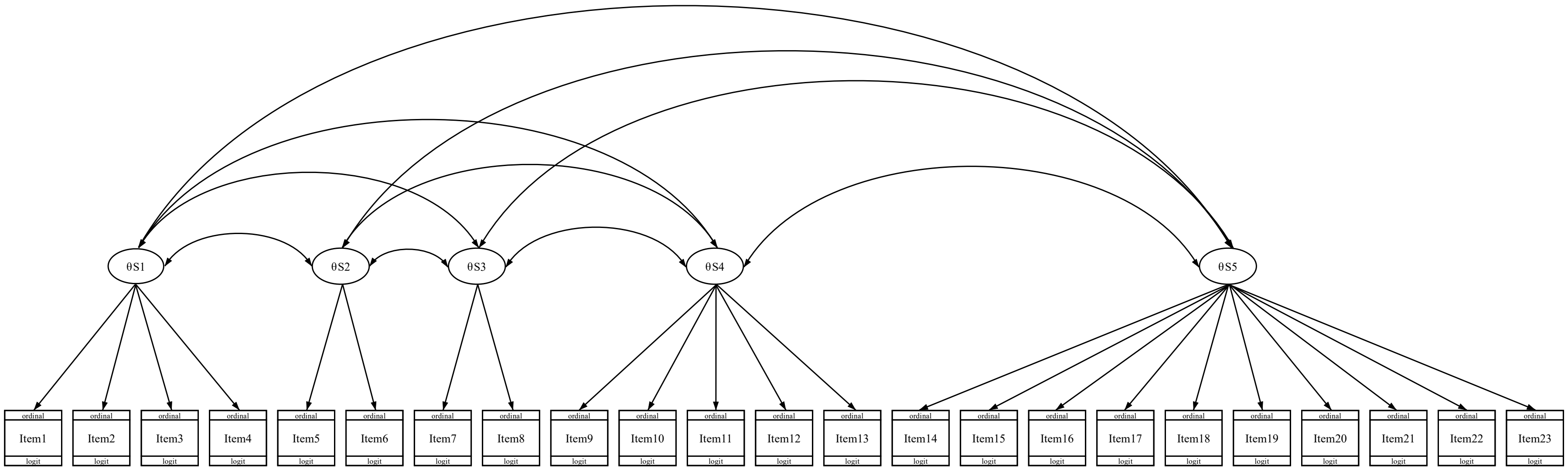
Competing *two-parameter logistic* (2-PL) confirmatory graded response models (GRMs) were estimated using the maximum marginal-likelihood (MML) estimation with the expectation maximisation (EM) algorithm (Cai et al. 2011; Gibbons et al. 2007; Reise 2012) and compared using the '*mirt*' package (Chalmers 2012) in *R* (R Core Team 2020). Figures 5.1 (a – d) depict path diagrams for the potential measurement models with observed indicators; *Item1 – Item23*; dependent on latent traits θ_G (*general factor*) measuring *overall QWE* and / or θ_{S1} – θ_{S5} (*specific factors*) measuring different *dimensions of QWE*; thus, θ_{S1} measuring *economic compensation*, θ_{S2} measuring *training and progression*, θ_{S3} measuring *employment security*, θ_{S4} measuring *working conditions*, and θ_{S5} measuring *work-time scheduling*.

[29] Correlation coefficients range from [–1, 1] with –1 indicating a perfect negative relationship and 1 a perfect positive relationship, while 0 indicates no relationship. As a rule of thumb, absolute coefficients < 0.30 indicate a weak relationship, between 0.30 and 0.49 a moderate relationship, and > 0.49 a strong relationship (Cohen 1988).

Figure 5.1: Path Diagrams for Measurement Models of Quality of Work and Employment



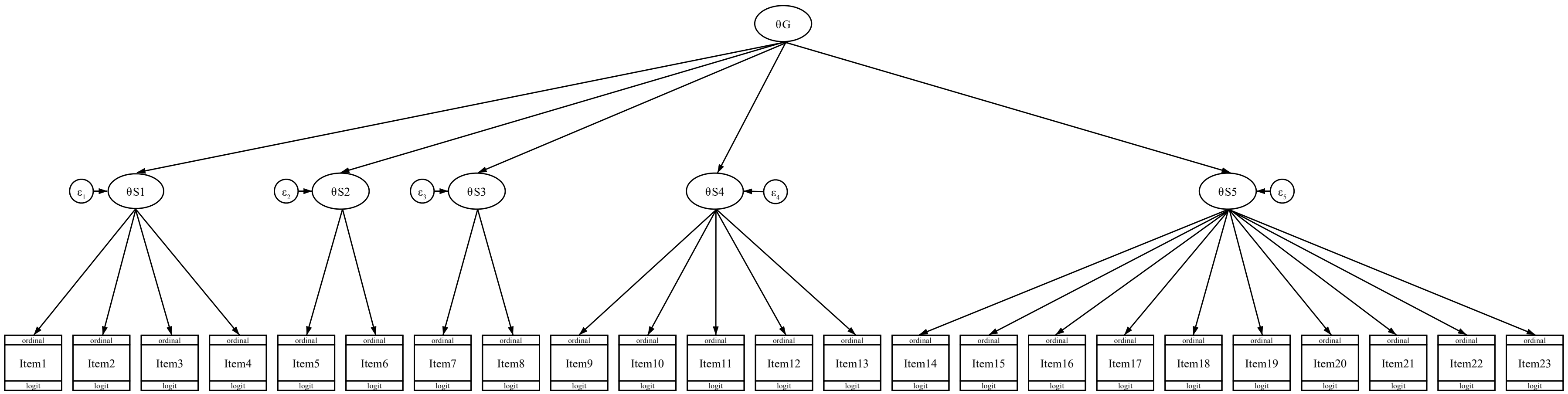
(a) Unidimensional model



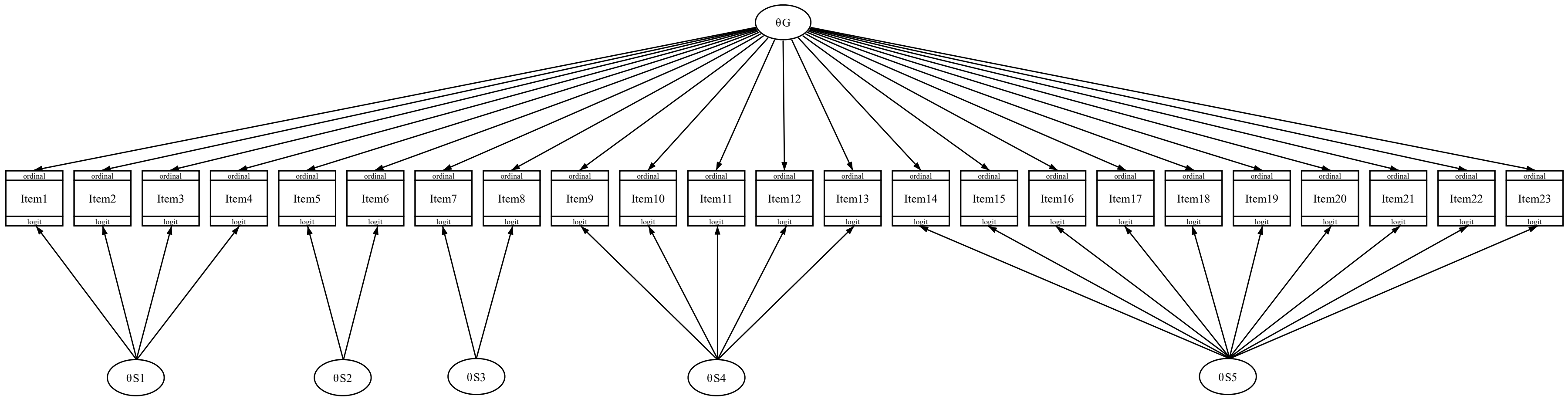
(b) Correlated-factors model

continued…

continued…



(c) Second-order factor model



(d) Bifactor model

For the *unidimensional model* (Figure 5.1 (a)), responses to observed indicators were hypothesised to be dependent on θ_G accounting for the common variance shared by all the indicators (Brown and Croudace 2015; Chen and Zhang 2018; Reise, Bonifay, and Haviland 2018). The 2-PL graded response *unidimensional IRT model* is represented mathematically by the formulations in Equations 3.5 and 3.6. In the case of the *correlated-factors model* (Figure 5.1 (b)), the model hypothesised that responses to the observed indicators were explained by the five dimensions of QWE, θ_{S1} – θ_{S5}, accounting for the common variance shared by indicators within that dimension. The dimensions were specified to be correlated to account for their shared common variance (Brown and Croudace 2015; Chen and Zhang 2018). Mathematically, the 2-PL graded response *correlated-factors IRT model* is represented by the formulations in Equations 3.7 and 3.6 with dimensions specified to be correlated in the estimation.

The *second-order factor model* (Figure 5.1 (c)) is similar to the correlated-factors model, but hypothesised that instead of the θ_{S1} – θ_{S5} (*first-order factors*) being correlated, there was a higher order factor, *second-order factor* (θ_G), accounting for the common variance shared by θ_{S1} – θ_{S5}, which are then conditionally independent (Brown and Croudace 2015; Chen et al. 2012; Chen, West, and Sousa 2006; Reise 2012). The 2-PL graded response *second-order IRT factor model* is expressed mathematically by Equations 5.1 and 5.2, with Equation 5.1 representing the measurement model for the observed indicators by their first-order factor or dimension of QWE, and Equation 5.2 representing the structure for each of the first-order factors by the second-order factor:

$$P\,(x_{iu} \geq u \mid \theta_S) \;=\; \frac{e^{(a_{iS}\theta_S \,+\, d_{iu})}}{1 \,+\, e^{(a_{iS}\theta_S \,+\, d_{iu})}} \tag{5.1}$$

$$\theta_S \;=\; \beta_S\theta_G +\, \varepsilon_S \tag{5.2}$$

where $P\,(x_{iu} \geq u \mid \theta_S)$ is the conditional probability of selecting response category u or higher to item i given a particular dimension of QWE (θ_S), a_{iS} is the slope for item x_i to θ_S, and d_{iu} is the intercept (or easiness) for response category u to item x_i. The conditional probability of selecting response category u given θ_S is expressed by the formulation in Equation 3.6. For Equation 5.2, β_S represent the factor loadings of a particular θ_S on θ_G, and ε_S is the residual of a particular θ_S representing the unique variance not accounted for by θ_G.

The *bifactor model* (Figure 5.1 (d)) hypothesised that responses to observed indicators were explained by θ_G given $\theta_{S1} - \theta_{S5}$, accounting for the common variance shared by all the indicators, but also $\theta_{S1} - \theta_{S5}$ accounting for any common variance among the indicators within that dimension over and above θ_G (Brown and Croudace 2015; Chen et al. 2012, 2006; Reise 2012; Reise et al. 2018). In terms of formulation, the *bifactor model* models within-item multidimensionality, as opposed to the simple structures of the other models (Desjardins and Bulut 2018; Paek and Cole 2020), with item responses being functions of θ_G and one other θ_S (Cai et al. 2011; DeMars 2013; Toland et al. 2017). In general, for a latent variable model to be identified, every latent variable must be scaled and the degrees of freedom, that is, the difference between the number of observed variances / covariances, and free parameters estimated by the model must be zero (just-identified) or greater (over-identified) (Kline 2016; Wang and Wang 2020). For the bifactor model in Figure 5.1 (d) and specifically in relation to *specific factors* defined by two items, θ_{S3} and θ_{S4}, for the model to be identified, slopes of the items within these factors were constrained to be equal (Cai et al. 2011). Thus, considering

Item 5 and Item 6 measuring θ_G and θ_{S3}, six free parameters would be estimated by the model.[30] However, the observed variance-covariance matrix from the two items would consist of three known values.[31] In the bifactor model, all the latent traits were assumed to be mutually orthogonal, and follow standard normal distributions with zero mean and unit variance (Cai et al. 2011; Reise 2012). This reduced the number of free parameters to four, thus, θ_G and θ_{S3} have unit variances. Constraining the slopes of the items within θ_{S3} further reduced the number of free parameters to three. This meant that the model related to the specific factors defined by two items had zero degrees of freedom and therefore just-identified (Cai et al. 2011).

Mathematically, the 2-PL graded response *bifactor IRT model* is represented as:

$$P\ (x_{iu} \geq u \mid \theta_G, \theta_S) = \frac{e^{(a_{iG}\theta_G + a_{iS}\theta_S + d_{iu})}}{1 + e^{(a_{iG}\theta_G + a_{iS}\theta_S + d_{iu})}} \tag{5.3}$$

where $P\ (x_{iu} \geq u \mid \theta_G, \theta_S)$ is the conditional probability of selecting response category u or higher to item i given overall QWE (θ_G) and dimension of QWE (θ_S), a_{iG} and a_{iS} are item x_i slopes to latent traits θ_G and θ_S respectively, and d_{iu} is the multidimensional intercept (or easiness) for response category u or higher to item x_i. The conditional probability of selecting response category u given θ_G and θ_S is the difference between adjacent cumulative logit functions as in Equation 3.6.

The unidimensional, correlated-factors and second-order factor models are all nested within the bifactor model (Chen and Zhang 2018; Reise 2012; Yung, Thissen, and McLeod 1999). Thus, the bifactor model can be reduced to a unidimensional model by constraining all factor loadings from θ_Ss to zero, while constraining factor loadings from θ_G in the bifactor

[30] Two slopes associated with θ_G, two slopes associated with θ_{S3} and variances for θ_G and θ_{S3}.

[31] Number of known values = $p\ (p + 1) / 2$, where p is the number of items in the model. For two items, there are three known values.

model to zero and relaxing the orthogonality constraint on the θ_Ss will reduce the bifactor model to a correlated-factors model (Chen and Zhang 2018; Reise 2012). The bifactor model can be reduced to a second-order factor model by constraining the direct effects of θ_G on the observed indicators in the bifactor model to zero and introducing indirect effects of θ_G on the indicators through the θ_Ss (Reise 2012; Yung et al. 1999). Hence, the *likelihood ratio test* (LRT) or *chi-square difference* ($\Delta\chi^2$) *test* was used to compare between nested models,[32] while relative information indices; thus the *Akaike's Information Criterion* (AIC), and the *Bayesian Information Criterion* (BIC); which do not require models to be nested were also used (Finch and French 2015; Kline 2016).

Model Estimation and Diagnostics

The second random sample were used to evaluate overall model-data correspondence based on the global fit statistics for the model that better fitted the data and to predict latent trait scores using the '*mirt*' package (Chalmers 2012). Model estimates were also obtained in M*plus* 8.8 (Muthén and Muthén 2017), estimated with the EM algorithm using the robust maximum likelihood (MLR) estimator (Finch and Bolin 2017). M*plus* estimates considered the complex sample design of the UKHLS, while estimates with the '*mirt*' package (Chalmers 2012) did not consider the clustering or stratification of the sample design and this can result in biased standard error estimates (Heeringa et al. 2017). In terms of model test statistics, the signed χ^2 statistics ($S-\chi^2$) was used to assess *item fit* (Orlando and Thissen 2000, 2003; Toland 2014). The Z_h statistic was used to evaluate *person fit* (Desjardins and Bulut 2018; Morizot et al. 2007; Paek and Cole 2020), and the M_2 limited information goodness-of-fit statistic was used to examine *model fit* (Cai et al. 2006; Maydeu-Olivares and Joe 2005, 2006, 2014).

[32] The LRT and the $\Delta\chi^2$ test are fundamentally the same as the χ^2 of a model is a function of the log likelihoods (Baldwin 2019).

Along with the model test statistics, local dependence (LD) pairwise residuals between items were estimated based on the signed G^2 statistics (G^2 *LD*) to evaluate the local independence assumption (Chen and Thissen 1997; Paek and Cole 2020). Standardised G^2 *LD* statistics (signed Cramer's V coefficients) were estimated to aid interpretation for polytomous items (Chalmers 2012). For approximate fit indices, the *root mean square error of approximation* (RMSEA) along with its 90% confidence interval, *standardised root mean square residual* (SRMSR), *comparative fit index* (CFI), and *Tucker-Lewis index* (TLI) were examined using the cut-off criteria outlined in Chapter 3 (Section 3.3.6).

Finally, in addition to assessing overall model-data correspondence, the properties of the measurement instrument, including the estimated latent trait scores, was evaluated using the bifactor statistical indices; thus, the *explained common variance*, *omega reliability coefficients*, *factor determinacy*, and *construct replicability* (Reise, Moore, and Haviland 2010; Rodriguez, Reise, and Haviland 2016a, 2016b). The indices were estimated using the '*Bifactor Indices Calculator*' package (Dueber 2021) in *R* (R Core Team 2020).

The explained common variance (ECV) is the proportion of the common variance between a set of items explained by a latent trait and measures the degree of unidimensionality between those items (Reise et al. 2018; Stucky and Edelen 2015). ECV values closer to one are indicative of a single factor accounting for the common variance between a set of items (Rodriguez et al. 2016a), with values approximately ≥ 0.85 suggesting the items are sufficiently unidimensional (Stucky and Edelen 2015).

Omega reliability coefficients are model-based measures of internal consistency and indicate the reliability of multidimensional measurement instruments (Stucky and Edelen

2015). Coefficient *omega* (ω)[33] reflects the proportion of total variance in the observed total score attributable to all modelled latent traits, while coefficient *omega hierarchical* (ω_H) reflects the variance attributable to a single latent trait (Reise, Bonifay, and Haviland 2013). While there is no threshold distinguishing between reliable and unreliable instruments, high ω values (closer to one) indicate high reliability of the instrument (Baldwin 2019; Rodriguez et al. 2016a), and high ω_H values (> 0.80) suggest dominant latent traits and the instruments could be considered unidimensional (Rodriguez et al. 2016a, 2016b).

The construct replicability (*H*) index informs how well specified the latent traits are in a measurement model, while the factor determinacy (*FD*) index determines the value of estimating the latent trait scores, especially for the specific factors, and using the scores in subsequent analysis (Rodriguez et al. 2016b). The *FD* index ranges between [0, 1] and is the correlation between latent trait scores and the latent trait score estimates, with high values indicating that the estimated scores are a good representation of the latent trait, and according to Gorsuch (1983) estimated scores should be used in subsequent analysis if the *FD* index > 0.90. On the other hand, the *H* index evaluates the feasibility of specifying a measurement model from a set of observed items, with values < 0.80 indicating poorly defined latent traits which would be expected to be unstable across studies (Rodriguez et al. 2016a, 2016b).

---

[33] Coefficient ω is analogous to the commonly used measure of internal consistency coefficient *alpha* (α, or Cronbach's α), but the model-based ω does not assume tau-equivalence; thus, equal factor loadings among items; as in α (Baldwin 2019; Reise, Bonifay, and Haviland 2013).

Table 5.1: Cut-off Criteria for Bifactor Statistical Indices

Criterion	**ECV**	**ω**	**ω_H**	***H* index**	***FD* index**
Desired values	≥ 0.85	Closer to 1	< 0.80	> 0.80	> 0.90

Notes: ECV: Explained common variance. ω: Omega. ω_H: Omega hierarchical. *H* index: Construct replicability. *FD* index: Factor determinacy.

5.2 Results

5.2.1 Assessment of Data Structure

A heat map of the weighted pairwise polychoric correlation matrix using the training data random sample for indicators of QWE, ordered by the *a priori* dimensions from the QWE conceptual framework, is displayed in Figure 5.2. The correlation coefficients describe the linear association between a pair of indicators and the different colour shades represent different magnitudes of the coefficients, with darker shades indicating higher correlations.

Figure 5.2**:** Heat Map of Polychoric Correlation Matrix for Indicators of QWE

Dimension	Indicator	Effective gross pay	Pension provision	Pay bonuses	Pay progression	Progression prospects	Training prospects	Employment type	Job security	Job tasks	Work pace	Work manner	Task order	Work hours	Part-time	Term-time	Job sharing	Flexi-time	Compressed hours	Annualised hours	Home working	Other flexibility	Informal flexibility	Working times	Weekend working	Collective bargaining
		Economic Compensation				Training and Progression		Employment Security		Working Conditions					Work-time Scheduling											Social Dialogue
Economic Compensation	Effective gross pay	1																								
	Pension provision	0.36	1																							
	Pay bonuses	0.18	0.14	1																						
	Pay progression	0.24	0.45	0.01	1																					
Training and Progression	Progression prospects	-0.02	0.05	0.08	0.14	1																				
	Training prospects	0.07	0.20	0.03	0.23	0.49	1																			
Employment Security	Employment type	0.17	0.29	0.37	0.13	-0.20	-0.03	1																		
	Job security	0.03	0.02	0.06	0.10	0.05	0.07	0.36	1																	
Working Conditions	Job tasks	0.20	-0.05	0.12	0.01	0.02	0.05	0.14	0.14	1																
	Work pace	0.15	-0.05	0.12	-0.04	0.02	0.02	0.15	0.13	0.66	1															
	Work manner	0.24	-0.01	0.10	0.00	0.03	0.06	0.14	0.13	0.71	0.72	1														
	Task order	0.24	0.01	0.10	0.05	0.01	0.05	0.13	0.11	0.70	0.68	0.77	1													
	Work hours	0.31	-0.05	0.12	0.02	0.03	0.03	-0.01	0.07	0.53	0.51	0.52	0.53	1												
Work-time Scheduling	Part-time	-0.09	0.07	-0.09	0.07	0.01	0.05	-0.02	0.09	-0.04	-0.03	-0.03	0.00	0.00	1											
	Term-time	0.02	0.27	-0.21	0.20	0.03	0.11	-0.09	0.04	0.03	-0.01	0.01	0.00	0.01	0.50	1										
	Job sharing	0.24	0.36	-0.10	0.29	0.02	0.12	0.07	0.06	0.07	0.04	0.12	0.14	0.14	0.65	0.57	1									
	Flexi-time	0.22	0.21	0.10	0.15	0.14	0.16	-0.03	0.02	0.18	0.17	0.19	0.22	0.46	0.31	0.24	0.45	1								
	Compressed hours	0.35	0.33	0.08	0.21	0.15	0.12	0.01	0.06	0.16	0.14	0.17	0.24	0.39	0.45	0.43	0.59	0.66	1							
	Annualised hours	0.24	0.25	0.01	0.10	0.11	0.15	0.05	0.10	0.14	0.14	0.14	0.16	0.23	0.34	0.49	0.53	0.41	0.64	1						
	Home working	0.50	0.20	0.25	0.09	0.12	0.13	0.02	-0.02	0.31	0.27	0.35	0.36	0.56	0.20	0.18	0.37	0.59	0.60	0.45	1					
	Other flexibility	0.15	0.12	0.13	0.12	0.12	0.21	-0.04	0.05	0.19	0.16	0.20	0.18	0.30	0.15	0.18	0.26	0.34	0.39	0.37	0.41	1				
	Informal flexibility	0.18	-0.01	0.23	0.01	0.07	0.11	-0.02	0.10	0.33	0.34	0.34	0.36	0.65	0.12	0.02	0.16	0.53	0.44	0.27	0.55	0.44	1			
	Working times	0.22	0.06	0.10	0.05	-0.04	-0.07	0.16	-0.02	0.14	0.14	0.19	0.20	0.23	-0.05	0.21	0.12	0.22	0.18	-0.03	0.35	0.01	0.12	1		
	Weekend working	0.22	0.11	-0.05	0.05	-0.03	-0.02	0.02	-0.07	0.07	0.06	0.14	0.15	0.17	-0.03	0.20	0.15	0.18	0.17	0.00	0.25	-0.04	0.07	0.61	1	
Social Dialogue	Collective bargaining	0.19	0.68	-0.26	0.50	-0.01	0.11	0.08	0.03	-0.14	-0.13	-0.10	-0.10	-0.14	0.27	0.39	0.47	0.13	0.33	0.30	-0.03	-0.01	-0.15	-0.07	0.01	1

Notes: First random sample, weighted correlation coefficients and unweighted sample size, n = 7,485. Excludes missing cases.

From the polychoric correlation matrix, coefficients were generally positive suggesting that the relationship between pairs of indicators of QWE tended to move in the same direction. If the coefficients were negative, they were generally close to zero, indicating that there was no linear relationship between the pair of indicators. Furthermore, indicators within each dimension tended to be more correlated with each other than those in other dimensions, suggesting that the indicators had stronger linear relationships within than between dimensions. However, there were exceptions where indicators between dimensions had stronger linear relationships than within dimensions, suggesting that there might be associations between dimensions and / or a common factor among all the indicators.

Economic compensation indicators were weak to moderately correlated, and the *pay bonuses* item did not particularly cohere with other indicators within the dimension. This was more correlated (positively or negatively) with items in other dimensions, for example, *employment type*, *term-time*, *home working*, *informal flexibility*, and *collective bargaining*. The *effective gross pay* item was also more positively correlated with some indicators in other dimensions, particularly *home working*, than within the economic compensation dimension.

Training and progression indicators were moderately correlated, as were employment security indicators. Indicators within the training and progression dimension were more correlated with each other than with indicators in other dimensions. On the other hand, for the employment security dimension, the *employment type* item was as equally correlated to *pay bonuses* as *job security*.

Working conditions indicators were all strongly correlated within the dimension. However, the *work hours* item was comparatively not as strongly correlated with other indicators within this dimension and was more correlated with some work-time scheduling indicators (*home working* and *informal flexibility*). In terms of work-time scheduling indicators, they were weak to strongly correlated, with some near zero negative coefficients, and as noted above, some items were more strongly correlated to items in other dimensions.

Lastly, the social dialogue dimension only had one indicator, *collective bargaining*. This was weakly correlated (positively or negatively) to some indicators within the economic compensation dimension but also strongly correlated positively to other indicators within this dimension (*pension provision* and *pay progression*). The *collective bargaining* item was also weakly correlated (positively or negatively) with indicators in other dimensions, although it was moderately positively correlated with some work-time scheduling indicators (*term-time*, *job sharing*, *compressed hours* and *annualised hours*).

Overall, the polychoric correlation matrix indicated that observed indicators of QWE cohered within the hypothesised dimensions of QWE. However, associations between indicators in different dimensions suggested there might also be an association between these dimensions, or an overall factor explaining the relationship between the indicators. This suggested that the measurement model of QWE needed to consider the hypothesised dimensions of QWE, the association between these dimensions and / or a general factor of QWE.

5.2.2 Comparison of Measurement Models

For a comparison of the competing measurement models; *unidimensional*, *correlated-factors*, *second-order factor*, and *bifactor models*; using the first random sample, results of the $\Delta\chi^2$ tests evaluating the null hypotheses of equal model fit to the data between each of the nested models and the bifactor model suggested statistically significant differences (p-values < 0.001) (Table 5.2). This indicated the null hypotheses of equal model fit between the models in the UK employee population should be rejected in favour of the bifactor model. Furthermore, the *AIC* and *BIC* values for the bifactor model were lower than those of the other models, indicating the bifactor model exhibited a better fit to the data and was, therefore, retained.

Table 5.2: Model Comparisons with $\Delta\chi^2$ Tests and Relative Information Indices

Model	***AIC***	***BIC***	***LL***	*Nested Model* v *Bifactor Model*	
				$\Delta\chi^2$	**Δ *df***
Bifactor	218292	218864	–109062.8		
Correlated-factors	256616	256665	–128006.9	37888***	11
Second-order factor	256360	256832	–128112.8	38100***	16
Unidimensional	263227	263664	–131551.4	44977***	21

Notes: Graded response models using the first random sample, unweighted sample size, n = 7,485. Tests exclude missing cases. *LL*: Log likelihood. *LRT* or $\Delta\chi^2 = -2\ (LL_N - LL_B)$, subscripts *N* and *B* represent nested and bifactor models, respectively. * $p <$ 0.05. ** $p < 0.01$. *** $p < 0.001$.

5.2.3 Graded Response Bifactor IRT Model

As the model was a confirmatory bifactor model, each indicator is associated with the *overall QWE* (θ_G) latent trait and the latent trait for one other *dimension of QWE* (θ_S), with other loadings constrained to zero. *Overall QWE* and each *dimension of QWE* were modelled to have a linear combination of latent trait coordinates, thus *compensatory model*, with the probability of responding in a particular response category to an item being the sum of the probabilities for each. Parameter estimates of the model from *R* and M*plus* using the second random sample are reported in both the IRT slope-intercept parameterisation (Tables 5.3 and 5.4) and factor-analytic metrics limited to standardised factor loadings (Appendix 5.1). To aid interpretation for the slope-intercept parameterisation, the estimates were transformed into *multidimensional discrimination* (A_i) (Equation 3.9) and *multidimensional difficulty* (B_i) (Equation 3.8) indices and presented in Appendix 5.2.

Table 5.3**:** Graded Response Bifactor IRT Model Slope Item Parameter Estimates

Item	**Conditional slopes**											
	R						M*plus*					
	a_{G}	a_{S1}	a_{S2}	a_{S3}	a_{S4}	a_{S5}	a_{G}	a_{S1}	a_{S2}	a_{S3}	a_{S4}	a_{S5}
Effective gross pay	0.820*** (0.033)	0.686*** (0.045)					0.827*** (0.043)	0.692*** (0.050)				
Pension provision	0.611* (0.242)	4.967* (2.046)					0.559*** (0.162)	4.090*** (0.838)				
Pay bonuses	0.363*** (0.033)	0.265*** (0.044)					0.364*** (0.041)	0.262*** (0.047)				
Pay progression	0.243*** (0.035)	0.893*** (0.064)					0.252*** (0.044)	0.899*** (0.069)				
Progression prospects	0.281*** (0.058)		1.764*** (0.071)				0.283*** (0.067)		1.727*** (0.088)			
Training prospects	0.350*** (0.046)		1.764*** (0.071)				0.354*** (0.057)		1.727*** (0.088)			
Employment type	0.280*** (0.069)			1.511*** (0.066)			0.283** (0.091)			1.421*** (0.106)		
Job security	0.178*** (0.038)			1.511*** (0.066)			0.185*** (0.050)			1.421*** (0.106)		
Job tasks	1.265*** (0.053)				2.228*** (0.058)		1.304*** (0.070)				2.208*** (0.075)	
Work pace	1.227*** (0.054)				2.375*** (0.061)		1.272*** (0.076)				2.367*** (0.085)	
Work manner	1.946*** (0.089)				3.901*** (0.127)		2.006*** (0.132)				3.841*** (0.181)	
Task order	1.560*** (0.061)				2.484*** (0.064)		1.604*** (0.082)				2.459*** (0.087)	
Work hours	2.105*** (0.067)				0.743*** (0.040)		2.113*** (0.090)				0.725*** (0.053)	

Continued…

Continued…

Item	**Conditional slopes**											
	R						M*plus*					
	a_G	a_{S1}	a_{S2}	a_{S3}	a_{S4}	a_{S5}	a_G	a_{S1}	a_{S2}	a_{S3}	a_{S4}	a_{S5}
Part-time	0.111* (0.053)					2.053*** (0.127)	0.127 (0.067)					2.014*** (0.142)
Term-time	0.040 (0.058)					1.819*** (0.090)	0.056 (0.074)					1.831*** (0.109)
Job sharing	0.586*** (0.072)					2.537*** (0.146)	0.612*** (0.096)					2.500*** (0.154)
Flexi-time	1.526*** (0.060)					0.941*** (0.053)	1.542*** (0.079)					0.936*** (0.071)
Compressed hours	1.574*** (0.081)					1.784*** (0.065)	1.584*** (0.111)					1.659*** (0.111)
Annualised hours	0.853*** (0.085)					1.541*** (0.092)	0.863*** (0.112)					1.539*** (0.129)
Home working	2.518*** (0.121)					0.754*** (0.066)	2.523*** (0.163)					0.743*** (0.089)
Other flexibility	0.885*** (0.049)					0.316*** (0.043)	0.885*** (0.059)					0.317*** (0.059)
Informal flexibility	1.816*** (0.064)					0.182*** (0.046)	1.808*** (0.083)					0.173** (0.058)
Working times	0.434*** (0.036)					0.034 (0.038)	0.439*** (0.042)					0.035 (0.046)

Notes: Second random sample, unweighted sample size, n = 8,490. a_G = conditional slope for *overall QWE* latent trait. a_{S1} – a_{S5} are conditional slopes for *dimensions of QWE* latent traits and estimates in parentheses are standard errors. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$. M*plus* estimates consider the complex sample design of the UKHLS data; thus, stratification, clustering, and probability sampling weights, while *R* estimates only account for the probability sampling weights.

Conditional Slopes and Multidimensional Discrimination Indices

For the conditional slopes of the items (Tables 5.3), estimates from *R* and M*plus* were comparable. However, the slope for θ_G conditional on θ_{S5} for part-time working was not statistically significant in the UK employee population based on the M*plus* estimates, while this was statistically significant based on the estimates from *R*. This might be due to M*plus* estimates accounting for the complex sample design of the UKHLS, while *R* estimates only accounted for the probability sampling weights. Outcomes of the tests of statistical significance for all other items in the model were consistent between *R* and M*plus*, and all items had statistically significant conditional slopes on θ_G and their θ_S, except for term-time and working times; thus *p*-values > 0.05 for the conditional slopes of θ_G given θ_{S5}, and θ_{S5} over and above θ_G, respectively; in addition to part-time based on the M*plus* estimates.

The statistically significant conditional slopes indicated non-zero estimates in the UK employee population, and substantively, these reflect the strength of association between the latent traits and the probability of responding positively (dichotomous items) or in a particular response category (polytomous items) to an item or how discriminating an item is for the latent traits (Bonifay 2020; Cai et al. 2011). Thus, for the bifactor model, the greater the conditional slope associated with one latent trait compared to another for an item, the greater the influence of that latent trait in the probability of responding positively or in a particular response category to the item.

Based on the M*plus* estimates and considering the economic compensation indicators, the conditional slopes for pension provision and pay progression items were greater for θ_{S1} over and above θ_G (4.09 and 0.90, respectively) than for θ_G given θ_{S1} (0.56 and 0.25, respectively). This suggested that the *economic compensation* latent trait had a greater influence in the probability of responding positively to these items than the *overall QWE* latent trait. This can be visualised in an item response surface (IRS) and for illustration, Figure 5.3 (a) displays the

IRS in the form of a contour plot for pay progression with the latent trait on the *x*-axis representing *overall QWE* and that on the *y*-axis representing *economic compensation*. As this item was dichotomously scored, the contour lines represent the probability of respondents reporting that their pay included annual increments. The rate of change of the probabilities was faster along the direction of the *y*-axis than the *x*-axis indicating that the pay progression item derived its discrimination, largely, along the direction of *economic compensation* than *overall QWE*. For effective gross pay and pay bonuses items, conditional slopes were slightly greater for θ_G given θ_{S1} (0.83 and 0.36, respectively) than θ_{S1} over and above θ_G (0.69 and 0.26, respectively), suggesting that they derived slightly more of their discrimination from *overall QWE* than *economic compensation*.

Figure 5.3: Item Response Contour Plots

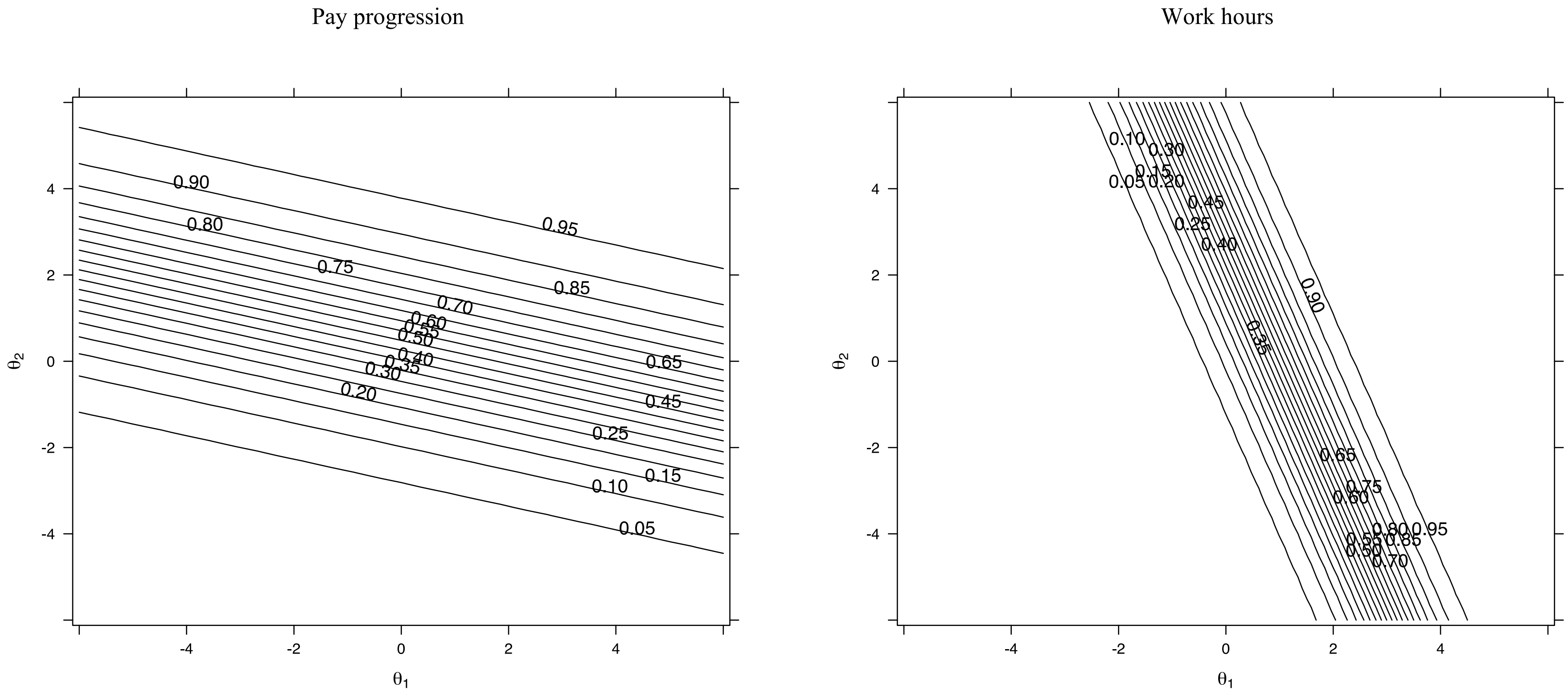


Figure 5.3 (a) Figure 5.3 (b)

Notes: The contour plots, with points along a straight line indicating combinations of (θ_1, θ_2) with an equal probability of selecting the positive response. The straight equiprobable contours demonstrate the compensatory nature of the model, with high levels of θ_2 compensating for low levels of θ_1 and vice versa, resulting in high probabilities of a positive response to the item. Furthermore, the directional impact of the slope associated with the IRS for MIRT models is apparent with the rate of change of the probability of a positive response greater along the direction of θ_2 than θ_1 from the point of origin (0, 0), indicating that θ_2 was more influential in responding to this item than θ_1 (Figure 5.3 (a)). This is in contrast to Figure 5.3 (b).

For training and progression, and employment security dimensions conditional slopes for items within each dimension were constrained to be equal for model identification purposes. Thus, in a bifactor model, all the latent traits were assumed to be mutually orthogonal, and follow standard normal distributions with zero mean and unit variance (Cai et al. 2011; Reise 2012). Where specific factors were defined by two items, slopes within these factors were constrained to be equal for the degrees of freedom to be zero and achieve model identification (Cai et al. 2011). In terms of indicators of training and progression, conditional slopes for progression prospects and training prospects items were greater for θ_{S2} over and above θ_G (both 1.73) than for θ_G given θ_{S2} (0.28 and 0.35, respectively). Similarly, for employment security indicators, conditional slopes for employment type and job security items were greater for θ_{S3} over and above θ_G (both 1.42) than for θ_G given θ_{S3} (0.28 and 0.19, respectively). This suggested that the items largely derived their discrimination from the *training and progression* or *employment security* latent traits than *overall QWE*.

Regarding working conditions indicators, conditional slopes for job tasks, work pace, work manner, and task order were greater for θ_{S4} over and above θ_G than for θ_G given θ_{S4}, while for work hours the conditional slope was greater for θ_G given θ_{S4} (2.11) than for θ_{S4} over and above θ_G (0.73). This indicated that while the *working conditions* latent trait had a greater influence in the probability of responding in a particular response category for job tasks, work pace, work manner, and task order than the *overall QWE* latent trait, the work hours item largely derived its discrimination from the *overall QWE* latent trait than *working conditions*.

In terms of the work-time scheduling indicators, conditional slopes for part-time, term-time, job sharing, compressed hours, and annualised hours were greater for θ_{S5} over and above θ_G than for θ_G given θ_{S5}. This suggested that these items derived their discrimination from the *work-time scheduling* latent trait than *overall QWE*. In the case of part-time and term-time items, the discrimination was entirely attributable to the *work-time scheduling* latent trait as the conditional slopes associated with the *overall QWE* latent trait were not statistically significant in the UK employee population. On the other hand, conditional slopes for flexi-time, home working, other flexibility, informal flexibility, and working times were greater for θ_G given θ_{S5} than for θ_{S5} over and above θ_G. This indicated that the *overall QWE* latent trait had a greater influence in the probability of responding positively or in a particular response category for these items than *work-time scheduling*. For working times item, the discrimination was entirely attributable to the *overall QWE* latent trait.

The conditional slopes were transformed into multidimensional discrimination indices (A_i) which, conceptually, are analogous to the discrimination parameter in unidimensional IRT models, however, they relate to a particular direction from the origin of the latent trait space. Thus, the A_i indicate the extent to which items discriminate or differentiate between employees with low and high levels of the latent traits around the point of the steepest slope in a particular direction, and the higher the value, the more discriminating the item (Bonifay 2020; Desjardins and Bulut 2018; Reckase 2009). From Appendix 5.2, pension provision (A_i = 5.00) and work manner (A_i = 4.36) were among the most discriminating items along the direction of the *economic compensation* and *work-time scheduling* latent traits, respectively, than *overall QWE*. This suggested they differentiated between employees with low and high levels of the latent traits well compared to other items. On the other hand, pay bonuses (A_i = 0.45) and working times (A_i = 0.44) items were among the least discriminating items both along the direction of the *overall QWE* latent trait than *economic compensation* and *work-time scheduling*, respectively. This indicated that they did not differentiate between employees with low and high levels of the latent traits well compared to other items. This is important in relation to the precision of an item in estimating the latent traits. In principle, items with higher A_i values provide more information about the latent traits, particularly along the direction of the steepest slope, than those with lower values (Bonifay 2020; Reckase 2009).

Multidimensional Intercepts and Difficulty Indices

In terms of multidimensional intercepts of the items (Tables 5.4), estimates from *R* and M*plus* were comparable and the outcomes of the tests of statistical significance were also consistent. All the estimates were statistically significant in the UK employee population, except for the intercept associated with responding 'yes' to the informal flexibility item, indicating that this intercept was not different from zero in the UK employee population. Substantively, the multidimensional intercepts indicate the easiness of responding positively to an item (dichotomous items) or in a particular response category (polytomous items), with higher values suggesting easier items or response categories in relation to the origin of the latent trait space.[34] Notably, for polytomous items, the multidimensional intercept estimates decrease with increasing categories indicating that higher categories are less easier to respond positively to than lower categories.

[34] The origin of the (θ_1, θ_2)-plane, (0, 0), represents the average difficulty relative to both latent traits.

Table 5.4: Graded Response Bifactor IRT Model Intercept Item Parameter Estimates

Item	Multidimensional intercepts							
	R				M*plus*			
	d_1	d_2	d_3	d_4	d_1	d_2	d_3	d_4
Effective gross pay	2.047*** (0.044)	0.664*** (0.032)	–0.417*** (0.031)	–1.673*** (0.039)	2.047*** (0.052)	0.663*** (0.039)	–0.419*** (0.036)	–1.676*** (0.046)
Pension provision	5.391** (1.971)				4.537*** (0.768)			
Pay bonuses	–0.979*** (0.029)				–0.979*** (0.035)			
Pay progression	–0.433*** (0.030)				–0.435*** (0.035)			
Progression prospects	–2.625*** (0.071)				–2.598*** (0.085)			
Training prospects	–0.609*** (0.041)				–0.606*** (0.048)			
Employment type	3.610*** (0.085)				3.512*** (0.119)			
Job security	4.950*** (0.126)	3.429*** (0.087)	0.161*** (0.035)		4.826*** (0.159)	3.336*** (0.112)	0.153*** (0.041)	
Job tasks	3.685*** (0.071)	1.969*** (0.055)	–0.807*** (0.048)		3.674*** (0.100)	1.961*** (0.069)	–0.810*** (0.055)	
Work pace	3.864*** (0.071)	2.163*** (0.057)	–0.490*** (0.048)		3.865*** (0.115)	2.162*** (0.080)	–0.495*** (0.058)	
Work manner	7.244*** (0.192)	4.399*** (0.131)	0.452*** (0.073)		7.179*** (0.289)	4.360*** (0.186)	0.440*** (0.084)	
Task order	5.059*** (0.100)	3.128*** (0.073)	0.278*** (0.051)		5.041*** (0.133)	3.117*** (0.093)	0.272*** (0.063)	
Work hours	1.093*** (0.048)	–0.349*** (0.042)	–2.056*** (0.056)		1.084*** (0.055)	–0.351*** (0.052)	–2.051*** (0.071)	
Part-time	0.579*** (0.046)				0.572*** (0.053)			
Term-time	–2.393*** (0.076)				–2.397*** (0.090)			
Job sharing	–2.918*** (0.122)				–2.887*** (0.135)			
Flexi-time	–1.184*** (0.044)				–1.186*** (0.055)			
Compressed hours	–3.320*** (0.097)				–3.315*** (0.127)			
Annualised hours	–4.105*** (0.105)				–4.096*** (0.179)			
Home working	–3.153*** (0.114)				–3.143*** (0.151)			
Other flexibility	–1.961*** (0.044)				–1.958*** (0.054)			
Informal flexibility	1.046*** (0.043)	0.060 (0.038)			1.038*** (0.050)	0.058 (0.045)		
Working times	1.094*** (0.029)				1.094*** (0.036)			

Notes: Second random sample, unweighted sample size, $n = 8{,}490$. $d_1 – d_4$ are thresholds between response categories and estimates in parentheses are standard errors. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$. M*plus* estimates consider the complex sample design of UKHLS data, while *R* estimates only consider the probability sampling weights.

Figure 5.4: Item Response Contour Plots

Employment type

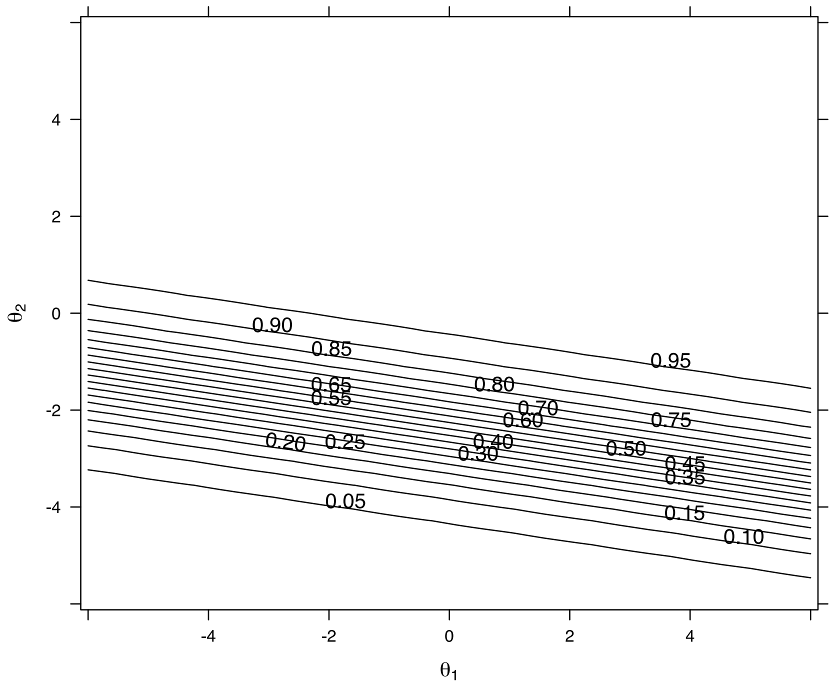


Figure 5.4 (a)

Other flexibility

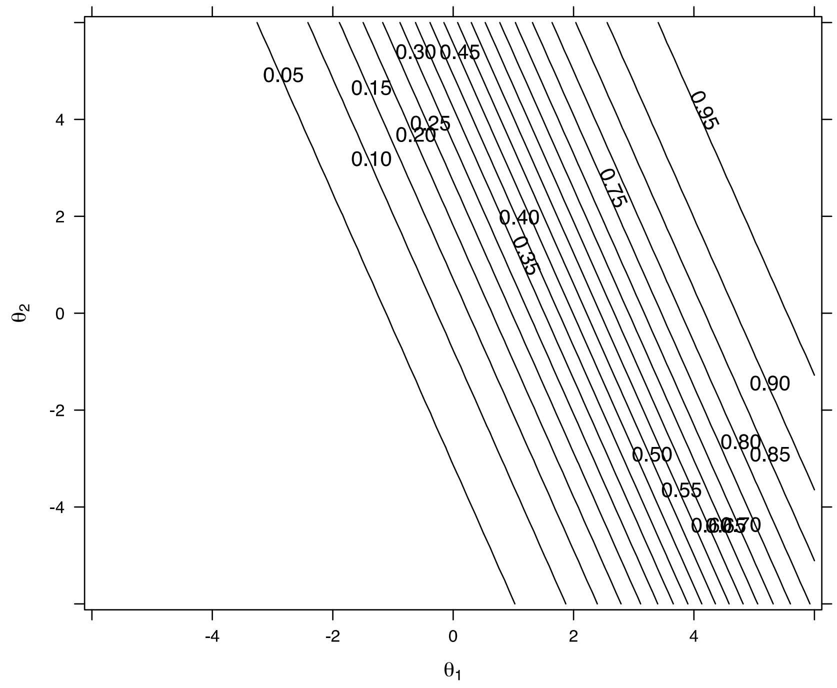


Figure 5.4 (b)

Notes: See notes in Figure 5.3.

The multidimensional intercepts were transformed into multidimensional difficulty indices (B_i), and these are interpreted similarly to difficulty parameters for unidimensional models but relate to a specific direction. Thus, higher values of the B_i indicate that items require higher levels of the latent traits for a respondent to have a probability > 0.5 of selecting a positive or particular response category along the direction of the steepest slope from the origin. From Appendix 5.2, annualised hours (B_{i1} = 2.33), pay bonuses (B_{i1} = 2.18), and other flexibility (B_{i1} = 2.09) had among the highest multidimensional difficult indices in the direction of their steepest slope. This indicated that UK employees required higher levels of the associated latent traits to have a high probability of responding positively to the items in the direction of the steepest slope from the origin. In contrast, employment type (B_{i1} = –2.35), working times (B_{i1} = –2.51), and job security (B_{i1} = –3.25), had among the lowest multidimensional difficult indices in the direction of their steepest slope. This indicated that UK employees required lower levels of the associated latent traits to have a high probability of responding positively or in a particular category to the items in the direction of the steepest slope from the origin. This can be observed in the location of the contour lines of the IRSs in Figure 5.4, with the probabilities > 0.5 located below the point of origin for the employment type item (Figure 5.4 (a)) and located above the point of origin for the other flexibility item (Figure 5.4 (b)).

5.2.4 Model Diagnostics

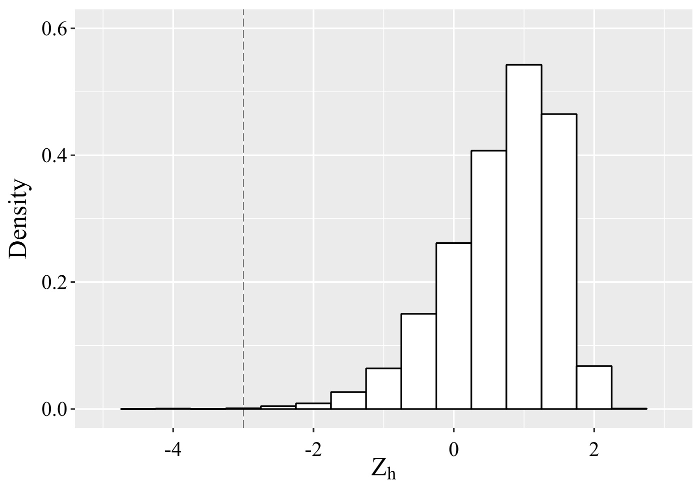
Item Fit

The $S - \chi^2$ statistics and associated *p*-values for item fit suggested no statistically significant difference between model-predicted and observed response proportions by item response category for five indicators; *pension provision, pay progression, employment type, compressed hours*, and *other flexibility* (*p*-values > 0.05); but differences were statistically significant for other indicators (*p*-values < 0.05). This test is, however, sensitive to sample size and RMSEA indices for all indicators were < 0.02, suggesting response proportions predicted by the bifactor IRT model closely fitted observed response proportions for the indicators (Appendix 5.3).

Person Fit

In terms of person fit, Figure 5.5 displays a histogram of the distribution of Z_h statistics with a vertical line at –3, below which respondents' response patterns were aberrant. A small number of cases, 4 (0.05%), had a Z_h statistic < –3, suggesting overall response patterns for respondents aligned with item parameters estimated by the bifactor IRT model.

Figure 5.5: Distribution of Z_h Statistic for Graded Response Bifactor IRT Model



Notes: Second random sample, unweighted sample size, n = 7,508. Test excludes missing cases. Values of Z_h < –3 indicate aberrant response patterns.

Model Fit

Global fit statistics for the bifactor model, as well as those for models nested in the bifactor model for reference, are presented in Table 5.5. The overall model fit test indicated the bifactor model did not fit the data, this was statistically significant (M_2 (193) = 2175.14, $p <$ 0.001) and the null hypothesis of no difference between observed and model-predicted estimates was rejected, but the test is also sensitive to sample size. Approximate fit indices, overall, suggested adequate model fit, the RMSEA was 0.037 (90% CI [0.036, 0.038]) indicating close fit (upper 90% CI was < 0.05), and the SRMSR was < 0.05 suggesting adequate fit. The CFI (0.953) was > 0.95, but the TFI (0.946) was slightly < 0.95, however this represents acceptable model fit.

Table 5.5: Global Fit Statistics for Graded Response Bifactor IRT Model

Model	M_2	*df*	**RMSEA** [90% CI]	**SRMSR**	**CFI**	**TLI**
Bifactor	2175.14***	193	0.037 [0.036, 0.038]	0.039	0.956	0.946
Correlated factors	7505.45***	204	0.069 [0.068, 0.070]	0.092	0.838	0.812
Second-order factor	7643.80***	209	0.069 [0.068, 0.070]	0.073	0.835	0.813
Unidimensional	8719.31***	214	0.073 [0.072, 0.074]	0.091	0.811	0.791

Notes: Second random sample, unweighted sample size, n = 7,508. Tests exclude missing cases. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

The standardised G^2 *LD* pairwise residuals between the items based on the bifactor model indicated absolute values < 0.15 and were distributed around zero (Appendix 5.5). These were | < 0.20 |, and according to Morizot et al. (2007) suggested the local independence assumption was tenable given *overall QWE* and *dimensions of QWE* latent traits. Considered together, the statistical evidence from the model diagnostics suggested the graded response bifactor IRT model adequately fitted the data, thus it performed well in accounting for the common variance shared by the indicators.

5.2.5 Predicting Latent Trait Scores

Based on the graded response bifactor IRT, latent trait scores and their associated standard errors were predicted for the UK employee population using the *expected a posterior* (EAP) estimator (Brown and Croudace 2015). As part of model specification for the bifactor model, all the latent traits were assumed to be mutually orthogonal, and follow standard normal distributions with zero mean and unit variance (Cai et al. 2011; Reise 2012). Where specific factors were defined by two items, slopes within these factors were constrained to be equal for the degrees of freedom to be zero and achieve model identification (Cai et al. 2011) (see Section 5.1.3). With the distributions of the latent traits assumed, latent trait scores and their associated standard errors were estimated for each employee for *overall QWE,* and the five *dimensions of QWE* in the measurement model. These along with the item response patterns for a sample of employees are shown in Appendix 5.6. The distributions of the scores are displayed in violin and box plots in Figure 5.6, while scatterplots of the latent trait scores for each employee and their associated standard errors are shown in Figure 5.7.

Figure 5.6: Distributions of Latent Trait Scores for All Employees

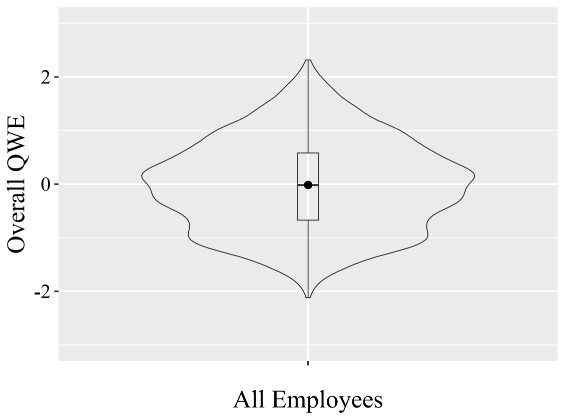


Figure 5.6 (a)

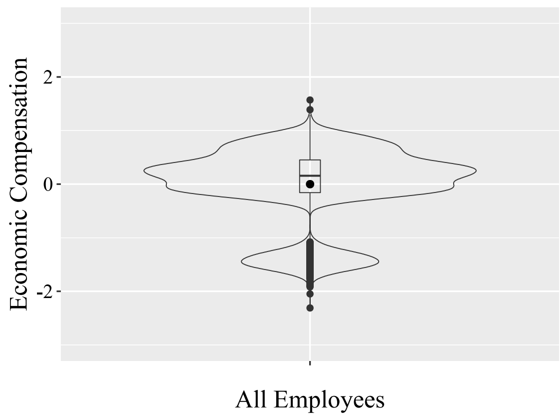


Figure 5.6 (b)

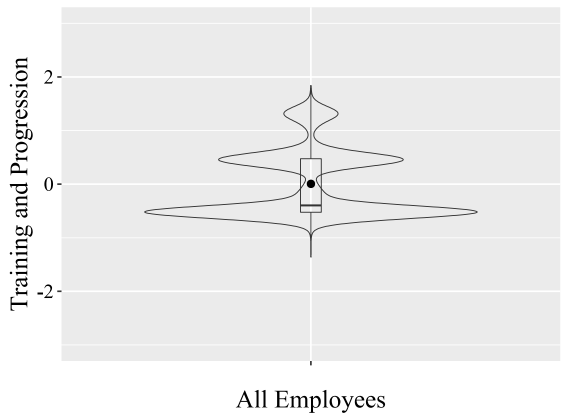


Figure 5.6 (c)

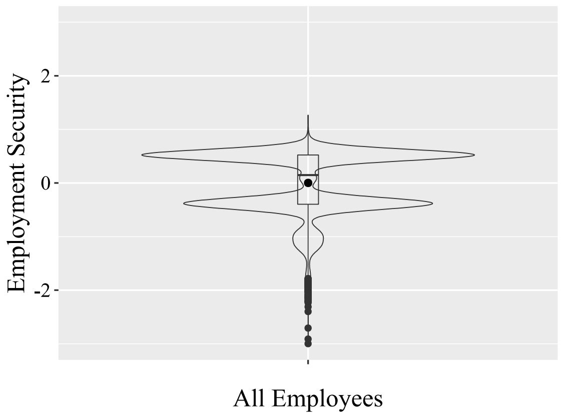


Figure 5.6 (d)

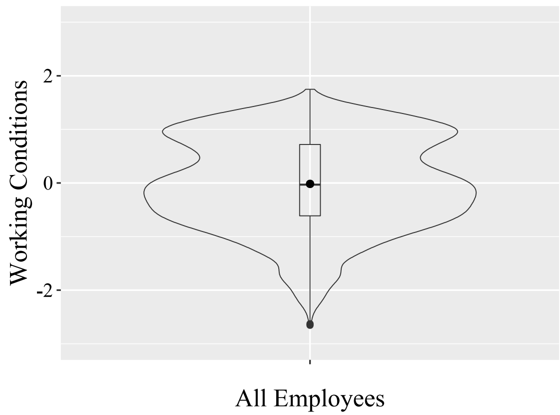


Figure 5.6 (e)

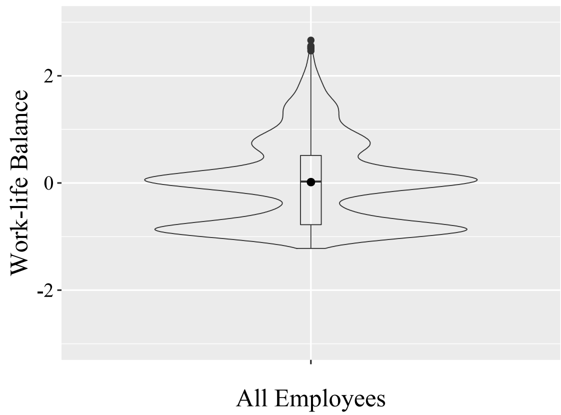


Figure 5.6 (f)

Notes: Second random sample, unweighted sample size, n = 8,490. Distribution of *overall QWE* latent trait scores are conditional on scores of other *dimensions of QWE* in the measurement model, while distributions of latent trait scores for each *dimension of QWE* are conditional on scores for *overall QWE*. Dot within each boxplot represent mean scores.

Figure 5.7: Scatterplot of Latent Trait Scores against Standard Errors for All Employees

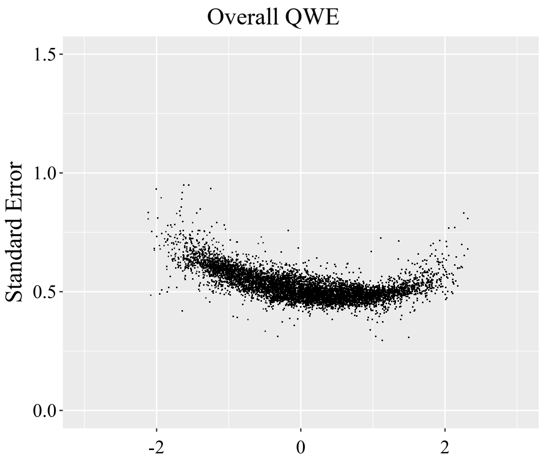


Figure 5.7 (a)

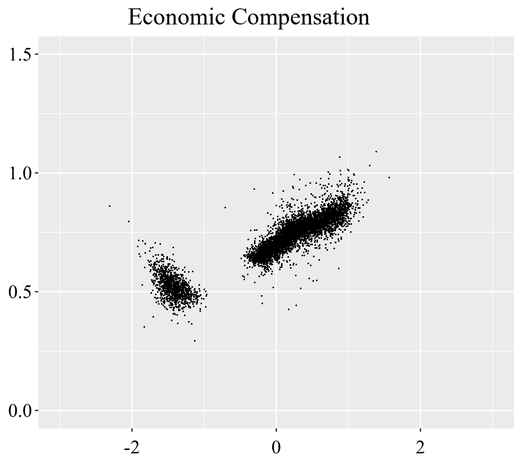


Figure 5.7 (b)

Figure 5.7 (c)

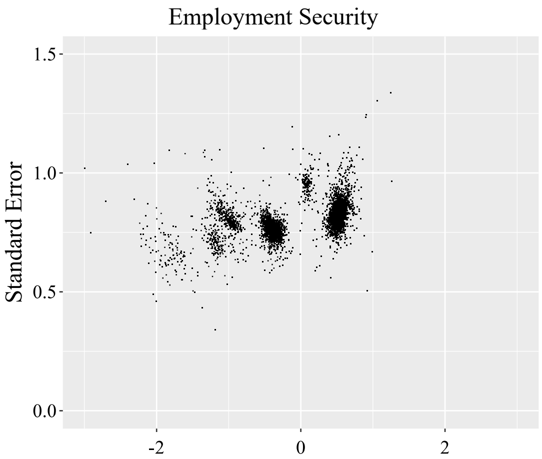


Figure 5.7 (d)

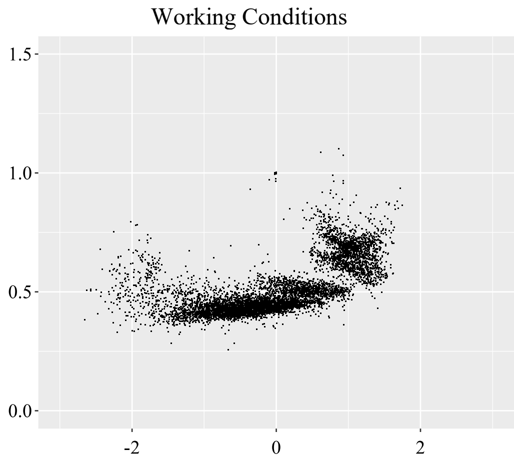


Figure 5.7 (e)

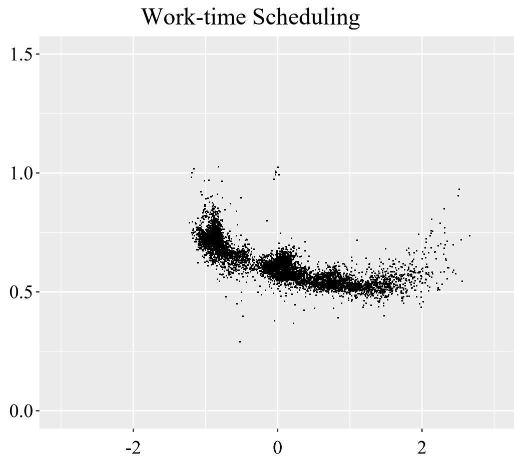


Figure 5.7 (f)

Notes: Second random sample, unweighted sample size, n = 8,490. Standard errors for *overall QWE* latent trait scores are conditional on latent trait scores for the other *dimensions of QWE* in the measurement model, while standard errors for latent trait scores for each *dimension of QWE* are conditional on scores for *overall QWE*.

For *overall QWE*, the distribution of the latent trait scores was based on the response patterns to 23 items in the measurement model. Given other *dimensions of QWE* in the measurement model, response patterns associated with slightly above average levels of *overall QWE* were the most common, while those associated with very high or low levels of *overall QWE* were not common (Figure 5.6 (a)). Considering the scatterplot in Figure 5.7 (a) there was little variation in standard errors of the *overall QWE* latent trait scores, given other *dimensions of QWE*, between approximately [–1.5, 1.5] of the latent trait continuum but more variation at the extreme ends. This suggested there was more uncertainty in the estimates of the latent trait scores at the extreme ends of the continuum and less so in the middle of the continuum.

The distribution of the latent trait scores for *economic compensation* was based on four items in the measurement model, over and above *overall QWE*. Response patterns associated with slightly above average levels of *economic compensation* were the most common, while there was also a cluster of response patterns associated with low levels of *economic compensation* (Figure 5.6 (b)). The scatterplot in Figure 5.7 (b) highlighted clustering of scores (gap in data points) along the *economic compensation* latent trait continuum. Conditional on *overall QWE*, the scatterplot indicated little variation in standard errors at low levels of the *economic compensation* latent trait continuum but increased with increasing latent trait scores from above average. This indicated that estimates of the *economic compensation* scores were more uncertain at higher end of the continuum.

For *training and progression*, the distribution of the latent trait scores was based on the response patterns to two items in the measurement model, over and above *overall QWE*. However, the slopes for these items on the latent trait were constrained to be equal to achieve model identification, and this meant these items were of equal importance in measuring *training and progression*. The response pattern associated with the lowest levels of *training and progression*, approximately half a standard deviation below the mean, was the most common. On the other hand, that associated with the highest levels, approximately 1.5 standard deviations above the mean, was the least common (Figure 5.6 (c)). The scatterplot for the *training and progression* scores and associated standard errors, over and above *overall QWE*, indicated standard errors that were more spread out along the range of the latent trait continuum (Figure 5.7 (c)).[35] This indicated that estimates of the *training and progression* scores had uncertainty across the range of the continuum.

Similarly to *training and progression*, the distribution of the *employment security* latent trait scores was based on the response patterns to two items in the measurement model, over and above *overall QWE*. The slopes for these items on the latent trait were also constrained to be equal to achieve model identification. The response pattern associated with the highest levels of *employment security*, approximately half a standard deviation above the mean, was the most common, while response patterns associated with lower levels were less common (Figure 5.6 (d)). The scatterplot for the *employment security* scores and associated standard errors, conditional on *overall QWE*, suggested that the standard errors were more spread out along the range of the latent trait continuum (Figure 5.7 (d)). This suggested that estimates of the *employment security* scores were uncertain across the range of the continuum.

The distribution of the latent trait scores for *working conditions* was based on five items in the measurement model, over and above *overall QWE*. Response patterns associated with approximately average levels of *working conditions* were the most common, while those associated with very low levels were not common (Figure 5.6 (e)). In terms of the scatterplot in Figure 5.7 (e) there was little variation in standard errors of the *working conditions* latent trait scores, over and above *overall QWE*, between approximately [–1.5, 0.5] of the latent trait

[35] This latent trait was measured by two items; however, there were more data points in the scatterplot than response patterns to the items as estimates of the latent trait scores and standard errors were conditional on *overall QWE*, which varied between respondents. The data points were clustered around corresponding response patterns ((0,0), (0,1), (1,0), and (1,1)) but the estimates varied conditional on *overall QWE*.

continuum, but there was more variation at the extreme ends. This suggested there was more uncertainty in the estimates of the latent trait scores at the extreme ends of the continuum than in the middle of the continuum.

Lastly, for *work-time scheduling*, the distribution of the latent trait scores was based on the response patterns to 10 items in the measurement model, conditional on *overall QWE*. Response patterns associated with approximately average levels of *work-time scheduling*, as well as approximately a standard deviation below the mean, were among the most common, while those associated with very high levels were not common (Figure 5.6 (f)). Finally, the scatterplot for the *work-time scheduling* scores and associated standard errors, over and above *overall QWE*, indicated standard errors had little variation approximately between [–0.5, 1.5] along the latent trait continuum but had more variation at the extreme ends of the continuum (Figure 5.7 (f)). This suggested that estimates of the *work-time scheduling* scores were uncertain at the extreme ends of the continuum.

5.2.6 Evaluation of Psychometric Properties of the Measurement Instrument

The properties of the measurement instruments were evaluated using the bifactor statistical indices based on the standardised factor loadings estimated by the bifactor GRM. The indices are shown in Table 5.7.

Table 5.6: Statistical Indices from Graded Response Bifactor IRT Model

Latent Trait	Bifactor Statistical Indices				
	ECV	ω	ω_H	*H* index	*FD* index
Overall QWE (θ_G)	0.317	0.900	0.552	0.875	0.925
Economic compensation (θ_{S1})	0.102	0.631	0.524	0.889	0.947
Training and progression (θ_{S2})	0.084	0.689	0.668	0.675	0.825
Employment security (θ_{S3})	0.072	0.617	0.603	0.607	0.781
Working conditions (θ_{S4})	0.197	0.926	0.602	0.855	0.937
Work-time scheduling (θ_{S5})	0.228	0.877	0.481	0.857	0.923

Notes: Second random sample, unweighted sample size, n = 8,490. ECV: Explained common variance. ω: Omega. ω_H: Omega hierarchical. *H* index: Construct replicability. *FD* index: Factor determinacy.

From the ECV indices, the *overall QWE* latent trait explained 32% (ECV = 0.32) of the common variance among the indicators of QWE. Since this is < 0.85, it suggests that a one factor solution did not sufficiently account for the common variance among the items. In terms of *dimensions of QWE*, *economic compensation* explained 10% (ECV = 0.10), *training and progression* explained 8% (ECV = 0.08), *employment security* explained 7% (ECV = 0.07), *working conditions* explained 20% (ECV = 0.20), and *work-time scheduling* explained 23% (ECV = 0.23) of the common variance among the items over and above *overall QWE*.

In terms of *omega* (ω) reliability coefficients, coefficients for *overall QWE* (ω = 0.90), *working conditions* (ω = 0.93), and *work-time scheduling* (ω = 0.88) suggested the instruments had high internal consistency reliability, while coefficients for *economic compensation* (ω = 0.63), *training and progression* (ω = 0.68), and *employment security* (ω = 0.61) suggested the instruments had acceptable internal consistency reliability. The *omega hierarchical* (ω_H) coefficients for *overall QWE* and *dimensions of QWE* were all < 0.80, beyond which reliable common variance between the items could be attributable to a single latent trait.

Relating to the *H* index, values for *overall QWE*, *economic compensation*, *working conditions*, and *work-time scheduling* latent traits were > 0.80 suggesting these latent traits were well defined and likely to be replicable across studies. *H* values for *training and progression* and *employment security* latent traits were < 0.80 suggesting these latent traits were poorly defined and expected to be unstable across studies. Lastly, values for the *FD* index for *overall QWE*, *economic compensation*, *working conditions*, and *work-time scheduling* latent traits were > 0.90, indicating that their estimated scores were a good representation of the latent trait. On the other hand, values for the *FD* index for *training and progression* and *employment security* were < 0.90, suggesting that estimated scores were not a good representation of the latent traits and could, therefore, not be used in subsequent analysis.

5.3 Discussion

There is agreement within the labour market research literature that, conceptually, QWE is a multidimensional concept, however there is no consensus on how the concept should be measured, as evidenced by the numerous indices developed to measure QWE (Muñoz de Bustillo et al. 2011b). These vary in terms of observed items used, the aggregation of those items, including whether to report scores for overall QWE and/or dimensions of QWE. In this chapter, IRT modelling was used to develop a measurement instrument of QWE, based on *a priori* conceptual framework, and four measurement models; *unidimensional*, *correlated-factors*, *second-order factor*, and *bifactor models*; were considered. Results suggested that a bifactor model exhibited a better fit to the data compared to other models.

The bifactor model has some advantages in modelling QWE compared to the other models. While the unidimensional model offers a useful representation of the overall or general factor, a unidimensional measure is clearly not representative of the concept of QWE as it does not consider the multidimensional nature of QWE. The correlated-factors model considers the multidimensional nature of QWE and would be a useful representation if interest was only in the dimensions or specific factors as it does not consider a general factor. On the other hand, both the second-order and bifactor models are useful representations if the interest is in the general and specific factors (Brown and Croudace 2015), as is the case in the research of QWE. Considering the second-order and bifactor models, the orthogonal nature of *overall QWE* and *dimensions of QWE* modelled by the bifactor model meant that the bifactor model had several advantages over the second-order factor model. Firstly, the *dimensions of QWE* in a bifactor model can be investigated independently of *overall QWE*, whereas in the second-order factor model, the dimensions are dependent on *overall QWE*. Secondly, the relationship between *overall QWE* or the *dimensions of QWE* and the observed items can be directly examined in a bifactor model, while in a second-order factor model there is no direct relationship between *overall QWE* and the observed items. Thirdly, the latent trait mean differences for both *overall QWE* and *dimensions of QWE* can be compared between groups in a bifactor model, subject to adequate measurement equivalence, but only the latent trait mean differences of *overall QWE* between groups can be compared in a second-order factor model (Chen et al. 2012).

A challenge, however, with bifactor models is the interpretation of scores for orthogonal latent traits and the within-item multidimensional latent structure (Bonifay 2020). Thus, the general factor (*overall QWE*) is interpreted conditional on all the specific factors (*dimensions of QWE*) in the model, while specific factors are interpreted conditional on the general factor (alternatively, over and above the general factor) (Bonifay 2020; Brown and Croudace 2015; Chen et al. 2012). Furthermore, bifactor models seemingly have a tendency to exhibit superior model fit which can lead to conclusions that are not generalisable to other scenarios (Bonifay and Cai 2017). Therefore, evaluating the quality or properties of the measurement instruments developed from the model is essential (Rodriguez et al. 2016b).

An evaluation of the properties of the measurement instrument based on the bifactor model supported the multidimensional latent structure of QWE. Thus, the ECV indices indicated that a one factor solution did not sufficiently account for the common variance among the indicators of QWE. Additionally, *omega* reliability coefficients suggested the *overall QWE*, *working conditions*, and *work-time scheduling* latent traits had high internal consistency reliability, while *economic compensation*, *training and progression*, and *employment security* latent traits had acceptable internal consistency reliability. The *H* index suggested that *overall QWE*, *economic compensation*, *working conditions*, and *work-time scheduling* latent traits were well defined, and the *FD* index indicated that estimated scores were a good representation of the latent traits. However, the *H* and *FD* indices for *training and progression*, and *employment security* latent traits suggested they were not well defined, and their scores should not be used in subsequent analysis. This can be attributed to that the *H* and *FD* indices are influenced by the number of indicators measuring a particular latent trait and the size of loadings (Rodriguez et al. 2016a). In the bifactor model, *training and progression*, and *employment security* latent traits were each measured by two items and in the model estimation, slopes for these items within the dimensions were constrained to be equal for model identification. This is because, in general, for a latent variable model to be identified, every latent variable must be scaled, and the degrees of freedom must be zero (just-identified) or greater (over-identified) (Kline 2016; Wang and Wang 2020). In the bifactor model, all the latent traits were assumed to be mutually orthogonal, and follow standard normal distributions with zero mean and unit variance (Cai et al. 2011; Reise 2012). Where specific factors were defined by two items, slopes within these factors were constrained to be equal for the degrees of freedom to be zero and achieve model identification (Cai et al. 2011). This could be addressed by including more appropriate items that measure these dimensions in social surveys or UKHLS.

This chapter has applied IRT modelling to construct a measure of QWE that addresses some of the limitations with existing measurement instruments, such as the weighting of observed items on the latent traits. The modelling enabled the evaluation of how items cohered within the measure and results suggested that the measurement of QWE was better represented by a bifactor model. This addressed the issue of whether to report *overall* and/or *dimensions of QWE* and an examination of the properties of these measurement instrument. The next chapter will evaluate the measurement equivalence of the instrument for different groups of employees, and conditional on adequate measurement equivalence, compare levels of QWE for different groups in the UK employee population.

5.4 Appendices

5.4.1 Appendix 5.1: Standardised Factor Loadings based on the Graded Response Bifactor IRT Model

Item	Factor Loadings											
	R						M*plus*					
	θ_G	θ_{S1}	θ_{S2}	θ_{S3}	θ_{S4}	θ_{S5}	θ_G	θ_{S1}	θ_{S2}	θ_{S3}	θ_{S4}	θ_{S5}
Effective gross pay	0.408	0.341					0.392	0.328				
Pension provision	0.115	0.940					0.124	0.907				
Pay bonuses	0.206	0.150					0.195	0.140				
Pay progression	0.126	0.461					0.124	0.441				
Progression prospects	0.114		0.715				0.112		0.685			
Training prospects	0.141		0.712				0.140		0.683			
Employment type	0.122			0.659			0.122			0.612		
Job security	0.078			0.662			0.080			0.615		
Job tasks	0.411				0.724		0.415				0.703	
Work pace	0.387				0.749		0.392				0.730	
Work manner	0.416				0.834		0.427				0.818	
Task order	0.460				0.732		0.465				0.713	
Work hours	0.750				0.265		0.734				0.252	

Continued…

Continued…

Item	**Factor Loadings**											
	R						M*plus*					
	θ_G	θ_{S1}	θ_{S2}	θ_{S3}	θ_{S4}	θ_{S5}	θ_G	θ_{S1}	θ_{S2}	θ_{S3}	θ_{S4}	θ_{S5}
Part-time	0.042					0.769	0.047					0.742
Term-time	0.016					0.730	0.022					0.710
Job sharing	0.188					0.816	0.194					0.794
Flexi-time	0.617					0.381	0.603					0.366
Compressed hours	0.552					0.582	0.542					0.567
Annualised hours	0.348					0.692	0.341					0.608
Home working	0.804					0.241	0.790					0.233
Other flexibility	0.455					0.163	0.433					0.155
Informal flexibility	0.728					0.073	0.704					0.067
Working times	0.247					0.019	0.235					0.019

Notes: Second random sample, unweighted sample size, n = 8,490. a_G = conditional slope for *overall QWE* latent trait. $a_{S1} – a_{S5}$ = conditional slopes for *dimensions of QWE* (S1 – S5) latent traits. M*plus* estimates consider the complex sample design of UKHLS data; thus, stratification, clustering, and probability sampling weights, while *R* estimates do not consider the complex sample design and only account for the probability sampling weights.

5.4.2 Appendix 5.2: Multidimensional Discrimination and Difficulty Indices from Graded Response Bifactor IRT Model

Item	A_i	B_{i1}	B_{i2}	B_{i3}	B_{i4}
Effective gross pay	1.069	–1.915	–0.621	0.390	1.565
Pension provision	5.004	–1.077			
Pay bonuses	0.449	2.179			
Pay progression	0.926	0.468			
Progression prospects	1.786	1.470			
Training prospects	1.798	0.339			
Employment type	1.536	–2.349			
Job security	1.521	–3.254	–2.254	–0.106	
Job tasks	2.562	–1.439	–0.769	0.315	
Work pace	2.673	–1.445	–0.809	0.183	
Work manner	4.360	–1.662	–1.009	–0.104	
Task order	2.933	–1.725	–1.066	–0.095	
Work hours	2.232	–0.490	0.156	0.921	
Part-time	2.056	–0.282			
Term-time	1.819	1.315			
Job sharing	2.603	1.121			
Flexi-time	1.793	0.661			
Compressed hours	2.287	1.451			
Annualised hours	1.761	2.330			
Home working	2.629	1.200			
Other flexibility	0.940	2.087			
Informal flexibility	1.825	–0.573	–0.033		
Working times	0.435	–2.512			

Notes: Second random sample, unweighted sample size, n = 8,490. A_i = multidimensional discrimination index. B_{i1} – B_{i4} are multidimensional difficulty indices for different response categories, with subscript (1) indicating the first threshold, (2) second threshold, (3) third threshold, and (4) fourth threshold.

5.4.3 Appendix 5.3: Item Fit Statistics for Graded Response Bifactor IRT Model

Item	$S-\chi^2$	*df*	**RMSEA**
Effective gross pay	169.59***	106	0.009
Pension provision	27.04	29	0.000
Pay bonuses	71.74***	29	0.014
Pay progression	31.21	29	0.003
Progression prospects	44.74*	30	0.008
Training prospects	65.89***	30	0.013
Employment type	26.75	29	0.000
Job security	109.54*	86	0.006
Job Tasks	119.36***	70	0.010
Work Pace	154.71***	69	0.013
Work Manner	137.44***	63	0.013
Task Order	106.26**	65	0.009
Work Hours	152.31***	68	0.013
Part-time	70.21***	29	0.014
Term-time	66.17***	28	0.013
Job sharing	55.25**	28	0.011
Flexi-time	47.35**	26	0.010
Compressed hours	28.76	24	0.005
Annualised hours	41.27*	24	0.010
Home working	41.11**	22	0.011
Other flexibility	24.68	27	0.000
Informal flexibility	88.56**	52	0.010
Working times	46.72*	29	0.009

Notes: Second random sample, unweighted sample size, $n = 7{,}508$. Test excludes missing cases. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

5.4.4 Appendix 5.5: Standardised and Unstandardised G^2 Local Dependence Residuals for the Bifactor Graded Response Model

	Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10	Item11	Item12	Item13	Item14	Item15	Item16	Item17	Item18	Item19	Item20	Item21	Item22	Item23
Item1		0.068	0.067	0.100	–0.060	0.062	0.064	–0.028	0.054	–0.038	0.055	0.053	–0.062	–0.094	0.048	0.122	–0.050	0.094	0.067	0.142	–0.045	–0.089	0.081
Item2	76.78		0.048	0.030	0.046	0.122	0.104	0.037	–0.055	–0.047	0.049	0.040	–0.065	0.024	0.103	0.138	0.094	0.109	0.060	0.084	0.058	–0.041	0.047
Item3	75.47	38.90		–0.047	0.040	–0.034	0.108	0.057	0.039	0.049	0.049	0.046	–0.065	–0.066	–0.109	–0.084	–0.052	–0.046	–0.050	0.043	0.041	0.060	0.038
Item4	170.24	13.82	37.16		0.079	0.138	0.060	0.068	–0.038	–0.046	0.041	0.039	–0.047	0.046	0.115	0.149	0.069	0.082	0.037	0.023	0.038	–0.042	0.016
Item5	57.81	35.08	26.13	101.47		0.021	–0.047	0.026	0.031	–0.045	–0.034	–0.026	–0.048	0.009	0.016	0.013	0.044	0.025	0.023	0.021	0.037	0.034	–0.039
Item6	61.84	243.76	19.39	312.30	6.94		–0.008	0.060	0.067	–0.052	0.048	0.047	–0.068	0.024	0.060	0.054	0.042	0.027	0.026	0.028	0.062	0.032	–0.059
Item7	69.37	186.15	199.27	60.57	35.96	1.19		–0.073	0.057	0.061	0.045	0.051	–0.030	–0.012	–0.023	0.021	–0.020	0.011	0.013	–0.014	0.017	–0.039	0.033
Item8	39.87	23.40	54.91	78.51	11.09	58.58	65.36		0.051	0.047	0.052	0.047	–0.042	0.038	0.038	0.034	–0.037	0.019	0.026	–0.055	0.024	-0.050	–0.037
Item9	149.37	51.82	26.06	24.09	15.40	74.01	55.15	130.69		0.142	–0.125	–0.117	0.090	–0.054	0.052	–0.058	–0.061	–0.051	0.025	–0.045	–0.022	–0.046	0.044
Item10	74.65	37.64	41.15	36.75	33.65	44.90	63.36	114.58	1037.24		–0.148	–0.138	0.101	–0.045	–0.037	–0.038	–0.052	–0.051	–0.029	-0.054	–0.032	0.047	0.047
Item11	154.23	41.28	41.88	29.26	19.17	36.91	35.20	137.10	797.68	1120.60		0.131	–0.101	–0.053	0.039	0.047	–0.047	–0.038	0.027	0.035	0.031	0.059	0.056
Item12	142.05	27.13	35.87	25.21	10.88	36.49	44.61	111.86	706.39	975.81	884.97		–0.094	–0.037	0.039	0.042	–0.031	–0.030	0.020	0.039	0.027	–0.045	0.062
Item13	197.54	72.80	72.15	37.53	37.40	75.28	15.53	91.74	418.55	519.76	523.74	456.03		–0.038	0.040	0.036	–0.084	–0.055	–0.029	–0.089	–0.069	0.124	0.058
Item14	150.30	10.01	74.35	36.39	1.26	9.69	2.51	24.52	49.51	35.32	48.47	23.58	25.30		–0.014	0.036	0.019	0.013	–0.040	–0.013	0.017	0.050	–0.047
Item15	39.26	181.43	202.91	226.95	4.39	58.34	9.43	24.87	46.94	23.73	26.14	26.14	27.83	3.61		0.017	–0.021	0.028	0.059	0.026	0.022	–0.067	0.095
Item16	250.48	327.73	120.18	379.06	2.77	47.44	7.49	19.55	57.57	24.56	37.76	30.61	21.78	22.07	4.97		0.024	0.017	0.012	0.028	0.026	–0.052	0.049
Item17	42.14	152.36	46.74	80.36	32.07	28.81	7.02	22.23	62.78	46.04	38.38	16.27	120.88	6.12	7.26	10.25		0.057	0.021	0.019	–0.022	0.042	0.026
Item18	150.35	204.83	35.70	114.68	10.18	12.37	2.02	6.19	44.04	44.25	25.20	15.07	51.83	2.76	13.64	4.88	55.71		0.064	0.028	0.023	–0.038	0.024
Item19	75.01	61.62	42.87	23.28	8.58	10.85	2.70	11.72	11.11	13.90	12.77	6.69	14.38	27.51	60.71	2.61	7.65	70.11		0.034	0.037	–0.037	–0.039
Item20	339.01	122.09	31.00	8.85	7.42	13.01	3.44	49.65	34.42	49.05	20.97	26.66	134.78	2.82	11.25	13.63	6.25	13.21	19.32		0.022	–0.050	0.053
Item21	34.74	58.35	29.32	24.46	22.31	62.58	5.25	10.27	7.99	17.05	16.93	12.72	80.83	5.04	8.38	11.94	8.43	8.83	23.19	8.05		0.070	–0.063
Item22	266.60	28.16	60.64	30.74	18.98	17.13	26.23	86.34	71.05	75.70	120.40	68.55	524.89	42.23	76.18	46.63	30.37	25.05	23.18	42.05	84.03		–0.055
Item23	109.33	37.29	24.63	4.37	25.36	56.70	18.78	23.52	33.73	38.37	52.91	66.52	58.19	38.33	153.90	40.85	11.34	10.04	25.97	47.27	68.14	50.81	

Notes: Second random sample, unweighted sample size, n = 7,508. Tests exclude missing cases. Standardised G^2 *LD* residuals in upper diagonal and unstandardised G^2 *LD* residuals in lower diagonal.

5.4.5 Appendix 5.6: Latent Trait Scores for Overall and Dimensions of Quality of Work and Employment

Item1	Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10	Item11	Item12	Item13	Item14	Item15	Item16	Item17	Item18	Item19	Item20	Item21	Item22	Item23	θ_G (SE)	θ_{S1} (SE)	θ_{S2} (SE)	θ_{S3} (SE)	θ_{S4} (SE)	θ_{S5} (SE)
0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	–1.984 (0.834)	–1.093 (0.424)	–0.254 (0.945)	–1.596 (0.676)	–0.785 (0.691)	–0.504 (0.834)
0	0	0	0	0	0	0	0	1	1	2	1	0	0	0	0	0	0	0	0	0	0	0	–1.244 (0.603)	–1.057 (0.581)	0.014 (0.768)	–1.789 (0.495)	–0.224 (0.505)	–1.310 (0.974)
0	0	0	0	0	0	0	0	2	2	3	2	0	1	0	0	0	0	0	0	0	0	0	–1.203 (0.578)	–1.395 (0.618)	–0.471 (0.684)	–1.679 (0.649)	0.567 (0.417)	–0.189 (0.771)
0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	–2.030 (0.526)	–1.128 (0.306)	–1.217 (1.136)	–1.188 (0.357)	–1.676 (0.503)	–0.481 (0.396)
0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	2	1	–0.708 (0.853)	–1.313 (0.485)	–0.315 (0.794)	–1.443 (0.856)	–2.069 (0.527)	–0.204 (0.688)
0	0	0	0	0	0	0	1	0	0	1	2	0	0	0	0	0	0	0	0	0	0	1	–1.528 (0.810)	–1.200 (0.499)	–0.174 (0.735)	–1.711 (0.634)	–0.620 (0.501)	–0.872 (1.017)
0	0	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	–1.841 (0.665)	–1.267 (0.427)	–0.933 (1.084)	–0.952 (0.554)	–1.672 (0.562)	–0.101 (0.550)
0	0	0	0	0	0	0	2	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	–1.848 (0.797)	–1.251 (0.530)	–0.319 (0.807)	–0.986 (0.806)	–0.652 (0.533)	–0.693 (0.859)
0	0	0	0	0	0	0	2	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	–1.135 (0.691)	–1.334 (0.513)	–0.387 (0.772)	–0.976 (0.806)	–1.402 (0.586)	0.073 (0.606)
0	0	0	0	0	0	0	2	0	1	1	1	0	1	0	0	0	0	0	0	0	0	1	–1.696 (0.644)	–1.323 (0.503)	–0.227 (0.680)	–1.267 (0.752)	–0.595 (0.397)	0.242 (0.552)

Chapter 6 Measurement Equivalence and Multiple Group Analysis

The objective of this chapter is to establish what independent variables affect *overall QWE* and *dimensions of QWE* in the UK employee population and compare levels between groups without controlling for any other characteristics. Measures of *overall QWE* and *dimensions of QWE* are based on the instrument developed in Chapter 5. The first section describes the methodology of conducting the analyses in terms of the data and sample, the dependent and independent variables used, as well as the methods applied. The second section presents the results of evaluating measurement equivalence or invariance of the instruments for different groups using *differential item functioning* (DIF), followed by results of the multiple group analyses based on the graded response bifactor model for each predictor of QWE. The final section discusses the findings considering what factors affected *overall QWE* and *dimensions of QWE*, and how the levels of the latent traits compared between groups.

6.1 Methodology

6.1.1 Data and Sample

The data used are as described in Section 4.1.1, thus Wave 8 (2016 – 2017) of *Understanding Society: The United Kingdom Household Longitudinal Study* (UKHLS) (University of Essex, Institute for Social and Economic Research 2018), with the sample limited to employees who were in a paid job, participated in full interviews, and aged 16 years old and over, who were in a paid job, and the base sample was 16,981.

6.1.2 Variables

The dependent variables used were based on the QWE of measurement instruments developed in Chapter 5, and these were limited to the *overall QWE*, *economic compensation*, *working conditions*, and *work-time scheduling* latent traits only as estimated scores for *training* *and progression* and *employment security* were not a good representation of these latent traits and could not be used in subsequent analysis (see Section 5.2.6). The latent traits are assumed to follow standard normal distributions; thus, $\theta \sim N(0, 1)$; and the scales of the scores represent standard deviations from the population mean.

The independent variables or predictors of QWE considered were classified into *demographic* (sex, ethnic group, and age group), *socio-demographic* (relationship status, parental status, illness or disability, and region), and *socio-economic* (education, occupational classification, full or part-time, organisational sector, and organisation size) characteristics (Section 4.1.2).

6.1.3 Methods

Analyses were conducted in multiple packages in *R* (R Core Team 2020). The *'lordif'* package (Choi et al. 2011) was used to conduct DIF analysis evaluating the measurement equivalence of the QWE measurement instruments for different demographic, socio-demographic, and socio-economic characteristics. DIF analysis was conducted separately for each set of indicators measuring *overall QWE*, *economic compensation*, *working conditions*, and *work-time scheduling* latent traits (Choi 2021) (Appendix 6.1).

The '*lordif*' package uses a flexible iterative hybrid ordinal logistic regression / IRT framework to perform DIF detection. This involves using Monte Carlo simulations to generate multiple DIF-free datasets with the same dimension as the observed data, while also preserving group differences in latent trait estimates. The DIF-free simulated datasets are then used to estimate purified latent trait estimates, while the observed data are used to estimate the initial single-group item parameters (ignoring DIF) (Choi et al. 2011). The algorithm used for conducting the DIF analysis is outlined in Appendix 6.2. This process involved fitting separate graded response models for each latent trait and obtaining non-group specific item parameters and estimates of the latent traits. For the set of items measuring a latent trait, three nested ordinal logistic regression models (Equation (3.11)) were fitted for each item.[36] The comparison between *Models* 1 and 2 evaluated uniform DIF, while the comparison between *Models* 2 and 3 evaluated non-uniform DIF, and the comparison between *Models* 1 and 3 evaluated total DIF (Choi et al. 2011; Mair 2018). These comparisons were based on the likelihood ratio χ^2 test to flag DIF items, but this test is sensitive to large sample sizes, and the magnitude of DIF was evaluated by examining the change in McFadden's pseudo R^2 statistic between the models (Choi et al. 2011) relative to Cohen's guideline of effect size (Cohen 1988; Sawilowsky 2009).

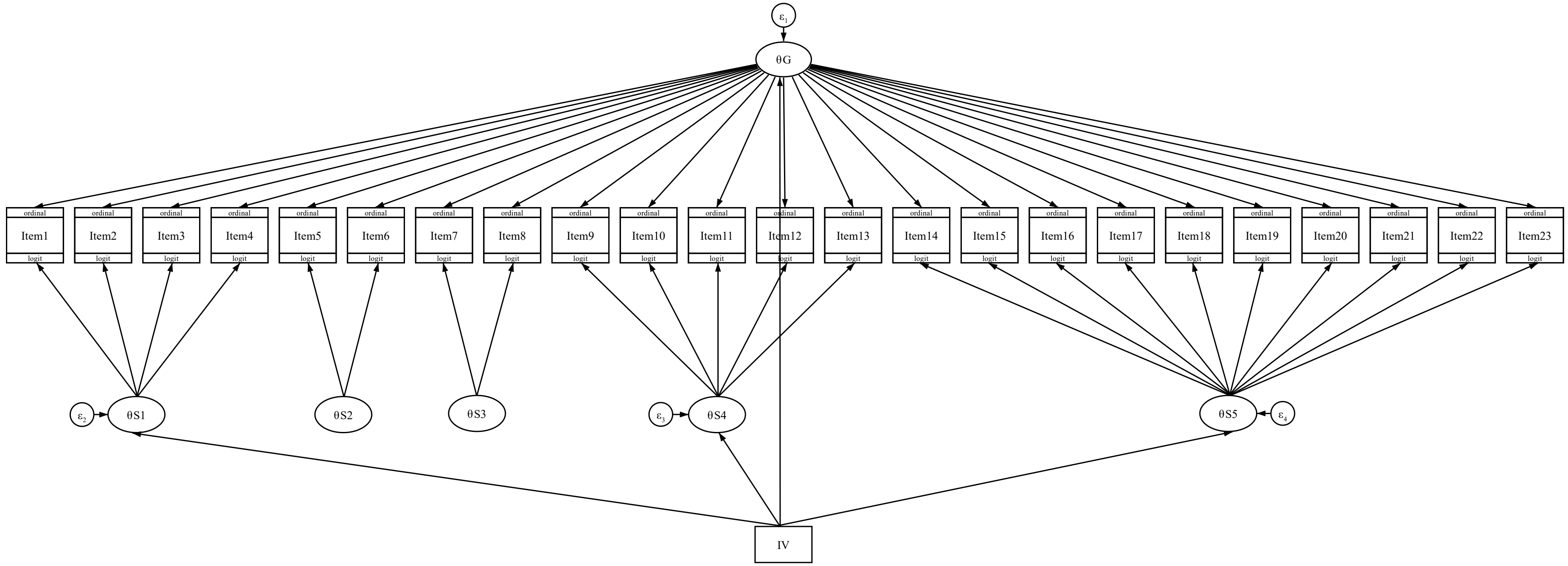
The algorithm then treated items flagged for DIF within a set measuring a latent trait as unique between groups and sparse variables, containing responses for a particular group but missing for other groups, were created. The graded response models were refitted using data with sparse variables and a single set of item parameter estimates obtained for non-DIF items (anchor items), while group-specific item parameters were obtained for DIF items. Item parameters from the anchor items and group-specific item parameters were used to obtain latent trait estimates that account for DIF, which were then used to refit the three nested ordinal logistic regression models. This process is automated in the '*lordif*' package and repeated until the same set of items were flagged for DIF over successive iterations or the maximum number of iterations reached (Choi et al. 2011).

Conditional on adequate measurement equivalence between groups of predictors of QWE, mean comparisons of *overall QWE* and *dimensions of QWE* were conducted between groups for each predictor of QWE based on group-specific item parameters. Multiple group bifactor IRT models were estimated for each independent variable as displayed in the path diagram in Figures 6.1 using the '*mirt*' package (Chalmers 2012). The path diagram shows the

[36] *Model 1* included the latent trait as the predictor to the item, *Model 2* included the latent trait and group membership as predictors, while *Model 3* included the latent trait, group membership, and an interaction term between the latent trait and group membership as predictors. (Choi et al. 2011; Mair 2018).

measurement model from Chapter 5, but also includes a structural model introducing the effect of an independent variable (IV) on *overall QWE* (θ_G), *economic compensation* (θ_{S1}), *working conditions* (θ_{S4}), and *work-time scheduling* (θ_{S5}). Group-specific item parameters were produced and used to predict latent trait scores to compare between groups.

Figure 6.1: Multiple Group Graded Response Bifactor IRT Model



Notes: The path diagram depicts the bifactor measurement model consisting of *overall QWE* (θ_G), *economic compensation* (θ_{S1}), *training and progression* (θ_{S2}), *employment security* (θ_{S3}), *working conditions* (θ_{S4}), and *work-time scheduling* (θ_{S5}) latent traits and the structural model introducing the effect of a predictor of QWE (IV) on θ_G, θ_{S1}, θ_{S4}, and θ_{S5} along with their respective error term (ε_i) capturing the variance of each latent trait not explained by the IV. θ_{S2} and θ_{S3} were excluded from the structural model as they were poorly defined.

To compare means (μ) between groups in the UK employee population, the null hypotheses (H_0) of equal means were tested, with the two-sample Student's *t*-test used for dichotomous predictors of QWE, thus:

$$H_0: \mu_1 = \mu_2 \text{ and } H_1: \mu_1 \neq \mu_2 \qquad \text{Hypothesis (1)}$$

while the one-way ANOVA test was used for polytomous predictors of QWE, thus, for a predictor with *k* groups:

$$\begin{gathered} H_0: \mu_1 = \mu_2 \ldots = \mu_k \\ H_1: \text{At least two of } \mu_1, \mu_2, \ldots \mu_k \text{ are unequal} \end{gathered} \qquad \text{Hypothesis (2)}$$

Tests were conducted at the 5% significance level, with *p*-values < 0.05 suggesting rejection of the H_0 and differences being statistically significant in the UK employee population, while *p*-values > 0.05 suggested statistically insignificant differences and retention of the H_0.

Prior to estimating the ANOVA or *t*-tests, assumptions about the data were checked to determine whether parametric or non-parametric methods were appropriate. Based on the sampling design, latent trait scores between groups for each predictor of QWE were independent and had an interval level of measurement. For the sampling distribution and based on the central limit theorem, sampling distributions tend to be normal in large samples (> 30), regardless of the shape of distribution of the sample data (Field et al. 2012). Due to large samples for all groups in this study, sampling distributions were assumed to be normal. Regarding homogeneity of variances between groups, Levene's tests were conducted for each predictor of QWE to establish whether this assumption was tenable or violated in the data. This tests the null hypothesis of equal population variances between groups, with *p*-values < 0.05

indicating violation of the assumption, while p-values > 0.05 indicated the assumption was tenable at the 5% significance level. In cases where homogeneity of variance was violated, ANOVA or t-tests with a Welch's correction were estimated.

In addition to comparing means between groups with the ANOVA or t-tests, effect sizes[37] were estimated to examine the magnitudes of the effect of the predictors on *overall QWE* and *dimensions of QWE* in the UK employee population. The omega-squared (ω^2) measure, which is a less biased estimator of population effect size (Baldwin 2019; Lakens 2013; Olejnik and Algina 2003), was used and indicated the proportion of the variance in the dependent variable explained by the independent variable in the population (Baldwin 2019; Olejnik and Algina 2003). The ω^2 estimate for two-sample t-tests is computed as:

$$\omega^2 = \frac{t^2 - 1}{t^2 + df + 1} \tag{6.1}$$

where t is the t-statistic and df is the degrees of freedom of the two-sample t-test. For the one-way ANOVA, the ω^2 estimate is computed as:

$$\omega^2 = \frac{df * (F - 1)}{df * (F - 1) + n} \tag{6.2}$$

where F is the F-statistic, df is the degrees of freedom of the one-way ANOVA, and n is the sample size excluding missing cases. In terms of interpretation of the effect sizes, this was based on Cohen's guidelines and the criteria are outlined in Table 6.1.

[37] Effect size measures can be classified into two families: thus, d family measures consisting of standardised mean differences (e.g. Cohen's d, Hedges' g), and r family measures consisting of measures of strength of association or proportion of explained variance (e.g. coefficient of determination (r^2), eta-squared (η^2), omega-squared (ω^2)).

Table 6.1: Guideline for Effect Size Interpretation

Size of effect	r	r^2
Negligible effect	< 0.1	< 0.01
Small effect	0.1 – 0.3	0.01 – 0.09
Moderate effect	0.3 – 0.5	0.09 – 0.25
Large effect	> 0.5	> 0.25

Adapted from Cohen (1988) and Sawilowsky (2009).

For the ANOVA tests, *post hoc* tests for pairwise mean comparisons between groups were conducted. The Tukey-Kramer post hoc procedure was used where the assumption of homogeneity of variances was tenable, while the Games-Howell post hoc procedure was used where the homogeneity of variances assumption was violated. For these tests, p-values < 0.05 indicated statistically significant mean differences between a pair of groups and p-values > 0.05 indicated statistically insignificant mean differences (Field et al. 2012).

6.2 Results

6.2.1 Differential Item Functioning

DIF analyses for each of the demographic, socio-demographic, and socio-economic characteristics affecting *overall QWE*, *economic compensation*, *working conditions*, and *work-time scheduling* latent traits were conducted. For illustration purposes, only detailed results by ethnic group are presented in Table 6.2 and Appendix 6.3. Diagnostic plots based on group-specific item parameters are also produced for items flagged for some form of DIF and Figure 6.3 display illustrative DIF diagnostic plots by ethnic group for items measuring the *overall QWE* latent trait.

Table 6.2: Differential Item Functioning by Ethnic Group

	DIF Model 2* v *DIF Model 1 (Uniform DIF)								***DIF Model 3* v *DIF Model 2*** (Non-uniform DIF)								***DIF Model 3* v *DIF Model 1*** (Total DIF effect)							
	θ_G		θ_{S1}		θ_{S4}		θ_{S5}		θ_G		θ_{S1}		θ_{S4}		θ_{S5}		θ_G		θ_{S1}		θ_{S4}		θ_{S5}	
	p (df)	R^2	*p* (df)	R^2	*p* (df)	R^2	*p* (df)	R^2	*p* (df)	R^2	*p* (df)	R^2	*p* (df)	R^2	*p* (df)	R^2	*p* (df)	R^2	*p* (df)	R^2	*p* (df)	R^2	*p* (df)	R^2
Item1	ns (3)	.0001	ns (3)	.0001					ns (3)	.0001	*** (3)	.0005					ns (6)	.0002	*** (6)	.0002				
Item2	*** (3)	.0024	ns (3)	.0001					ns (3)	.0002	ns (3)	.0000					*** (6)	.0026	ns (6)	.0026				
Item3	*** (3)	.0010	*** (3)	.0009					ns (3)	.0001	*** (3)	.0026					** (6)	.0010	*** (6)	.0010				
Item4	ns (3)	.0002	ns (3)	.0002					ns (3)	.0003	*** (3)	.0036					ns (6)	.0005	*** (6)	.0005				
Item5	*** (3)	.0075							ns (3)	.0003							*** (6)	.0079						
Item6	*** (3)	.0039							ns (3)	.0004							*** (6)	.0043						
Item7	*** (3)	.0032							ns (3)	.0003							*** (6)	.0035						
Item8	*** (3)	.0010							ns (3)	.0003							*** (6)	.0012						
Item9	*** (3)	.0008			*** (3)	.0008			ns (3)	.0001			ns (3)	.0001			*** (6)	.0008			*** (6)	.0009		
Item10	** (3)	.0003			** (3)	.0003			** (3)	.0003			** (3)	.0003			*** (6)	.0006			*** (6)	.0006		
Item11	ns (3)	.0000			ns (3)	.0001			ns (3)	.0000			ns (3)	.0000			ns (6)	.0001			ns (6)	.0001		
Item12	*** (3)	.0007			*** (3)	.0009			ns (3)	.0001			** (3)	.0003			*** (6)	.0008			*** (6)	.0012		
Item13	*** (3)	.0018			*** (3)	.0014			ns (3)	.0001			ns (3)	.0000			*** (6)	.0019			*** (6)	.0015		
Item14	*** (3)	.0008					** (3)	.0005	ns (3)	.0001					ns (3)	.0002	** (6)	.0009					** (6)	.0008
Item15	ns (3)	.0006					ns (3)	.0002	ns (3)	.0001					ns (3)	.0000	ns (6)	.0007					ns (6)	.0002
Item16	*** (3)	.0026					*** (3)	.0015	ns (3)	.0001					** (3)	.0009	*** (6)	.0027					*** (6)	.0024
Item17	ns (3)	.0001					*** (3)	.0014	*** (3)	.0011					ns (3)	.0004	*** (6)	.0013					*** (6)	.0018
Item18	*** (3)	.0018					ns (3)	.0003	*** (3)	.0013					ns (3)	.0001	*** (6)	.0031					ns (6)	.0004
Item19	ns (3)	.0005					ns (3)	.0001	ns (3)	.0005					ns (3)	.0006	ns (6)	.0009					ns (6)	.0006
Item20	** (3)	.0008					ns (3)	.0004	*** (3)	.0019					ns (3)	.0001	*** (6)	.0027					ns (6)	.0005
Item21	ns (3)	.0001					ns (3)	.0002	ns (3)	.0004					ns (3)	.0004	ns (6)	.0005					ns (6)	.0006
Item22	*** (3)	.0013					*** (3)	.0009	*** (3)	.0014					ns (3)	.0001	*** (6)	.0028					*** (6)	.0010
Item23	*** (3)	.0019					*** (3)	.0018	*** (3)	.0011					ns (3)	.0003	*** (6)	.0029					*** (6)	.0021

Notes: θ_G = *Overall QWE*. θ_{S1} = *Economic compensation*. θ_{S4} = *Working conditions*. θ_{S5} = *Work-time scheduling*. *p* represents *p*-values of the $\Delta\chi^2$ test between DIF models tested at the 1% significance level, with statistically significant differences indicating presence of DIF. ns represents 'not significant' ($p \geq .01$). ** $p < .01$. *** $p < .001$. *df* = Degrees of freedom. R^2 = McFadden Pseudo R^2.

Figure 6.2: Distributions of Overall QWE Latent Trait Scores by Ethnic Group

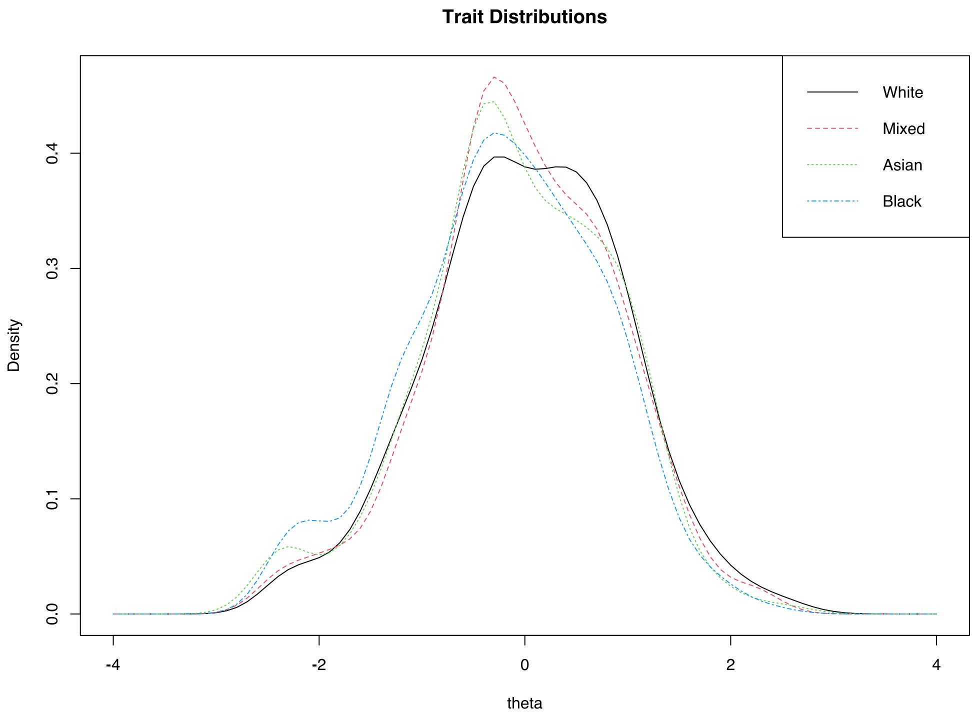


Figure 6.3: DIF Diagnostic Plots for Item 9 (Job tasks) by Ethnic Group for *Overall QWE*

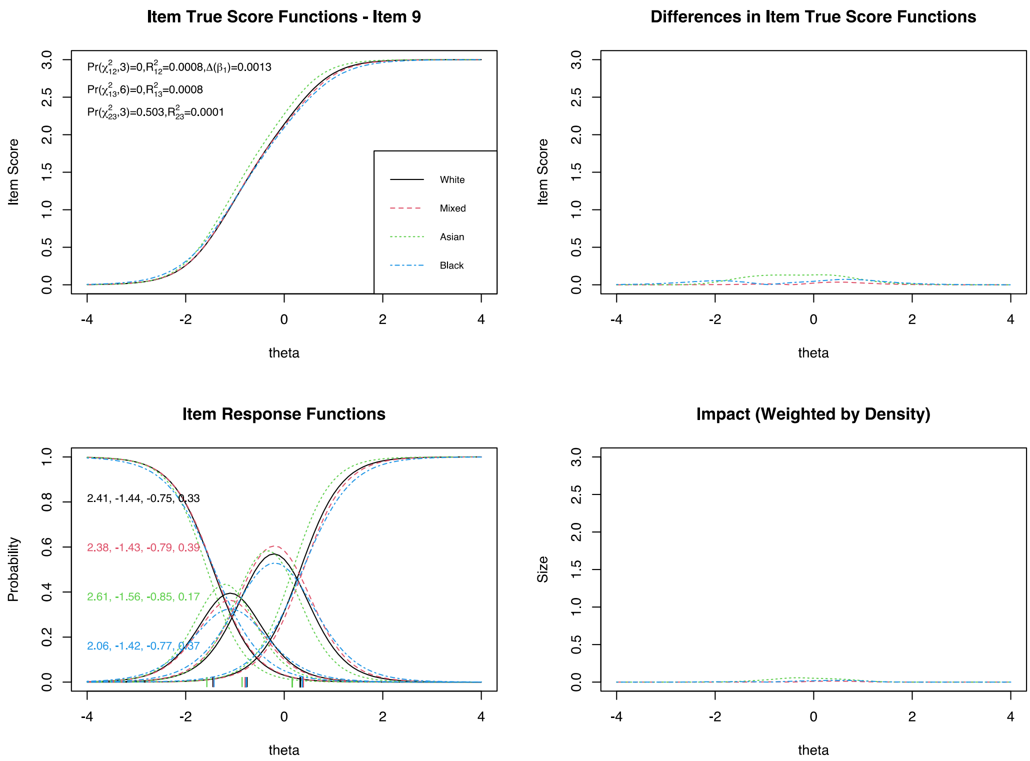


Figure 6.2 shows group-specific kernel density plots of the distributions of *overall QWE* latent trait scores (*theta*), these were normally distributed and broadly overlapped. However, compared to employees from other ethnic backgrounds, above-average scores were slightly more common for employees from a White ethnic background, while below-average scores were slightly more common for employees from other ethnic backgrounds.

At item-level, for items measuring *overall QWE*, 17 were flagged for some form of DIF by ethnic group (Table 6.2) and as an example, diagnostic plots for Item 9 (*job tasks*) are displayed in Figure 6.3, while those for other items are shown in Appendix 6.3. The upper-left plot in Figure 6.3 depicts functions of the expected or true scores (y-axis) given overall QWE (*theta* on x-axis) to Item 9 by ethnic group. The curve for employees from Asian or Asian British ethnic group was slightly above those for employees from White or Mixed ethnic backgrounds across the overall QWE continuum, suggesting Asian or Asian British employees had a higher differential performance on this item,[38] indicating presence of uniform DIF. This was observed in the differences in difficulty parameters by ethnic group in the lower-left plot in Figure 6.3 which shows group-specific category response functions, including the discrimination and difficulty parameters printed within the plot. For example, considering the highest response category, estimates of difficulty parameters were 0.33 for White, 0.38 for Mixed, 0.17 for Asian or Asian British, and 0.37 for Black or Black British which suggested employees from different ethnic groups required different levels of overall QWE to have a 0.5 probability of selecting this category, indicating the presence of uniform DIF.

On the other hand, the curve of expected scores for employees from Black or Black British ethnic group intersected curves for employees from other ethnic groups (upper-left plot, Figure 6.3), suggesting these employees had a higher differential performance for part of the

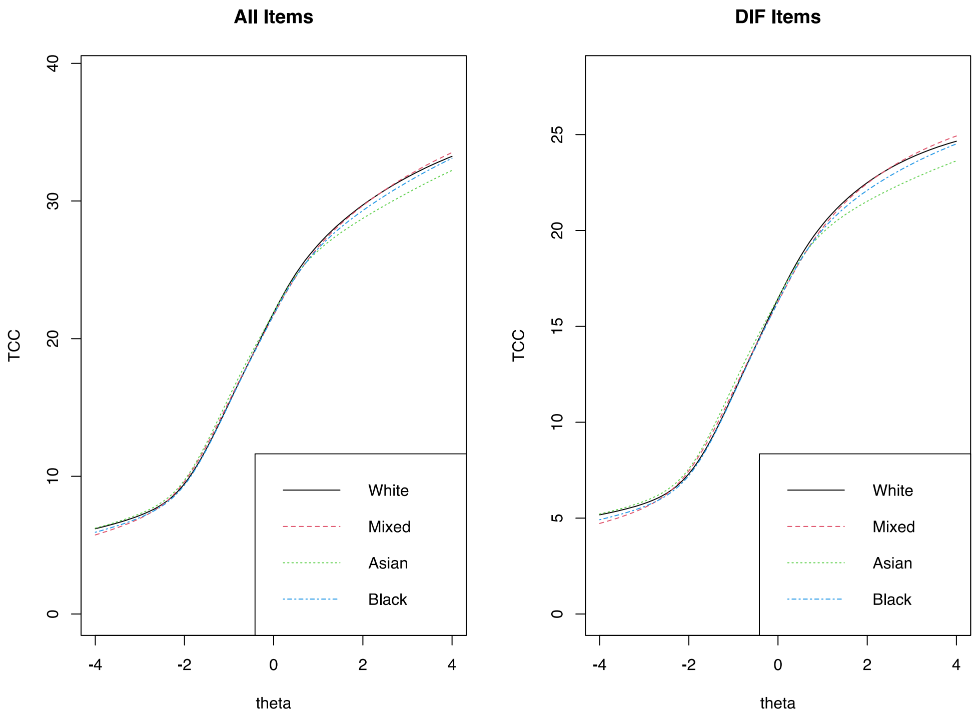
[38] For the same level of the latent trait, Asian or Asian British employees had slightly higher expected scores on this item than employees from other ethnic backgrounds.

overall QWE continuum (at low levels) and a lower differential performance in other parts (at high levels) on this item, compared to employees from other ethnic groups, indicating presence of non-uniform DIF. Again, this was observed in the differences in discrimination parameters by ethnic group in the lower-left plot in Figure 6.3; thus, 2.41 for White, 2.40 for Mixed, 2.62 for Asian or Asian British, and 2.06 for Black or Black British; which suggested the item differentiated between employees with different levels of overall QWE around the difficulty parameter differently by ethnic group, indicating the presence of non-uniform DIF.

A comparison between DIF models in Equation (3.11) evaluating different forms of DIF for Item 9 by ethnic group in measuring overall QWE indicated that uniform DIF (*DIF Model 2* v *DIF Model 1*, *Pr* ($\Delta\chi^2_{12}$ (3) < 0.001) and total DIF effect (*DIF Model 3* v *DIF Model 1*, *Pr* ($\Delta\chi^2_{13}$ (6) < 0.001) were statistically significant, while non-uniform DIF (*DIF Model 3* v *DIF Model 2*, *Pr* ($\Delta\chi^2_{23}$ (3) = 0.503) was not statistically significant in the UK employee population at the 1% significant level (Table 6.2 and upper-left plot in Figure 6.3).

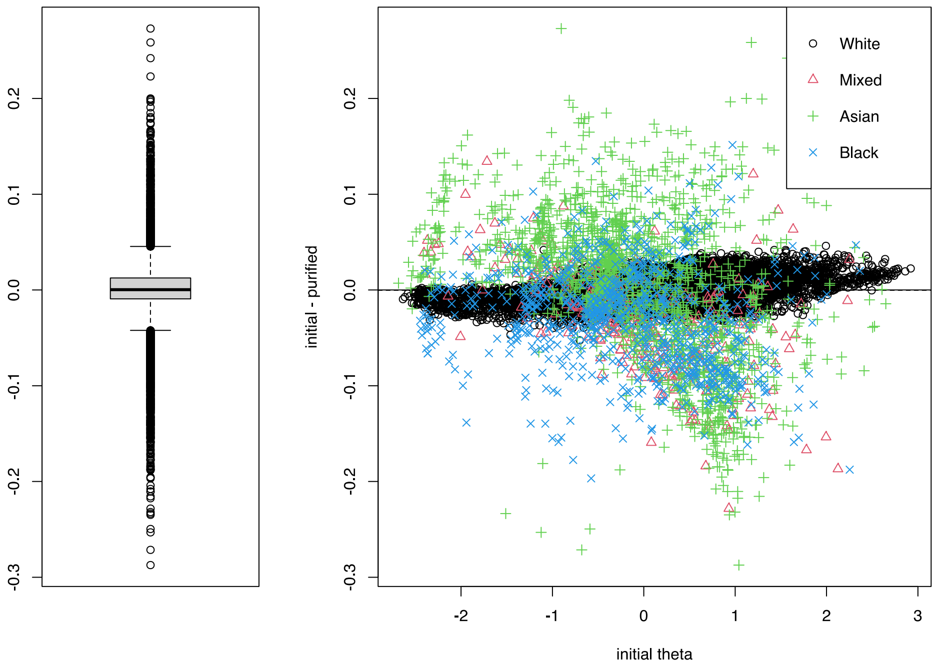
Continuing to focus on Figure 6.3, the upper-right plot displays absolute differences in expected or true scores between employees from a White ethnic background and those from other ethnic groups for Item 9, given overall QWE. These differences were small, and when weighted by the White ethnic group score distribution, the expected impact was negligible (lower-right plot, Figure 6.3). This was indicated by the negligible effect size of the magnitude of DIF relative to Cohen's rule of thumb (<0.01) based on McFadden's pseudo R^2 statistics for all the DIF models; thus R^2_{12} = 0.008 for uniform DIF, R^2_{23} = 0.001 for non-uniform DIF, and R^2_{13} = 0.008 for total DIF effect (Table 6.2 and upper-left plot in Figure 6.3). Results of DIF analyses for other items measuring *overall QWE* by ethnic group are presented in Table 6.2, along with diagnostic plots for items flagged for DIF in Appendix 6.3.

Figure 6.4: Test Characteristics Functions



Moving on to test-level plots, Figure 6.4 displays test characteristic functions (TCFs) showing the impact of DIF items on group-specific expected total scores given overall QWE. The left plot depicts the expected total scores for all the items and the plot on the right depicts the expected total scores for items that exhibited DIF given overall QWE latent trait level by ethnic group. While there were notably differences in expected total scores for employees from Asian or Asian British ethnic backgrounds compared to those from other ethnic groups at high levels of overall QWE, and small differences in expected total scores for employees from a Mixed ethnic background compared to those from other ethnic groups at low levels of overall QWE for items exhibiting DIF, when all items were considered, the differences were slight.

Figure 6.5: Plots of Individual-level DIF Impact for Overall QWE by Ethnic Group



Notes: The plots display a scatterplot (right) of the difference in the scores ignoring DIF (initial latent trait score) and accounting for DIF (purified latent trait score) and a boxplot (left) showing the distribution of the difference by ethnic group.

Lastly, Figure 6.5 displays plots of individual-level DIF impact, and these show the difference in overall QWE scores ignoring DIF (*initial theta*) and accounting for DIF (*purified theta*). *Purified theta* scores are obtained by using items that did not exhibit any DIF as anchor items and then using group-specific item parameters for items that exhibited DIF (Choi et al. 2011; Kim et al. 2007). The scatterplot displays differences in scores between *initial* and *purified theta* (*y*-axis) against *initial theta* (*x*-axis) by ethnic group. The solid horizontal line at zero represents no difference, while the dotted line represents the mean difference between *initial* and *purified* scores. The scatter plot suggested that at low levels of overall QWE, accounting for DIF resulted in slightly lower scores for employees from Mixed and Asian or Asian British ethnic backgrounds (*initial theta* > *purified theta*)[39] and slightly higher scores for

[39] Thus, there were very few data points for these groups of employees below zero on the *y*-axis at low levels of *initial theta*, indicating that *initial* minus *purified theta* scores were largely positive.

employees from Black or Black British ethnic backgrounds (*initial theta* < *purified theta*). On the other hand, at high levels of overall QWE, accounting for DIF resulted in slightly lower scores for employees from Mixed, Asian or Asian British and Black or Black British ethnic backgrounds (*initial theta* < *purified theta*). However, the mean difference in scores was zero (dotted horizontal line), while the boxplot shows the distribution of the differences for all employees and the median was also zero. This suggested that accounting for DIF by ethnic group had minimal impact on individual overall QWE scores. Results of DIF analyses by ethnic group for the other latent traits (economic compensation, working conditions, and work-time scheduling) suggested that where items were flagged for DIF, the magnitudes of DIF had a negligible effect size relative to Cohen's rule of thumb (< 0.01) (Table 6.2).[40] This indicated that between-group comparison by ethnic group based on these measurement instruments was feasible.

An alternative approach uses MIMIC modelling to evaluate DIF by investigating whether a predictor directly affects the observed items given the model, with a statistically significant direct effect indicating the presence of uniform DIF (Wang and Wang 2020). Refer to Appendix 6.4 for results of DIF analysis by ethnic group using the MIMIC model. While this method accounted for the multidimensional latent structure of the observed items, it only evaluated uniform DIF and did not provide estimates of the magnitude of DIF.

In terms of DIF analyses by other demographic, socio-demographic, and socio-economic characteristics in Table 4.2 for the economic compensation, working conditions, work-time scheduling, and overall QWE measurement instruments, results suggested that where items were flagged for DIF, the magnitudes of DIF had either a small (< 0.09) or a negligible (< 0.01) effect size relative to Cohen's rule of thumb.[41] This indicated that between-

[40] Diagnostic plots for items flagged for DIF are, however, not presented due to the numerous numbers of plots.

[41] Tables of results are available but are not included.

group comparison by these characteristics based on the measurement instruments was feasible. Again, diagnostics plots for items flagged for DIF are not presented due to the numerous numbers of plots.

6.2.2 Multiple Group Analysis

Results of multiple group analyses for the graded response bifactor model by each predictor of QWE are presented in Appendices 6.5 – 6.8. Distributions of *overall QWE* and *dimensions of QWE* latent trait scores by each predictor are displayed in Appendix 6.9. These are not described in this section but are interpreted similarly to distributions in Section 5.2.5, however in comparing distributions, the focus is on similarities and / or differences in the shapes of the distributions between groups. For measures of central tendency, mean (*M*) and median (*Mdn*) estimates along with their associated standard deviations (*SD*) and interquartile ranges (*IQR*), respectively, are reported. Plots of group means with their 95% confidence intervals error bars are presented to aid comparison between groups, with the *y*-axes representing the latent trait scores (Figures 6.6 – 6.17). In terms of interpretation of the latent trait scores, *overall QWE* is interpreted conditional on other *dimensions of QWE* in the measurement model, while each *dimension of QWE* is interpreted over and above *overall QWE*. For the Levene's test, the test statistics and *p*-values are reported, while the degrees of freedom for the test statistics are available but not reported.

Demographic Characteristics

Table 6.3 summaries results of effect size estimates and Levene's tests evaluating the homogeneity of variances assumption for demographic characteristics. For Levene's tests, if homogeneity of variances was tenable, a *t*-test or ANOVA test assuming equal variances between groups was estimated, and if untenable, tests with a Welch's correction were estimated. In relation to post hoc procedures for ANOVA tests, the Tukey-Kramer's procedure was used where equality of variances was tenable, and the Games-Howell procedure was used where equality of variances was untenable. The table also includes an indication of whether the association between these predictors and the latent traits were significant or not.

Table 6.3: Summary Table of Levene's Tests and Effect Size Estimates for Demographic Characteristics

Characteristic	Overall QWE		Economic compensation		Working conditions		Work-time scheduling	
	Levene's test	Effect size	Levene's test	Effect size	Levene's test	Effect size	Levene's test	Effect size
Sex[a]	Untenable	Small***	Tenable	Negligible***	Tenable	Negligible	Untenable	Small***
Ethnic group[b]	Untenable	Negligible***	Untenable	Negligible***	Untenable	Negligible**	Untenable	Negligible***
Age group[b]	Untenable	Small***	Untenable	Small***	Untenable	Negligible***	Untenable	Negligible***

Notes: [a]: *t*-test. [b]: ANOVA test. Levene's test: If equality of variances is tenable, a *t*-test or ANOVA test assuming equal variances between groups is estimated, and if untenable, a *t*-test or ANOVA test with a Welch's correction is estimated. For post hoc procedures relating to ANOVA tests, the Tukey-Kramer is used where equality of variances is tenable, and the Games-Howell procedure is used where equality of variances is untenable. Negligible: Predictor explains less than 1% of the variation in the latent trait. Small: Predictor explains between 1% and 9% of the variation in the latent trait. Moderate: Predictor explains between 9% and 25% of the variation in the latent trait. Large: Predictor explains more than 25% of the variation in the latent trait. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

Figure 6.6: Mean Comparison of Latent Trait Scores by Sex

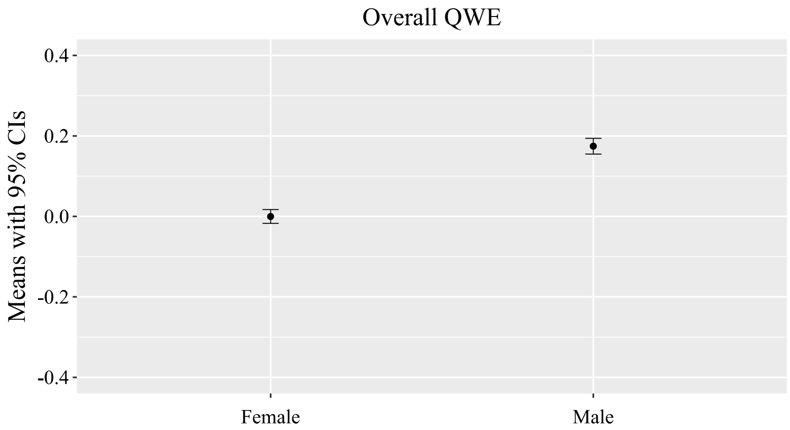


Figure 6.6 (a)

Figure 6.6 (b)

Figure 6.6 (c)

Figure 6.6 (d)

Sex

Overall QWE by Sex

On average, female employees in the sample had lower levels of *overall QWE* compared to male employees based on either the mean (M = –0.000, SD = 0.830 for females v M = 0.174, SD = 0.861 for males), or the median (Mdn = –0.028, IQR = 1.220 for females v Mdn = 0.182, IQR = 1.230 for males) (Appendix 6.5).

The homogeneity of variances assumption in the scores by sex was untenable (F = 7.77, p = 0.005), and a two-sample t-test with a Welch's correction indicated statistically significant mean differences between males and females (Welch's t (15636.80) = –13.15, $p < 0.001$, ω^2 = 0.011) (Figure 6.6 (a)). This suggested that females had poorer *overall QWE* than males in the UK employee population, however sex had a small effect size and explained approximately 1.1% of the variation.

Economic Compensation by Sex

For *economic compensation*, female employees in the sample had slightly lower levels compared to males in terms of the mean or median. Thus M = 0.008, SD = 0.670 for females v M = 0.073, SD = 0.678 for males, and Mdn = 0.133, IQR = 0.706 for females v Mdn = 0.205, IQR = 0.701 for males (Appendix 6.6).

There was homogeneity of variances in the scores by sex (F = 0.06, p = 0.811), and a two-sample t-test with equal variances suggested differences were statistically significant (t (16418) = –6.13, $p < 0.001$, ω^2 = 0.002) (Figure 6.6 (b)). This indicated that females had poorer *economic compensation* than males in the UK employee population, but the effect size was negligible with sex explaining approximately 0.2% of the variation.

Working Conditions by Sex

In terms of *working conditions*, in this sample and on average, levels were comparable by sex when considering either the mean (M = –0.030, SD = 0.854 for females v M = –0.043, SD = 0.855 for males), or the median (Mdn = –0.040, IQR = 1.307 for females v Mdn = –0.076, IQR = 1.351 for males) (Appendix 6.7).

The assumption of homogeneity of variances of the scores by sex was tenable (F = 0.55, p = 0.460), and a two-sample t-test with equal variances indicated statistically insignificant differences between males and females (t (16418) = 0.96, p = 0.339, ω^2 = –0.00001)[42] (Figure 6.6 (c)). This suggested that there was no difference in *working conditions* by sex in the UK employee population.

Work-time Scheduling by Sex

Regarding *work-time scheduling*, female employees in the sample had on average, higher levels than male employees when considering either the mean (M = 0.011, SD = 0.757 for females v M = –0.359, SD = 0.757 for males), or the median (Mdn = –0.076, IQR = 0.943 for females v Mdn = –0.477, IQR = 1.036 for males) (Appendix 6.8).

The homogeneity of variances assumption in the scores by sex was untenable (F = 25.32, p < 0.001). A two-sample t-test with a Welch's correction indicated statistically significant differences in means between males and females (Welch's t (15868.82) = 31.21, p < 0.001, ω^2 = 0.058) (Figure 6.6 (d)). This suggested that females were more aware of and had access to other forms of better *work-time scheduling* than males in the UK employee population, however, sex had a small effect size and explained approximately 5.8% of the variation.

[42] Negative effect size estimates are related to sampling fluctuations and correction for bias of ω^2 (Okada 2017).

Figure 6.7: Mean Comparison of Latent Trait Scores by Ethnic Group

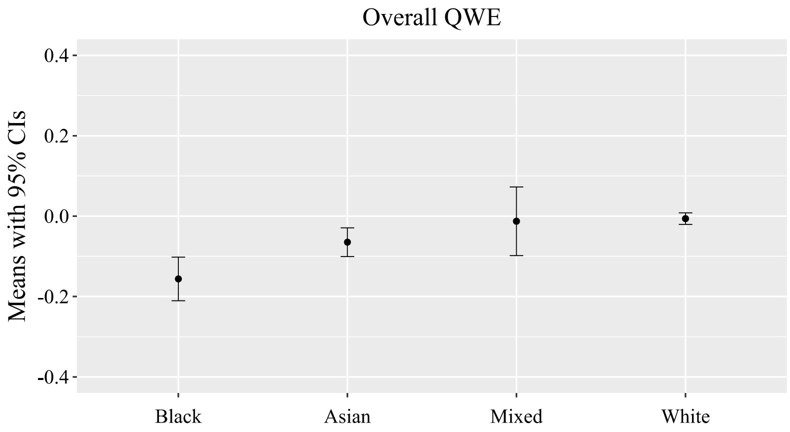


Figure 6.7 (a)

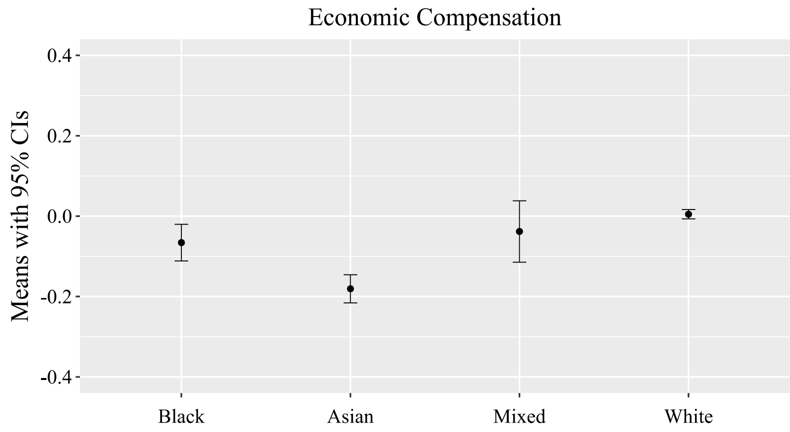


Figure 6.7 (b)

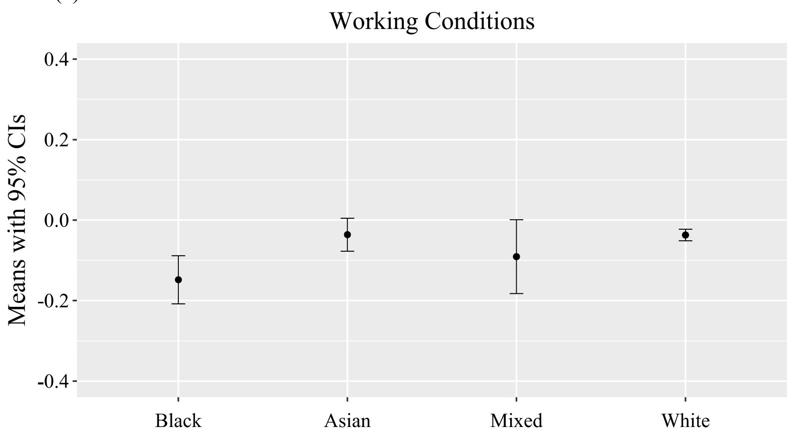


Figure 6.7 (c)

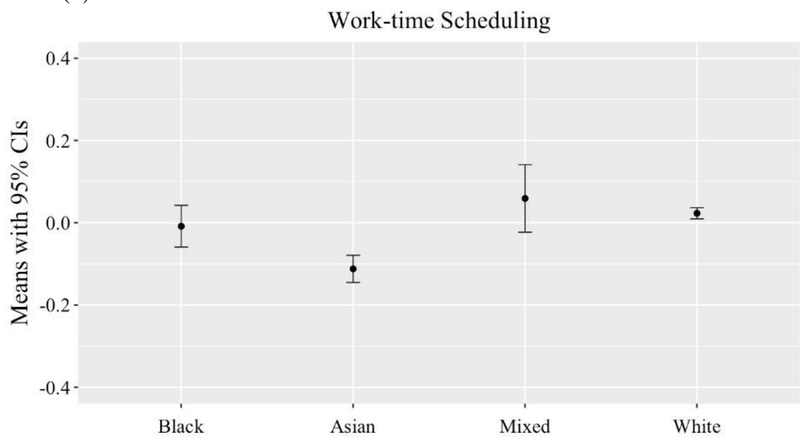


Figure 6.7 (d)

Ethnic Group

Overall QWE by Ethnic Group

Considering *overall QWE*, employees from a Black or Black British ethnic background ($M = –0.156$, $SD = 0.803$ and $Mdn = –0.177$, $IQR = 1.200$) in this sample had on average, the lowest levels than employees from other ethnic backgrounds, while those from a White ethnic background ($M = –0.006$, $SD = 0.844$ and $Mdn = –0.008$, $IQR = 1.245$) had the highest (Appendix 6.5).

The assumption of homogeneity of variances of the scores by ethnic group was untenable ($F = 18.55$, $p < 0.001$). A one-way ANOVA test with a Welch's correction indicated statistically significant differences between at least two groups (Welch's F (3, 1131.76) = 11.20, $p < 0.001$, $\omega^2 = 0.002$). However, ethnicity had a negligible effect size and explained approximately 0.2% of the variation.

The Games-Howell procedure indicated significant differences in *overall QWE* between employees from Black or Black British compared to those from other ethnic backgrounds (p-values < 0.05), and Asian or Asian British v White ($p = 0.015$) ethnic backgrounds, while other pairwise comparisons were statistically insignificant (p-values > 0.05) (Figure 6.7 (a)). Thus, Black or Black British employee had the poorest *overall QWE* in the UK employee population.

Economic Compensation by Ethnic Group

On average, employees from an Asian or Asian British ethnic background ($M = –0.181$, $SD = 0.751$ and $Mdn = –0.017$, $IQR = 0.826$) in this sample had the lowest levels of *economic compensation* than employees from other ethnic backgrounds, while those from a White ethnic background ($M = 0.005$, $SD = 0.678$ and $Mdn = 0.129$, $IQR = 0.715$) had the highest (Appendix 6.6).

The homogeneity of variances assumption in the scores by ethnic group was untenable ($F = 22.34$, $p < 0.001$). A one-way ANOVA test with a Welch's correction indicated statistically significant mean differences between at least two ethnic groups (Welch's F (3, 1116.37) = 34.10, $p < 0.001$, $\omega^2 = 0.006$). However, the magnitude of the differences was negligible, and ethnicity explained approximately 0.6% of the variation.

The Games-Howell procedure suggested mean differences in *economic compensation* between employees from Asian or Asian British compared to those from other ethnic backgrounds (p-values < 0.01), and Black or Black British v White ($p = 0.017$) ethnic backgrounds were statistically significant, while other pairwise comparisons were insignificant (p-values > 0.05) (Figure 6.7 (b)). Employees from Asian or Asian British backgrounds had the poorest *economic compensation* in the UK employee population.

Working Conditions by Ethnic Group

Regarding *working conditions*, in this sample and on average, employees from a Black or Black British ethnic background had the lowest levels ($M = –0.148$, $SD = 0.884$ and $Mdn = –0.147$, $IQR = 1.283$), while those from White ($M = –0.037$, $SD = 0.839$ and $Mdn = –0.054$, $IQR = 1.315$) and Asian or Asian British ($M = –0.036$, $SD = 0.881$ and $Mdn = –0.067$, $IQR = 1.284$) ethnic backgrounds had among the highest levels (Appendix 6.7).

The assumption of homogeneity of variances in the scores by ethnic group was untenable ($F = 2.92$, $p = 0.033$). From the one-way ANOVA test with a Welch's correction, mean differences between at least two groups were statistically significant (Welch's F (3, 1116.91) = 4.57, $p = 0.003$, $\omega^2 = 0.001$). Ethnicity had a negligible effect size and explained approximately 0.1% of the variation.

From the Games-Howell procedure, mean differences in *working conditions* between employees from Asian or Asian British v Black or Black British ($p = 0.013$), and Black or Black British v White ($p = 0.002$) ethnic backgrounds were statistically significant, while other pairwise comparisons were statistically insignificant (p-values > 0.05) (Figure 6.7 (c)). Black or Black British employees had among the poorest *working conditions* in the UK employee population.

Work-time Scheduling by Ethnic Group

In terms of *work-time scheduling*, on average, employees from an Asian or Asian British ethnic background ($M = –0.112$, $SD = 0.709$ and $Mdn = –0.060$, $IQR = 1.061$) had the lowest levels, while levels for employees from other ethnic backgrounds were similar (Appendix 6.8).

The homogeneity of variances assumption in the scores by ethnic group was untenable ($F = 12.87$, $p < 0.001$), and a one-way ANOVA test with a Welch's correction indicated that mean differences between at least two groups were statistically significant (Welch's F (3, 1130.11) = 19.04, $p < 0.001$, $\omega^2 = 0.003$). However, ethnicity had a negligible effect size and explained approximately 0.3% of the variation.

From the Games-Howell procedure, mean differences in *work-time scheduling* between employees from Asian or Asian British compared to those from any other ethnic background were statistically significant (p-values < 0.05), while other pairwise comparisons were not statistically significant (p-values > 0.05) (Figure 6.7 (d)). Asian or Asian British employees were less aware of and had poorer access to other forms of *work-time scheduling* in the UK employee population.

Figure 6.8: Mean Comparison of Latent Trait Scores by Age Group

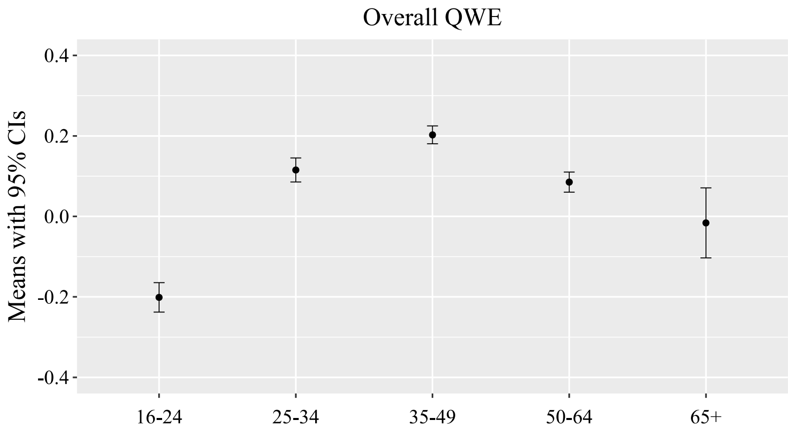


Figure 6.8 (a)

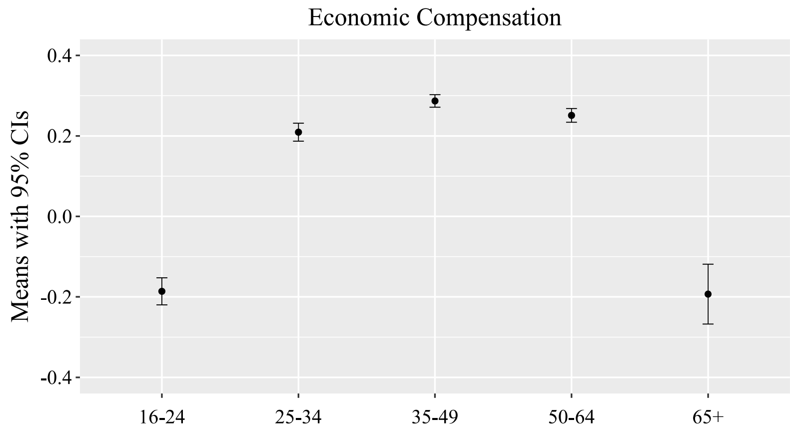


Figure 6.8 (b)

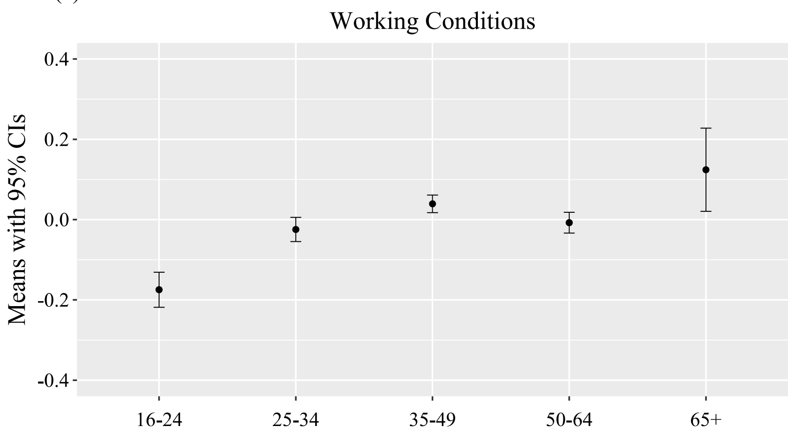


Figure 6.8 (c)

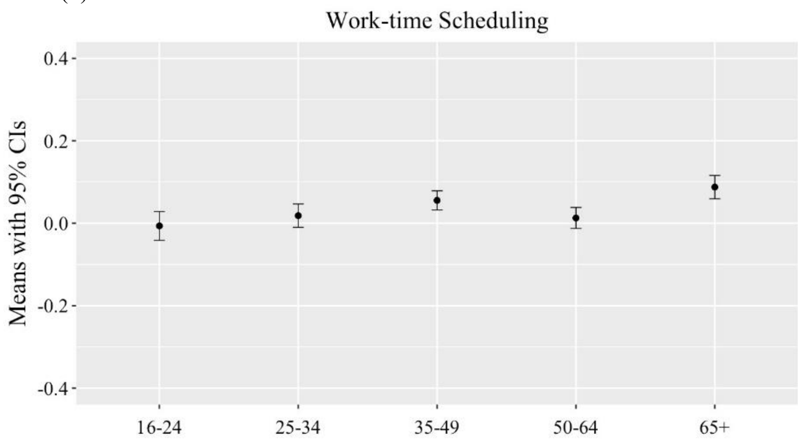


Figure 6.8 (d)

Age Group

Overall QWE by Age Group

The relationship between *overall QWE* and age group in this sample was concave curvilinear. On average, employees aged 16 – 24 years old (M = –0.201, SD = 0.754 and Mdn = –0.216, IQR = 1.026) and 65+ years old (M = –0.016, SD = 0.816 and Mdn = –0.003, IQR = 1.338) had lower levels, while those aged 35 – 49 years old (M = 0.203, SD = 0.904 and Mdn = 0.198, IQR = 1.355) had the highest (Appendix 6.5).

The homogeneity of variances assumption in the scores by age group was untenable (F = 32.65, $p < 0.001$), and a one-way ANOVA test with a Welch's correction suggested that mean differences between at least two groups were statistically significant (Welch's F (4, 2389.15) = 88.13, $p < 0.001$, $\omega^2 = 0.021$). However, age had a small effect size and explained approximately 2.1% of the variation.

Based on the Games-Howell procedure, mean differences in *overall QWE* by age group were statistically significant for all pairwise comparisons (p-values < 0.05), except for those aged 25 – 34 v 50 – 64 years old (p = 0.550), and 50 – 64 v 65+ years old (p = 0.181) (Figure 6.8 (a)). Thus, those aged 16 – 24 years old had the poorest *overall QWE* in the UK employee population.

Economic Compensation by Age Group

The association between *economic compensation* and age group was also concave curvilinear. Employees aged 65+ years old (M = –0.193, SD = 0.736 and Mdn = –0.040, IQR = 1.310) and 16 – 24 years old (M = –0.186, SD = 0.696 and Mdn = –0.025, IQR = 1.268) having on average, lower levels, while those aged 35 – 49 years old (M = 0.287, SD = 0.638 and Mdn = 0.402, IQR = 0.660) had the highest (Appendix 6.6).

The assumption of homogeneity of variances in the scores by age group was untenable (F = 46.60, $p < 0.001$). A one-way ANOVA test with a Welch's correction suggested that mean differences between at least two groups were statistically significant (Welch's F (4, 2336.03) = 189.93, $p < 0.001$, $\omega^2 = 0.044$). In terms of the magnitude of differences, age had a small effect size and explained approximately 4.4% of the variation.

The Games-Howell procedure indicated that mean differences in *economic compensation* by age group were statistically significant for all pairwise comparisons (p-values < 0.05), except for those aged 16 – 24 v 65+ years old ($p > 0.05$) (Figure 6.8 (b)). Thus, 16 – 24 and 65+ years old had the lowest *economic compensation* in the UK employee population.

Working Conditions by Age Group

Considering *working conditions*, on average, employees aged 16 – 24 years old (M = –0.175, SD = 0.902 and Mdn = –0.147, IQR = 1.252) had the lowest levels, while those aged 65+ years old (M = 0.124, SD = 1.055 and Mdn = 0.270, IQR = 1.566) had the highest levels (Appendix 6.7).

The homogeneity of variances assumption in the scores by age group was untenable (F = 12.66, $p < 0.001$). A one-way ANOVA test with a Welch's correction indicated that mean differences between at least two groups were statistically significant (Welch's F (4, 2349.86) = 20.31, $p < 0.001$, $\omega^2 = 0.005$). Age had a negligible effect size and explained approximately 0.5% of the variation.

From the Games-Howell procedure, mean differences in *working conditions* between employees aged 16 – 24 years old compared to employees in other age groups (all p-values $<$ 0.001), and those aged 25 – 34 v 35 – 49 years old ($p = 0.007$) were statistically significant. Other pairwise comparisons were not statistically significant (p-values > 0.05) (Figure 6.8 (c)). Employees aged 16 – 24 years old had the poorest *working conditions* in the UK employee population.

Work-time Scheduling by Age Group

On average, there were slight differences in *work-time scheduling* levels by age group in this sample. Based on the mean, employees aged 16 – 24 years old (M = –0.006, SD = 0.721) had the lowest levels, while those aged 35 – 49 years old (M = 0.077, SD = 0.826) had the highest. However, based on the median, employees aged 50 – 44 years old (Mdn = 0.027, IQR = 1.287) had the lowest levels, while those aged 65+ years old (Mdn = 0.051, IQR = 0.631) had the highest (Appendix 6.8).

The assumption of homogeneity of variances in the scores by age group was untenable ($F = 30.82$, $p < 0.001$). A one-way ANOVA test with a Welch's correction indicated that mean differences between at least two groups were statistically significant (Welch's F (4, 2433.66) = 5.25, $p < 0.001$, $\omega^2 = 0.001$). Age had a negligible effect size and explained approximately 0.1% of the variation.

From the Games-Howell procedure, mean differences in *work-time scheduling* between employees aged 16 – 24 v 35 – 49 years old ($p < 0.001$), and 25 – 34 v 35 – 49 years old (p = 0.036) were statistically significant in the UK employee population, while other pairwise comparisons were statistically insignificant (p-values > 0.05) (Figure 6.8 (d)).

Socio-demographic Characteristics

Table 6.4 summaries results of effect size estimates and Levene's tests evaluating the homogeneity of variances assumption for socio-demographic characteristics. For Levene's tests, if homogeneity of variances was tenable, a t-test or ANOVA test assuming equal variances between groups was estimated, and if untenable, tests with a Welch's correction were estimated. In relation to post hoc procedures for ANOVA tests, the Tukey-Kramer's procedure was used where equality of variances was tenable, and the Games-Howell procedure was used where equality of variances was untenable. The table also includes an indication of whether the association between these procedure characteristics and the latent traits were significant or not.

Table 6.4: Summary Table of Levene's Tests and Effect Size Estimates for Socio-demographic Characteristics

Characteristic	Overall QWE		Economic compensation		Working conditions		Work-time scheduling	
	Levene's test	Effect size	Levene's test	Effect size	Levene's test	Effect size	Levene's test	Effect size
Relationship status[b]	Untenable	Small***	Untenable	Small***	Untenable	Negligible***	Untenable	Negligible**
Parental status[b]	Untenable	Negligible***	Untenable	Negligible***	Tenable	Negligible**	Untenable	Negligible**
Illness or disability[a]	Untenable	Negligible**	Untenable	Negligible	Untenable	Negligible**	Tenable	Negligible***
Region[b]	Untenable	Small***	Untenable	Negligible***	Untenable	Negligible*	Tenable	Negligible***

Notes: [a]: *t*-test. [b]: ANOVA test. Levene's test: If equality of variances is tenable, a *t*-test or ANOVA test assuming equal variances between groups is estimated, and if untenable, a *t*-test or ANOVA test with a Welch's correction is estimated. For post hoc procedures relating to ANOVA tests, the Tukey-Kramer is used where equality of variances is tenable, and the Games-Howell procedure is used where equality of variances is untenable. Negligible: Predictor explains less than 1% of the variation in the latent trait. Small: Predictor explains between 1% and 9% of the variation in the latent trait. Moderate: Predictor explains between 9% and 25% of the variation in the latent trait. Large: Predictor explains more than 25% of the variation in the latent trait. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

Figure 6.9: Mean Comparison of Latent Trait Scores by Relationship Status

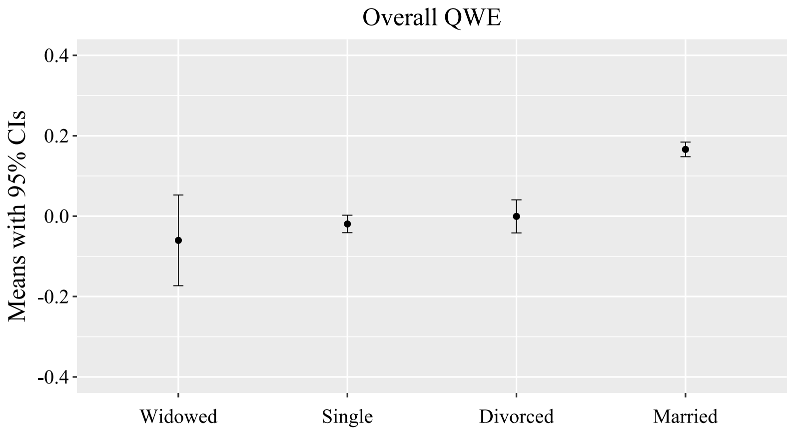


Figure 6.9 (a)

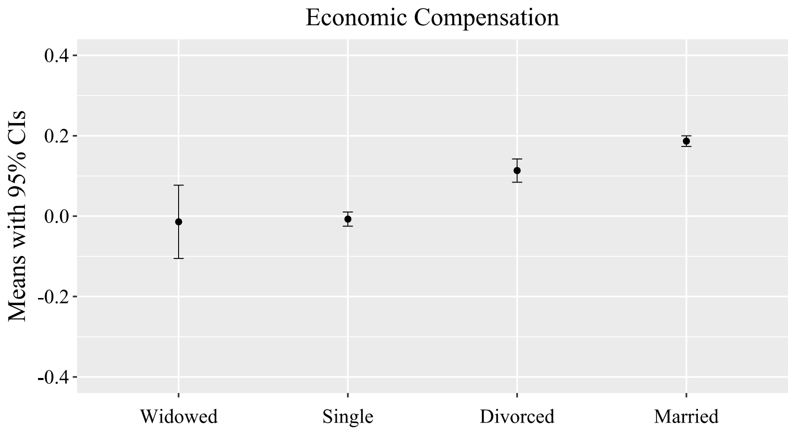


Figure 6.9 (b)

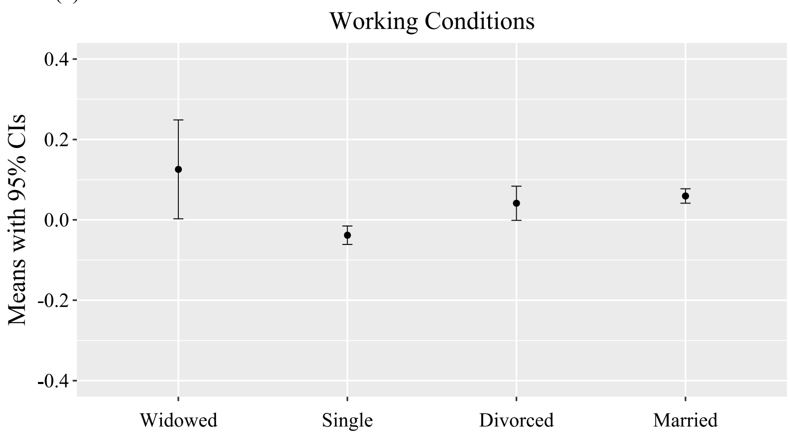


Figure 6.9 (c)

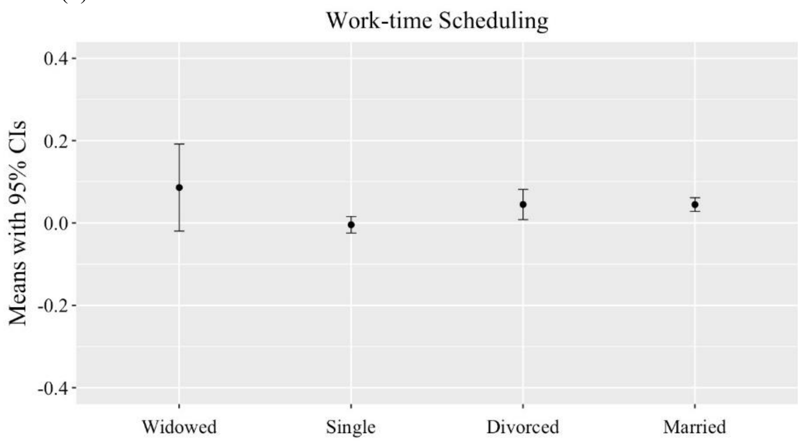


Figure 6.9 (d)

Relationship Status

Overall QWE by Relationship Status

On average, in this sample, married or cohabiting employees (M = 0.166, SD = 0.882 and Mdn = 0.184, IQR = 1.316) had the highest *overall QWE*, while levels for other groups were similar, with widowed employees (M = –0.060, SD = 0.838 and Mdn = –0.055, IQR = 1.329) having the lowest (Appendix 6.5).

The homogeneity of variances assumption in the scores by relationship status was untenable ($F = 19.63$, $p < 0.001$). A one-way ANOVA test with a Welch's correction suggested statistically significant mean differences between at least two groups (Welch's F (3, 976.26) = 62.01, $p < 0.001$, $\omega^2 = 0.011$). The magnitude of the differences was small and relationship status explained approximately 1.1% of the variation.

Based on the Games-Howell procedure, mean differences in *overall QWE* between married or cohabiting employees compared to other groups were statistically significant (p-values < 0.001), while other pairwise comparisons were not statistically significant (p-values > 0.05) (Figure 6.9 (a)). Thus, married or cohabiting employees had better *overall QWE* than employees in any other relationship status in the UK employee population.

Economic Compensation by Relationship Status

For *economic compensation*, widowed (M = –0.014, SD = 0.677 and Mdn = 0.092, IQR = 0.745) and single (M = –0.007, SD = 0.670 and Mdn = 0.135, IQR = 0.708) employees in this sample had, on average, lower levels, while married or cohabiting employees (M = 0.187, SD = 0.641 and Mdn = 0.314, IQR = 0.662) had the highest (Appendix 6.6).

The assumption of homogeneity of variances in the scores by relationship status was untenable ($F = 16.10$, $p < 0.001$). A one-way ANOVA test with a Welch's correction indicated means differences between at least two groups were statistically significant (Welch's F (3, 974.44) = 101.02, $p < 0.001$, $\omega^2 = 0.018$). Relationship status had a small effect size and explained approximately 1.8% of the variation.

The Games-Howell procedure indicated mean differences in *economic compensation* by relationship status were statistically significant for all pairwise comparisons (p-values < 0.05), except for single v widowed employees (p-values > 0.05) (Figure 6.9 (b)). Single and widowed employees had the poorest *economic compensation* in the UK employee population.

Working Conditions by Relationship Status

Considering *working conditions*, in this sample and on average, single employees (M = –0.038, SD = 0.873 and Mdn = –0.054, IQR = 1.306) had the lowest levels, while widowed employees (M = 0.126, SD = 0.912 and Mdn = 0.184, IQR = 1.302) had the highest (Appendix 6.7).

The homogeneity of variances assumption in the scores by relationship status was untenable ($F = 3.84$, $p = 0.009$). A one-way ANOVA test with a Welch's correction suggested statistically significant mean differences between at least two groups (Welch's F (3, 971.53) = 15.44, $p < 0.001$, $\omega^2 = 0.003$). However, relationship status had a negligible effect size and explained approximately 0.3% of the variation.

From the Games-Howell procedure, mean differences in *working conditions* between single compared to divorced/separated employees ($p = 0.007$) or married/cohabiting employees ($p < 0.001$) were statistically significant, while other pairwise comparisons were statistically insignificant (p-values > 0.05) (Figure 6.9 (c)). Single employees had among the poorest *working conditions* in the UK employee population.

Work-time Scheduling by Relationship Status

Regarding *work-time scheduling*, there were slight differences in levels by relationship status in the sample. Single employees (M = –0.004, SD = 0.761 and Mdn = 0.007, IQR =

1.173) had the lowest levels, while widowed employees (*M* = 0.086, *SD* = 0.784 and *Mdn* = 0.043, *IQR* = 0.993) had the highest (Appendix 6.8).

The assumption of homogeneity of variances in the scores by relationship status was untenable ($F = 13.69$, $p < 0.001$), and a one-way ANOVA test with a Welch's correction indicated statistically significant mean differences between at least two groups (Welch's F (3, 975.83) = 5.24, $p = 0.001$, $\omega^2 = 0.001$). The magnitude of the differences was negligible and relationship status explained approximately 0.1% of the variation.

The Games-Howell procedure indicated that mean differences in *work-time scheduling* between single v married or cohabiting employees ($p = 0.001$) were statistically significant in the UK employee population, while differences in other pairwise comparisons were statistically insignificant (*p*-values > 0.05) (Figure 6.9 (d)).

Figure 6.10: Mean Comparison of Latent Trait Scores by Parental Status

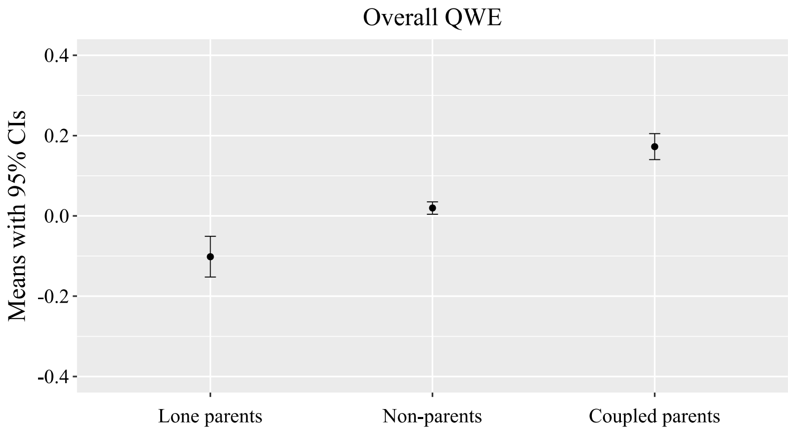


Figure 6.10 (a)

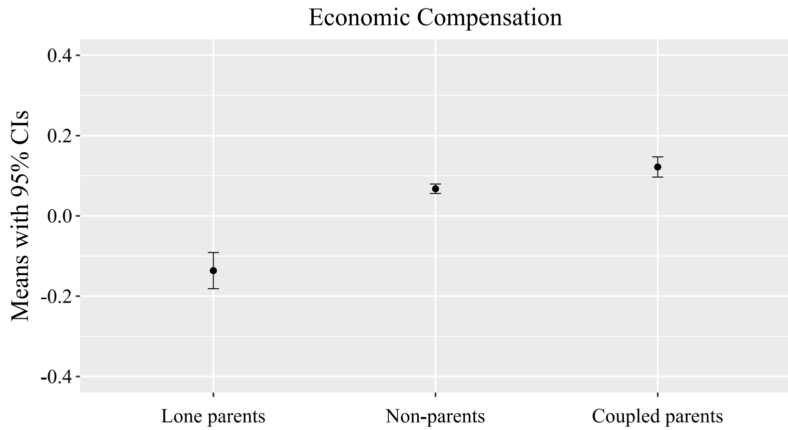


Figure 6.10 (b)

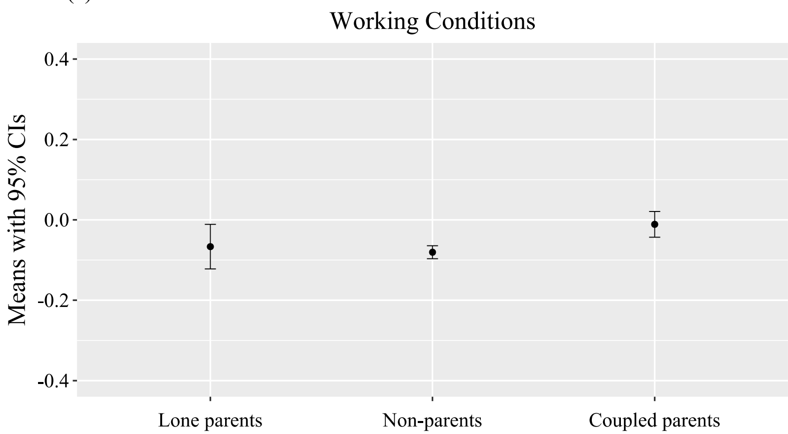


Figure 6.10 (c)

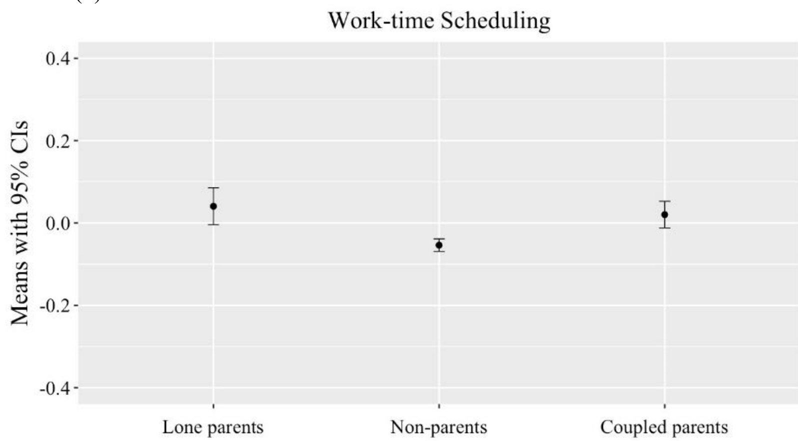


Figure 6.10 (d)

Parental Status

Overall QWE by Parental Status

On average, in this sample, lone parents with school age children (M = –0.102, SD = 0.843 and Mdn = –0.094, IQR = 1.185) had the lowest levels of *overall QWE* compared to those in other parental groups, with coupled parents with school age children (M = 0.172, SD = 0.890 and Mdn = 0.168, IQR = 1.304) having the highest (Appendix 6.5).

The homogeneity of variances assumption in the scores by parental status was untenable (F = 3.51, p = 0.030). A one-way ANOVA test with a Welch's correction indicated statistically significant mean differences between at least two groups (Welch's F (2, 2528.33) = 50.59, p < 0.001, ω^2 = 0.006). Parental status had a negligible effect size and explained approximately 0.6% of the variation.

Based on the Games-Howell post hoc procedure, mean differences in *overall QWE* between all pairwise comparisons by parental status were statistically significant (p-values < 0.001) (Figure 6.10 (a)). Lone parents with school age children had the poorest *overall QWE* in the UK employee population.

Economic Compensation by Parental Status

In terms of *economic compensation*, lone parents with school age children in this sample (M = –0.136, SD = 0.749 and Mdn = 0.031, IQR = 0.760) had on average, the lowest levels compared to other parental groups, while coupled parents with school age children (M = 0.122, SD = 0.689 and Mdn = 0.260, IQR = 0.700) had the highest levels (Appendix 6.6).

The assumption of homogeneity of variances in the scores by parental status was untenable (F = 23.67, p < 0.001), and a one-way ANOVA test with a Welch's correction suggested mean differences between at least two groups were statistically significant (Welch's F (2, 2456.95) = 48.25, p < 0.001, ω^2 = 0.006). However, parental status had a negligible effect size and explained approximately 0.6% of the variation.

The Games-Howell procedure suggested that mean differences in *economic compensation* between all pairwise comparisons by parental status were statistically significant (p-values < 0.001) (Figure 6.10 (b)). Thus, lone parents with school age children had the poorest *economic compensation* in the UK employee population.

Working Conditions by Parental Status

Regarding *working conditions*, there were slight differences in levels by parental status in the sample, with coupled parents with school age children ($M = –0.011$, $SD = 0.884$ and $Mdn = –0.018$, $IQR = 1.405$) having the highest levels. Lone parents with school age children ($M = –0.067$, $SD = 0.921$ and $Mdn = –0.093$, $IQR = 1.472$) and employees with no school age children ($M = –0.081$, $SD = 0.913$ and $Mdn = –0.064$, $IQR = 1.408$) had similar levels (Appendix 6.7).

The homogeneity of variances in the scores by parental status was tenable ($F = 1.73$, $p = 0.177$). A one-way ANOVA test with equal variances suggested statistically significant mean differences between at least two groups were (F (2, 16324) = 6.87, $p = 0.001$, $\omega^2 = 0.001$). However, the magnitude of the differences was negligible and parental status explained approximately 0.1% of the variation.

The Tukey-Kramer procedure, mean differences in *working conditions* between coupled parents with school age children v employees with no school age children were statistically significant ($p < 0.001$), while other pairwise comparisons were statistically insignificant (p-values > 0.05) (Figure 6.10 (c)). Coupled parents with school age children had better *working conditions* than employees in other parental groups in the UK employee population.

Work-time Scheduling by Parental Status

For *work-time scheduling*, there were also slight differences between groups in the sample on average, with employees with no school age children (M = 0.004, SD = 0.761 and Mdn = 0.007, IQR = 1.173) having the lowest levels. Levels for lone parents with school age children (M = 0.041, SD = 0.745 and Mdn = 0.025, IQR = 1.000) and coupled parents with school age children (M = 0.050, SD = 0.811 and Mdn = –0.003, IQR = 1.346) were similar (Appendix 6.8).

The homogeneity of variances assumption in the scores by parental status was untenable (F = 14.14, $p < 0.001$). A one-way ANOVA test with a Welch's correction suggested statistically significant mean differences between at least two groups (Welch's F (2, 2539.32) = 5.08, p = 0.006, ω^2 = 0.001). However, parental status had a negligible effect size and explained approximately 0.1% of the variation.

From the Games-Howell procedure, mean differences in *work-time scheduling* between coupled parents with school age children v employees with no school age children (p = 0.010) were statistically significant, while other pairwise comparisons were statistically insignificant (p-values > 0.05) (Figure 6.10 (d)). Thus, coupled parents with school age children were more aware of and had better access to other forms of *work-time scheduling* than employees in other parental groups in the UK employee population.

Figure 6.11: Mean Comparison of Latent Trait Scores by Illness or Disability

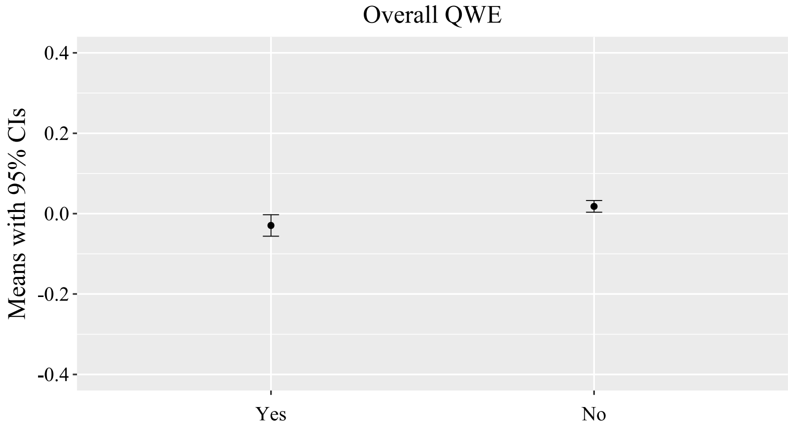


Figure 6.11 (a)

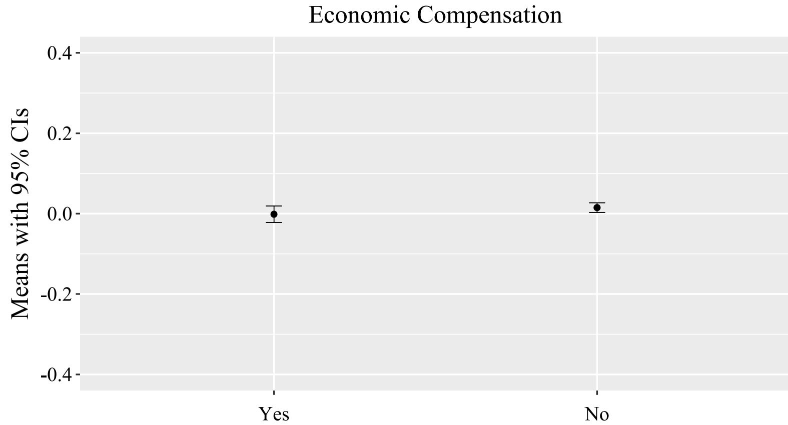


Figure 6.11 (b)

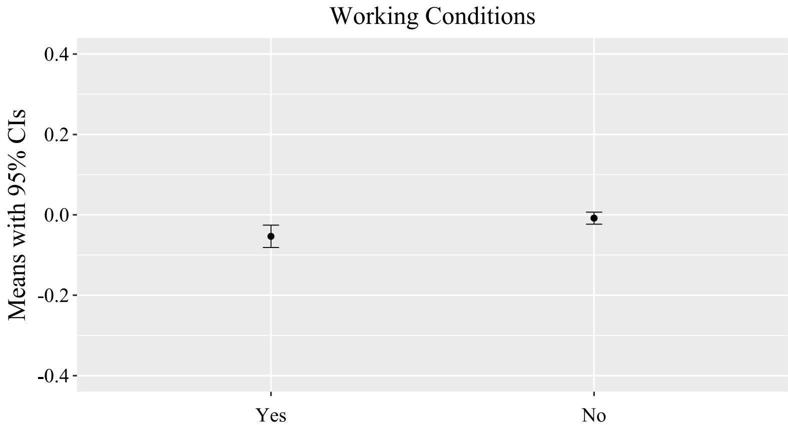


Figure 6.11 (c)

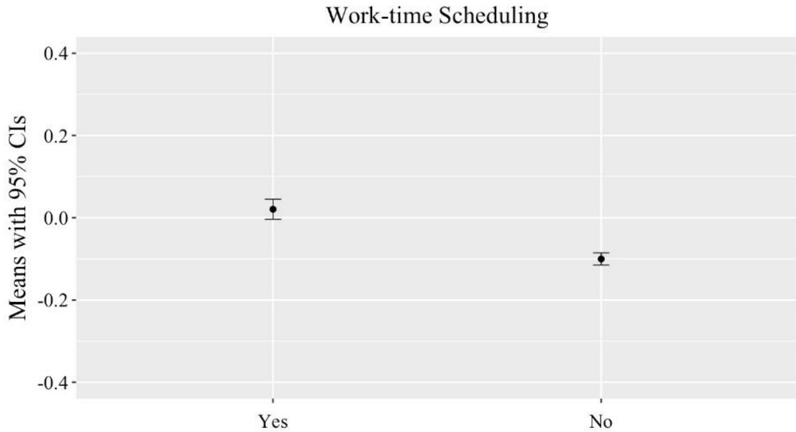


Figure 6.11 (d)

Longstanding Illness or Disability

Overall QWE by Illness or Disability

On average, employees with longstanding illness or disability in this sample had slightly lower *overall QWE* (M = –0.029, SD = 0.860 and Mdn = –0.047, IQR = 1.292) than those without (M = 0.018, SD = 0.833 and Mdn = 0.009, IQR = 1.205) (Appendix 6.5).

The homogeneity of variances assumption in the scores by longstanding illness or disability was untenable (F = 13.02, $p < 0.001$). A two-sample t-test with a Welch's correction indicated statistically significant mean differences by longstanding illness or disability (Welch's t (6490.15) = 3.05, p = 0.002, ω^2 = 0.001) (Figure 6.11 (a)). This suggested that those with a longstanding illness or disability had slightly poorer *overall QWE* than those without in the UK employee population. However, longstanding illness or disability had a negligible effect size and explained approximately 0.1% of the variation.

Economic Compensation by Illness or Disability

In this sample, *economic compensation* levels were, on average, similar by longstanding illness or disability based on either the mean or median (M = –0.002, SD = 0.662 and Mdn = 0.119, IQR = 0.658 for those with a longstanding illness or disability v M = 0.015, SD = 0.683 and Mdn = 0.156, IQR = 0.692 for those without) (Appendix 6.6).

The homogeneity of variances assumption in the scores by longstanding illness or disability was untenable (F = 7.08, p = 0.005). A two-sample t-test with a Welch's correction suggested mean differences between the groups were statistically insignificant (Welch's t (6842.59) = 1.36, p = 0.173, ω^2 = 0.0001) (Figure 6.11 (b)). This indicated that there was no difference in levels of *economic compensation* by longstanding illness or disability in the UK employee population.

Working Conditions by Illness or Disability

On average, employees with a longstanding illness or disability in this sample had slightly lower levels of *working conditions* (M = –0.053, SD = 0.893 and Mdn = –0.042, IQR = 1.391) than those without (M = –0.008, SD = 0.850 and Mdn = –0.026, IQR = 1.315) (Appendix 6.7).

The homogeneity of variances assumption in the scores by longstanding illness or disability was untenable (F = 20.57, $p < 0.001$). A two-sample t-test with a Welch's correction indicated statistically significant mean differences (Welch's t (6400.85) = 2.81, p = 0.005, ω^2 = 0.001) (Figure 6.11 (c)). Employees with a longstanding illness or disability had slightly poorer *working conditions* than those without in the UK employee population. However, this had negligible effect size and explained approximately 0.1% of the variation.

Work-time Scheduling by Illness or Disability

For *work-time scheduling*, employees with a longstanding illness or disability in this sample had slightly higher levels (M = 0.021, SD = 0.783 and Mdn = 0.004, IQR = 1.316), than those without (M = –0.040, SD = 0.782 and Mdn = –0.043, IQR = 1.280) (Appendix 6.8).

The assumption of homogeneity of variances in the scores by longstanding illness or disability was tenable (F = 0.61, p = 0.436), and a two-sample t-test with equal variances indicated statistically significant mean differences (t (16359) = –4.27, $p < 0.001$, ω^2 = 0.0011) (Figure 6.11 (d)). This suggested that employees with a longstanding illness or disability were more aware of and had better access to other forms of *work-time scheduling* than those without in the UK employee population. However, longstanding illness or disability had negligible effect size and explained approximately 0.1% of the variation.

Figure 6.12: Mean Comparison of Latent Trait Scores by Region

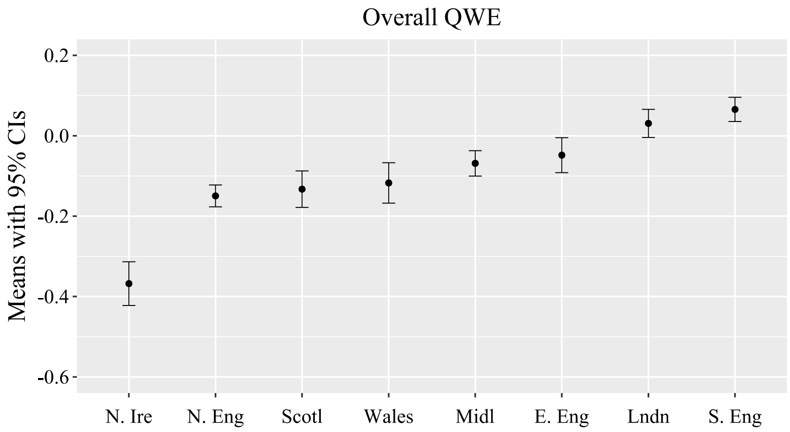


Figure 6.12 (a)

Figure 6.12 (b)

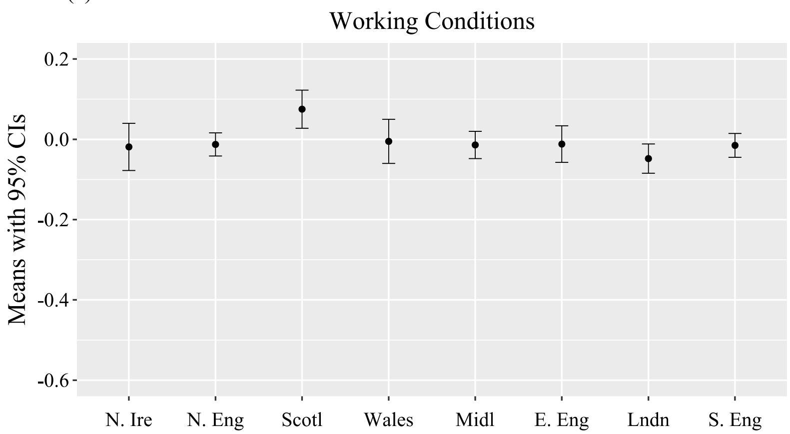


Figure 6.12 (c)

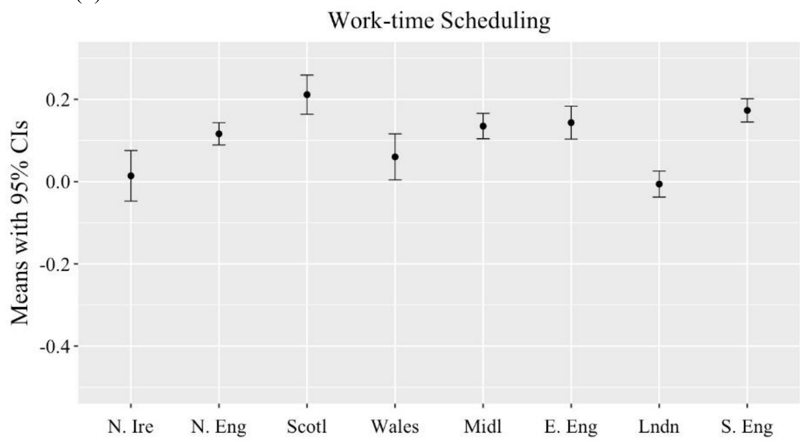


Figure 6.12 (d)

Region

Overall QWE by Region

Considering *overall QWE* by region in this sample, on average, employees in Northern Ireland (M = –0.368, SD = 0.872 and Mdn = –0.439, IQR = 1.392) had the lowest levels. On the other hand, those in London (M = 0.031, SD = 0.847 and Mdn = 0.013, IQR = 1.192) and Southern England (M = 0.066, SD = 0.876 and Mdn = 0.046, IQR = 1.284) had higher levels than those in other regions (Appendix 6.5).

The assumption of homogeneity of variances in the scores by region was untenable (F = 4.00, p < 0.001). A one-way ANOVA test with a Welch's correction indicated statistically significant mean differences between at least two groups (Welch's F (7, 5433.86) = 38.60, p < 0.001, ω^2 = 0.016). In terms of magnitude, region had a small effect size and explained approximately 1.6% of the variation in *overall QWE*.

The Games-Howell procedure indicated that mean differences in *overall QWE* between employees in London v Southern England, or East of England, East of England v The Midlands, Wales, or Scotland, The Midlands v Wales, or Scotland, Northern England v Wales, or Scotland, and Wales v Scotland were statistically insignificant (p-values > 0.05), while all other pairwise comparisons were statistically significant (p-values < 0.01) (Figure 6.12 (a)). Thus, employees in Northern Ireland had the poorest *overall QWE* in the UK employee population.

Economic Compensation by Region

On average, there were slight differences in levels of *economic compensation* by region in this sample, with employees in Northern Ireland (M = –0.051, SD = 0.747 and Mdn = 0.124, IQR = 0.832) having the lowest levels, while those in Scotland (M = 0.090, SD = 0.666 and Mdn = 0.205 IQR = 0.697) had the highest (Appendix 6.6).

The homogeneity of variances assumption in the scores by region was untenable (F = 6.29, p < 0.001). A one-way ANOVA test with a Welch's correction suggested mean differences between at least two groups were statistically significant (Welch's F (7, 5420.09) = 9.98, $p <$ 0.001, $\omega^2 = 0.004$). However, the magnitude of the differences was negligible, and region explained approximately 0.4% of the variation.

The Games-Howell procedure indicated statistically significant mean differences for employees from Scotland or East of England compared to those in other regions, except for Southern England. On the other hand, mean differences between employees in Southern England, London, and Wales were statistically insignificant, while for those in London, Wales, the Midlands, Northern England, and Northern Ireland were also statistically insignificant (Figure 6.12 (b)). This suggested that employees in Scotland, East of England or Southern England had among the better *economic compensation* in the UK employee population.

Working Conditions by Region

In terms of *working conditions*, there were slight differences in levels by region in the sample. Employees in Scotland ($M = 0.075$, $SD = 0.913$ and $Mdn = 0.104$, $IQR = 1.446$) had the highest levels on average, while for those in London ($M = –0.048$, $SD = 0.887$ and $Mdn = –0.041$, $IQR = 1.344$) had the least (Appendix 6.7).

The assumption of homogeneity of variances in the scores by region was untenable ($F = 2.43$, $p = 0.017$), and a one-way ANOVA test with a Welch's correction indicated statistically significant mean differences between at least two groups were (Welch's F (7, 5417.92) = 2.40, $p = 0.019$, $\omega^2 = 0.001$). However, region had a negligible effect size and explained approximately 0.1% of the variation.

From the Games-Howell procedure, mean differences in *working conditions* between employees in London v Scotland ($p = 0.002$), Southern England v Scotland ($p = 0.035$), and Northern England v Scotland ($p = 0.042$) were statistically significant, while all other pairwise comparisons were statistically insignificant (p-values > 0.05) (Figure 6.12 (c)). Employees in Scotland had among the better *working conditions* in the UK employee population.

Work-time Scheduling by Region

For *work-time scheduling*, in this sample and on average, employees in London ($M = -0.006$, $SD = 0.769$ and $Mdn = 0.006$, $IQR = 1.174$) and Northern Ireland ($M = 0.002$, $SD = 0.822$ and $Mdn = 0.029$, $IQR = 1.169$) had lower levels than those in other regions. Those in Scotland ($M = 0.111$, $SD = 0.825$ and $Mdn = 0.086$, $IQR = 1.393$) and Southern England ($M = 0.119$, $SD = 0.800$ and $Mdn = 0.091$, $IQR = 1.330$) had higher levels (Appendix 6.8).

The homogeneity of variances assumption in the scores by region was tenable ($F = 1.99$, $p = 0.053$). A one-way ANOVA test with equal variances suggested statistically significant mean differences between at least two groups (F (7, 16758) = 6.40, $p < 0.001$, $\omega^2 = 0.002$). Region had a negligible effect size and explained approximately 0.2% of the variation.

Based on the Tukey-Kramer procedure, mean differences in *work-time scheduling* between employees in London v Southern England ($p < 0.001$), The Midlands ($p = 0.006$), Northern England ($p = 0.022$), or Scotland ($p < 0.001$), Southern England v Northern Ireland ($p = 0.002$), and Scotland v Northern Ireland ($p = 0.021$) were statistically significant. All other pairwise comparisons were statistically insignificant (p-values > 0.05) (Figure 6.12 (d)). Employees in London or Northern Ireland had among the least awareness of and poorest access to other forms of *work-time scheduling* in the UK employee population.

Socio-economic Characteristics

Table 6.5 summaries results of effect size estimates and Levene's tests evaluating the homogeneity of variances assumption for socio-economic characteristics. For Levene's tests, if homogeneity of variances was tenable, a t-test or ANOVA test assuming equal variances between groups was estimated, and if untenable, tests with a Welch's correction were estimated. In relation to post hoc procedures for ANOVA tests, the Tukey-Kramer's procedure was used where equality of variances was tenable, and the Games-Howell procedure was used where equality of variances was untenable. The table also includes an indication of whether the association between these procedure characteristics and the latent traits were significant or not.

Table 6.5: Summary Table of Levene's Tests and Effect Size Estimates for Socio-economic Characteristics

Characteristic	Overall QWE		Economic compensation		Working conditions		Work-time scheduling	
	Levene's test	Effect size	Levene's test	Effect size	Levene's test	Effect size	Levene's test	Effect size
Education[b]	Untenable	Small***	Untenable	Small***	Untenable	Small**	Untenable	Small***
Occupational classification[b]	Untenable	Moderate***	Untenable	Large***	Untenable	Small***	Untenable	Small***
Full or part-time[a]	Untenable	Small***	Untenable	Moderate***	Untenable	Negligible***	Untenable	Small***
Organisational sector[a]	Untenable	Negligible	Untenable	Moderate***	Untenable	Negligible	Untenable	Moderate***
Organisation size[b]	Untenable	Small***	Untenable	Moderate***	Untenable	Small***	Untenable	Small***

Notes: [a]: *t*-test. [b]: ANOVA test. Levene's test: If equality of variances is tenable, a *t*-test or ANOVA test assuming equal variances between groups is estimated, and if untenable, a *t*-test or ANOVA test with a Welch's correction is estimated. For post hoc procedures relating to ANOVA tests, the Tukey-Kramer is used where equality of variances is tenable, and the Games-Howell procedure is used where equality of variances is untenable. Negligible: Predictor explains less than 1% of the variation in the latent trait. Small: Predictor explains between 1% and 9% of the variation in the latent trait. Moderate: Predictor explains between 9% and 25% of the variation in the latent trait. Large: Predictor explains more than 25% of the variation in the latent trait. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$.

Figure 6.13: Mean Comparison of Latent Trait Scores by Education

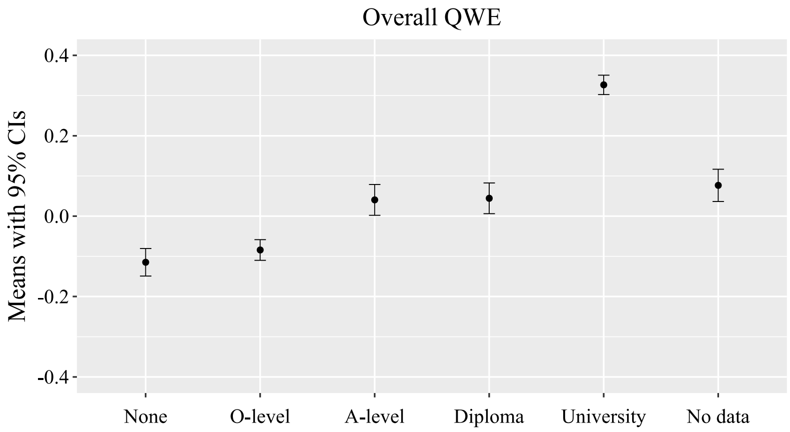


Figure 6.13 (a)

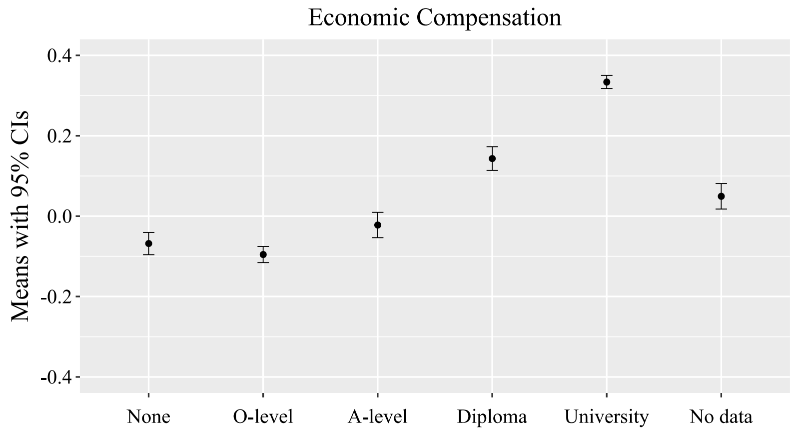


Figure 6.13 (b)

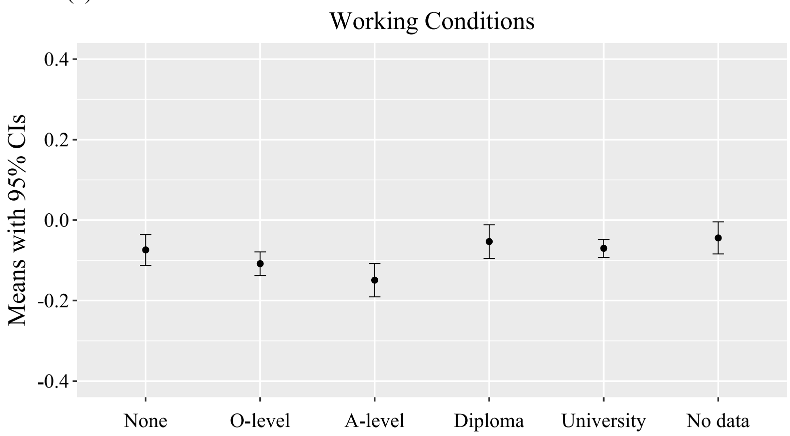


Figure 6.13 (c)

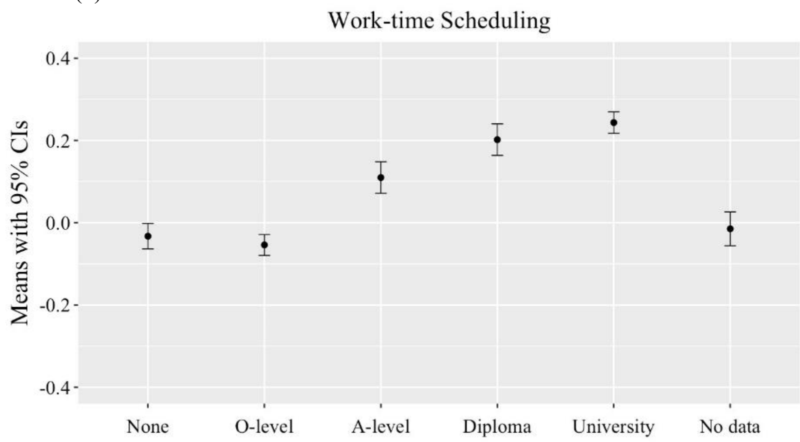


Figure 6.13 (d)

Education

Overall QWE by Education

In relation to *overall QWE* by education in this sample, on average, employees with no qualifications ($M = –0.115$, $SD = 0.829$ and $Mdn = –0.107$, $IQR = 1.233$) had the lowest levels compared to those in other educational groups, while those with a university or higher degree ($M = 0.327$, $SD = 0.884$ and $Mdn = 0.361$, $IQR = 1.375$) having the highest (Appendix 6.5).

The assumption of homogeneity of variances in the scores by education was untenable ($F = 16.45, p < 0.001$). A one-way ANOVA test with a Welch's correction suggested statistically significant mean differences in *overall QWE* between at least two groups (Welch's F (5, 6166.52) = 137.81, $p < 0.001$, $\omega^2 = 0.039$). However, education had a small effect size and explained approximately 3.9% of the variation.

From the Games-Howell procedure, mean differences in *overall QWE* between employees with up to A-level, or up to diploma in higher education (p-values > 0.05), and no qualifications v GCSE / O-level or lower (p-values > 0.05) were not statistically significant. All other pairwise comparisons were statistically significant in the UK employee population (p-values < 0.001) (Figure 6.13 (a)).

Economic Compensation by Education

For *economic compensation* in this sample, employees with GCSE / O-level or lower ($M = –0.096$, $SD = 0.638$ and $Mdn = –0.033$, $IQR = 0.835$), or no qualifications ($M = –0.068$, $SD = 0.667$ and $Mdn = 0.000$, $IQR = 0.892$) had on average, lower levels than those with other educational qualifications. Those with a university or higher degree ($M = 0.334$, $SD = 0.600$ and $Mdn = 0.444$, $IQR = 0.722$) had the highest (Appendix 6.6).

The homogeneity of variances assumption in the scores by education was untenable ($F = 26.96$, $p < 0.001$). A one-way ANOVA test with a Welch's correction indicated statistically significant mean differences between at least two groups (Welch's F (5, 6054.29) = 273.40, p

< 0.001, $\omega^2 = 0.075$). However, the magnitude of the differences was small, and education explained approximately 7.5% of the variation.

The Games-Howell procedure indicated statistically significant mean differences in *economic compensation* between all pairwise comparisons (all *p*-values < 0.05), except for employees with up to A-level v no qualifications, and those with no qualifications v GCSE / O-level or lower (*p*-values > 0.05) (Figure 6.13 (b)). Employees with a university or higher degree had better *economic compensation* than those in other educational groups in the UK employee population.

Working Conditions by Education

On average, there were slight differences in *working conditions* by educational groups in the sample. Employees with up to A-level ($M = –0.149$, $SD = 0.885$ and $Mdn = –0.150$, $IQR = 1.317$) had the lowest levels, while those with up to a diploma in higher education had higher levels ($M = –0.053$, $SD = 0.894$ and $Mdn = –0.047$, $IQR = 1.458$) (Appendix 6.7).

The assumption of homogeneity of variances in the scores by education was untenable ($F = 17.34$, $p < 0.001$). A one-way ANOVA test with a Welch's correction suggested mean differences between at least two groups were statistically significant (Welch's F (5, 6093.89) = 3.95, $p = 0.001$, $\omega^2 = 0.001$), but education had a negligible effect size. Thus, it explained approximately 0.1% of the variation.

From the Games-Howell procedure, mean differences between employees with up to A-level v up to a diploma in higher education ($p = 0.017$) and up to A-level v university or higher degree ($p = 0.013$) were statistically significant, while other pairwise comparisons were statistically insignificant (*p*-values > 0.05) (Figure 6.13 (c)). Employees with up to A-level qualifications had among the poorest *working conditions* in the UK employee population.

Work-time Scheduling by Education

Regarding *work-time scheduling*, in this sample, employees with no qualifications ($M = -0.033$, $SD = 0.747$ and $Mdn = 0.012$, $IQR = 1.090$) or GCSE / O-level or lower ($M = -0.012$, $SD = 0.765$ and $Mdn = 0.032$, $IQR = 0.892$) had on average, lower levels than those in other educational groups. On the other hand, those with a university or higher degree had the highest ($M = 0.194$, $SD = 0.859$ and $Mdn = 0.099$, $IQR = 1.404$) (Appendix 6.8).

The assumption of homogeneity of variances in the scores by education was untenable ($F = 24.93$, $p < 0.001$). A one-way ANOVA test with a Welch's correction indicated that mean differences between at least two groups were statistically significant (Welch's F (5, 6181.44) = 41.43, $p < 0.001$, $\omega^2 = 0.012$). Education had a small effect size and explained approximately 1.2% of the variation.

From the Games-Howell procedure, mean differences between all pairwise comparisons by education were statistically significant (p-values < 0.05), except for employees with up to A-level v up to a diploma in higher education, no qualifications v GCSE / O-level or lower (p-values > 0.05) (Figure 6.13 (d)). Thus, employees with no qualifications or GCSE / O-level or lower had among the least awareness of and poorest access to other forms of *work-time scheduling* in the UK employee population.

Figure 6.14: Mean Comparison of Latent Trait Scores by Occupational Classification

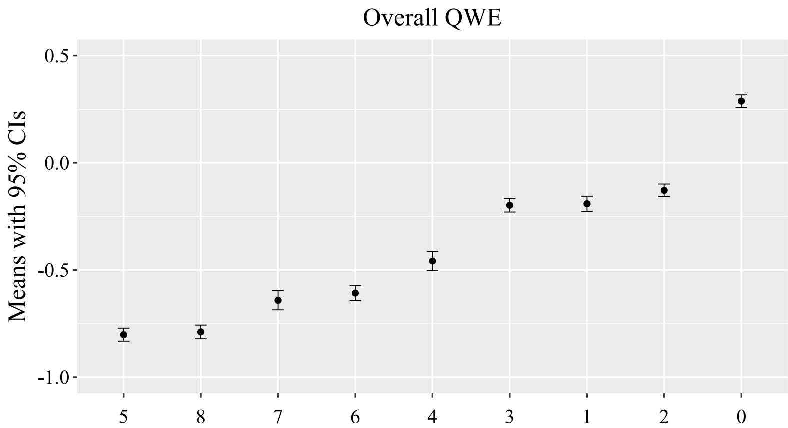


Figure 6.14 (a)

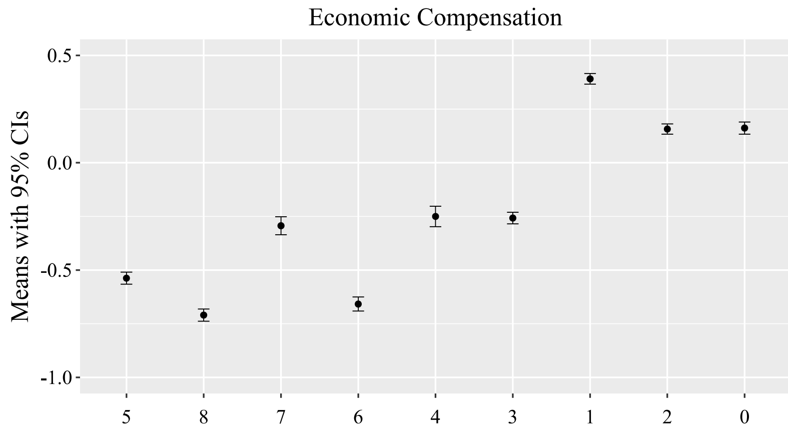


Figure 6.14 (b)

Figure 6.14 (c)

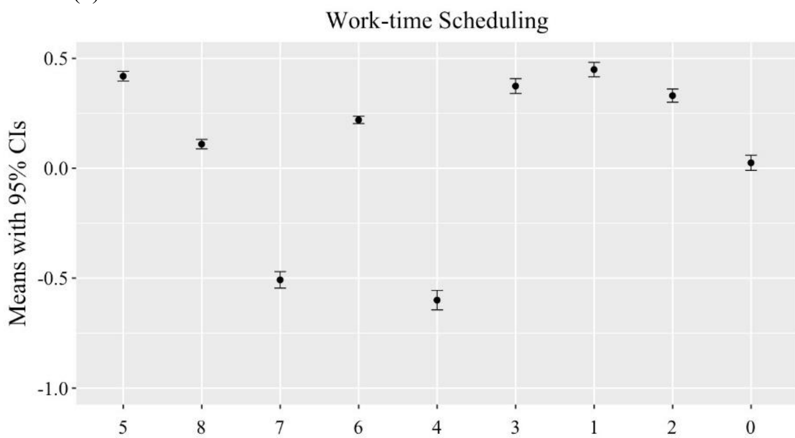


Figure 6.14 (d)

Notes: 0 = managers and senior officials, 1 = professional occupations, 2 = associate professional and technical occupations, 3 = administrative and secretarial occupations, 4 = skilled trades occupations, 5 = personal service occupations, 6 = sales and customer service occupations, 7 = process, plant and machine operatives, and 8 = elementary occupations.

Occupational Classification

Overall QWE by Occupational Classification

On average, in this sample, employees in personal service (M = –0.802, SD = 0.660 and Mdn = –0.808, IQR = 1.076) and elementary (M = –0.789, SD = 0.678 and Mdn = –0.750, IQR = 1.057) occupations had among the lowest levels of *overall QWE* than those in other occupational groups. In contrast managers and senior officials (M = 0.288, SD = 0.735 and Mdn = 0.324, IQR = 1.033) had the highest (Appendix 6.5).

The homogeneity of variances assumption in the scores by occupational classification was untenable ($F = 57.56$, $p < 0.001$). A one-way ANOVA test with a Welch's correction indicated that mean differences between at least two groups were statistically significant (Welch's F (8, 5800.56) = 525.96, $p < 0.001$, $\omega^2 = 0.203$). This indicated that occupational classification had a moderate effect size and explained approximately 20.3% of the variation.

The Games-Howell procedure indicated that mean differences between employees in professional v associate professional and technical, or administrative and secretarial occupations, personal service v elementary occupations, and sales and customer services occupations v process, plant and machine operatives were not statistically significant (all p-values > 0.05). All other pairwise comparisons were statistically significant (p-values < 0.01). This indicated that managers and senior officials had better *overall QWE* than employees in other occupational groups in the UK employee population (Figure 6.14 (a)).

Economic Compensation by Occupational Classification

In terms of *economic compensation*, in this sample and on average, employees in elementary occupations (M = –0.710, SD = 0.606 and Mdn = –0.754, IQR = 0.797) had the lowest levels compared to employees in other occupational groups, while those in professional occupations (M = 0.391, SD = 0.632 and Mdn = 0.490, IQR = 0.821) had the highest levels (Appendix 6.6).

Variances of the scores by occupational classification were not homogenous ($F = 26.07$, $p < 0.001$) and a one-way ANOVA test with a Welch's correction indicated that mean differences between at least two groups were statistically significant (Welch's F (8, 5739.67) = 768.68, $p < 0.001$, $\omega^2 = 0.272$). This suggested that occupational classification had a large effect size and explained approximately 27.2% of the variation.

Based on the Games-Howell procedure, mean differences between managers and senior officials v associate professional and technical occupations, administrative and secretarial occupations v skilled trades occupations, or process, plant and machine operatives, skilled trades occupations v process, plant and machine operatives, and sales and customer services occupations v elementary occupations were not statistically significant (p-values > 0.05). All other pairwise comparisons were statistically significant (p-values < 0.001) (Figure 6.14 (b)). Employees in professional occupations had better economic compensation compared to those in other occupational groups in the UK employee population.

Working Conditions by Occupational Classification

For *working conditions*, in this sample and on average, employees working as process, plant and machine operatives ($M = –0.424$, $SD = 0.986$ and $Mdn = –0.450$, $IQR = 1.402$) and those in sales and customer service occupations ($M = –0.423$, $SD = 0.932$ and $Mdn = –0.467$, $IQR = 1.268$) had lower levels than those in other occupational groups, while managers and senior officials ($M = 0.172$, $SD = 0.782$ and $Mdn = 0.408$, $IQR = 1.213$) had the highest (Appendix 6.7).

The homogeneity of variances assumption in the scores by occupational classification was untenable ($F = 17.99$, $p < 0.001$). A one-way ANOVA test with a Welch's correction suggested statistically significant mean differences between at least two groups (Welch's F (8, 5699.24) = 90.72, $p < 0.001$, $\omega^2 = 0.042$). However, occupational classification had a small effect size and explained approximately 4.2% of the variation.

The Games-Howell procedure indicated that mean differences in *working conditions* between employees in professional v skilled trades occupations, associate professional and technical v administrative and secretarial, skilled trades, or personal services occupations, administrative and secretarial v skilled trades, personal services, or elementary occupations, skilled trades v personal services occupations, sales and customer services v process, plant and machine operatives, or elementary occupations, and process, plant and machine operatives v elementary occupations were not statistically significant (all p-values > 0.05). All other pairwise comparisons were statistically significant (p-values < 0.01), and managers and senior officials had better *working conditions* than employees in other occupational groups in the UK employee population (Figure 6.14 (c)).

Work-time Scheduling by Occupational Classification

On average, in this sample, employees in skilled trades occupations (M = –0.344, SD = 0.633 and Mdn = –0.665, IQR = 0.931) and those working as process, plant and machine operatives (M = –0.334, SD = 0.617 and Mdn = –0.646, IQR = 0.969) had among the lowest levels of *work-time scheduling* than employees in other occupations. On the other hand, those in personal services (M = 0.241, SD = 0.705 and Mdn = 0.182, IQR = 0.754), administrative and secretarial (M = 0.243, SD = 0.818 and Mdn = 0.172, IQR = 1.339), and professional occupations (M = 0.267, SD = 0.846 and Mdn = 0.178, IQR = 1.435) had among the highest levels (Appendix 6.8).

The assumption of homogeneity of variances in the scores by occupational classification was untenable (F = 82.83, p < 0.001). A one-way ANOVA test with a Welch's correction indicated that mean differences between at least two groups were statistically significant (Welch's F (8, 5873.98) = 158.74, p < 0.001, ω^2 = 0.071). However, occupational classification had a small effect size and explained approximately 7.1% of the variation.

Based on the Games-Howell procedure, mean differences between employees in elementary occupations v managers and senior officials, professional v associate professional and technical, administrative and secretarial, or personal services occupations, associate professional and technical v administrative and secretarial, or personal services occupations, administrative and secretarial v personal services occupations, and skilled trades occupations v process, plant and machine operatives, were not statistically significant (all p-values > 0.05), while all other pairwise comparisons were statistically significant (p-values < 0.05). Employees in skilled trades occupations and those who worked as process, plant and machine operatives had among the least awareness of and poorest access to other forms of *work-time scheduling* than employees in other occupations in the UK employee population (Figure 6.14 (d)).

Figure 6.15: Mean Comparison of Latent Trait Scores by Full or Part-time

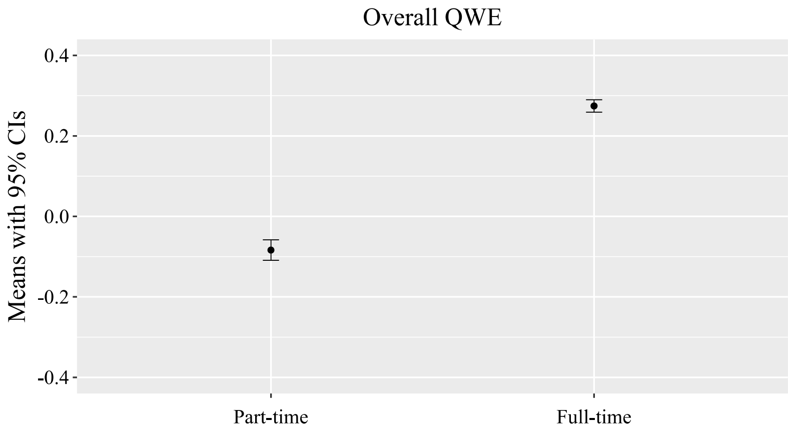


Figure 6.15 (a)

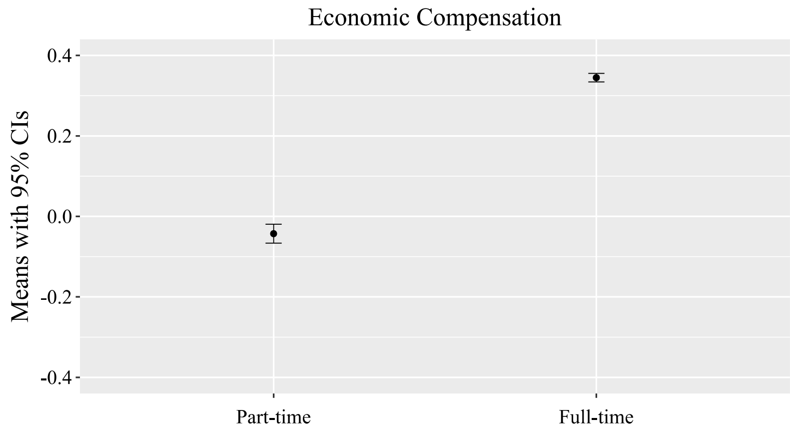


Figure 6.15 (b)

Figure 6.15 (c)

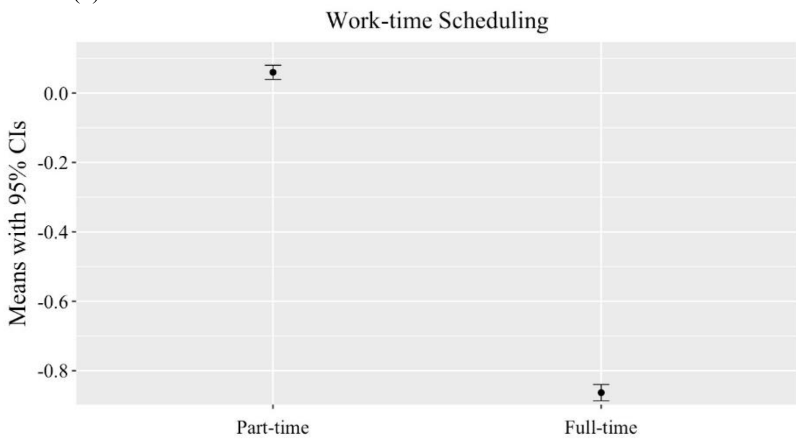


Figure 6.15 (d)

Full or Part-time

Overall QWE by Full or Part-time

In this sample, on average, employees in part-time employment (M = –0.084, SD = 0.798 and Mdn = –0.086, IQR = 1.160) had lower levels *overall QWE* than those in full-time employment (M = 0.274, SD = 0.875 and Mdn = 0.264, IQR = 1.304) (Appendix 6.5).

The homogeneity of variances assumption in the scores by full or part-time employment was untenable (F = 65.55, p < 0.001), and a two-sample t-test with a Welch's correction indicated statistically significant mean differences (Welch's t (6736.79) = 23.55, p < 0.001, ω^2 = 0.076) (Figure 6.15 (a)). This suggested that part-time employees had poorer *overall QWE* than full-time employees in the UK employee population. However, the magnitude of the differences was small, and full or part-time employment explained approximately 7.6% of the variation.

Economic Compensation by Full or Part-time

In terms of *economic compensation*, part-time employees in this sample had on average, lower levels than full-time employees (M = –0.043, SD = 0.732 and Mdn = 0.102, IQR = 1.286 for part-time v M = 0.597, SD = 0.678 and Mdn = 0.425, IQR = 0.667 for full-time) (Appendix 6.6).

The assumption of homogeneity of variances in the scores by full or part-time employment was untenable (F = 457.78, p < 0.001). A two-sample t-test with a Welch's correction suggested mean differences were statistically significant (Welch's t (5366.19) = 29.61, p < 0.001, ω^2 = 0.140) (Figure 6.15 (b)), and part-time employees had poorer *economic compensation* than full-time employees in the UK employee population. Full or part-time employment had a moderate effect size and explained approximately 14.0% of the variation.

Working Conditions by Full or Part-time

Regarding *working conditions*, in this sample and on average, employees in part-time employment (M = –0.116, SD = 0.909 and Mdn = –0.130, IQR = 1.341) had lower levels than those in full or part-time employment (M = –0.020, SD = 0.854 and Mdn = –0.076, IQR = 1.336) (Appendix 6.7).

The homogeneity of variances assumption in the scores by full or part-time employment was untenable (F = 15.23, p < 0.001), and a two-sample t-test with a Welch's correction suggested statistically significant mean differences (Welch's t (5916.24) = 5.75, p < 0.001, ω^2 = 0.005) (Figure 6.15 (c)). This indicated that part-time employees had poorer *working conditions* than full-time employees in the UK employee population; however, the effect size was negligible. Thus, full or part-time employment explained approximately 0.5% of the variation.

Work-time Scheduling by Full or Part-time

On average, part-time employees in the sample had higher levels of *work-time scheduling* than full-time employees (M = 0.060, SD = 0.643 and Mdn = –0.029, IQR = 0.576 for part-time v M = –0.863, SD = 1.338 and Mdn = –0.613, IQR = 2.286 for full-time) (Appendix 6.8).

The assumption of homogeneity of variances in the scores by full or part-time employment was untenable (F = 1099.8, p < 0.001). A two-sample t-test with a Welch's correction indicated statistically significant mean differences (Welch's t (8245.03) = –26.44, p < 0.001, ω^2 = 0.078) (Figure 6.15 (d)). This suggested that part-time employees were more aware of and had better access to other forms of *work-time scheduling* than full-time employees in the UK employee population. However, full or part-time employment had a small effect size and explained approximately 7.8% of the variation.

Figure 6.16: Mean Comparison of Latent Trait Scores by Organisational Sector

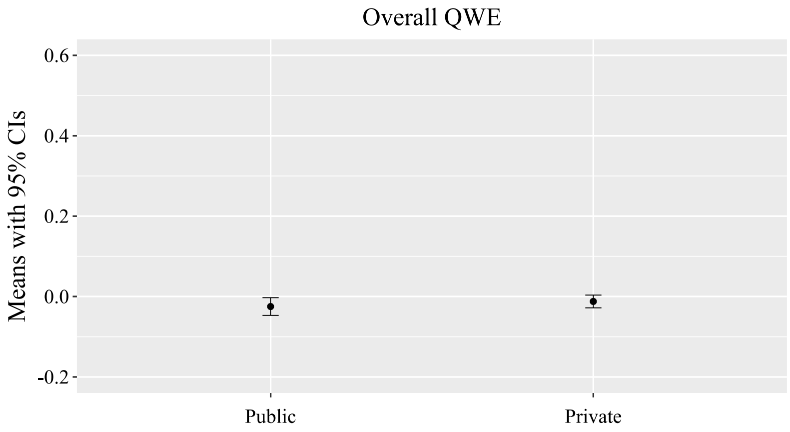


Figure 6.16 (a)

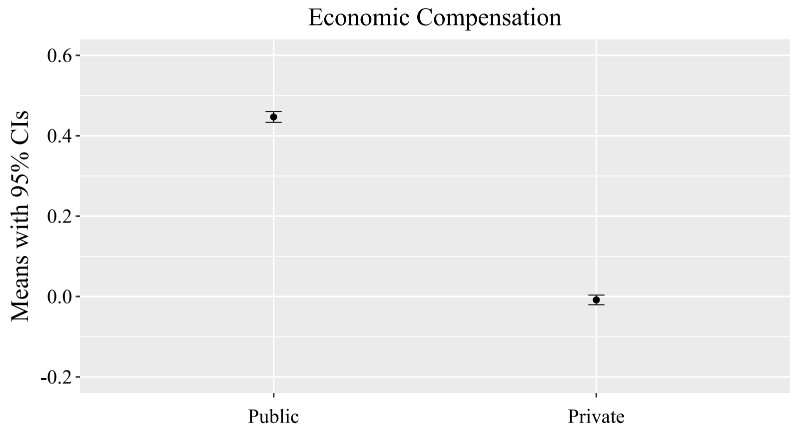


Figure 6.16 (b)

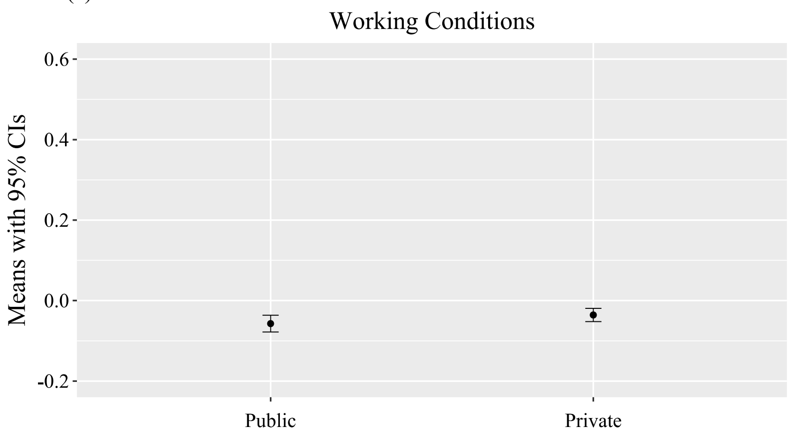


Figure 6.16 (c)

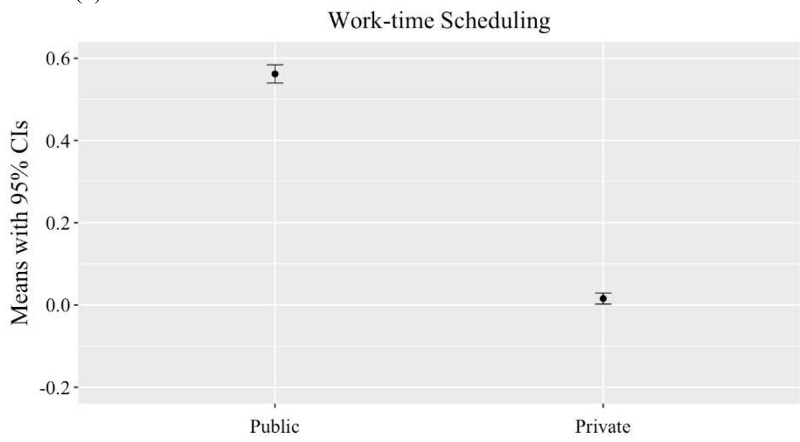


Figure 6.16 (d)

Organisational Sector

Overall QWE by Organisational Sector

On average, in this sample, *overall QWE* was comparable by organisational sector; thus, $M = –0.025$, $SD = 0.865$ and $Mdn = –0.054$, $IQR = 1.352$ for public sector employees compared to $M = –0.012$, $SD = 0.825$ and $Mdn = –0.025$, $IQR = 1.165$ for private sector employees (Appendix 6.5).

There was a violation of the assumption of homogeneity of variances in the scores by organisational sector ($F = 47.37$, $p < 0.001$), and a two-sample *t*-test with a Welch's correction suggested mean differences were statistically insignificant (Welch's t (12030.36) = 0.91, p = 0.361, $\omega^2 = –0.00001$) (Figure 6.16 (a)). This indicated that there was no difference in *overall QWE* by organisational sector in the UK employee population.

Economic Compensation by Organisational Sector

For *economic compensation*, in this sample, employees in the public sector ($M = 0.447$, $SD = 0.527$ and $Mdn = 0.521$, $IQR = 0.739$) had on average, higher levels than those in the private sector ($M = –0.008$, $SD = 0.624$ and $Mdn = 0.063$, $IQR = 0.729$) (Appendix 6.6).

The assumption of homogeneity of variances in the scores by organisational sector was untenable ($F = 159.23$, $p < 0.001$). A two-sample *t*-test with a Welch's correction indicated statistically significant mean differences (Welch's t (14190.07) = –49.48, $p < 0.001$, $\omega^2 = 0.147$) (Figure 6.16 (b)). This suggested that private sector employees had poorer *economic compensation* than public sector employees in the UK employee population. Organisational sector had a moderate effect size and explained approximately 14.7% of the variation.

Working Conditions by Organisational Sector

On average, *working conditions* by organisational sector in this sample were comparable; thus, $M = –0.057$, $SD = 0.821$ and $Mdn = –0.084$, $IQR = 1.238$ for public sector employees compared to $M = -0.036$, $SD = 0.851$and $Mdn = -0.055$, $IQR = 1.329$ for private sector employees (Appendix 6.7).

The homogeneity of variances assumption in the scores by organisational sector was untenable ($F = 15.82$, $p < 0.001$). A two-sample t-test with a Welch's correction suggested mean differences were statistically insignificant (Welch's t (12857.42) = 1.59, $p = 0.112$, $\omega^2 = 0.0001$) (Figure 6.16 (c)). This indicated that there was no difference in *working conditions* by organisational sector in the UK employee population.

Work-time Scheduling by Organisational Sector

In terms of *work-time scheduling*, on average, public sector employees ($M = 0.562$, $SD = 0.874$ and $Mdn = 0.415$, $IQR = 1.221$) in this sample had higher levels than private sector employees ($M = 0.016$, $SD = 0.700$ and $Mdn = 0.042$, $IQR = 0.967$) (Appendix 6.8).

The homogeneity of variances assumption in the scores by organisational sector was untenable ($F = 712.10$, $p < 0.001$). A two-sample t-test with a Welch's correction indicated statistically significant mean differences (Welch's t (10425.98) = –41.18, $p < 0.001$, $\omega^2 = 0.140$) (Figure 6.16 (d)). This suggested that private sector employees were less aware of and had poorer access to other forms of *work-time scheduling* than public sector employees in the UK employee population. Organisational sector had a moderate effect size and explained approximately 14.0% of the variation.

Figure 6.17: Mean Comparison of Latent Trait Scores by Organisation Size

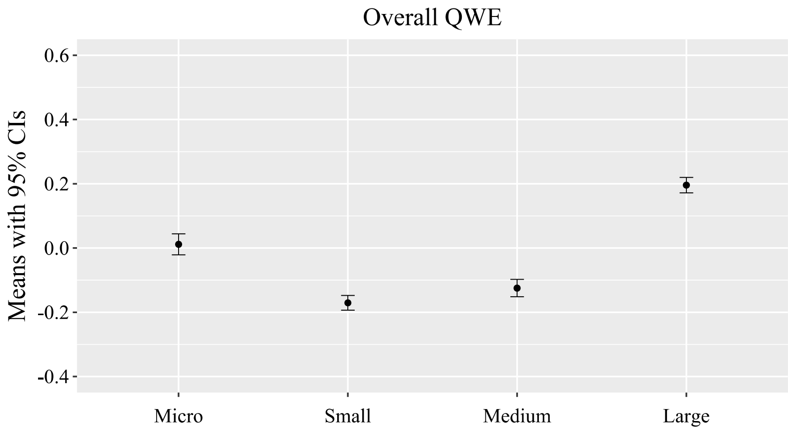


Figure 6.17 (a)

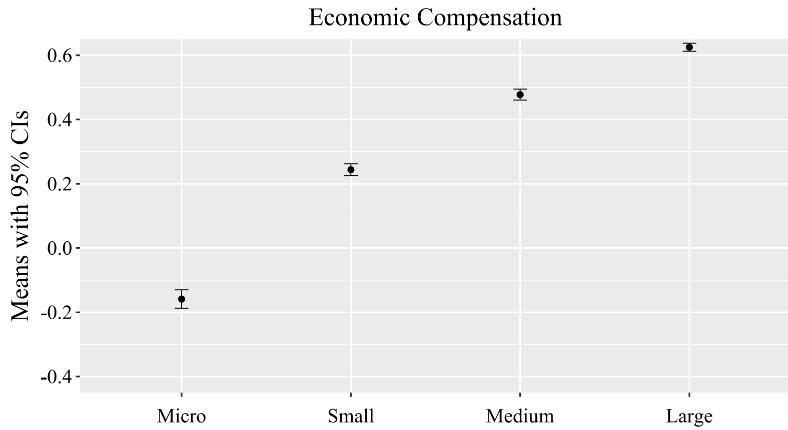


Figure 6.17 (b)

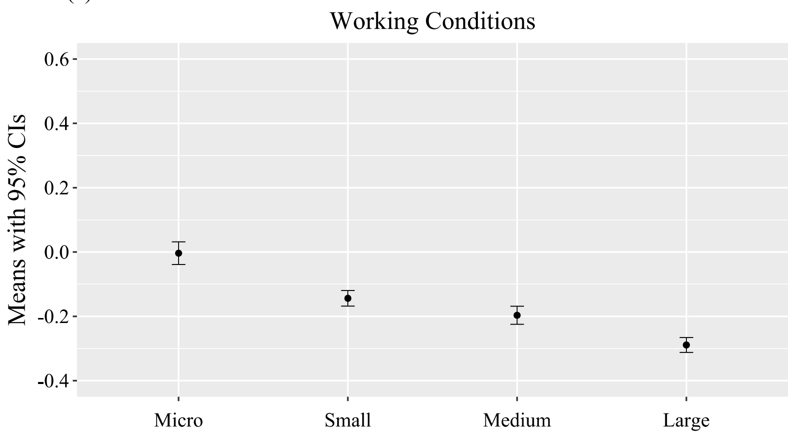


Figure 6.17 (c)

Figure 6.17 (d)

Organisation Size

Overall QWE by Organisation Size

In this sample, the association between *overall QWE* and organisation size was convex curvilinear, with employees in small size organisations having on average, the lowest levels ($M = –0.171$, $SD = 0.820$ and $Mdn = –0.195$, $IQR = 1.182$). On the other hand, those in micro ($M = 0.012$, $SD = 0.825$ and $Mdn = 0.020$, $IQR = 1.177$) and large size organisations ($M = 0.196$, $SD = 0.893$ and $Mdn = 0.216$, $IQR = 1.367$) had among the highest (Appendix 6.5).

The assumption of homogeneity of variances in the scores by organisation size was untenable ($F = 22.52$, $p < 0.001$). A one-way ANOVA test with a Welch's correction suggested statistically significant mean differences between at least two groups (Welch's F (3, 7957.15) = 179.75, $p < 0.001$, $\omega^2 = 0.031$). However, organisation size had a small effect size and explained approximately 3.1% of the variation.

From the Games-Howell procedure, mean differences between all pairwise comparisons by organisation size were statistically significant (p-values < 0.001), except for employees in small v medium size organisations ($p > 0.05$) (Figure 6.17 (a)). This suggested that employees in small or medium size organisations had among the poorest *overall QWE* in the UK employee population.

Economic Compensation by Organisation Size

On average, in this sample, levels of *economic compensation* increased with increasing organisation size. Thus, employees in micro size organisations ($M = –0.159$, $SD = 0.734$ and $Mdn = 0.038$, $IQR = 1.297$) had the lowest levels, while those in large size organisations ($M = 0.625$, $SD = 0.469$ and $Mdn = 0.638$ $IQR = 0.540$) had the highest (Appendix 6.6).

The homogeneity of variances assumption in the scores by organisation size was untenable ($F = 398.78$, $p < 0.001$), and a one-way ANOVA test with a Welch's correction suggested statistically significant mean differences between at least two groups (Welch's F (3, 7508.70) = 987.27, $p < 0.001$, $\omega^2 = 0.152$). Organisation size had a moderate effect size and explained approximately 15.2% of the variation.

The Games-Howell procedure suggested that mean differences in between all pairwise comparisons by organisation size were statistically significant (p-values < 0.001) (Figure 6.17 (b)). Thus, employees in micro size organisations had the poorest *economic compensation* in the UK employee population.

Working Conditions by Organisation Size

Regarding *working conditions*, in this sample and on average, levels decreased with increasing organisation size. Thus, employees in micro size organisations (M = –0.004, SD = 0.896 and Mdn = 0.045, IQR = 1.493) had the highest levels, while those in large size organisations (M = –0.289, SD = 0.865 and Mdn = –0.294, IQR = 1.325) had the lowest (Appendix 6.7).

The assumption of homogeneity of variances in the scores by organisation size was untenable ($F = 7.20$, $p < 0.001$). A one-way ANOVA test with a Welch's correction suggested mean differences between at least two groups were statistically significant (Welch's F (3, 7871.07) = 63.60, $p < 0.001$, $\omega^2 = 0.011$). However, organisation size had a small effect size and explained approximately 1.1% of the variation.

From the Games-Howell procedure, mean differences between all pairwise comparisons by organisation size were statistically significant (p-values < 0.05) (Figure 6.17 (c)). This indicated that employees in micro size organisations had the poorest *working conditions* in the UK employee population.

Work-time Scheduling by Organisation Size

On average, levels of *work-time scheduling* in this sample increased with increasing organisation size. Thus, employees in micro size organisations (M = –0.043, SD = 0.708 and

Mdn = 0.051, *IQR* = 0.989) had the lowest levels, while those in large size organisations (*M* = 0.327, *SD* = 0.903 and *Mdn* = 0.192, *IQR* = 1.474) had the highest (Appendix 6.8).

The homogeneity of variances assumption in the scores by organisation size was untenable ($F = 91.20, p < 0.001$). A one-way ANOVA test with a Welch's correction suggested statistically significant mean differences between at least two groups (Welch's F (3, 8124.32) = 135.53, $p < 0.001$, $\omega^2 = 0.024$). However, organisation size had a small effect size and explained approximately 2.4% of the variation.

The Games-Howell procedure suggested mean differences between all pairwise comparisons by organisation size were statistically significant (p-values < 0.05) (Figure 6.17 (d)). This indicated that employees in micro size organisations had the least awareness of and poorest access to other forms of *work-time scheduling* in the UK employee population.

6.3 Discussion

This chapter used the measurement instruments of QWE developed in Chapter 5 to investigate which predictors affected *overall QWE* and different *dimensions of QWE* in the UK employee population and examined between-group comparisons for each of these predictors. Prior to conducting between-group comparisons, DIF was used to evaluate measurement equivalence of the instruments to establish whether the items measured the same latent traits between groups. While DIF analyses indicated that some items measuring *overall* and different *dimensions of QWE* exhibited differential performance between some groups, the magnitude of DIF was negligible and between-group comparison was feasible. As the latent trait scores were estimated from a bifactor model, scores for *overall QWE* were interpreted given other *dimensions of QWE* in the measurement model, while scores for each of the *dimensions of QWE* were interpreted over and above *overall QWE*.

Regarding the relationship between the predictors and the latent traits, firstly, in terms of demographic characteristics and considering sex, results from the analysis supported previous literature that highlighted inequalities in the labour market by sex (Korpi 2018). Thus, in the UK employee population, on average, females had poorer *overall QWE* and *economic compensation* than males. On the other hand, males had on average, less awareness of and had poorer access to other forms of *work-time scheduling* compared to females, while there were no differences in *working conditions* by sex. However, sex had a small effect size in explaining variations in *overall QWE* or *work-time scheduling*, and a negligible effect size on *economic compensation*.

The study also supported evidence from literature about disparities in the labour market by ethnic group (Dillon 2020; Korpi 2018). However, while there were statistically significant differences in *overall QWE* and other *dimensions of QWE* in the UK employee population by ethnic group, this had a negligible effect size in explaining variations in these latent traits. On average, employees from a Black or Black British ethnic background had the poorest *overall QWE* and *working conditions*, while those from an Asian or Asian British ethnic background had the poorest *economic compensation* and *work-time scheduling*. On the other hand, employees from a Mixed (except for *working conditions*) or White ethnic background had among the best levels. Estimates for employees from a Mixed ethnic background were partly affected by the small sample size, which led to broad 95% CIs for the means and had an impact on the pairwise comparison relative to other ethnic groups.

In terms of age, results from this study were consistent with evidence that suggested variations in the different forms of precarious work by age (Kim and Kurz 2001), particularly for young workers (Arranz et al. 2019) but also for older workers. Mean differences in *overall QWE*, and other *dimensions of QWE* by age group were statistically significant in the UK employee population. The relationships between age group and *overall QWE* as well as *economic compensation* were concave curvilinear in nature, highlighting the disparities experienced by young (16 – 24 years old) and older (65+ years old) workers in the labour market. Furthermore, young workers had the poorest *working conditions* and also among the least awareness of and poorest access to other forms of *work-time scheduling*. However, age group had a small effect size in explaining variations in *overall QWE* or *economic compensation*, and a negligible effect size on *working conditions* or *work-time scheduling*. The relatively small sample size for employees aged 65+ years old also had an impact on the 95% CIs for the means and affected the pairwise comparison relative to other age groups.

Secondly, regarding socio-demographic characteristics and considering relationship status, evidence from this analysis supported the notion of a marriage premium in the labour market (Bardasi and Taylor 2008; Ribar 2004). Thus, associations between relationship status and *overall QWE*, and other *dimensions of QWE* were statistically significant in the UK employee population. Results suggested that married / cohabiting employees had better outcomes in the labour market; thus on average, they had higher levels of *overall QWE* and *economic compensation* compared to employees in other relationships, while they were also more aware of and had better access to other forms of *work-time scheduling* than single employees. On the other hand, single employees had poorer *working conditions* compared to married / cohabiting or divorced / separated employees. The relatively small sample size for widowed employees had an impact on the 95% CIs for the means and affected the pairwise comparison relative to other groups. In terms of practical significance, relationship status had a small effect size in explaining variations in *overall QWE* or *economic compensation*, and a negligible effect size on *working conditions* or *work-time scheduling*.

In terms of parental status, results from this analysis highlighted the disadvantages experienced by lone parents in the labour market (Esser and Olsen 2018; Klett-Davies 2016; Nieuwenhuis and Maldonado 2018). On average, coupled parents with children of primary school age had better *overall QWE* and *economic compensation*, while lone parents with primary school age children had the poorest. Coupled parents with children of primary school age also had better *working conditions* and were more aware of and had better access to other forms of *work-time scheduling* than employees with no primary school age children. While the associations in *overall QWE* and other *dimensions of QWE* were statistically significant in the UK employee population, the practical significance of parental status in explaining variations in these latent traits was negligible.

For longstanding illness or disability, differences in *overall QWE*, *working conditions* and *work-time scheduling* were statistically significant in the UK employee population. The findings supported evidence from literature that highlighted the challenges employees with a longstanding illness or disability encounter in the labour market (Grover and Piggott 2015; TUC 2021a). On average, employees with a longstanding illness or disability had poorer *overall QWE* and *working conditions* but were more aware of and had better access to other forms of *work-time scheduling* compared to employees without a longstanding illness or disability. However, longstanding illness or disability had a negligible effect size in explaining variations in *overall QWE* and other *dimensions of QWE*.

Considering regional differences, this analysis partly supported evidence from other literature that suggested disparities in the labour market within and across regions and nations of the UK (Jones and Green 2009). However, advantages were not limited to London and the South East regions. While there were statistically significant differences in *overall QWE* and other *dimensions of QWE* in the UK employee population, region had a small effect size in explaining variations in *overall QWE* and a negligible effect size on *economic compensation*, *working conditions* or *work-time scheduling*. On average, employees in Northern Ireland had the poorest *overall QWE*, with those in the East of England, London or Southern England having among the highest levels. On the other hand, employees in Scotland, East of England or Southern England had among the better *economic compensation*, while those in Scotland also had among the better *working conditions*. However, employees in London or Northern

Ireland had among the least awareness of and poorest access to other forms of *work-time scheduling*.

Lastly, in terms of socio-economic characteristics and firstly, considering education, there were statistically significant differences in *overall QWE* and other *dimensions of QWE* in the UK employee population. Results supported evidence from previous studies that higher levels of education resulted in human capital (Okay-Somerville and Scholarios 2013; Solomon et al. 2022) that afforded individuals greater job resources (Solomon et al. 2022). Thus, the analysis found that employees with a university or higher degree had better *overall QWE*, *economic compensation*, and were more aware of and had better access to other forms of *work-time scheduling* than those in other educational groups. There were slight differences in *working conditions* by education, with employees with GCSE / O-level or lower as well as those with up to A-level qualifications having among the poorest *working conditions*. However, in terms of practical significance, education had a small effect size in explaining variations in *overall QWE*, *economic compensation* or *work-time scheduling*, and a negligible effect size on *working conditions*.

In terms of occupational classification, results from this analysis generally supported evidence from previous literature which suggested better outcomes for high-skilled employees, partly attributed to skills differentials in the occupational hierarchy (Gallie 2015; Wheatley 2022). Mean differences of *overall QWE* and other *dimensions of QWE* by occupational classification were statistically significant in the UK employee population. On average, managers and senior officials had better *overall QWE* and *working conditions*, while those in professional occupations had better *economic compensation*. On the other hand, employees who worked in skilled trades occupations or as process, plant and machine operatives had the least awareness of and poorest access to other forms of *work-time scheduling*. While the practical significance of occupational classification explaining variations in *working conditions* or *work-time scheduling* was small, it had a moderate effect size explaining variations in *overall QWE* and a large effect size explaining variations in *economic compensation*.

Regarding full or part-time employment, results from this analysis supported previous literature highlighting the precarious nature of part-time compared to full-time employment (Lyonette et al. 2010; Warren and Lyonette 2015). There were statistically significant differences in *overall QWE* and other *dimensions of QWE* in the UK employee population. The results highlighted the disadvantages of non-standard forms of employment compared to standard forms. Thus, employees in part-time employment had poorer *overall QWE*, *economic compensation*, and *working conditions*, but were more aware of and had better access to other forms of *work-time scheduling* compared to employees in full-time employment. However, in terms of practical significance, full or part-time employment had a small effect size in explaining variations in *overall QWE* or *work-time scheduling*, a moderate effect size explaining variations in *economic compensation*, and a negligible effect size explaining variations in *working conditions*.

Considering organisational sector, results indicated statistically significant differences in *economic compensation* and *work-time scheduling* in the UK employee population. Results highlighted better *economic compensation* and more awareness of and better access to other forms of *work-time scheduling* for employees in public than private sector organisations. The study supported evidence from previous literature that highlighted better outcomes for employees by organisational sector and this included the public sector pay premium relative to the private sector (Cribb et al. 2014; Murphy et al. 2020). However, differences in *overall QWE* and *working conditions* by organisational sector were not statistically significant. Organisational sector had a moderate effect size in explaining variations in *economic compensation* or *work-time scheduling*.

Finally, in terms of organisation size, this analysis partly supported evidence from other literature that suggested better outcomes in some aspects of QWE for employees in large size organisations and better outcomes in some aspects for employees in micro size organisations (Bryson et al. 2021; Forth et al. 2006). Results suggested better *overall QWE* for employees in micro or large size organisations than those in small or medium size organisations. On the other hand, *economic compensation* and, the awareness of and access to other forms of *work-time scheduling* increased with increasing organisation size, while *working conditions* decreased with increasing organisation size. The differences were statistically significant differences in *overall QWE* and other *dimensions of QWE* in the UK employee population. However, organisation size had a small effect size in explaining variations in *overall QWE*, *working conditions* or *work-time scheduling*, while it had a moderate effect size on *economic compensation*.

In summary, this chapter has evaluated the measurement equivalence of the QWE measurement instrument for different group of employees. While some items exhibited differential performance between different groups, the magnitude of DIF was negligible. This meant that between-group comparison based on the measurement instruments was feasible. Group mean comparisons were conducted and effect sizes for each of the predictors were estimated. Results suggested that socio-economic characteristics explained more of the variation in *overall* or other *dimensions of QWE* than demographic or socio-demographic characteristics. While the analysis in this chapter focused on the relationship between each predictor of QWE, and *overall* and other *dimensions of QWE*, it did not capture the influence of other predictors on *overall* and other *dimensions of QWE*. The next chapter will model the relationship between *overall* and other *dimensions of QWE*, and each predictor while also controlling for other predictors.

6.4 Appendices

6.4.1 Appendix 6.1: Differential Item Functioning for Multidimensional Constructs

Re: DIF for Multidimensional Constructs

CS Choi, Seung W [redacted]
To Ndebele, Nhlanhla

Reply | Reply All | Forward

Mon 30/08/2021 00:21

Dear Nhlanhla,

If the bifactor model fits your data substantially better than the unidimensional model, I don't think you should use lordif as it assumes unidimensionality. With that said, you may want to run a separate DIF analysis for each specific factor, e.g., S4 and S5. The other factors are too small for a separate analysis.

Best wishes,

Seung

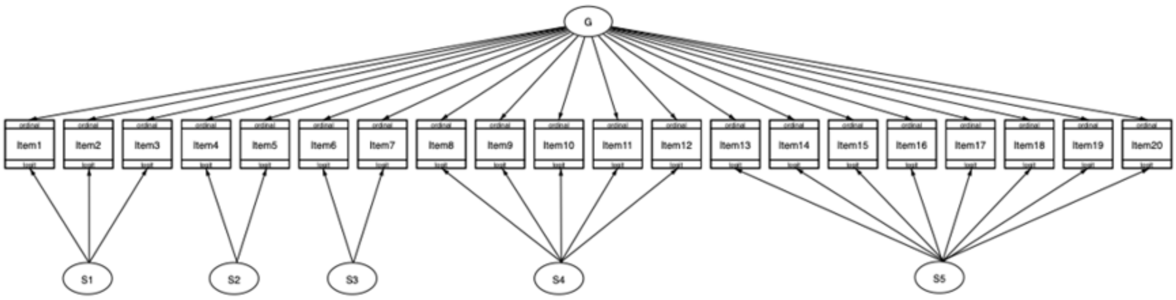
Seung W. Choi

[redacted]

From: "Ndebele, Nhlanhla" <Nhlanhla.Ndebele.2@city.ac.uk>
Date: Saturday, August 28, 2021 at 3:13 PM
To: "Choi, Seung W" [redacted]
Subject: DIF for Multidimensional Constructs

Dear Prof. Choi,

You helped me a couple of months ago with DIF for longitudinal data and would appreciate your expertise again but in terms of DIF for multidimensional constructs. I have estimated a bifactor IRT model (path diagram below) to measure some latent constructs and the model fits the data well.



I use your ***lordif*** package in R and have conducted DIF analyses using all the 20 items by different groups, however, this assumes the model is unidimensional. Is it conceptually accurate to analyse DIF in this way for such a model when each item is associated with 2 latent constructs? If this not accurate, how might I account for the specific factors (S_xs)?

I would really appreciate your advice?

Regards,

Nhlanhla

6.4.2 Appendix 6.2: Algorithm for DIF Detection with the '*lordif*' Package

3.2. Algorithm

In what follows, we will describe the algorithm used in the **lordif** package in more detail.

1. Data preparation: Check for sparse cells (rarely observed response categories; determined by a minimum cell count specified by the user (e.g., `minCell = 5`); collapse/recode response categories as needed based on the minimum cell size requirement specified.

2. IRT calibration: Fit the graded response model (using the `grm` function in **ltm**) to obtain a single set of item parameters for all groups combined.

3. Trait estimation: Obtain trait (ability) estimates using the expected a posteriori (EAP) estimator with omitted responses treated as not presented.

4. Logistic regression: Fit three (ordinal) logistic models (Models 1, 2 and 3) on each item using the `lrm` function in **Design** (observe these are item-wise regressions); generate three likelihood-ratio χ^2 statistics for comparing three nested logistic regression models (Models 1 vs. 3, Models 1 vs. 2, and Models 2 vs. 3); compute three pseudo R^2 measures – Cox & Snell (Cox and Snell 1989), Nagelkerke (Nagelkerke 1991), and McFadden (Menard 2000) – for three nested models and compute differences between them; compute the absolute proportional change in point estimates for β_1 from Model 1 to Model 2 as follows: $|(\beta_1 - \beta_1^*)/\beta_1^*|$, where β_1^* is the regression coefficient for the matching criterion (ability) from Model 1 and β_1 is the same term from Model 2.

5. Detecting DIF: Flag DIF items based on the detection criterion (`"Chisqr"`, `"R2"`, or `"Beta"`) and a corresponding flagging criterion specified by the user (e.g., `alpha = 0.01` for `criterion = "Chisqr"`); for `criterion = "Chisqr"` an item is flagged if any one of the three likelihood ratio χ^2 statistics is significant (the 2-*df* test for non-uniform DIF, χ^2_{13} , as a sole criterion may lack power if DIF is attributable primarily to uniform DIF, although inflated Type I error might be of concern).

6. Sparse matrix: Treat DIF items as unique to each group and prepare a sparse response matrix by splitting the response vector for each flagged item into a set of sparse vectors containing responses for members of each group (e.g., males and females if DIF was found related to gender). In other words, each DIF item is split into multiple sparse variables such that each variable corresponds to the data of just one group and missing for all other groups. Note that sparse matrices are to account for DIF in the trait estimate; (ordinal) logistic regression DIF detection is performed on the original data matrix.

7. IRT recalibration: Refit the graded response model on the sparse matrix data and obtain a single set of item parameter estimates for non-DIF items and group-specific item parameter estimates for DIF items.

8. Scale transformation: Equate Stocking and Lord (1983) item parameter estimates from the matrix calibration to the original (single-group) calibration by using non-DIF items as anchor items (this step is necessary only when looking at DIF impact and can be deferred until the iterative cycles have concluded).

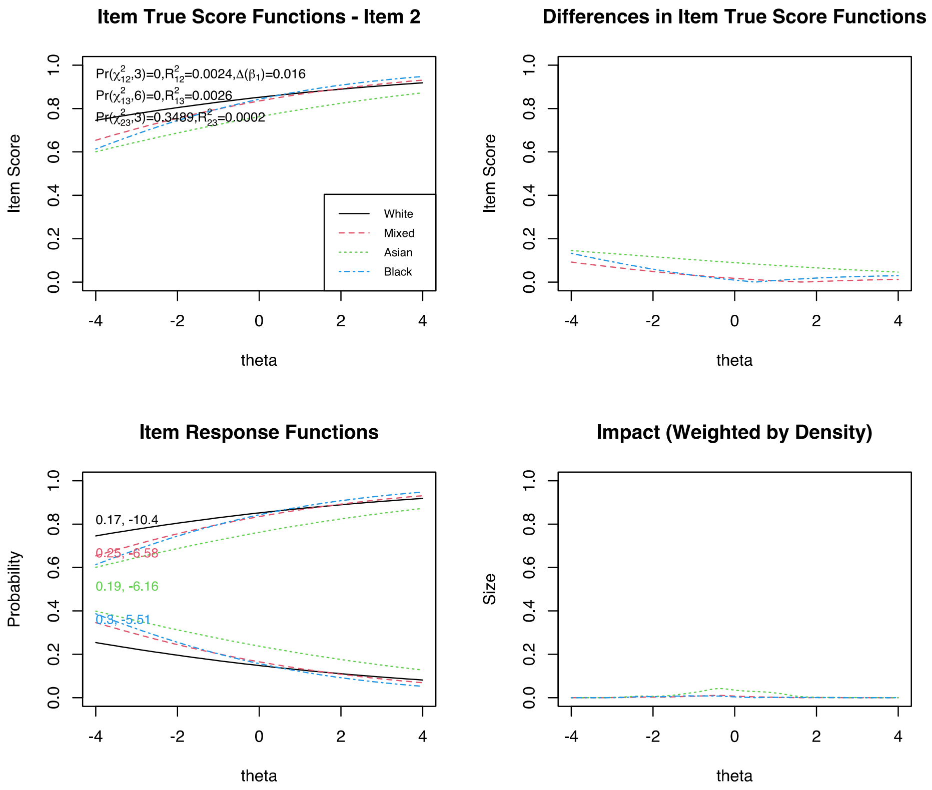
9. Trait re-estimation: Obtain EAP trait (ability) estimates based on item parameter estimates from the entire sample for items that did not have DIF and group-specific item parameter estimates for items that had DIF.

10. Iterative cycle: Repeat Steps 4 through 9 until the same items are flagged for DIF or a preset maximum number of iterations has been reached. Using the trait estimates from the previous round that account for DIF detected to that point, (ordinal) logistic regression DIF detection is repeated on all items including previously flagged items.

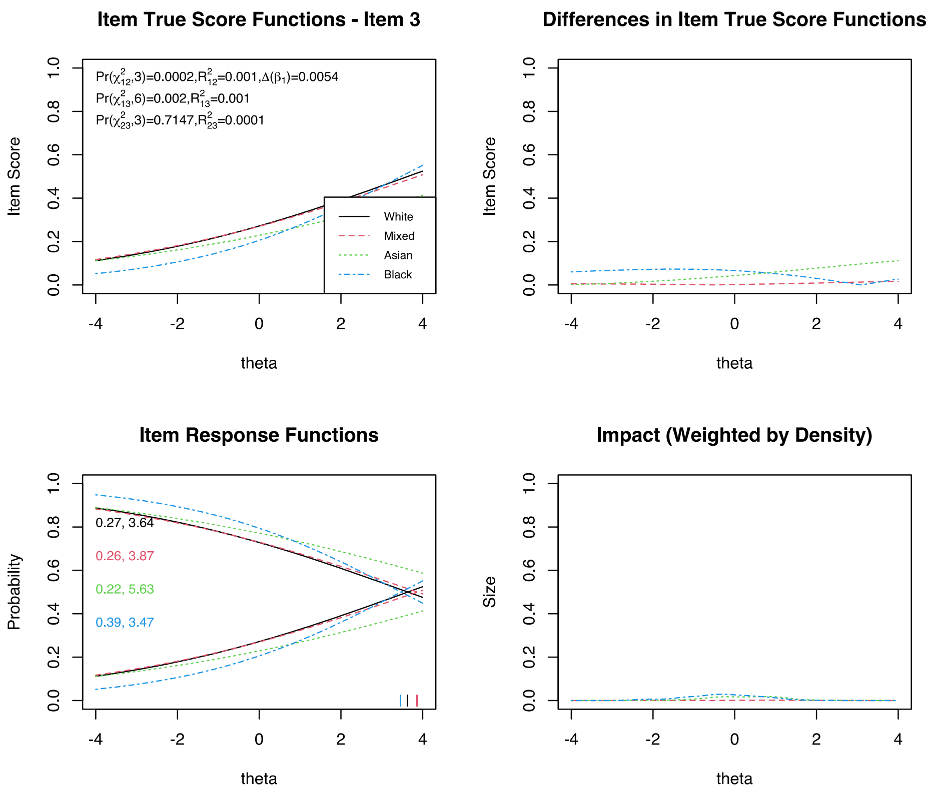
11. Monte Carlo simulation: Generate DIF-free datasets `nr` number of times (e.g., `nr = 1000`), using the final trait estimates accounting for DIF (Step 10) and the initial single-group item parameter estimates (Step 2). Each simulated dataset contains the same number of cases by group as the empirical dataset and reflects observed group differences in trait estimates. For each simulated dataset, obtain trait (ability) estimates based on the single-group item parameter estimates and run the OLR/IRT procedure. Compute the DIF statistics and magnitude measures for each simulated dataset and store the results for all replications. Identify a threshold value for each statistic/magnitude measure that cuts off the most extreme (defined by α) end of its cumulative distribution.

(Choi et al. 2011:8-9)

6.4.3 Appendix 6.3: Diagnostic Plots for Items Flagged for DIF by Ethnic Group for Overall QWE



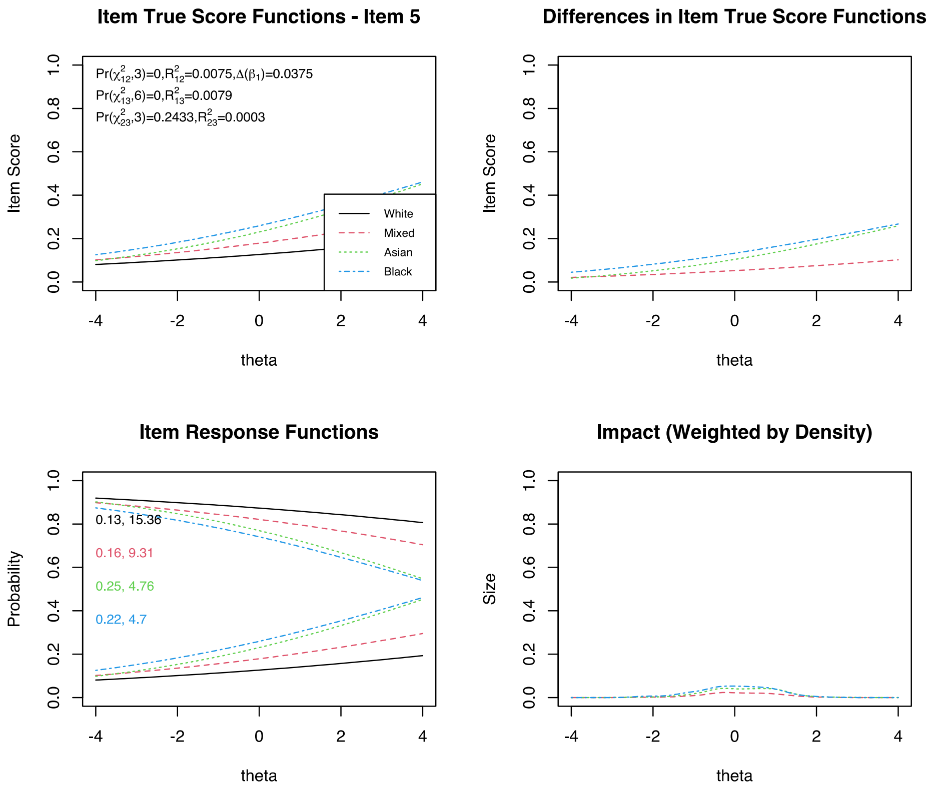
Item 2: Pension provision



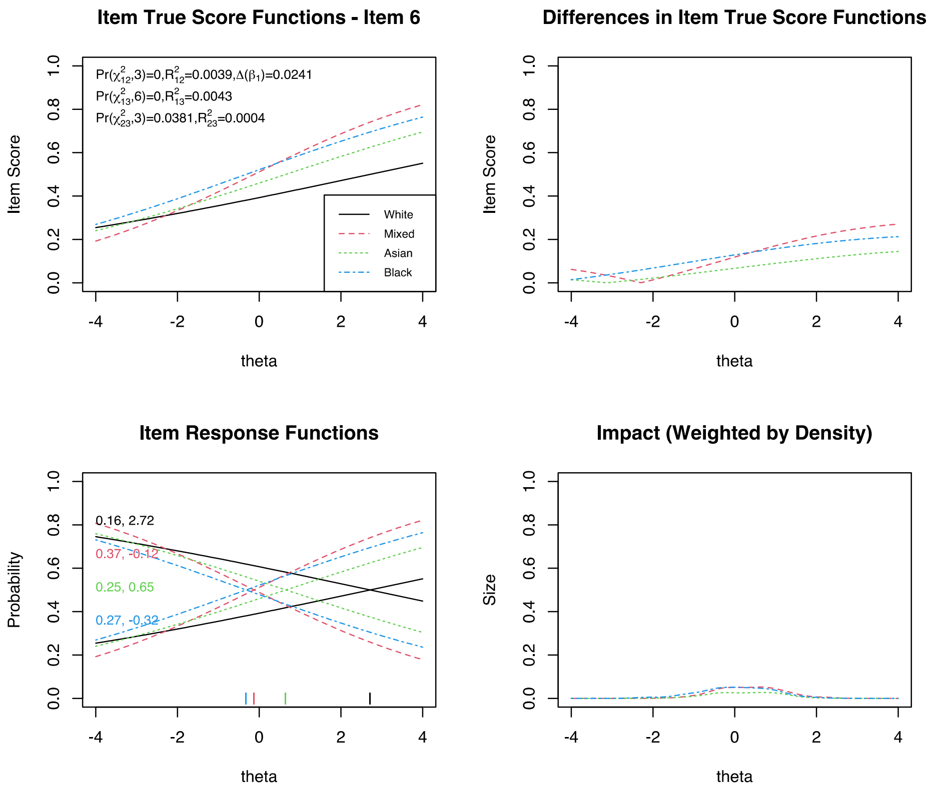
Item 3: Pay bonuses

Continued…

Continued…

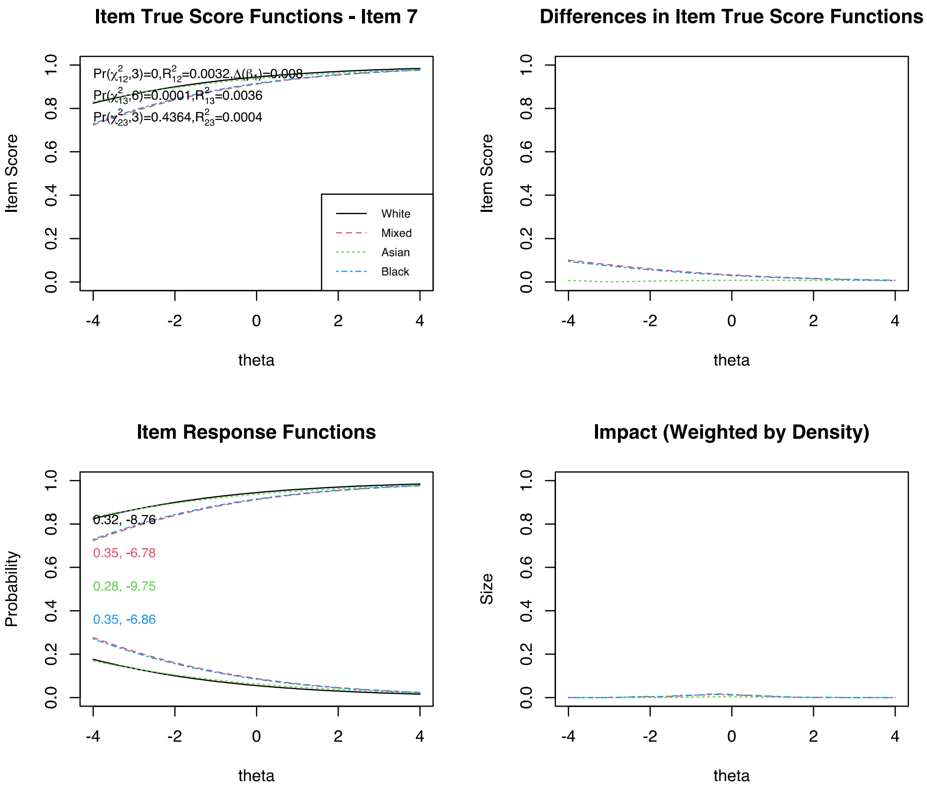


Item 5: Progression prospects

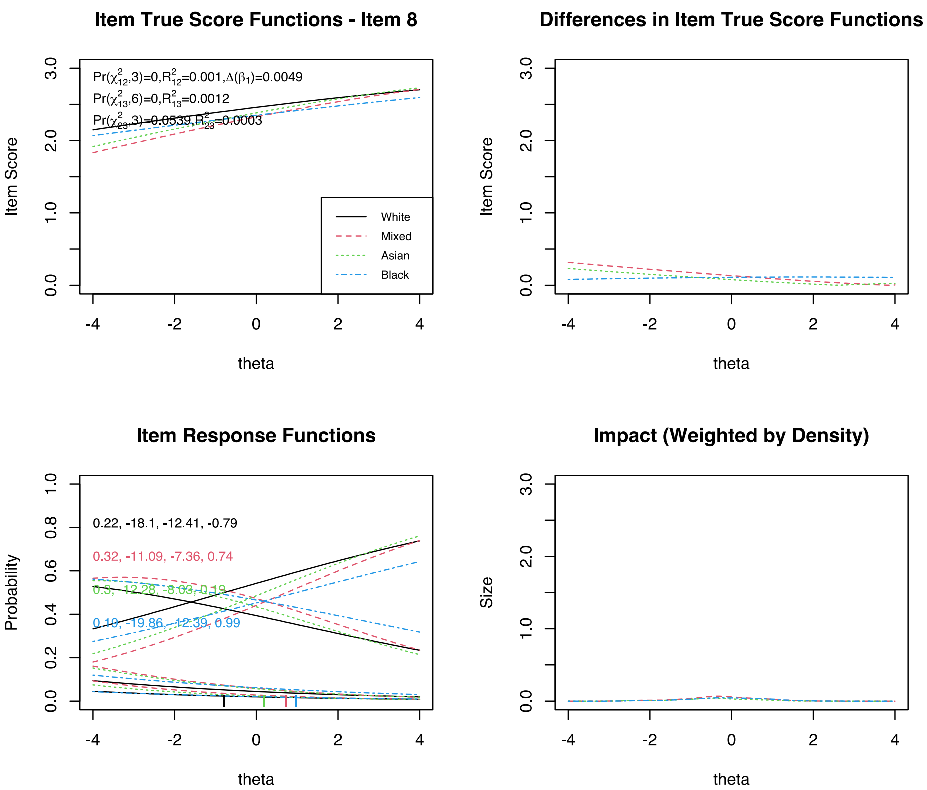


Item 6: Training prospects

Continued…



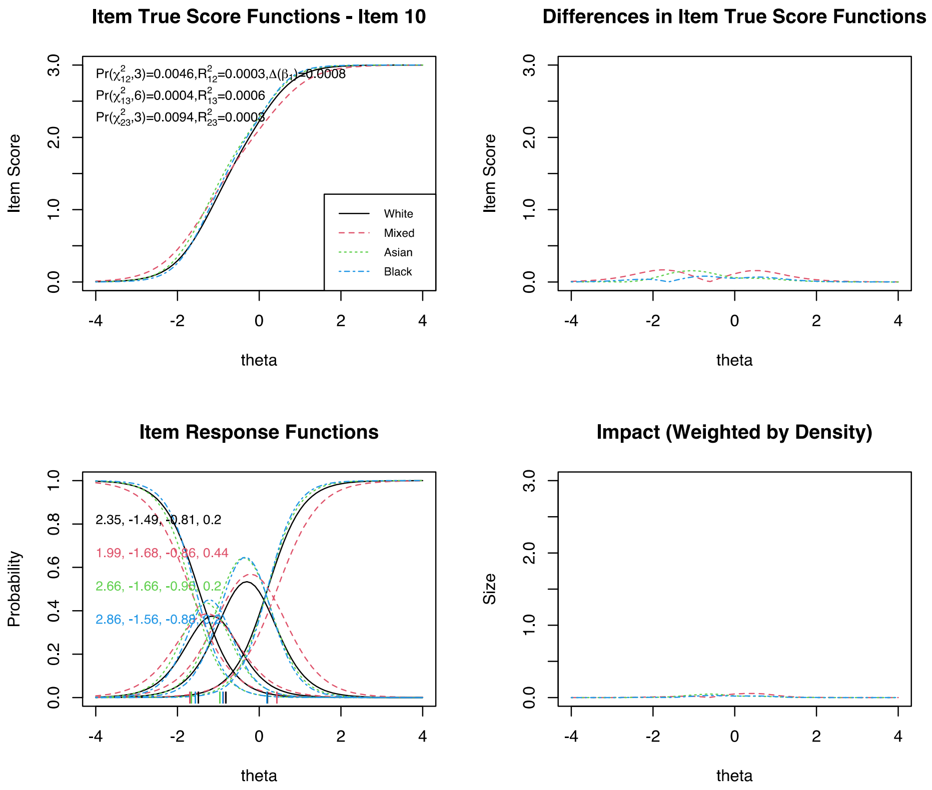
Item 7: Employment type



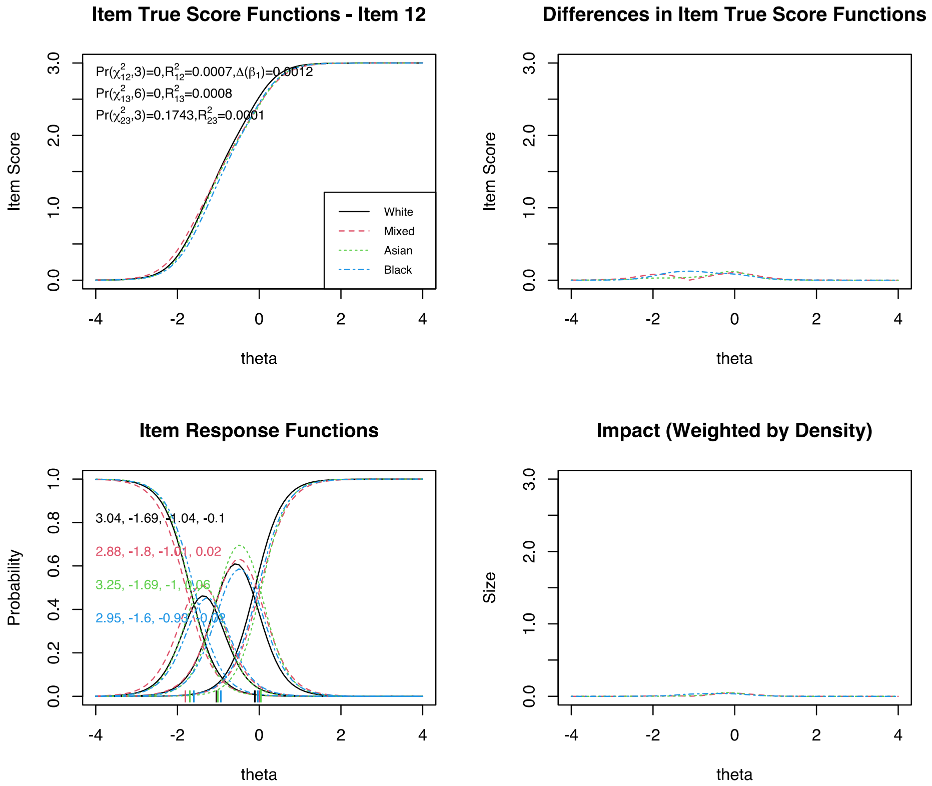
Item 8: Job security

Continued…

Continued…

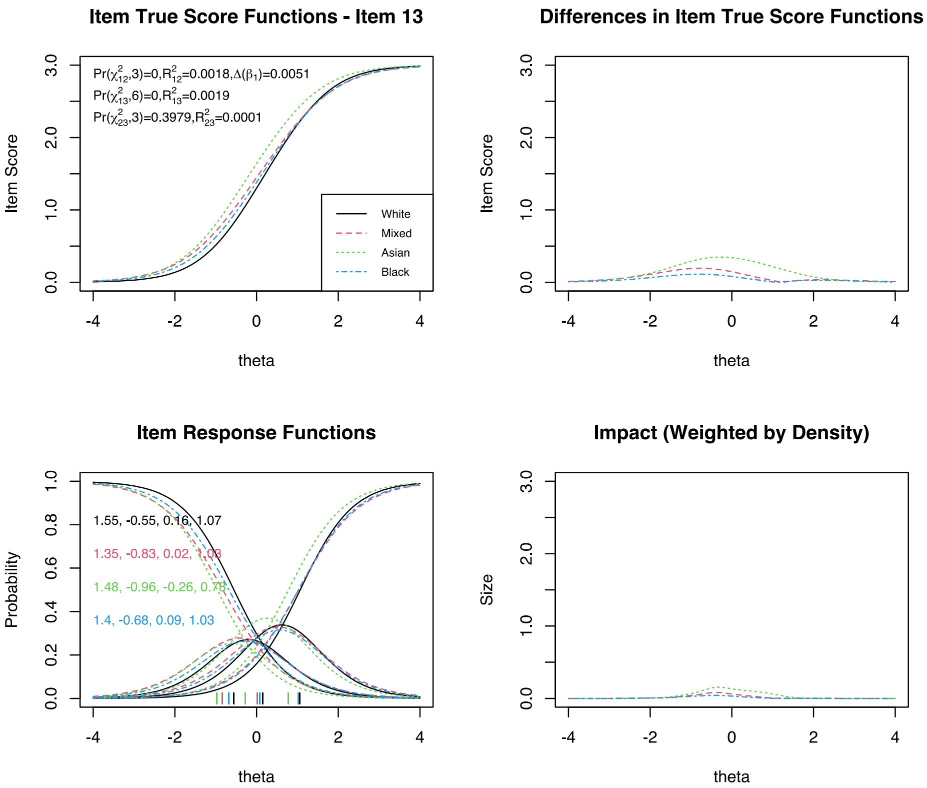


Item 10: Work pace

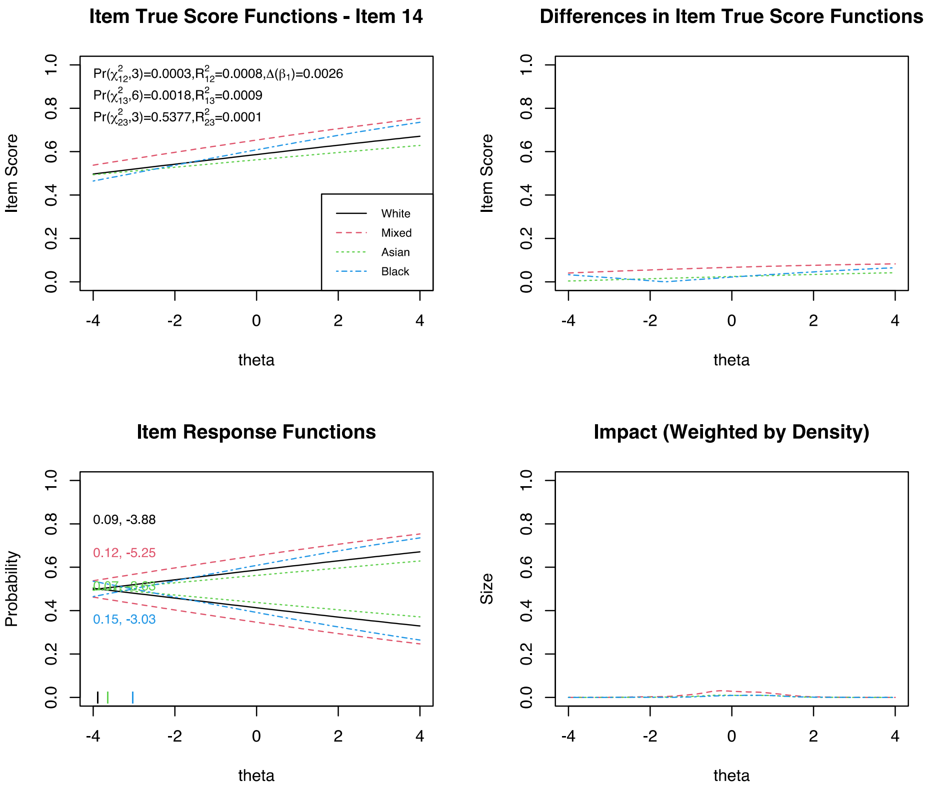


Item 12: Task order

Continued…

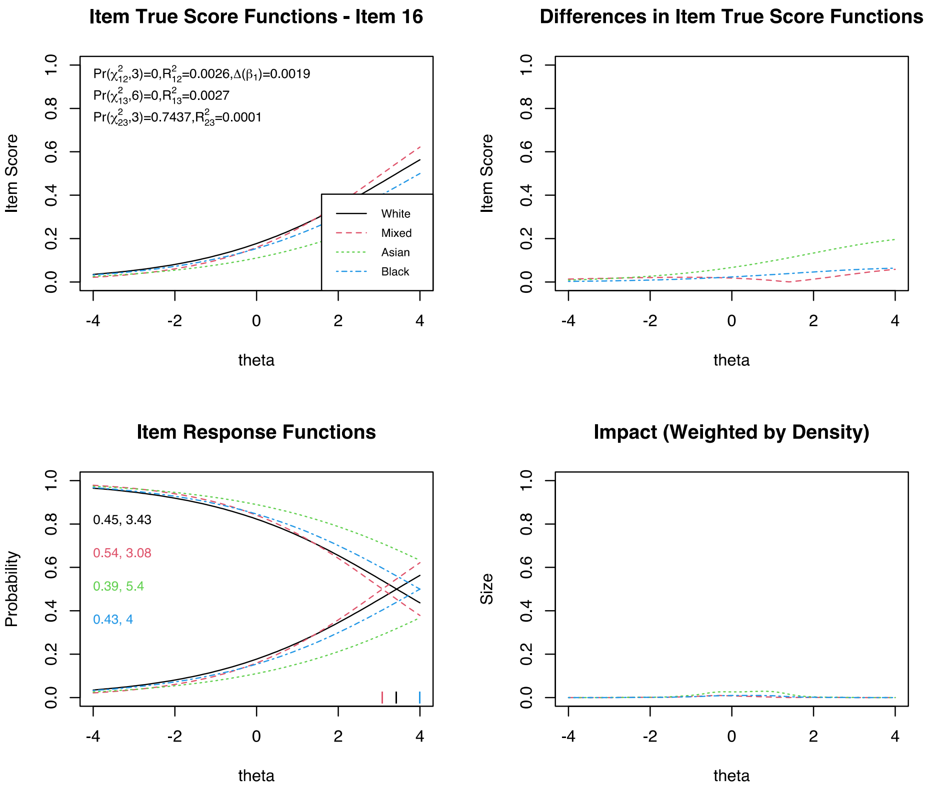


Item 13: Work hours

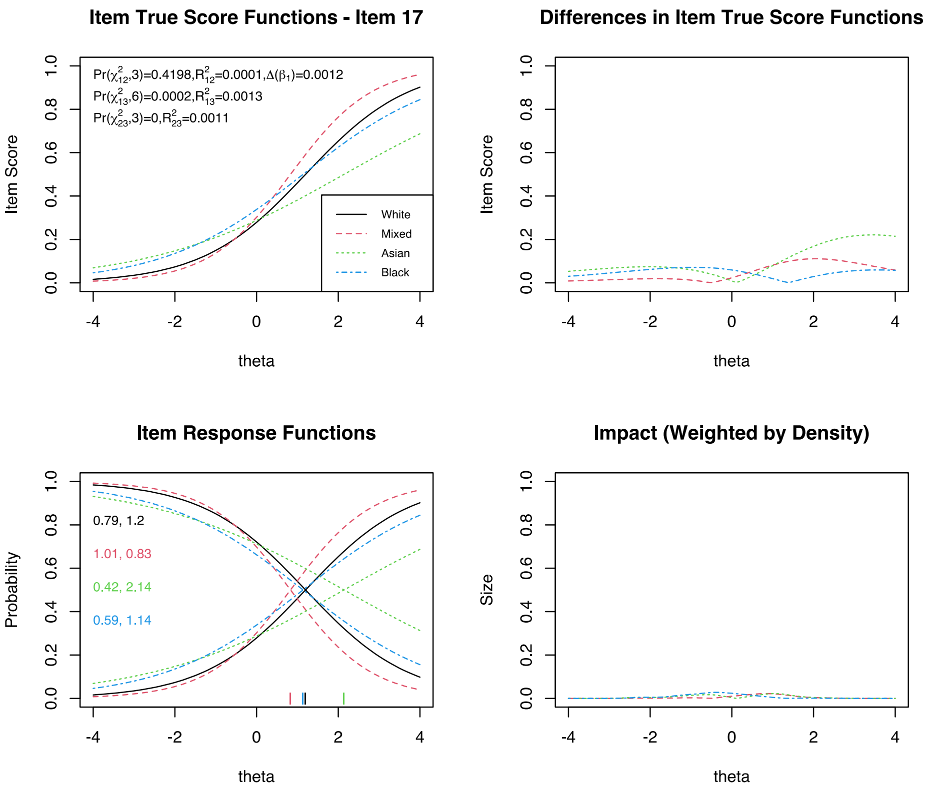


Item 14: Part-time

Continued…



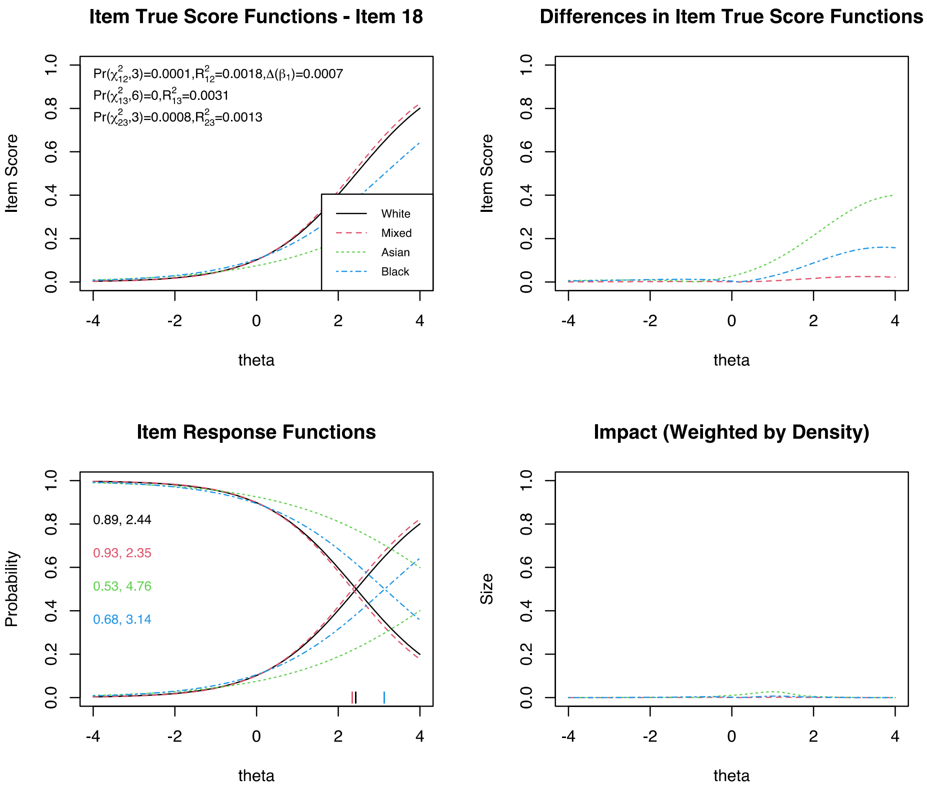
Item 16: Job sharing



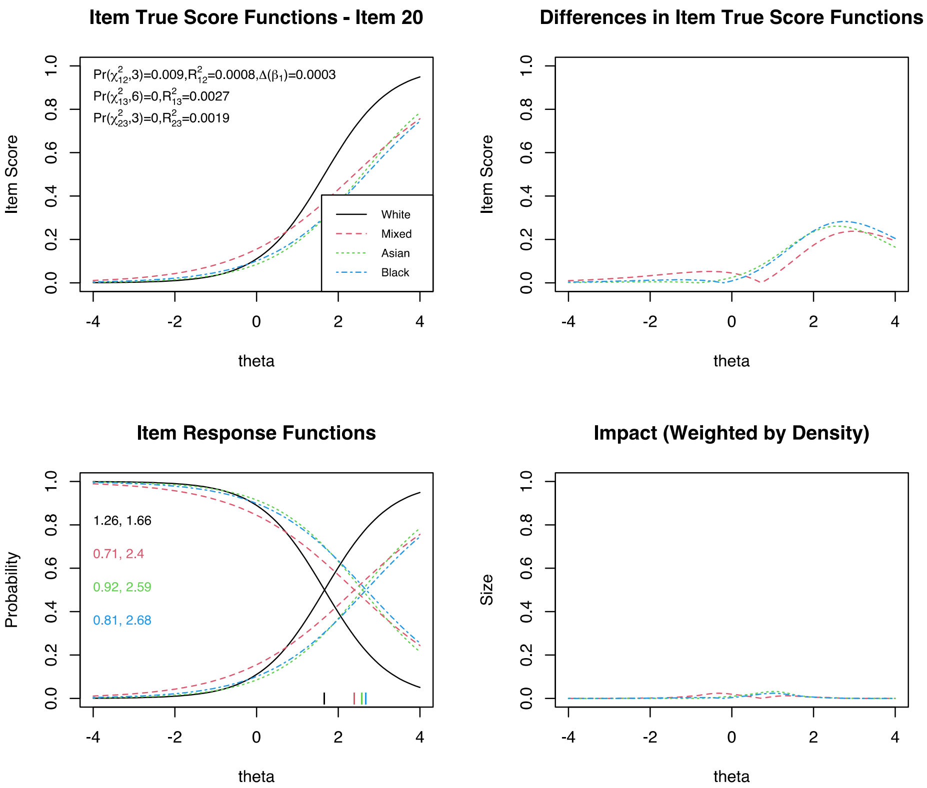
Item 17: Flexi-time

Continued…

Continued…



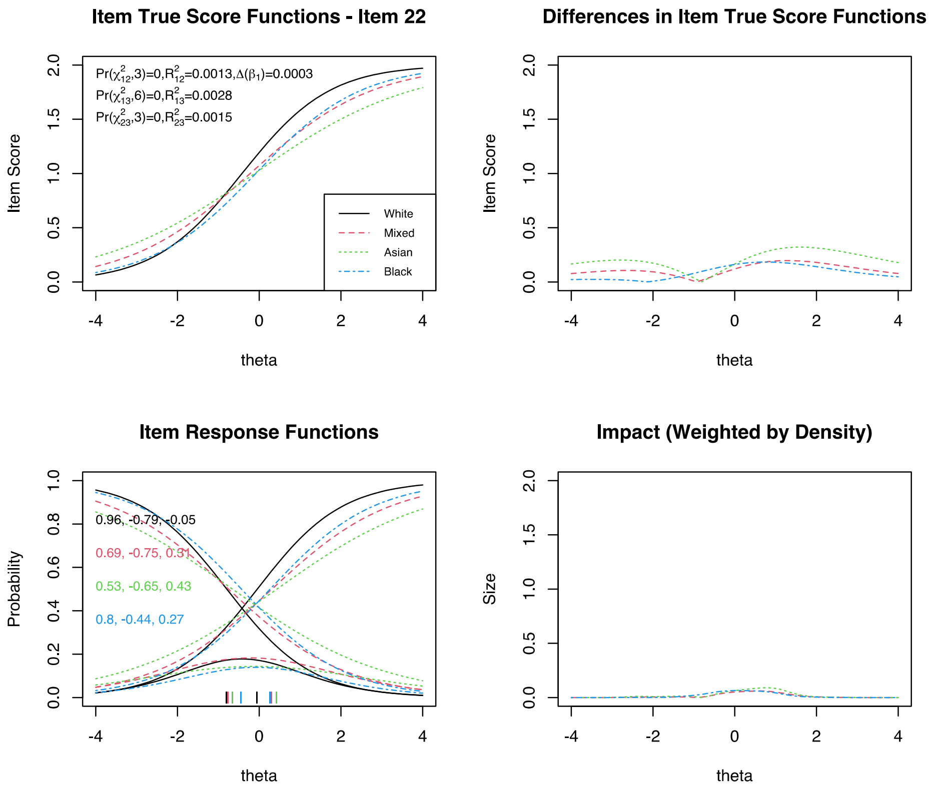
Item 18: Compressed hours



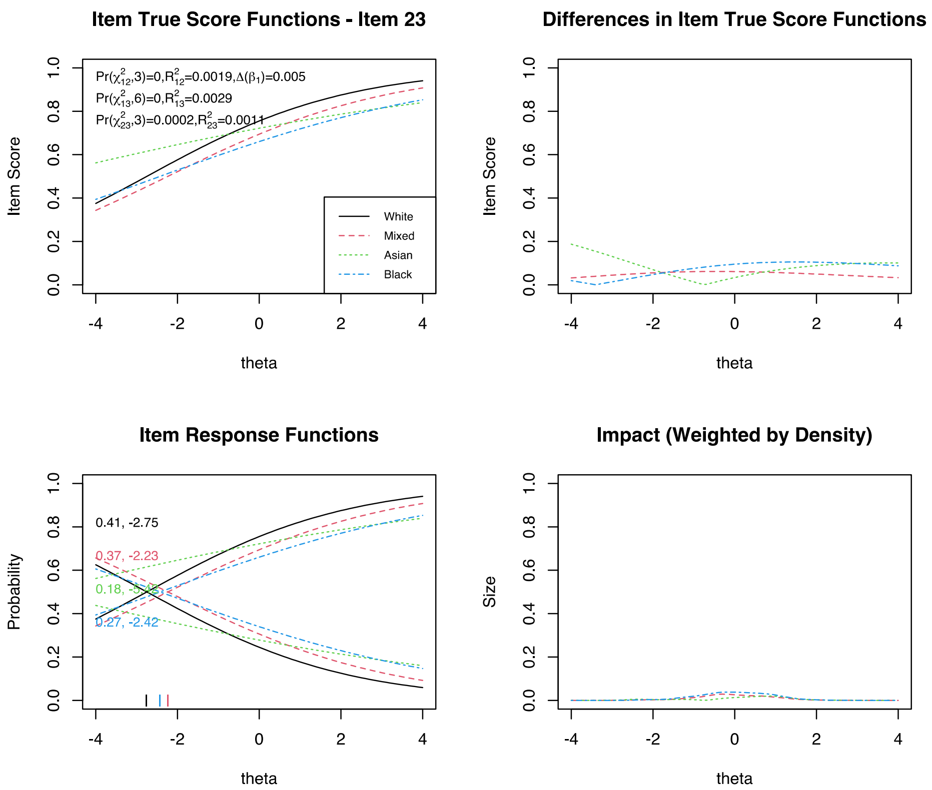
Item 20: Home working

Continued…

Continued…



Item 22: Informal flexibility



Item 23: Working times

6.4.4 Appendix 6.4: DIF by Ethnic Group using the MIMIC Model

STANDARDIZED MODEL RESULTS

STDY Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
ITEM1 ON				
ETH_GRP2	-0.196	0.160	-1.228	0.220
ETH_GRP3	-0.044	0.181	-0.245	0.806
ETH_GRP4	-0.130	0.176	-0.738	0.460
ITEM2 ON				
ETH_GRP2	-0.865	0.347	-2.495	0.013
ETH_GRP3	-0.390	0.286	-1.363	0.173
ETH_GRP4	-0.428	0.318	-1.346	0.178
ITEM3 ON				
ETH_GRP2	-0.037	0.121	-0.302	0.763
ETH_GRP3	-0.144	0.100	-1.440	0.150
ETH_GRP4	-0.272	0.096	-2.829	0.005
ITEM4 ON				
ETH_GRP2	-0.330	0.183	-1.799	0.072
ETH_GRP3	-0.104	0.158	-0.656	0.512
ETH_GRP4	-0.235	0.172	-1.370	0.171
ITEM5 ON				
ETH_GRP2	0.083	0.130	0.639	0.523
ETH_GRP3	0.411	0.073	5.646	0.000
ETH_GRP4	0.507	0.076	6.684	0.000
ITEM6 ON				
ETH_GRP2	0.359	0.109	3.299	0.001
ETH_GRP3	0.311	0.065	4.787	0.000
ETH_GRP4	0.379	0.070	5.404	0.000
ITEM7 ON				
ETH_GRP2	-0.531	0.165	-3.226	0.001
ETH_GRP3	0.012	0.082	0.149	0.881
ETH_GRP4	-0.164	0.097	-1.700	0.089
ITEM8 ON				
ETH_GRP2	-0.350	0.125	-2.799	0.005
ETH_GRP3	-0.111	0.056	-1.967	0.049
ETH_GRP4	-0.167	0.070	-2.390	0.017
ITEM9 ON				
ETH_GRP2	-0.081	0.140	-0.582	0.560
ETH_GRP3	0.034	0.073	0.474	0.635
ETH_GRP4	-0.133	0.165	-0.803	0.422
ITEM10 ON				
ETH_GRP2	-0.164	0.137	-1.201	0.230
ETH_GRP3	-0.053	0.083	-0.636	0.525
ETH_GRP4	-0.028	0.164	-0.171	0.864
ITEM11 ON				
ETH_GRP2	-0.165	0.149	-1.107	0.268
ETH_GRP3	-0.128	0.095	-1.354	0.176
ETH_GRP4	-0.129	0.188	-0.683	0.494
ITEM12 ON				
ETH_GRP2	-0.099	0.156	-0.631	0.528
ETH_GRP3	-0.205	0.074	-2.770	0.006
ETH_GRP4	-0.197	0.164	-1.203	0.229

Continued…

ITEM13 ON				
ETH_GRP2	0.428	0.144	2.983	0.003
ETH_GRP3	0.182	0.148	1.229	0.219
ETH_GRP4	0.051	0.076	0.666	0.505
ITEM14 ON				
ETH_GRP2	0.301	0.160	1.884	0.060
ETH_GRP3	0.115	0.228	0.506	0.613
ETH_GRP4	0.122	0.202	0.606	0.544
ITEM15 ON				
ETH_GRP2	0.214	0.178	1.206	0.228
ETH_GRP3	0.119	0.223	0.534	0.594
ETH_GRP4	0.075	0.185	0.403	0.687
ITEM16 ON				
ETH_GRP2	-0.033	0.165	-0.203	0.839
ETH_GRP3	-0.019	0.223	-0.084	0.933
ETH_GRP4	0.062	0.182	0.339	0.734
ITEM17 ON				
ETH_GRP2	0.483	0.184	2.632	0.008
ETH_GRP3	0.164	0.145	1.134	0.257
ETH_GRP4	0.165	0.065	2.538	0.011
ITEM18 ON				
ETH_GRP2	0.304	0.170	1.784	0.074
ETH_GRP3	-0.023	0.147	-0.155	0.877
ETH_GRP4	0.021	0.098	0.213	0.831
ITEM19 ON				
ETH_GRP2	0.182	0.176	1.031	0.302
ETH_GRP3	0.112	0.170	0.660	0.510
ETH_GRP4	0.025	0.130	0.192	0.847
ITEM20 ON				
ETH_GRP2	0.396	0.244	1.627	0.104
ETH_GRP3	0.024	0.179	0.135	0.893
ETH_GRP4	-0.065	0.119	-0.548	0.584
ITEM21 ON				
ETH_GRP2	0.261	0.158	1.651	0.099
ETH_GRP3	0.065	0.113	0.571	0.568
ETH_GRP4	-0.050	0.083	-0.604	0.546
ITEM22 ON				
ETH_GRP2	0.356	0.210	1.697	0.090
ETH_GRP3	-0.203	0.192	-1.059	0.290
ETH_GRP4	-0.153	0.155	-0.989	0.323
ITEM23 ON				
ETH_GRP2	-0.178	0.122	-1.460	0.144
ETH_GRP3	-0.145	0.073	-1.978	0.048
ETH_GRP4	-0.243	0.075	-3.224	0.001

Notes: The table shows results of the DIF analysis by ethnic group using the MIMIC model (multiple group analysis model in this instance) and a statistically significant direct effects of the predictor on the observed item given the model indicate the presence of uniform DIF. Results suggests a presence of uniform DIF for some items; however, this does not account for non-uniform, nor does it provide estimates of the effect size of the DIF.

6.4.5 Appendix 6.5: Multiple Group Analysis Descriptive Statistics for *Overall QWE*

	Sample	Mean (S.D)	Median	Quartiles		Min.	Max.
				Q1	*Q3*		
Sex							
Female	8973	–0.000 (0.830)	–0.028	–0.641	0.579	–2.066	2.546
Male	7447	0.174 (0.861)	0.182	–0.459	0.771	–2.414	2.589
Ethnic group							
White	13047	–0.006 (0.844)	–0.008	–0.654	0.591	–2.304	2.483
Mixed	327	–0.013 (0.785)	–0.070	–0.555	0.518	–1.842	1.897
Asian or Asian British	1774	–0.065 (0.768)	–0.060	–0.576	0.439	–2.368	2.216
Black or Black British	842	–0.156 (0.803)	–0.177	–0.784	0.416	–2.052	1.976
Age group							
16 – 24	1643	–0.201 (0.754)	–0.216	–0.748	0.278	–2.224	2.421
25 – 34	3189	0.115 (0.856)	0.087	–0.530	0.710	–2.199	2.816
35 – 49	6403	0.203 (0.904)	0.198	–0.500	0.855	–2.314	2.677
50 – 64	4978	0.085 (0.900)	0.090	–0.627	0.744	–2.302	2.419
65 +	379	–0.016 (0.816)	–0.003	–0.660	0.678	–2.106	2.171
Relationship status							
Single	5526	–0.019 (0.825)	–0.048	–0.635	0.524	–2.181	2.722
Married or cohabiting	8963	0.166 (0.882)	0.174	–0.510	0.806	–2.283	2.492
Divorced or separated	1767	–0.001 (0.880)	–0.045	–0.708	0.648	–2.015	2.590
Widowed	214	–0.060 (0.838)	–0.055	–0.774	0.555	–1.694	2.256
Parental status							
Lone parents with school age children	1062	–0.102 (0.843)	–0.094	–0.732	0.453	–2.349	2.137
Coupled parents with school age children	2917	0.172 (0.890)	0.168	–0.493	0.811	–2.455	2.367
Employees without school age children	12348	0.020 (0.874)	0.007	–0.652	0.638	–2.485	2.683

Continued…

Continued…

	Sample	Mean (S.D)	Median	Quartiles Q1	Q3	Min.	Max.
Illness or Disability							
Yes	3956	–0.029 (0.860)	–0.047	–0.707	0.585	–2.461	2.537
No	12405	0.018 (0.833)	0.009	–0.607	0.598	–2.368	2.504
Region							
London	2262	0.031 (0.847)	0.013	–0.557	0.635	–2.255	2.470
Southern England	3256	0.066 (0.876)	0.046	–0.582	0.702	–2.100	2.394
East of England	1425	–0.048 (0.837)	–0.061	–0.651	0.522	–2.298	2.480
The Midlands	2643	–0.069 (0.828)	–0.060	–0.695	0.489	–2.350	2.335
Northern England	3707	–0.150 (0.841)	–0.171	–0.792	0.436	–2.543	2.229
Wales	1063	–0.117 (0.834)	–0.121	–0.756	0.504	–2.163	2.433
Scotland	1421	–0.133 (0.871)	–0.151	–0.862	0.484	–2.257	2.317
Northern Ireland	989	–0.368 (0.872)	–0.439	–1.101	0.291	–2.432	2.162
Occupational classification							
Managers & senior officials	2451	0.288 (0.735)	0.324	–0.207	0.826	–2.085	2.065
Professional occupations	2409	–0.191 (0.884)	–0.177	–0.961	0.502	–2.273	2.178
Associate professional & technical occupations	2895	–0.128 (0.799)	–0.114	–0.699	0.453	–2.460	2.046
Administrative & secretarial occupations	1935	–0.198 (0.718)	–0.159	–0.663	0.298	–2.277	1.852
Skilled trades occupations	933	–0.458 (0.701)	–0.451	–0.980	0.003	–2.535	1.936
Personal service occupations	1844	–0.802 (0.660)	–0.808	–1.387	–0.311	–2.486	1.510
Sales & customer service occupations	1250	–0.608 (0.636)	–0.582	–1.031	–0.178	–2.326	1.691
Process, plant & machine operatives	1010	–0.641 (0.722)	–0.633	–1.250	–0.079	–2.326	1.684
Elementary occupations	1727	–0.789 (0.678)	–0.750	–1.351	–0.294	–2.677	1.631

Continued…

Continued…

	Sample	Mean (S.D)	Median	Quartiles *Q1*	*Q3*	Min.	Max.
Education							
No qualifications	2242	–0.115 (0.829)	–0.107	–0.752	0.481	–2.064	2.425
GCSE / O-level or lower	3883	–0.084 (0.817)	–0.085	–0.713	0.485	–2.325	2.458
Up to A-level	1753	0.040 (0.818)	0.025	–0.570	0.606	–2.146	2.553
Up to Diploma in HE	1777	0.044 (0.819)	0.033	–0.586	0.618	–1.864	2.682
University or higher degree	5166	0.327 (0.884)	0.361	–0.360	1.015	–1.999	2.432
No recorded data	1899	0.077 (0.890)	0.086	–0.632	0.719	–2.329	2.600
Full or Part-time							
Part-time	3760	–0.084 (0.798)	–0.086	–0.705	0.455	–2.241	2.217
Full-time	12382	0.274 (0.875)	0.264	–0.396	0.908	–2.249	2.883
Organisational sector							
Private sector	10233	–0.012 (0.825)	–0.025	–0.623	0.542	–2.373	2.478
Public sector	5973	–0.025 (0.865)	–0.054	–0.737	0.615	–2.082	2.263
Organisation size							
Micro	2494	0.012 (0.830)	0.020	–0.584	0.593	–2.296	2.317
Small	4938	–0.171 (0.820)	–0.195	–0.799	0.383	–2.326	2.366
Medium	3793	–0.125 (0.851)	–0.164	–0.806	0.463	–2.221	2.522
Large	5344	0.196 (0.893)	0.216	–0.476	0.891	–2.286	2.617

6.4.6 Appendix 6.6: Multiple Group Analysis Descriptive Statistics for *Economic Compensation*

	Sample	Mean (S.D)	Median	Quartiles		Min.	Max.
				Q1	*Q3*		
Sex							
Female	8973	0.008 (0.670)	0.133	–0.247	0.459	–2.241	1.414
Male	7447	0.073 (0.678)	0.205	–0.187	0.514	–2.185	2.205
Ethnic group							
White	13047	0.005 (0.678)	0.129	–0.262	0.453	–2.236	2.234
Mixed	327	–0.038 (0.702)	0.109	–0.308	0.394	–1.762	1.255
Asian or Asian British	1774	–0.181 (0.751)	–0.017	–0.483	0.343	–2.199	1.863
Black or Black British	842	–0.066 (0.673)	0.035	–0.294	0.373	–1.847	1.232
Age group							
16 – 24	1643	–0.186 (0.696)	–0.025	–0.916	0.352	–1.949	2.142
25 – 34	3189	0.209 (0.645)	0.328	–0.051	0.631	–1.876	2.450
35 – 49	6403	0.287 (0.638)	0.402	0.046	0.706	–1.878	2.225
50 – 64	4978	0.251 (0.611)	0.348	0.014	0.623	–1.876	1.748
65 +	379	–0.193 (0.736)	–0.040	–0.953	0.357	–1.664	1.359
Relationship status							
Single	5526	–0.007 (0.670)	0.135	–0.256	0.452	–2.025	2.300
Married or cohabiting	8963	0.187 (0.641)	0.314	–0.063	0.599	–2.025	2.241
Divorced or separated	1767	0.113 (0.619)	0.213	–0.130	0.505	–1.802	1.390
Widowed	214	–0.014 (0.677)	0.092	–0.289	0.456	–1.614	1.307
Parental status							
Lone parents with school age children	1062	–0.136 (0.749)	0.031	–0.362	0.398	–1.973	1.291
Coupled parents with school age children	2917	0.122 (0.689)	0.260	–0.114	0.586	–2.055	1.829
Employees without school age children	12348	0.067 (0.666)	0.196	–0.164	0.494	–2.149	2.354

Continued…

Continued…

	Sample	Mean (S.D)	Median	Quartiles *Q1*	*Q3*	Min.	Max.
Illness or Disability							
Yes	3956	–0.002 (0.662)	0.119	–0.226	0.432	–2.117	1.222
No	12405	0.015 (0.683)	0.156	–0.217	0.475	–2.136	2.244
Region							
London	2262	0.000 (0.715)	0.146	–0.254	0.458	–1.953	2.300
Southern England	3256	0.032 (0.665)	0.146	–0.233	0.461	–2.256	1.883
East of England	1425	0.078 (0.660)	0.208	–0.185	0.523	–1.879	1.357
The Midlands	2643	–0.034 (0.689)	0.099	–0.282	0.424	–1.967	1.359
Northern England	3707	–0.028 (0.672)	0.090	–0.276	0.412	–2.261	1.830
Wales	1063	–0.017 (0.686)	0.106	–0.275	0.441	–2.225	1.308
Scotland	1421	0.090 (0.666)	0.205	–0.154	0.543	–1.860	1.490
Northern Ireland	989	–0.051 (0.747)	0.124	–0.343	0.489	–2.256	1.268
Occupational classification							
Managers & senior officials	2451	0.162 (0.717)	0.238	–0.303	0.734	–2.152	2.277
Professional occupations	2409	0.391 (0.632)	0.490	0.041	0.862	–1.953	2.083
Associate professional & technical occupations	2895	0.157 (0.655)	0.184	–0.240	0.633	–1.976	1.768
Administrative & secretarial occupations	1935	–0.258 (0.603)	–0.242	–0.623	0.113	–2.142	1.455
Skilled trades occupations	933	–0.250 (0.743)	–0.201	–0.752	0.279	–2.399	1.347
Personal service occupations	1844	–0.538 (0.616)	–0.548	–0.993	–0.116	–2.192	1.351
Sales & customer service occupations	1250	–0.658 (0.593)	–0.691	–1.087	–0.280	–2.681	1.332
Process, plant & machine operatives	1010	–0.294 (0.680)	–0.260	–0.724	0.174	–2.717	1.409
Elementary occupations	1727	–0.710 (0.606)	–0.754	–1.105	–0.308	–2.729	1.325

Continued…

Continued…

	Sample	Mean (S.D)	Median	Quartiles		Min.	Max.
				Q1	*Q3*		
Education							
No qualifications	2242	–0.068 (0.667)	0.000	–0.486	0.406	–1.799	1.718
GCSE / O-level or lower	3883	–0.096 (0.638)	–0.033	–0.489	0.346	–2.041	1.676
Up to A-level	1753	–0.022 (0.671)	0.064	–0.443	0.455	–2.158	2.463
Up to Diploma in HE	1777	0.143 (0.636)	0.220	–0.253	0.604	–1.739	1.793
University or higher degree	5166	0.334 (0.600)	0.444	0.006	0.728	–2.138	2.056
No recorded data	1899	0.049 (0.703)	0.174	–0.446	0.560	–2.085	1.829
Full or Part-time							
Part-time	3760	–0.043 (0.732)	0.102	–0.793	0.493	–2.025	1.698
Full-time	12382	0.345 (0.597)	0.425	0.081	0.748	–2.187	2.363
Organisational sector							
Private sector	10233	–0.008 (0.624)	0.063	–0.303	0.426	–1.882	1.817
Public sector	5973	0.447 (0.527)	0.521	0.115	0.854	–1.641	2.370
Organisation size							
Micro	2494	–0.159 (0.734)	0.038	–0.849	0.448	–1.632	1.611
Small	4938	0.244 (0.647)	0.361	0.033	0.668	–1.632	1.801
Medium	3793	0.477 (0.539)	0.524	0.228	0.807	–1.563	2.210
Large	5344	0.625 (0.469)	0.638	0.393	0.933	–1.632	2.528

6.4.7 Appendix 6.7: Multiple Group Analysis Descriptive Statistics for *Working Conditions*

	Sample	Mean (S.D)	Median	Quartiles *Q1*	Quartiles *Q3*	Min.	Max.
Sex							
Female	8973	–0.030 (0.854)	–0.040	–0.615	0.692	–2.624	1.619
Male	7447	–0.043 (0.855)	–0.076	–0.645	0.706	–2.610	1.633
Ethnic group							
White	13047	–0.037 (0.839)	–0.054	–0.627	0.688	–2.659	1.637
Mixed	327	–0.091 (0.843)	–0.097	–0.631	0.534	–2.349	1.515
Asian or Asian British	1774	–0.036 (0.881)	–0.067	–0.617	0.667	–2.497	1.645
Black or Black British	842	–0.148 (0.884)	–0.147	–0.754	0.529	–2.522	1.546
Age group							
16 – 24	1643	–0.175 (0.902)	–0.147	–0.785	0.467	–2.678	1.678
25 – 34	3189	–0.025 (0.866)	–0.038	–0.599	0.690	–2.883	1.725
35 – 49	6403	0.039 (0.897)	0.051	–0.578	0.822	–2.776	1.740
50 – 64	4978	–0.008 (0.930)	0.023	–0.665	0.809	–2.938	1.739
65 +	379	0.124 (1.055)	0.270	–0.522	1.044	–2.555	1.647
Relationship status							
Single	5526	–0.038 (0.873)	–0.054	–0.636	0.670	–2.561	1.664
Married or cohabiting	8963	0.059 (0.863)	0.059	–0.543	0.813	–2.722	1.732
Divorced or separated	1767	0.041 (0.910)	0.055	–0.605	0.838	–2.478	1.728
Widowed	214	0.126 (0.912)	0.184	–0.372	0.930	–2.471	1.548
Parental status							
Lone parents with school age children	1062	–0.067 (0.921)	–0.093	–0.711	0.761	–2.829	1.559
Coupled parents with school age children	2917	–0.011 (0.884)	–0.018	–0.633	0.772	–2.755	1.702
Employees without school age children	12348	–0.081 (0.913)	–0.064	–0.713	0.695	–2.990	1.702

Continued…

Continued…

	Sample	Mean (S.D)	Median	Quartiles Q1	Q3	Min.	Max.
Illness or Disability							
Yes	3956	–0.053 (0.893)	–0.042	–0.686	0.705	–2.596	1.690
No	12405	–0.008 (0.850)	–0.026	–0.597	0.718	–2.673	1.708
Region							
London	2262	–0.048 (0.887)	–0.041	–0.664	0.680	–2.641	1.624
Southern England	3256	–0.015 (0.863)	–0.003	–0.633	0.734	–2.727	1.652
East of England	1425	–0.012 (0.875)	–0.041	–0.628	0.777	–2.746	1.660
The Midlands	2643	–0.014 (0.892)	–0.006	–0.622	0.755	–2.644	1.691
Northern England	3707	–0.013 (0.897)	–0.020	–0.614	0.765	–2.827	1.731
Wales	1063	–0.005 (0.913)	0.017	–0.697	0.825	–2.799	1.629
Scotland	1421	0.075 (0.913)	0.104	–0.570	0.876	–2.755	1.746
Northern Ireland	989	–0.019 (0.941)	0.008	–0.659	0.812	–2.655	1.660
Occupational classification							
Managers & senior officials	2451	0.172 (0.782)	0.408	–0.393	0.820	–2.530	1.479
Professional occupations	2409	–0.130 (0.821)	–0.166	–0.745	0.583	–2.728	1.517
Associate professional & technical occupations	2895	–0.225 (0.870)	–0.255	–0.845	0.533	–3.020	1.454
Administrative & secretarial occupations	1935	–0.259 (0.906)	–0.238	–0.909	0.515	–3.066	1.388
Skilled trades occupations	933	–0.157 (0.918)	–0.125	–0.789	0.682	–2.913	1.446
Personal service occupations	1844	–0.222 (0.887)	–0.275	–0.834	0.471	–2.612	1.554
Sales & customer service occupations	1250	–0.423 (0.932)	–0.467	–1.026	0.242	–3.026	1.446
Process, plant & machine operatives	1010	–0.424 (0.986)	–0.450	–1.092	0.310	–2.944	1.454
Elementary occupations	1727	–0.345 (0.980)	–0.396	–1.015	0.422	–2.693	1.497

Continued…

Continued…

	Sample	Mean (S.D)	Median	Quartiles		Min.	Max.
				Q1	*Q3*		
Education							
No qualifications	2242	–0.074 (0.922)	–0.066	–0.706	0.740	–2.661	1.618
GCSE / O-level or lower	3883	–0.108 (0.928)	–0.097	–0.741	0.715	–2.668	1.635
Up to A-level	1753	–0.149 (0.885)	–0.150	–0.739	0.578	–2.687	1.543
Up to Diploma in HE	1777	–0.053 (0.894)	–0.047	–0.706	0.752	–2.817	1.564
University or higher degree	5166	–0.070 (0.824)	–0.083	–0.664	0.642	–2.785	1.613
No recorded data	1899	–0.044 (0.885)	–0.074	–0.693	0.740	–2.676	1.563
Full or Part-time							
Part-time	3760	–0.116 (0.909)	–0.130	–0.729	0.612	–2.721	1.651
Full-time	12382	–0.020 (0.854)	–0.028	–0.622	0.714	–2.685	1.651
Organisational sector							
Private sector	10233	–0.036 (0.851)	–0.055	–0.621	0.708	–2.613	1.608
Public sector	5973	–0.057 (0.821)	–0.084	–0.624	0.614	–2.719	1.581
Organisation size							
Micro	2494	–0.004 (0.896)	0.045	–0.672	0.821	–2.583	1.494
Small	4938	–0.144 (0.873)	–0.153	–0.753	0.608	–3.062	1.559
Medium	3793	–0.197 (0.878)	–0.234	–0.810	0.560	–2.689	1.469
Large	5344	–0.289 (0.865)	–0.294	–0.891	0.434	–2.910	1.499

6.4.8 Appendix 6.8: Multiple Group Analysis Descriptive Statistics for *Work-time Scheduling*

	Sample	Mean (S.D)	Median	Quartiles		Min.	Max.
				Q1	*Q3*		
Sex							
Female	8973	0.011 (0.757)	–0.076	–0.427	0.516	–1.510	2.463
Male	7447	–0.359 (0.757)	–0.477	–1.010	0.026	–1.743	2.541
Ethnic group							
White	13047	0.023 (0.787)	0.017	–0.778	0.559	–1.707	2.890
Mixed	327	0.059 (0.756)	0.042	–0.627	0.543	–1.235	2.343
Asian or Asian British	1774	–0.112 (0.709)	–0.060	–0.813	0.248	–1.313	2.480
Black or Black British	842	–0.009 (0.750)	0.001	–0.741	0.370	–1.256	2.443
Age group							
16 – 24	1643	–0.006 (0.721)	0.036	–0.727	0.300	–1.363	2.726
25 – 34	3189	0.028 (0.783)	0.029	–0.765	0.523	–1.317	2.923
35 – 49	6403	0.077 (0.826)	0.035	–0.741	0.637	–1.203	2.820
50 – 64	4978	0.046 (0.809)	0.027	–0.758	0.592	–1.267	2.944
65 +	379	0.009 (0.613)	0.051	–0.354	0.277	–1.357	1.937
Relationship status							
Single	5526	–0.004 (0.761)	0.007	–0.761	0.412	–1.485	2.788
Married or cohabiting	8963	0.045 (0.806)	0.006	–0.760	0.602	–1.237	2.818
Divorced or separated	1767	0.045 (0.788)	0.022	–0.731	0.532	–1.207	2.598
Widowed	214	0.086 (0.784)	0.043	–0.577	0.416	–1.103	2.723
Parental status							
Lone parents with school age children	1062	0.041 (0.745)	0.025	–0.614	0.386	–1.337	2.951
Coupled parents with school age children	2917	0.050 (0.811)	–0.003	–0.753	0.593	–1.242	2.785
Employees without school age children	12348	0.001 (0.792)	–0.010	–0.785	0.517	–1.513	2.841

Continued…

Continued…

	Sample	Mean (S.D)	Median	Quartiles		Min.	Max.
				Q1	*Q3*		
Illness or Disability							
Yes	3956	0.021 (0.783)	0.004	–0.776	0.540	–1.298	2.603
No	12405	–0.040 (0.782)	–0.043	–0.821	0.459	–1.639	2.835
Region							
London	2262	–0.006 (0.769)	0.006	–0.764	0.410	–1.216	2.484
Southern England	3256	0.119 (0.800)	0.091	–0.664	0.666	–1.554	2.853
East of England	1425	0.075 (0.781)	0.074	–0.717	0.620	–1.157	2.590
The Midlands	2643	0.077 (0.794)	0.066	–0.716	0.616	–1.214	2.812
Northern England	3707	0.064 (0.800)	0.061	–0.733	0.537	–1.278	3.120
Wales	1063	0.046 (0.822)	0.051	–0.758	0.540	–1.222	2.757
Scotland	1421	0.111 (0.825)	0.086	–0.722	0.671	–1.175	2.629
Northern Ireland	989	0.002 (0.822)	0.029	–0.789	0.380	–1.160	2.462
Occupational classification							
Managers & senior officials	2451	0.025 (0.867)	–0.013	–0.774	0.585	–1.250	2.553
Professional occupations	2409	0.267 (0.846)	0.178	–0.533	0.902	–1.279	2.907
Associate professional & technical occupations	2895	0.208 (0.841)	0.152	–0.595	0.800	–1.325	2.920
Administrative & secretarial occupations	1935	0.243 (0.818)	0.172	–0.483	0.856	–1.141	2.903
Skilled trades occupations	933	–0.344 (0.633)	–0.665	–0.820	0.111	–1.289	2.407
Personal service occupations	1844	0.241 (0.705)	0.182	–0.039	0.715	–1.207	2.705
Sales & customer service occupations	1250	0.128 (0.618)	0.161	–0.063	0.367	–1.122	2.462
Process, plant & machine operatives	1010	–0.334 (0.617)	–0.646	–0.824	0.145	–1.143	2.491
Elementary occupations	1727	–0.000 (0.669)	0.116	–0.713	0.293	–1.650	2.621

Continued…

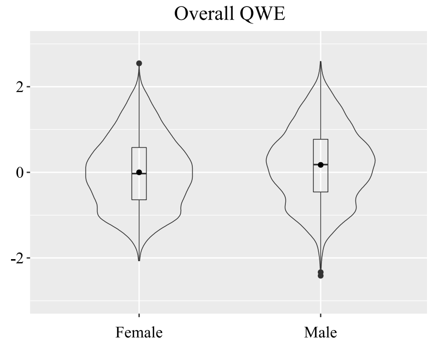
Continued…

	Sample	Mean (S.D)	Median	Quartiles		Min.	Max.
				Q1	*Q3*		
Education							
No qualifications	2242	–0.033 (0.747)	0.012	–0.782	0.308	–1.147	2.720
GCSE / O-level or lower	3883	–0.012 (0.765)	0.032	–0.770	0.372	–1.192	2.982
Up to A-level	1753	0.086 (0.792)	0.072	–0.690	0.558	–1.486	2.860
Up to Diploma in HE	1777	0.123 (0.794)	0.079	–0.618	0.660	–1.124	2.819
University or higher degree	5166	0.194 (0.859)	0.099	–0.618	0.786	–1.301	3.083
No recorded data	1899	0.036 (0.809)	0.048	–0.785	0.495	–1.179	2.656
Full or Part-time							
Part-time	3760	0.060 (0.643)	–0.029	–0.160	0.416	–1.542	2.338
Full-time	12382	–0.285 (0.862)	–0.217	–1.114	0.289	–1.685	2.754
Organisational sector							
Private sector	10233	0.016 (0.700)	0.042	–0.636	0.331	–1.182	3.119
Public sector	5973	0.562 (0.874)	0.415	–0.011	1.210	–1.229	3.109
Organisation size							
Micro	2494	–0.043 (0.708)	0.051	–0.726	0.263	–1.158	3.395
Small	4938	0.106 (0.762)	0.125	–0.659	0.546	–1.187	2.859
Medium	3793	0.158 (0.807)	0.135	–0.639	0.748	–1.345	2.971
Large	5344	0.327 (0.903)	0.192	–0.514	0.960	–1.283	3.091

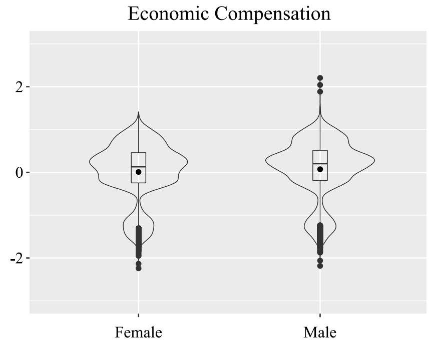
6.4.9 Appendix 6.9: Distributions of Latent Trait Scores based on Multiple Group Graded Response Bifactor IRT Model

Sex

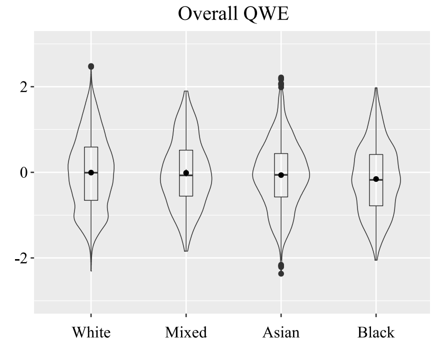
Ethnic group



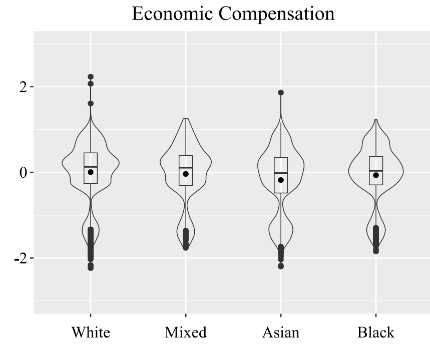
1 (a)



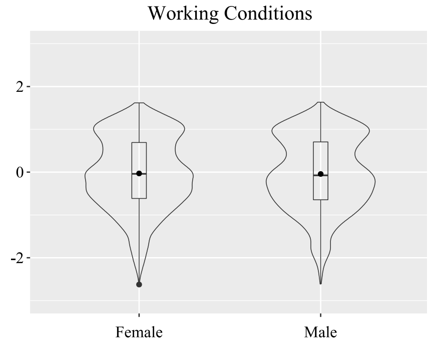
1 (b)



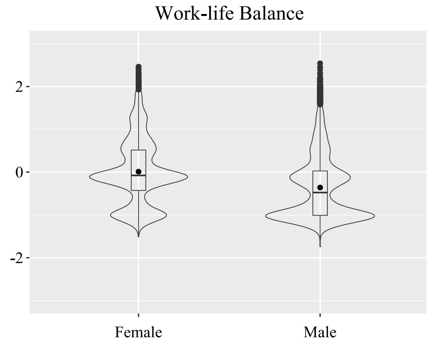
2 (a)



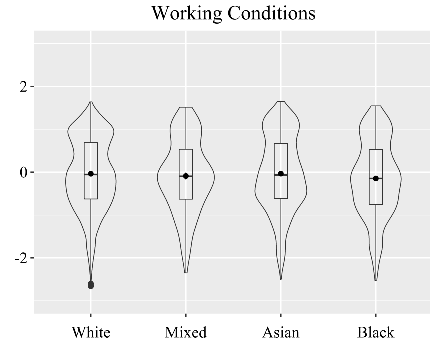
2 (b)



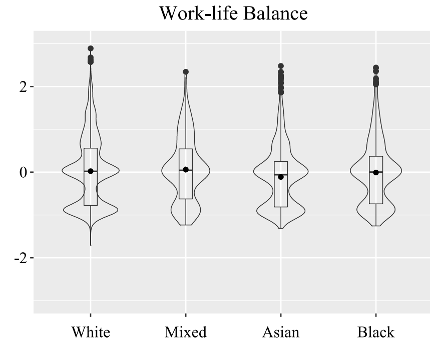
1 (c)



1 (d)

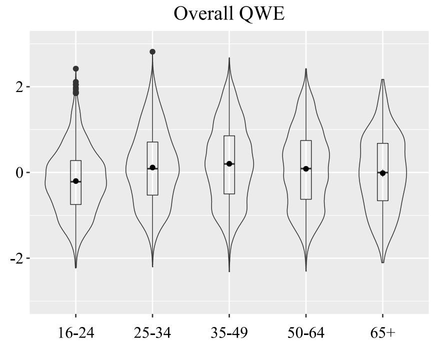


2 (c)

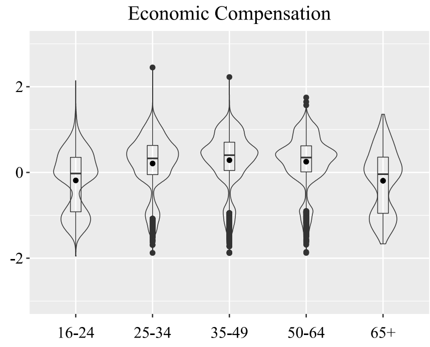


2 (d)

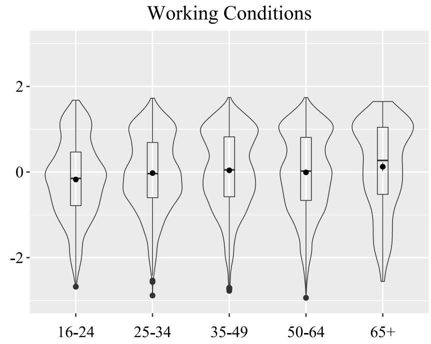
Age group



3 (a)



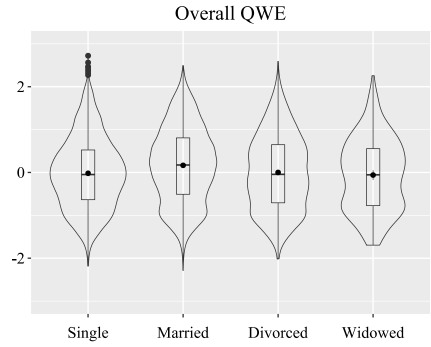
3 (b)



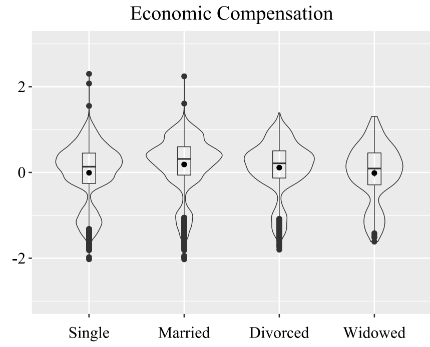
3 (c)

3 (d)

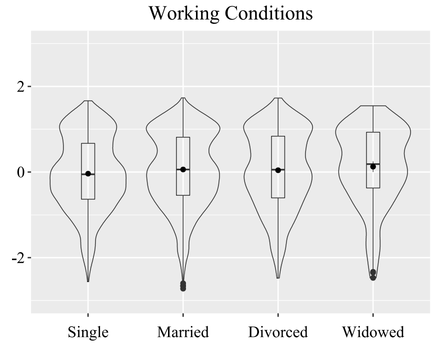
Relationship status



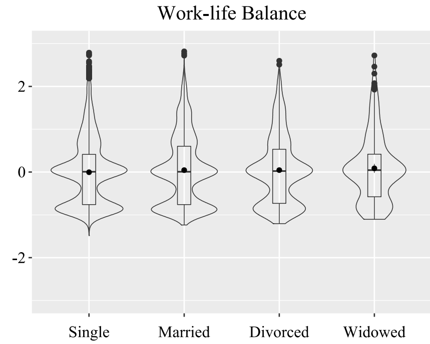
4 (a)



4 (b)

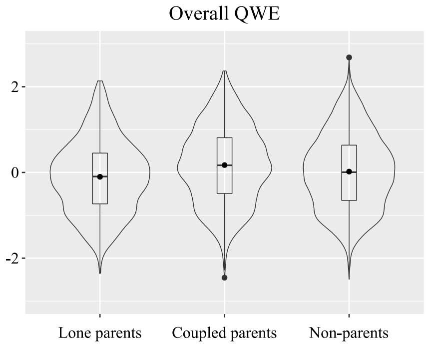


4 (c)

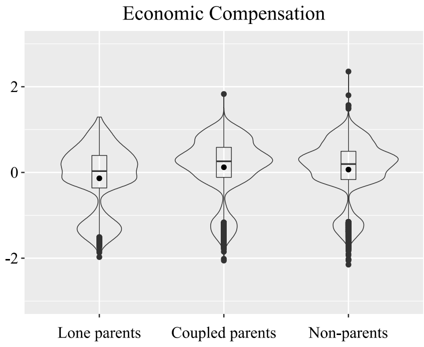


4 (d)

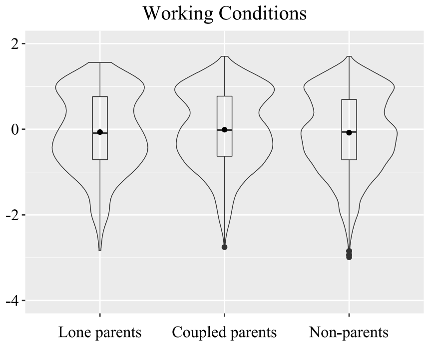
Parental status



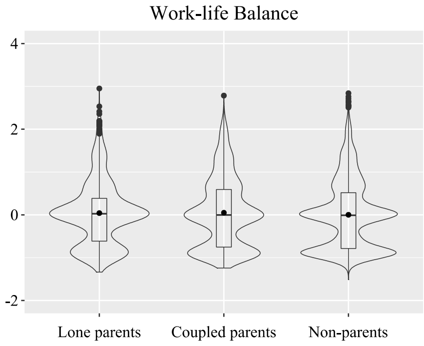
5 (a)



5 (b)

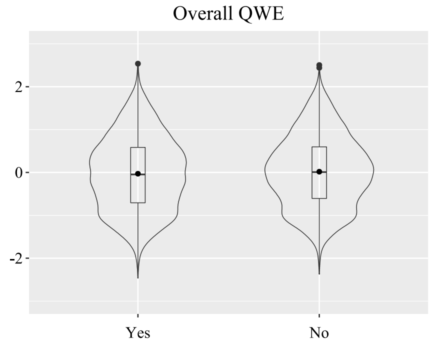


5 (c)

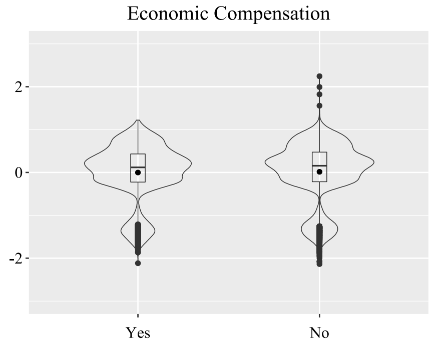


5 (d)

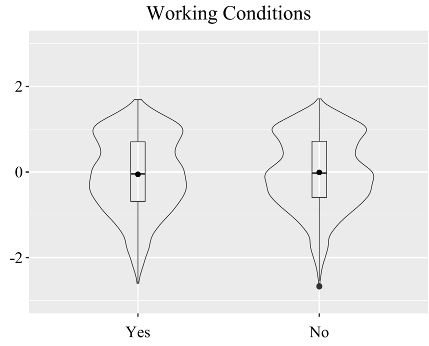
Illness or disability



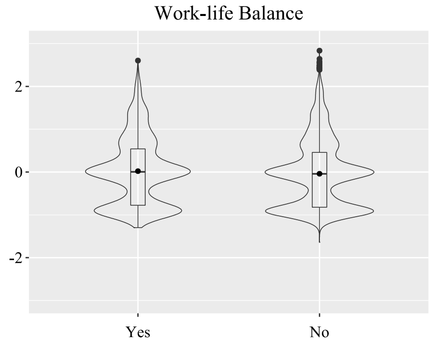
6 (a)



6 (b)



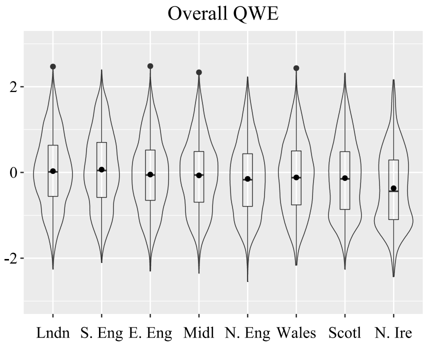
6 (c)



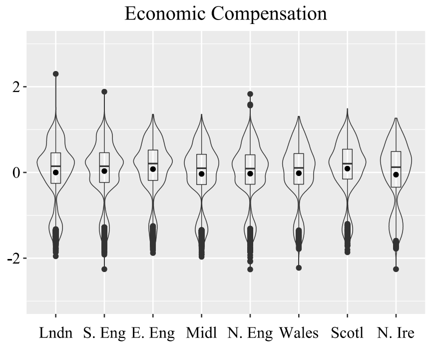
6 (d)

Region

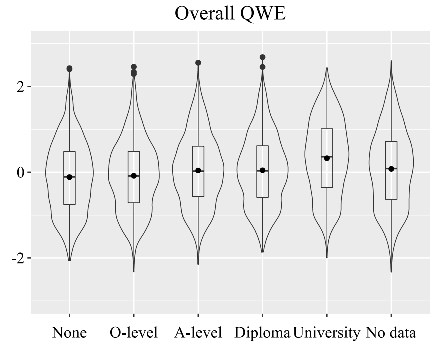
Education



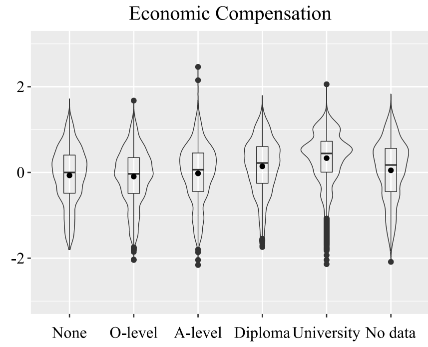
7 (a)



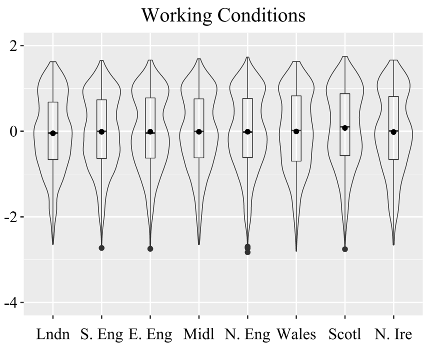
7 (b)



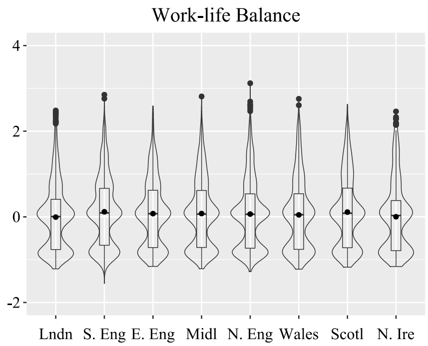
8 (a)



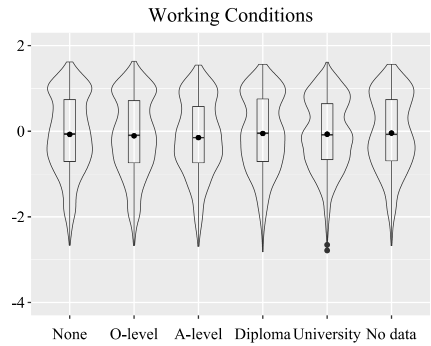
8 (b)



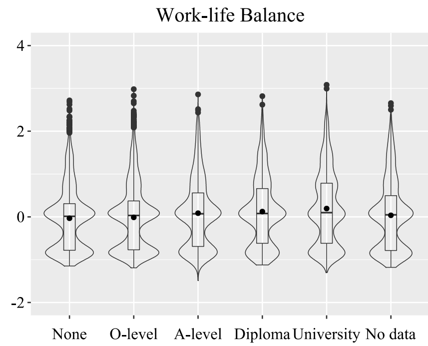
7 (c)



7 (d)



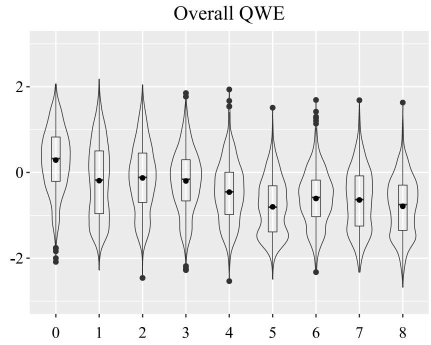
8 (c)



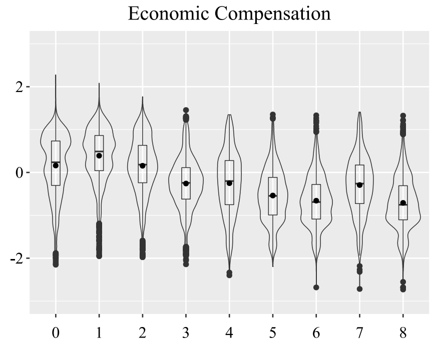
8 (d)

Occupational classification

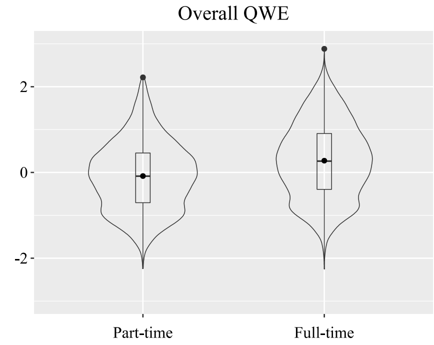
Full or part-time



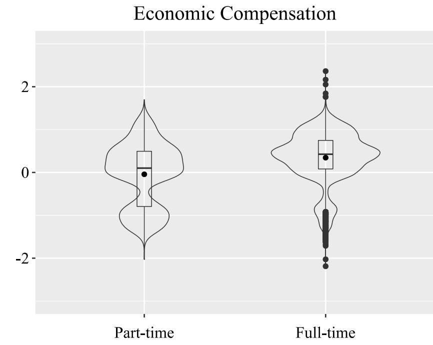
9 (a)



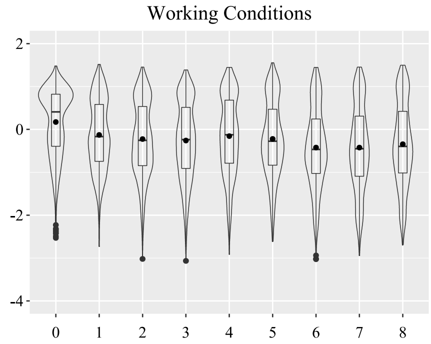
9 (b)



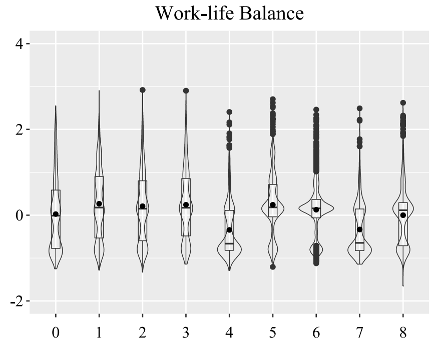
10 (a)



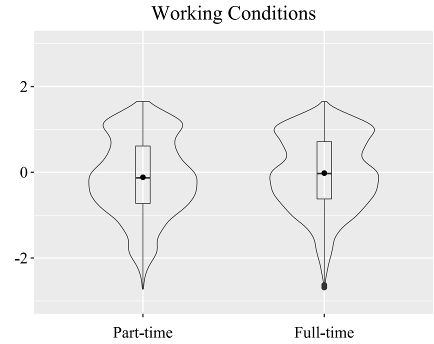
10 (b)



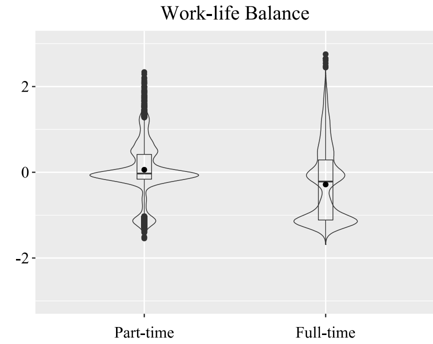
9 (c)



9 (d)



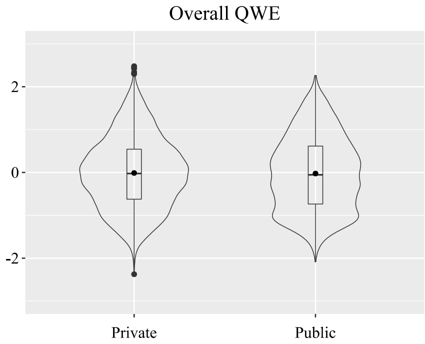
10 (c)



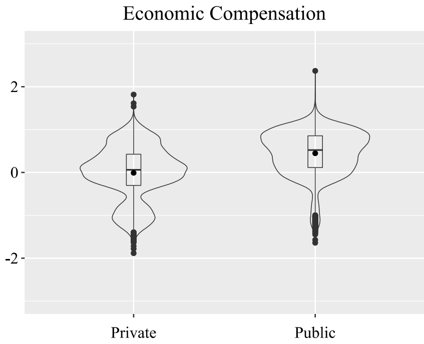
10 (d)

Notes for occupational classification

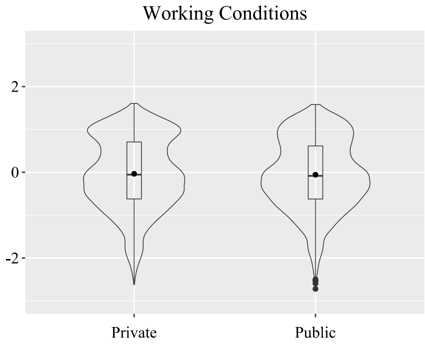
Organisational sector



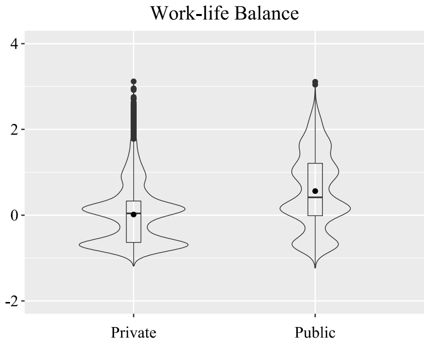
11 (a)



11 (b)

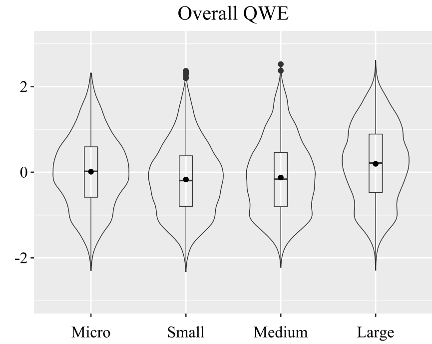


11 (c)

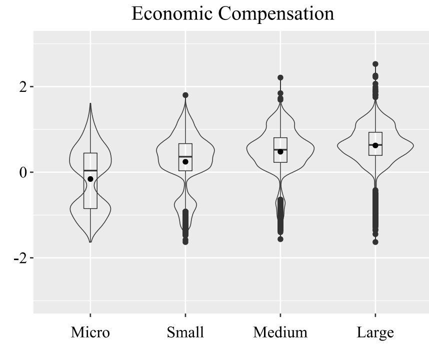


11 (d)

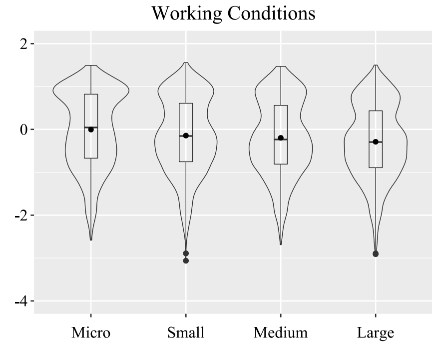
Organisation size



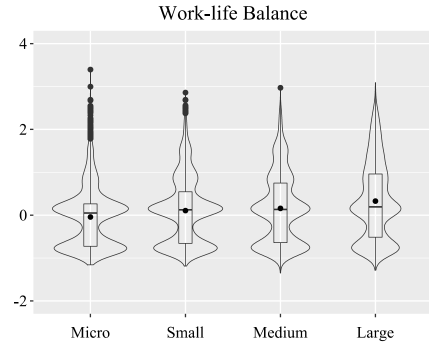
12 (a)



12 (b)



12 (c)



12 (d)

Chapter 7 Modelling Quality of Work and Employment

This chapter aims to investigate which employee characteristics predict *overall QWE* and *dimensions of QWE* in the UK employee population controlling for others. Measures of *overall QWE* and *dimensions of QWE* are based on the instrument developed in Chapter 5, while employee characteristics are classified into demographic, socio-demographic, and socio-economic characteristics. The first section describes the methodology of conducting the analyses regarding the data and sample, the dependent variables and predictors used, and the methods applied. The second section presents the results of the multiple indicators multiple causes models using the robust maximum likelihood and Bayesian estimators and the interpretation of the parameter estimates by each latent trait. The final section discusses the findings considering which predictors affect *overall QWE* and different *dimensions of QWE* and how the results compare with those from other literature.

7.1 Methodology

7.1.1 Data and Sample

The data from Wave 8 (2016 – 2017) of *Understanding Society: The United Kingdom Household Longitudinal Study* (UKHLS) (University of Essex, Institute for Social and Economic Research 2018) were used, and this is described in Section 4.1.1. The sample was limited to employees aged 16 years old and over who had a paid job and participated in full interviews, and the base sample amounted to 16,981 respondents.

7.1.2 Variables

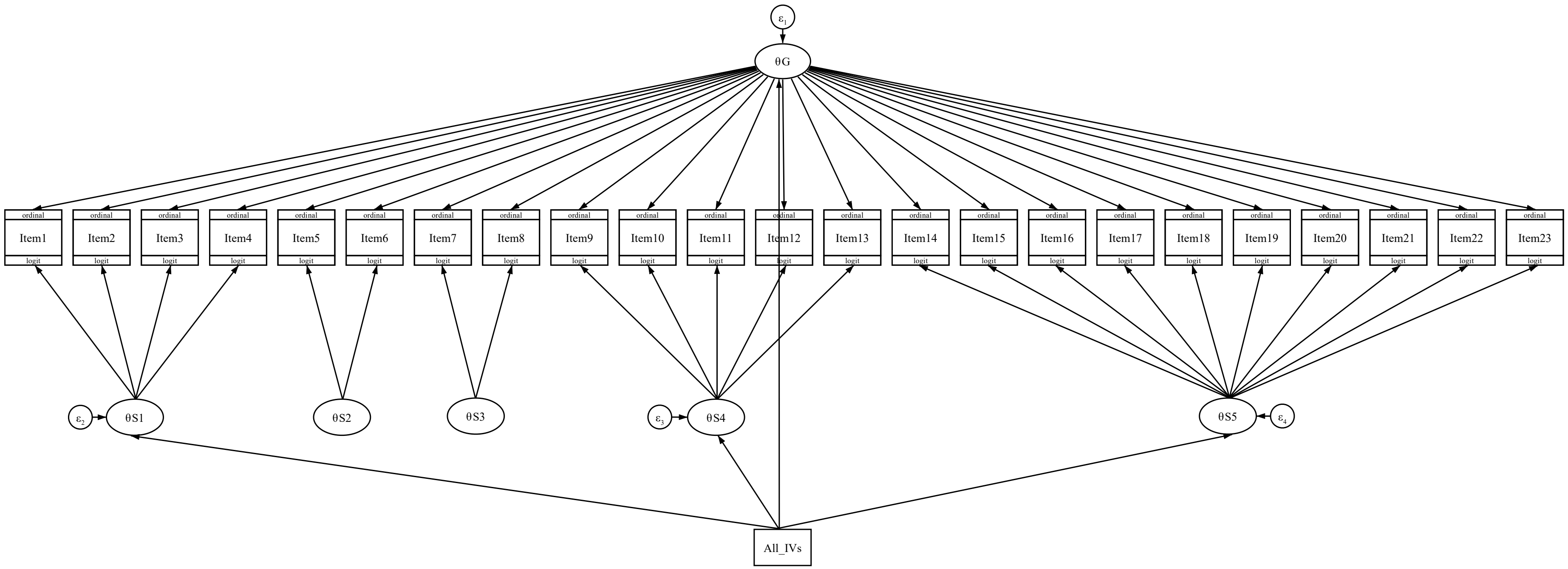
The dependent variables used were based on the measurement instruments of QWE developed in Chapter 5. These were limited to the *overall QWE*, *economic compensation*, *working conditions*, and *work-time scheduling* latent traits. *Training and progression* and *employment security* latent traits were excluded as estimated scores were not a good representation of the latent traits (see Section 5.2.6). The latent traits are assumed to follow standard normal distributions; thus, $\theta \sim N(0, 1)$, and the scales of the scores represent standard deviations from the population mean.

The predictors (exogenous or independent variables) considered were classified into *demographic* (sex, ethnic group, and age group), *socio-demographic* (relationship status, parental status, longstanding illness or disability, and region), and *socio-economic* (education, occupational classification, full or part-time, organisational sector, and organisation size) characteristics. These were all categorical and are described in detail in Section 4.1.2.

7.1.3 Methods

M*plus* 8.8 (Muthén and Muthén 2017) was used to estimate multiple indicators multiple causes (MIMIC) models. These consisted of 2-PL graded response bifactor measurement models (see Chapter 5) and structural models using a sequential regression approach to introduce a set of predictors into the model (Tabachnick and Fidell 2014). For each of the latent traits (*overall QWE*, *economic compensation*, *working conditions*, and *work-time scheduling*), three nested models were estimated. Model 1 included the effects of demographic characteristics, Model 2 introduced the effects of socio-demographic characteristics, and Model 3 introduced the effects of socio-economic characteristics. The path diagram for the final model with all the predictors is displayed in Figure 7.1

Figure 7.1: Path Diagram for MIMIC Model



Notes: The path diagram depicts the bifactor measurement model and a structural model modelling the effects of the predictors of QWE on the latent traits. The measurement model consisted of *overall QWE* (θ_G), *economic compensation* (θ_{S1}), *training and progression* (θ_{S2}), *employment security* (θ_{S3}), *working conditions* (θ_{S4}), and *work-time scheduling* (θ_{S5}) latent traits. The structural model introduced the effects of predictors of QWE, with the final model including al the predictors of QWE (All_IVs), on θ_G, θ_{S1}, θ_{S4}, and θ_{S5} along with their respective error term (ε_i) capturing the variance of each latent trait not explained by the IVs. θ_{S2} and θ_{S3} were excluded from the structural model as the estimated scores for these dimensions were not a good representation of the latent traits.

Generally, the null hypotheses tested postulated that there was no partial linear association between a predictor and a latent trait in the UK employee population, controlling for other predictors in the model. However, as the predictors were categorical, the null hypotheses (H_0) postulated that there was no statistically significant difference between the regression coefficients of a reference category (β) compared to another category, thus:

$$H_0: \beta = \beta_i = 0.$$

$$H_1: \beta_i \neq 0. \qquad \text{Hypothesis (7.1)}$$

where β_i is the regression coefficient for a category being compared to the reference category. Tests were conducted at the 5% significance level, with p-values < 0.05 suggesting a rejection of the H_0, indicating a statistically significant difference between the reference category and a particular category and an inference that there was a partial association between the predictor and the latent trait in the UK employee population. On the other hand, p-values $\geq$ 0.05 suggested failure to reject the H_0 and that the difference between the reference category and a particular category was statistically insignificant in the UK employee population.

The robust maximum likelihood (MLR) estimator in M*plus* 8.8 (Muthén and Muthén 2017) was used to estimate the MIMIC models. This provides parameter estimates with standard errors that are robust to non-normality and non-independence of observations associated with data obtained from complex sampling designs (Muthén and Muthén 2017; Wang and Wang 2020). However, the MIMIC models were computationally cumbersome to estimate with the MLR estimator due to the categorical observed items and the number of latent traits in the measurement model (Muthén and Asparouhov 2012). Therefore, the models were also estimated with a Bayesian estimator using the Markov Chain Monte Carlo (MCMC) algorithm with non-informative or diffuse priors to obtain parameter estimates analogous to those based on a MLR estimator (Johnson and Sinharay 2016; Muthén and Asparouhov 2012).

The MCMC algorithm uses an iterative process to estimate posterior distributions of model parameters based on the prior distribution and observed data (Finch and Bolin 2017). Default non-informative priors set in M*plus* were used; thus, *N* (0, 5) for parameter estimates of the slopes and thresholds of the measurement model and *N* (0, infinity) for parameter estimates of the regression coefficients of the structural model (Asparouhov and Muthén 2021; Muthén and Muthén 2017).

In terms of model fit evaluation, the potential scale reduction (PSR) was used to check the convergence of the MCMC algorithm for the Bayesian estimator, with values close to one, for example [1.0, 1.1] (Gelman et al. 2013), indicating convergence was achieved (Muthén and Asparouhov 2012; Wang and Wang 2020). Posterior parameter trace plots and posterior parameter distributions of estimated parameters also offer a visual inspection of convergence and how well the posterior distribution was simulated between different chains, with well mixed MCMC sequences indicating convergence was achieved (Muthén and Muthén 2017). After confirmation of convergence of the MCMC algorithm, the Bayesian posterior predictive checking (PPC) was used to evaluate model fit based on the posterior predictive *p*-value (PPP) (Asparouhov and Muthén 2021; Wang and Wang 2020). The PPP is a model χ^2 fit function comparing the fit statistic computed from the model posterior distribution given the observed data and that computed from replicated data generated from each MCMC iteration given the model posterior distribution. The distribution of the difference in the model χ^2 fit functions, along with the Bayesian 95% CI, are produced (Levy and Mislevy 2016; Muthén and Asparouhov 2012; Wang and Wang 2020). PPP values around 0.5 indicate excellent model fit; with observed data just as probable as data replicated by the model and the fit statistic difference of zero would fall close to the centre of the Bayesian 95% CI, while PPP values close to zero or one indicate model-data misfit, thus poor model fit or model over-fit, respectively (Muthén and Asparouhov 2012; Wang and Wang 2020).[43] There is no cut-off criterion for how low the PPP value can be before a model is considered significantly ill-fitting, however simulation studies have indicated values of 0.1, 0.05 or 0.01 exhibit reasonable fit (Muthén and Asparouhov 2012). Additionally, a positive lower limit of the Bayesian 95% CI (zero not being covered by the 95% CI) of the distribution of the difference in the model χ^2 fit functions is indicative of poor model fit (Muthén and Asparouhov 2012; Wang and Wang 2020).

For model fit using the MLR estimator, the Satorra-Bentler scaled $\Delta\chi^2$ test[44] and the relative information indices were used to compare which model better fitted the data between the sequential MIMIC models and the bifactor measurement model without predictors. The relative information indices used were the *Akaike's Information Criterion* (AIC), the *Bayesian Information Criterion* (BIC), and the sample-size adjusted *Bayesian Information Criterion* (ABIC) (Finch and French 2015; Kline 2016; Wang and Wang 2020). These are parsimony-adjusted measures of fit that penalise a model with more predictors, as opposed to R^2 which will always increase with the addition of more predictors (Field et al. 2012). While M*plus* provides estimates for R^2, adjusted R^2 estimates which penalise a model for non-significant predictors added to the model were calculated using Equation 7.1 to evaluate the variance in the latent traits explained by the demographic, socio-demographic, and socio-economic characteristics:

[43] The PPP value is not the same as the *p*-value for a χ^2 test of model fit but is rather akin to the frequentist conceptualisation of approximate fit indices, for example, the RMSEA and is the proportion of times from a set of iterations where the model χ^2 fit function based on the observed data is smaller than that of the replicated data (Muthén and Asparouhov 2012; Wang and Wang 2020).

[44] This is a Satorra-Bentler $\Delta\chi^2$ using a scaling correction to better approximate the $\Delta\chi^2$ difference test under non-normality (Bryant and Satorra 2012; Satorra and Bentler 2010) based on the log-likelihood and scaling correction factor estimated with the MLR estimator.

$$Adjusted\ R^2 = \frac{(n-1) * R^2 - k}{(n-k-1)} \tag{7.1}$$

where R^2 is the unadjusted value, n is the total sample size, and k is the number of predictors in the model. The R^2 statistic is, however, a measure of explanatory power rather than a measure of goodness-of-fit of the model (De Boeck and Wilson 2016).

7.2 Results

7.2.1 Multiple Indicators Multiple Causes Models Estimation

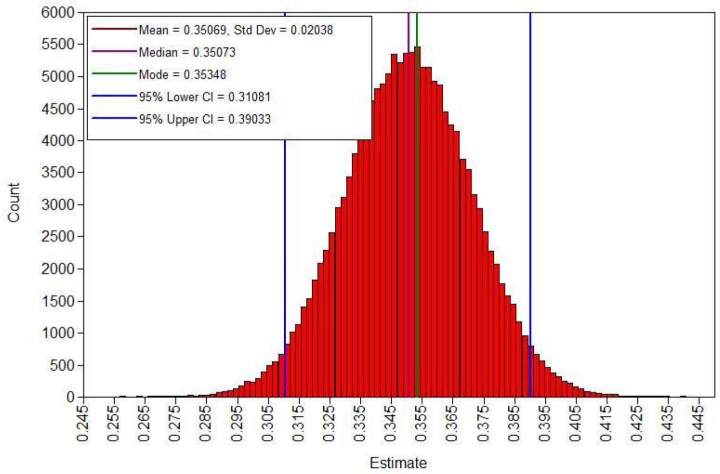
The MIMIC models based on the MLR and Bayesian estimators converged and terminated normally. However, model estimation with the MLR estimator was cumbersome and the models took longer to converge, particularly Model 3. This took approximately 113 times longer to estimate than with the Bayesian estimator (see model estimation times in Tables 7.2 – 7.5).

The Bayesian estimator fulfilled the convergence criterion of the PSR < 1.10. Model 1 terminated with the PSR = 1.017 after 100,000 iterations, while Model 2 terminated with the PSR = 1.023 after 150,000 iterations, and Model 3 terminated with the PSR = 1.039 after 200,000 iterations, estimated with three MCMC chains. Graphical displays of the Bayesian posterior trace plots and distributions for estimated parameters were produced for visual inspection of convergence and how well the posterior distributions were simulated. For illustrative purposes, plots of the effect of sex on *overall QWE* are displayed for the MIMIC models in Figures 7.2 (a – c). The histograms show normal posterior parameter distributions for the effect of sex on *overall QWE*, while the posterior parameter trace plots show that the chains mixed well, indicating that the MCMC algorithm reached equilibria in estimating these parameters and model convergence was achieved.

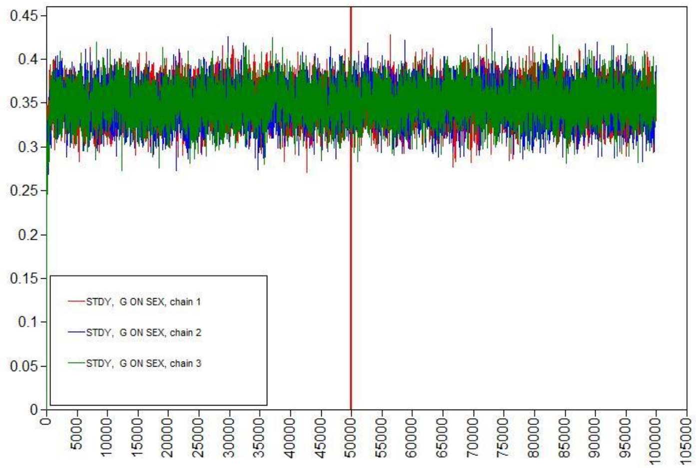
Based on the differences in the model χ^2 fit functions, the models had adequate fit to the data. The lower limit of the Bayesian 95% CI of the difference between observed and model-replicated data χ^2 fit functions were below zero for each of the three models: thus, [–10.16, 168.73] for Model 1, [–42.82, 180.43] for Model 2, and [–18.39, 263.25] for Model 3. Furthermore, the posterior predictive p-values were 0.043 for Models 1 and 3, and 0.118 for Model 2, indicating adequate fit using a cut-off criterion of PPP value > 0.01. The estimates are printed within the histograms and scatterplots in Figures 7.3 (a – c).[45]

[45] Model fit with the Bayesian estimator can be improved by specifying small-variance normal priors for cross-loadings; for example, N (0, 0.01); to reflect that cross-loadings in the model assumed to be zero are approximately, but not exactly zero (Asparouhov and Muthén 2021; Muthén and Asparouhov 2012); however, these were constrained to be zero for the estimates to remain analogous to those based on the MLR estimator.

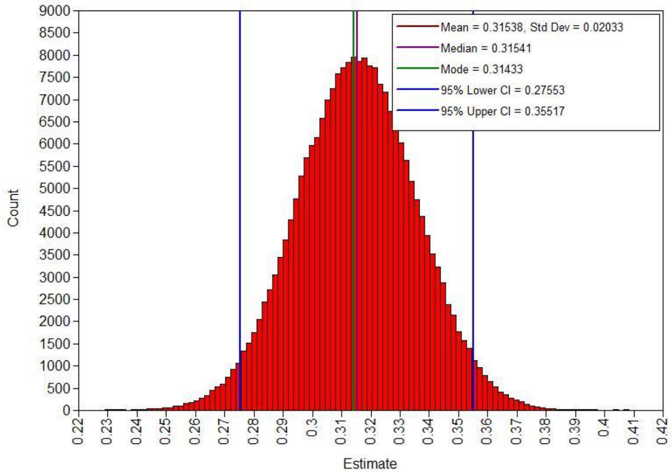
Figure 7.2: Posterior Parameter Distributions and Trace Plots for Overall QWE by Sex based on the Bayesian Estimator



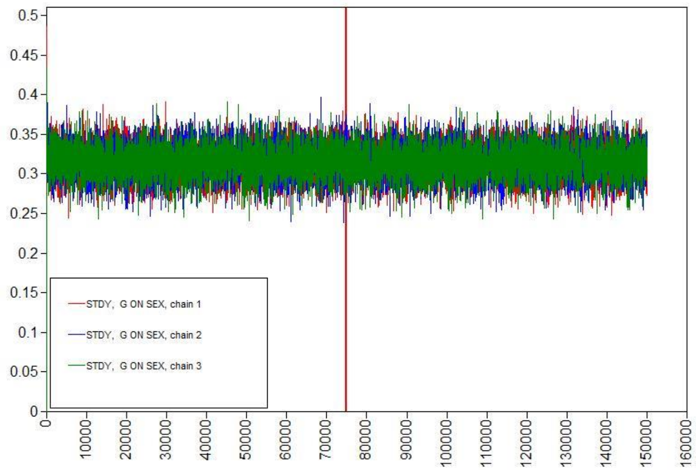
a) Model 1 Posterior Parameter Distribution



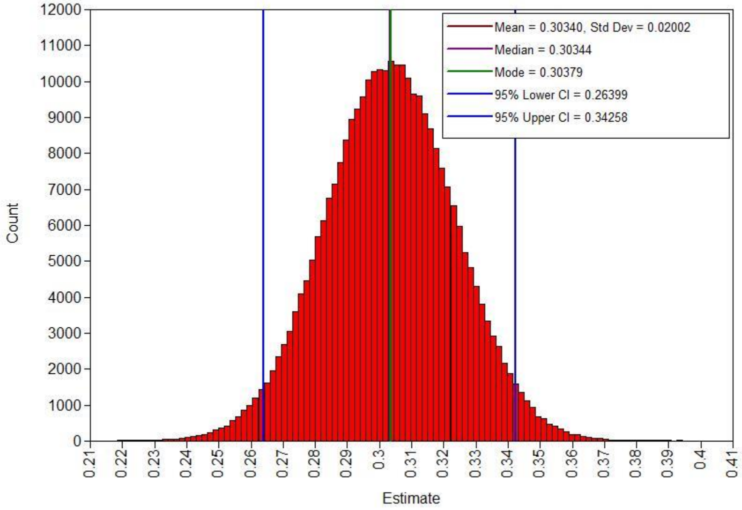
a) Model 1 Posterior Parameter Trace Plot



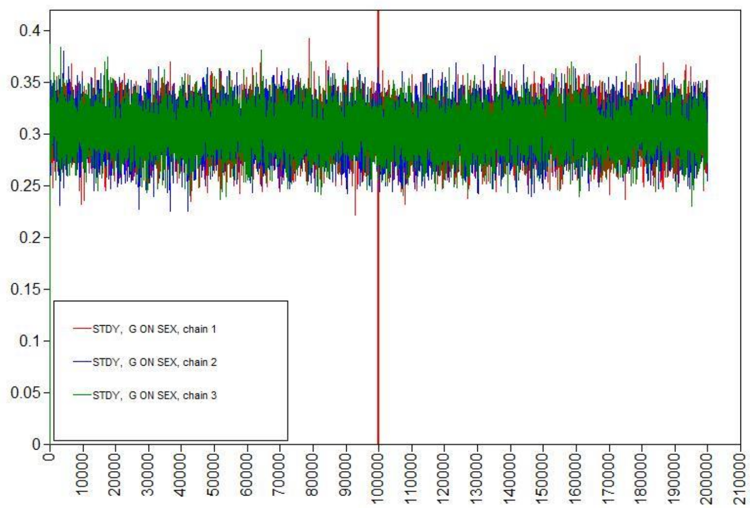
b) Model 2 Posterior Parameter Distribution



b) Model 2 Posterior Parameter Trace Plot



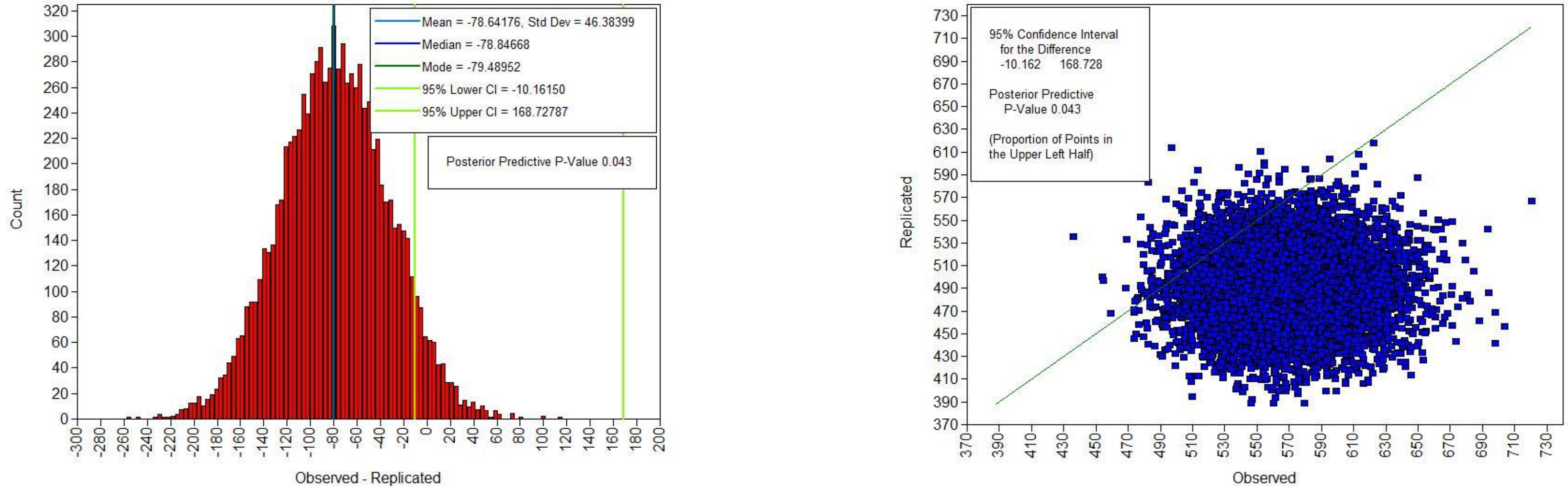
c) Model 3 Posterior Parameter Distribution



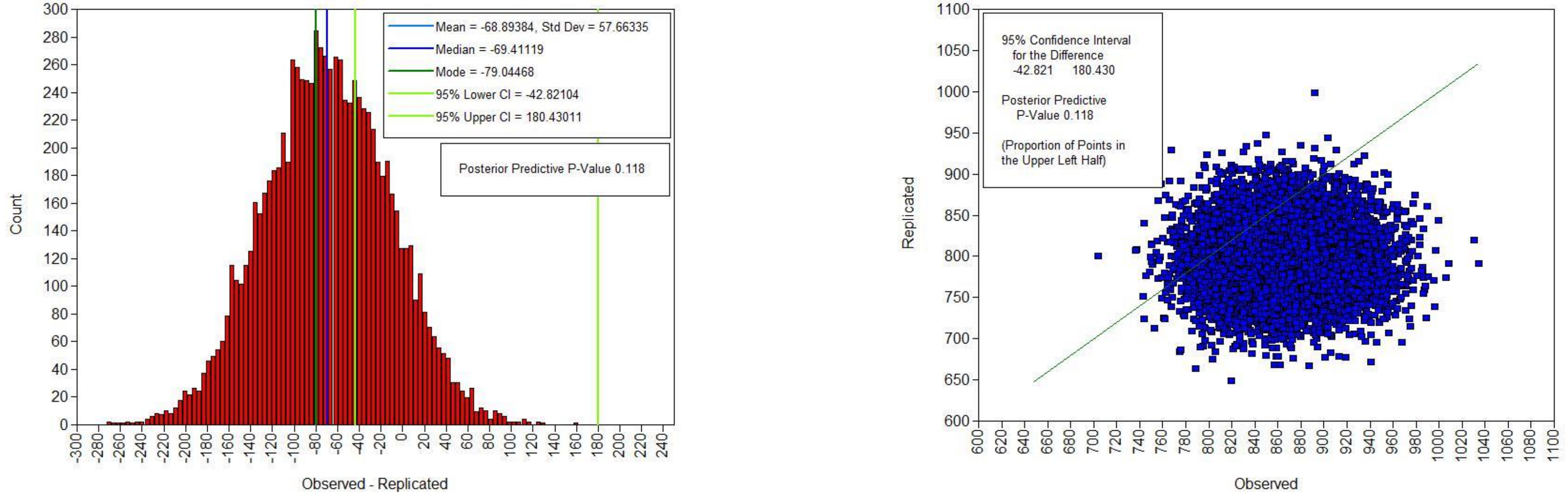
c) Model 3 Posterior Parameter Trace Plot

Notes: The estimates are standardised parameter estimates. The histograms represent the posterior parameter distributions estimated by the model and the posterior means, posterior standard deviations, and 95% credible intervals printed within the charts correspond to the estimates for the MIMIC Models for *overall QWE* (labelled as *G* in the legend of the trace plots) by sex (coefficients for males) presented in Table 7.2. For the trace plots, the vertical axes represent the scale for the posterior parameter estimates, while the horizontal axes represent the number of iterations ran for the MCMC algorithm to estimate each model; thus 100,000 iterations for Model 1, 150,000 iterations for Model 2, and 200,000 iterations for Model 3. The first half of the iterations were considered as 'burn-in' iterations and discarded, with the second half of the iterations used to estimate the posterior parameter distributions. The three chains in each trace plot mixed well, indicating that the MCMC algorithm reached equilibria in the estimating the posterior parameter distributions.

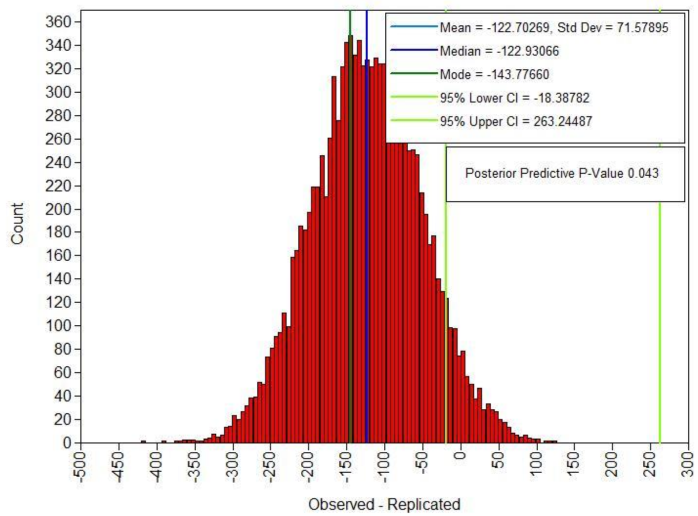
Figure 7.3: Distributions and Scatterplots for Posterior Predictive Checks of the MIMIC Models

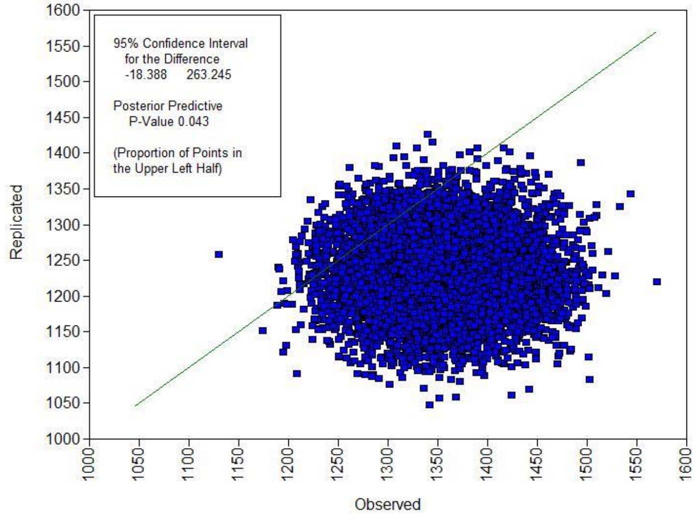


Model 1



Model 2





Model 3

Notes: The histograms display the posterior parameter distributions of the difference in χ^2 fit functions between observed and model-replicated data along with their Bayesian 95% CIs. The lower limits of the 95% CIs were below zero for each of the three models. However, the difference of zero did not fall close to the centre of the distributions, suggesting the models had adequate, but not close or excellent, fit to the data. This was reflected in the asymmetrical scatterplots showing the observed and model-replicated data, with a small proportion of the observed data replicated by the model.

For the MLR estimator, a comparison of the sequential MIMIC models and the bifactor measurement model without predictors using the $\Delta\chi^2$ and the relative information indices suggested that the model with demographic, socio-demographic, and socio-economic characteristics (Model 3) exhibited better fit to the data (Table 7.1). This was indicated by the large values of the $\Delta\chi^2$ differences between Model 3 compared to Model 0, Model 1, or Model 2, as well as and the lower *AIC*, *BIC*, and *ABIC* values for Model 3 compared to other models.

Table 7.1: Model Comparisons with $\Delta\chi^2$ Tests and Relative Information Indices

Model	*AIC*	*BIC*	*ABIC*	*LL*	*df*	*c*	c_d	$\Delta\chi^2$
Model 3	462497	464334	463574	–231009.6	239	1.757		
Model 2	490365	491654	491123	–245015.5	167	1.798	1.662	16855
Model 1	494212	495100	494734	–246990.8	115	1.750	1.763	18124
Model 0	506885	507528	507264	–253359.6	83	1.812	1.728	25872

Notes: *Model 0* is the bifactor measurement model with no predictors. *Model 1* consists of the bifactor measurement model and a structural model with demographic predictors. *Model 2* consists of the bifactor measurement model and a structural model with demographic and socio-demographic predictors. *Model 3* consists of the bifactor measurement and structural models with demographic, socio-demographic, and socio-economic predictors. *LL*: log-likelihood. c is the scaling correction factor obtained with the MLR estimator. c_d is the Satorra-Bentler scaled difference test, $c_d = (df_N * c_N - df_F * c_F) / (df_N - df_F)$, where subscripts *N* and *F* represent nested and full models, respectively. The Satorra-Bentler scaled $\Delta\chi^2$ tests compare each of the nested models to the full model, $\Delta\chi^2 = -2 * (LL_N - LL_F) / c_d$ (Wang and Wang 2020).

7.2.2 Multiple Indicators Multiple Causes Models

Results of the MIMIC models are presented in Tables 7.2 – 7.5 by each latent trait and include estimates based on the MLR and Bayesian estimators. Each table reports nested models with Model 1 modelling the effects of demographic characteristics on the latent traits and nested in the Model 2, which introduced the effects of socio-demographic characteristics, and was in turn nested in Model 3, which introduced the effects of socio-economic characteristics.

As the measurement model is based on a bifactor model, latent trait scores for *overall QWE* are interpreted conditional on other *dimensions of QWE* in the measurement model, while those for each *dimension of QWE* are interpreted over and above *overall QWE*. In terms of reporting coefficients, results are based on the MLR estimator as these account for the complex sample design of the UKHLS. However, a general comparison of the estimates based on the MLR and Bayesian estimators is provided.

Table 7.2: MIMIC Model Results for Overall QWE

	MLR Estimator			Bayesian Estimator		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
	Estimates	Estimates	Estimates	Estimates	Estimates	Estimates
Sex (Reference: Female)						
Male	0.326*** (0.027)	0.316*** (0.027)	0.311*** (0.029)	0.351*** (0.020) [0.311, 0.390]	0.315*** (0.020) [0.276, 0.355]	0.303*** (0.020) [0.264, 0.343]
Ethnic group (Reference: White)						
Mixed	–0.055 (0.107)	–0.071 (0.107)	–0.127 (0.102)	0.038 (0.067) [–0.093, 0.170]	–0.080 (0.068) [–0.212, 0.053]	–0.063 (0.062) [–0.185, 0.059]
Asian or Asian British	–0.001 (0.043)	–0.142** (0.046)	–0.108* (0.046)	–0.054 (0.031) [–0.115, 0.007]	–0.196*** (0.033) [–0.260, –0.132]	–0.159*** (0.031) [–0.220, –0.099]
Black or Black British	–0.214*** (0.060)	–0.319*** (0.066)	–0.167** (0.061)	–0.145*** (0.043) [–0.230, –0.060]	–0.302*** (0.046) [–0.392, –0.212]	–0.169*** (0.043) [–0.253, –0.084]
Age group (Reference: 16 – 24)						
25 – 34	0.210*** (0.047)	0.201*** (0.050)	0.062 (0.050)	0.324*** (0.038) [0.251, 0.397]	0.250*** (0.039) [0.175, 0.326]	0.144*** (0.037) [0.071, 0.217]
35 – 49	0.306*** (0.042)	0.224*** (0.051)	0.136* (0.056)	0.387*** (0.035) [0.320, 0.457]	0.271*** (0.039) [0.194, 0.348]	0.217*** (0.038) [0.142, 0.292]
50 – 64	0.135** (0.045)	0.068 (0.057)	0.090 (0.060)	0.238*** (0.036) [0.168, 0.310]	0.128** (0.042) [0.045, 0.210]	0.170*** (0.041) [0.090, 0.249]
65 +	0.103 (0.074)	0.005 (0.085)	0.187* (0.084)	0.125 (0.070) [–0.012, 0.262]	0.008 (0.075) [–0.139, 0.154]	0.193*** (0.070) [0.057, 0.330]

Continued…

	MLR Estimator			Bayesian Estimator		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
	Estimates	Estimates	Estimates	Estimates	Estimates	Estimates
Relationship status (Reference: Single)						
Married or cohabiting		0.165*** (0.038)	0.103** (0.033)		0.092*** (0.027) [0.038, 0.146]	0.050* (0.025) [0.001, 0.100]
Divorced or separated		0.007 (0.049)	0.000 (0.044)		–0.026 (0.037) [–0.099, 0.046]	–0.018 (0.034) [–0.084, 0.048]
Widowed		0.113 (0.103)	0.077 (0.094)		–0.025 (0.088) [–0.198, 0.147]	–0.020 (0.080) [–0.177, 0.136]
Parental status (Ref: Lone parents with school children)						
Coupled parents with school age children		0.051 (0.063)	–0.087 (0.058)		0.073 (0.049) [–0.022, 0.169]	–0.014 (0.045) [–0.103, 0.075]
Employees without school age children		0.025 (0.051)	–0.154** (0.049)		0.052 (0.040) [–0.027, 0.131]	–0.072 (0.038) [–0.146, 0.003]
Illness or Disability (Reference: Yes)						
No		0.054 (0.028)	0.004 (0.026)		0.044* (0.022) [0.001, 0.088]	0.011 (0.020) [–0.029, 0.051]
Region (Reference: London)						
Southern England		–0.063 (0.053)	–0.058 (0.048)		–0.065 (0.036) [–0.135, 0.005]	–0.059 (0.033) [–0.124, 0.006]
East of England		–0.258*** (0.059)	–0.195*** (0.052)		–0.243*** (0.042) [–0.326, –0.160]	–0.195*** (0.039) [–0.271, –0.118]
The Midlands		–0.204*** (0.056)	–0.122* (0.051)		–0.205*** (0.036) [–0.275, –0.134]	–0.130*** (0.034) [–0.196, –0.063]
Northern England		–0.331*** (0.053)	–0.276*** (0.049)		–0.355*** (0.034) [–0.422, –0.288]	–0.290*** (0.032) [–0.353, –0.227]
Wales		–0.311*** (0.064)	–0.170** (0.062)		–0.312*** (0.048) [–0.405, –0.218]	–0.199*** (0.045) [–0.287, –0.111]
Scotland		–0.494*** (0.064)	–0.369*** (0.056)		–0.444*** (0.044) [–0.529, –0.358]	–0.337*** (0.041) [–0.416, –0.257]
Northern Ireland		–0.779*** (0.080)	–0.496*** (0.072)		–0.753*** (0.050) [–0.850, –0.655]	–0.519*** (0.047) [–0.611, –0.428]

Continued…

	MLR Estimator			Bayesian Estimator		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
	Estimates	Estimates	Estimates	Estimates	Estimates	Estimates
Education (Reference: No qualifications)						
GCSE / O-level or lower			–0.039 (0.037)			0.002 (0.031) [–0.058, 0.062]
Up to A-level			0.054 (0.047)			0.075* (0.037) [0.002, 0.148]
Up to diploma in HE			–0.012 (0.046)			0.010 (0.036) [–0.061, 0.081]
University or higher degree			0.182*** (0.042)			0.204*** (0.031) [0.143, 0.265]
No recorded data			0.009 (0.047)			0.084* (0.036) [0.014, 0.154]
Occupational classification (Ref: Managers & senior officials)						
Professional occupations			–0.512*** (0.044)			–0.551*** (0.032) [–0.614, –0.488]
Associate professional & technical occupations			–0.399*** (0.036)			–0.379*** (0.030) [–0.438, –0.321]
Administrative & secretarial occupations			–0.416*** (0.042)			–0.406*** (0.034) [–0.472, –0.340]
Skilled trades occupations			–0891*** (0.055)			–0.900*** (0.043) [–0.984, –0.814]
Personal service occupations			–1.294*** (0.050)			–1.282*** (0.036) [–1.352, –1.212]
Sales & customer service occupations			–1.070*** (0.056)			–1.014*** (0.040) [–1.092, –0.935]
Process, plant & machine operatives			–1.121*** (0.054)			–1.105*** (0.042) [–1.188, –1.022]
Elementary occupations			–1.302*** (0.050)			–1.306*** (0.036) [–1.376, –1.234]

Continued…

	MLR Estimator			Bayesian Estimator		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
	Estimates	Estimates	Estimates	Estimates	Estimates	Estimates
Full or Part-time (Reference: Part-time)						
Full-time			0.183*** (0.036)			0.214*** (0.024) [0.167, 0.262]
Organisational sector (Reference: Private sector)						
Public sector			–0.448*** (0.047)			–0.453*** (0.023) [–0.498, –0.408]
Organisation size (Reference: Micro)						
Small			–0.265*** (0.034)			–0.252*** (0.028) [–0.306, –0.198]
Medium			–0.325*** (0.039)			–0.283*** (0.030) [–0.342, –0.225]
Large			0.013 (0.040)			0.067* (0.029) [0.009, 0.124]
Intercept (constrained)	0.000	0.000	0.000	0.000	0.000	0.000
R^2	0.037*** (0.005)	0.074*** (0.007)	0.427*** (0.029)	0.045*** (0.004) [0.037, 0.054]	0.077*** (0.005) [0.067, 0.087]	0.433*** (0.011) [0.412, 0.455]
Adjusted R^2	0.037	0.073	0.426			
Unweighted sample size	16,678	16,582	16,068	16,678	16,582	16,068
Model estimation time (hours)	51.40	321.66	1,017.28	3.78	8.41	9.22

Notes: Data from UKHLS, Wave 8 (2016 – 2017). Standardised coefficients and estimates in parentheses are standard errors (MLR estimator) or posterior standard deviations (Bayesian estimator) and ones in square brackets indicate 95% credible intervals. Significance tests for the MLR estimator are based on two-tailed p-values, while those for the Bayesian estimator are based on one-tailed p-values. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$. Estimates based on the MLR estimator consider the complex sample design of the UKHLS data, while estimates based on the Bayesian estimator are unweighted.

Overall QWE

Model parameter estimates based on the MLR and Bayes estimators yielded similar results in predicting *overall QWE*. Posterior parameter trace plots for all predictors for Model 3 for *overall QWE* (excluding sex) indicated that for the Bayes estimator, the MCMC algorithm reached equilibria in estimating the posterior parameter distributions (Appendix 7.1).[46] Considering the full model, point estimates based on the MLR estimator were within the Bayesian 95% CIs of the estimates from the Bayes estimator, except for coefficients for the '25 – 34' and '35 – 49' age groups, 'married/cohabiting' relationship status, and 'employees without primary school age children' in terms of parental status (Table 7.2). Based on the adjusted R^2 estimates,[47] demographic characteristics explained approximately 4% of the variation in *overall QWE*, while the addition of socio-demographic characteristics resulted in the model explaining approximately 7% of the variation. However, when the socio-economic characteristics were introduced, the model explained approximately 43% of the variation in *overall QWE*.

The effect of sex on *overall QWE* was statistically significant when controlling for other predictors in the models. Considering Model 3, expected *overall QWE* was 0.311 units (*SE* = 0.029) higher for male than female employees. For ethnic group and Model 1, only the difference between employees from White and Black or Black British ethnic backgrounds was statistically significant. However, when socio-demographic and socio-economic characteristics were introduced, the difference between employees from White and Asian or Asian British ethnic backgrounds also became statistically significant. From Model 3, average *overall QWE*

[46] Bayesian posterior parameter trace plots associated with predictors for the MIMIC Model 3 for *economic compensation*, *working conditions*, and *work-life balance* are displayed in Appendices 7.3, 7.5 and 7.7, respectively. Posterior parameter distributions for the parameter estimates for *overall QWE*, *economic compensation*, *working conditions*, and *work-life balance* are available but not included.

[47] Refer to Appendix 7.2 for the posterior parameter distributions and trace plots for R^2 estimates (unadjusted) for o*verall QWE* based on the Bayesian estimator. Estimates based on the MLR and Bayesian estimators yielded similar results.

was lower for employees from Asian or Asian British (–0.108 units, *SE* = 0.046) and Black or Black British (–0.167 units, *SE* = 0.061) than for those from White ethnic backgrounds. Regarding age group, differences in expected *overall QWE* between employees aged 25 – 34, 35 – 49 or 50 – 64 years old compared to those aged 16 – 24 years old were statistically significant (Model 1). When socio-demographic characteristics were introduced, the difference between employees aged 16 – 24 and 50 – 64 years old became statistically insignificant. On the other hand, the introduction of socio-economic characteristics resulted in statistically significant differences between those aged 16 – 24 years old and those aged 35 – 49 or 65+ years old. Based on the full model, average *overall QWE* was lower for employees aged 16 – 24 years old compared to those aged 35 – 49 (0.136 units, *SE* = 0.056) or 65+ (0.187 units, *SE* = 0.084) years old. Notably, based on the Bayes estimator, differences in *overall QWE* between 16 – 24 years old and employees in other age groups were statistically significant (Model 3).

For relationship status, differences in *overall QWE* between married/cohabiting and single employees were statistically significant, while differences between single employees and those in any other relationship status were not statistically significant. Expected *overall QWE* was higher for married/cohabiting employees (0.103 units, *SE* = 0.033) than for single employees (Model 3). While from Model 2 the effect of parental status on *overall QWE* was not statistically significant, when socio-economic characteristics were introduced, differences between lone parents with and employees without primary school age children became statistically significant. Employees without primary school age children had lower expected *overall QWE* (–0.154 units, *SE* = 0.049) than lone parents with primary school age children. On the other hand, the effect of longstanding illness or disability on *overall QWE* was not statistically significant. In terms of region, there was no statistically significant difference in *overall QWE* between employees in London and Southern England. However, there were statistically significant differences between employees in London and those in any other region, with those in London having higher expected *overall QWE*. Controlling for socio-economic characteristics had the effect of reducing the magnitude of the average difference in *overall QWE* between employees in London and other regions.

Moving on to education, differences in *overall QWE* between employees with no qualifications and other educational attainment were not statistically significant, except for those with a university or higher degree. Expected *overall QWE* for employees with a university or higher degree was 0.182 units (*SE* = 0.042) higher than for those with no qualifications. For occupational classification, differences in *overall QWE* between managers and senior officials compared to employees in other occupational groups were statistically significant. On average, *overall QWE* for managers and senior officials was higher than that for employees in any other occupational group. Considering full or part-time employment, differences were statistically significant and expected *overall QWE* was 0.183 units (*SE* = 0.036) higher for employees in full-time compared to those in part-time employment. The effect of organisational sector on *overall QWE* was also statistically significant. On average, *overall QWE* for public sector employees was 0.448 units (*SE* = 0.047) lower than that for private sector employees. There were statistically significant differences in *overall QWE* for employees in micro size organisations compared to those in small or medium size organisations, but not so when compared to those in large size organisations. Expected *overall QWE* was 0.265 units (*SE* = 0.034) and 0.325 units (*SE* = 0.039) lower for employees in small and medium size organisations, respectively, than for those in micro size organisations.

Table 7.3: MIMIC Model Results for Economic Compensation

	MLR Estimator			Bayesian Estimator		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
	Estimates	Estimates	Estimates	Estimates	Estimates	Estimates
Sex (Reference: Female)						
Male	0.284*** (0.035)	0.277*** (0.031)	0.145*** (0.031)	0.233*** (0.027) [0.179, 0.286]	0.277*** (0.022) [0.233, 0.320]	0.164*** (0.023) [0.119, 0.209]
Ethnic group (Reference: White)						
Mixed	–0.077 (0.121)	–0.196 (0.128)	–0.178 (0.098)	0.053 (0.083) [–0.109, 0.215]	0.017 (0.076) [–0.133, 0.166]	–0.129 (0.067) [–0.258, 0.005]
Asian or Asian British	–0.145* (0.067)	–0.370*** (0.066)	–0.294*** (0.053)	–0.396*** (0.037) [–0.468, –0.324]	–0.483*** (0.037) [–0.556, –0.410]	–0.379*** (0.034) [–0.447, –0.312]
Black or Black British	–0.186* (0.080)	–0.321*** (0.076)	–0.252** (0.078)	–0.162** (0.052) [–0.263, –0.061]	–0.277*** (0.051) [–0.378, –0.176]	–0.177*** (0.047) [–0.265, –0.085]
Age group (Reference: 16 – 24)						
25 – 34	0.903*** (0.049)	0.741*** (0.053)	0.389*** (0.047)	0.870*** (0.045) [0.783, 0.958]	0.758*** (0.044) [0.672, 0.845]	0.390*** (0.039) [0.316, 0.467]
35 – 49	1.308*** (0.040)	1.071*** (0.054)	0.600*** (0.050)	1.154*** (0.045) [1.066, 1.244]	1.045*** (0.044) [0.958, 1.131]	0.605*** (0.040) [0.527, 0.683]
50 – 64	1.234*** (0.041)	0.944*** (0.055)	0.606*** (0.052)	1.036*** (0.047) [0.946, 1.131]	0.910*** (0.046) [0.819, 1.000]	0.586*** (0.043) [0.503, 0.669]
65 +	0.082 (0.095)	–0.070 (0.109)	–0.033 (0.083)	0.077 (0.084) [–0.087, 0.242]	0.058 (0.084) [–0.107, 0.221]	0.024 (0.075) [–0.122, 0.175]

Continued…

	MLR Estimator			Bayesian Estimator		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
	Estimates	Estimates	Estimates	Estimates	Estimates	Estimates
Relationship status (Reference: Single)						
Married or cohabiting		0.355*** (0.040)	0.236*** (0.032)		0.271*** (0.030) [0.211, 0.331]	0.222*** (0.027) [0.169, 0.277]
Divorced or separated		0.111* (0.051)	0.096* (0.044)		0.078 (0.041) [–0.002, 0.158]	0.122** (0.037) [0.051, 0.197]
Widowed		–0.025 (0.138)	0.088 (0.099)		–0.053 (0.096) [–0.242, 0.136]	0.069 (0.084) [–0.095, 0.237]
Parental status (Ref: Lone parents with school children)						
Coupled parents with school age children		0.285*** (0.072)	–0.016 (0.062)		0.252*** (0.055) [0.143, 0.360]	–0.048 (0.049) [–0.147, 0.048]
Employees without school age children		0.357*** (0.056)	0.016 (0.048)		0.335*** (0.046) [0.244, 0.425]	–0.002 (0.040) [–0.081, 0.076]
Illness or Disability (Reference: Yes)						
No		0.177*** (0.031)	0.083** (0.027)		0.174*** (0.025) [0.125, 0.223]	0.094*** (0.022) [0.050, 0.136]
Region (Reference: London)						
Southern England		–0.507*** (0.060)	–0.181** (0.053)		–0.382*** (0.041) [–0.463, –0.301]	–0.155*** (0.037) [–0.229, –0.083]
East of England		–0.420*** (0.071)	–0.142* (0.060)		–0.301*** (0.050) [–0.398, –0.203]	–0.107* (0.044) [–0.193, –0.020]
The Midlands		–0.633*** (0.063)	–0.262*** (0.056)		–0.509*** (0.042) [–0.591, –0.427]	–0.241*** (0.037) [–0.314, –0.166]
Northern England		–0.568*** (0.060)	–0.260*** (0.054)		–0.486*** (0.040) [–0.565, –0.408]	–0.233*** (0.035) [–0.303, –0.164]
Wales		–0.643*** (0.080)	–0.224*** (0.063)		–0.537*** (0.054) [–0.643, –0.431]	–0.160** (0.050) [–0.257, –0.060]
Scotland		–0.321*** (0.074)	–0.043 (0.067)		–0.197*** (0.051) [–0.296, –0.097]	0.043 (0.047) [–0.051, 0.136]
Northern Ireland		–0.675*** (0.103)	–0.333*** (0.089)		–0.392*** (0.057) [–0.503, –0.282]	–0.129* (0.052) [–0.230, –0.026]

Continued…

	MLR Estimator			Bayesian Estimator		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
	Estimates	Estimates	Estimates	Estimates	Estimates	Estimates
Education (Reference: No qualifications)						
GCSE / O-level or lower			0.033 (0.038)			0.046 (0.032) [–0.014, 0.110]
Up to A-level			0.123* (0.048)			0.127** (0.041) [0.046, 0.206]
Up to Diploma in HE			0.245*** (0.045)			0.249*** (0.040) [0.172, 0.327]
University or higher degree			0.358*** (0.040)			0.372*** (0.034) [0.307, 0.441]
No recorded data			0.047 (0.049)			0.009 (0.039) [–0.068, 0.086]
Occupational classification (Ref: Managers & senior officials)						
Professional occupations			0.299*** (0.048)			0.369*** (0.039) [0.289, 0.444]
Associate professional & technical occupations			–0.036 (0.043)			0.005 (0.036) [–0.066, 0.075]
Administrative & secretarial occupations			–0.475*** (0.048)			–0.460*** (0.038) [–0.535, –0.387]
Skilled trades occupations			–0.333*** (0.059)			–0.327*** (0.051) [–0.431, –0.231]
Personal service occupations			–0.683*** (0.061)			–0.637*** (0.046) [–0.730, –0.548]
Sales & customer service occupations			–0.772*** (0.063)			–0.775*** (0.046) [–0.865, –0.685]
Process, plant & machine operatives			–0.490*** (0.059)			–0.544*** (0.049) [–0.638, –0.448]
Elementary occupations			–0.935*** (0.058)			–0.926*** (0.045) [–1.016, –0.838]

Continued…

	MLR Estimator			Bayesian Estimator		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
	Estimates	Estimates	Estimates	Estimates	Estimates	Estimates
Full or Part-time (Reference: Part-time)						
Full-time			0.258*** (0.035)			0.254*** (0.026) [0.202, 0.305]
Organisational sector (Reference: Private sector)						
Public sector			0.595*** (0.058)			0.755*** (0.030) [0.696, 0.815]
Organisation size (Reference: Micro)						
Small			0.608*** (0.036)			0.628*** (0.028) [0.573, 0.684]
Medium			0.936*** (0.039)			0.986*** (0.030) [0.928, 1.044]
Large			1.182*** (0.040)			1.236*** (0.028) [1.181, 1.291]
Intercept (constrained)	0.000	0.000	0.000	0.000	0.000	0.000
R^2	0.196*** (0.011)	0.263*** (0.014)	0.870*** (0.008)	0.149*** (0.010) [0.131, 0.169]	0.217*** (0.009) [0.199, 0.236]	0.988*** (0.006) [0.974, 0.996]
Adjusted R^2	0.196	0.262	0.870			
Unweighted sample size	16,678	16,582	16,068	16,678	16,582	16,068
Model estimation time (hours)	51.40	321.66	1,017.28	3.78	8.41	9.22

Notes: Data from UKHLS, Wave 8 (2016 – 2017). Standardised coefficients and estimates in parentheses are standard errors (MLR estimator) or posterior standard deviations (Bayesian estimator) and ones in square brackets indicate 95% credible intervals. Significance tests for the MLR estimator are based on two-tailed p-values, while those for the Bayesian estimator are based on one-tailed p-values. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$. Estimates based on the MLR estimator consider the complex sample design of the UKHLS data, while estimates based on the Bayesian estimator are unweighted.

Economic Compensation

As with the model parameter estimates predicting *overall QWE*, estimates for the MIMIC models using the MLR and Bayes estimators yielded similar results in predicting *economic compensation*. The posterior parameter trace plots for predictors of *economic compensation* for Model 3 are presented in Appendix 7.3 and these suggested that the MCMC algorithm reached equilibria in estimating the posterior parameter distributions. Point estimates based on the MLR estimator for Model 3 were within the Bayesian 95% CIs of the estimates based on the Bayes estimator, except for coefficients for the 'Asian or Asian British' ethnic group, employees in the 'Northern Ireland' region, and that for employees in the 'public sector' (Table 7.3). Based on the adjusted R^2 estimates using the MLR estimator,[48] demographic characteristics explained approximately 20% of the variation in *economic compensation*. The addition of socio-demographic characteristics resulted in the model explaining approximately 26% of the variation, while the introduction of socio-economic characteristics resulted in the model explaining approximately 87% of the variation.

Differences in *economic compensation* by sex were statistically significant in all three models, but the magnitude of these differences decreased, especially with the introduction of socio-economic characteristics. Thus, average *economic compensation* was 0.145 units (*SE* = 0.031) higher for male than female employees (Model 3). On average, employees from a White ethnic background had higher *economic compensation* than those from Asian or Asian British (–0.294 units, *SE* = 0.053), or Black or Black British (–0.252 units, *SE* = 0.078) ethnic backgrounds and differences were statistically significant (Model 3). However, no significant differences were found between those from White and Mixed ethnic backgrounds in any of the

[48] Refer to Appendix 7.4 for the posterior parameter distributions and trace plots for R^2 estimates (unadjusted) for *economic compensation* based on the Bayes estimator. The MLR and Bayesian estimators yielded different R^2 estimates, particularly for Model 3, and the trace plot suggested that the MCMC algorithm did not converge and reach an equilibrium in estimating the posterior distribution.

three models. For age group, differences in e*conomic compensation* between employees aged 16 – 24 and 65+ year olds were not statistically significant in any of the models. On the other hand, differences between those aged 16 – 24 years old and employees in other age groups were statistically significant. Employees aged 25 – 34 (0.389 units, *SE* = 0.047), 35 – 49 (0.600 units, *SE* = 0.050) or 50 – 64 (0.606 units, *SE* = 0.052) years old had higher expected *economic compensation* than for those aged 16 – 24 years old (Model 3). However, the magnitude of these differences decreased with the introduction of socio-demographic and socio-economic characteristics.

In terms of relationship status, in both models, there were statistically significant differences in *economic compensation* between single employees and married/cohabiting or divorced/separated employees. However, differences between single and widowed employees were not statistically significant. On average, single employees had lower *economic compensation* compared to married/cohabiting (0.236 units, *SE* = 0.032) or divorced/separated (0.096 units, *SE* = 0.044) (Model 3). For parental status, while the effects on *economic compensation* were statistically significant when controlling for demographic and other socio-demographic characteristics, in Model 3 the effect of parental status was not statistically significant. The effect of longstanding illness or disability on *economic compensation* was statistically significant; however, the magnitude of the effect decreased with the introduction of socio-economic characteristics. Expected *economic compensation* was slightly higher for employees without (0.083 units, *SE* = 0.027) compared to those with a longstanding illness or disability (Model 3). From Model 2, differences in *economic compensation* between employees in London compared to those in other regions were statistically significant. However, when socio-economic characteristics were introduced (Model 3), the difference between employees in London and those in Scotland was not statistically significant. On average, employees in

London had higher *economic compensation* compared to those in other regions, although the magnitude of the effect decreased with the introduction of socio-economic characteristics.

Considering education, differences in *economic compensation* between employees with no qualifications and those with other educational qualifications were statistically significant, except for those with GCSE / O-level or lower. Expected *economic compensation* was lower for employees with no qualifications and the magnitude of the difference increased with increasing educational attainment. Thus, 0.123 units (*SE* = 0.048) higher for those with up to A-level, 0.245 units (*SE* = 0.045) higher for those with up to a diploma in higher education, and 0.358 units (*SE* = 0.040) higher for those with a university or higher degree. Regarding occupational classification, there were statistically significant differences in *economic compensation* between managers and senior officials compared to employees in other occupational groups, except for those in associate professional and technical occupations. On average, employees in professional occupations (0.299 units, *SE* = 0.048) had better *economic compensation* than managers and senior officials, while employees in other occupational groups had poorer *economic compensation* than managers and senior officials. Employees in full-time employment (0.285 units, *SE* = 0.035) had higher expected *economic compensation* than those in part-time employment, while expected *economic compensation* was higher for public sector (0.595 units, *SE* = 0.058) than the private sector employees, with differences statistically significant. Differences in *economic compensation* between employees in micro size organisations compared to those in organisations of other size were statistically significant, with expected *economic compensation* lower for employees in micro size organisations. The magnitude of the differences increased with increasing organisation size; thus, 0.608 units (*SE* = 0.036) higher for those in small size organisations, 0.936 units (*SE* = 0.039) higher for those in medium size organisations, and 1.182 units (*SE* = 0.040) higher for those in large size organisations.

Table 7.4: MIMIC Model Results for Working Conditions

	MLR Estimator			Bayesian Estimator		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
	Estimates	Estimates	Estimates	Estimates	Estimates	Estimates
Sex (Reference: Female)						
Male	–0.089** (0.026)	–0.091** (0.027)	–0.171*** (0.032)	–0.100*** (0.020) [–0.139, –0.061]	–0.085*** (0.020) [–0.124, –0.046]	–0.164*** (0.023) [–0.208, –0.120]
Ethnic group (Reference: White)						
Mixed	–0.178 (0.093)	–0.185 (0.095)	–0.163 (0.107)	–0.093 (0.066) [–0.222, 0.035]	–0.055 (0.066) [–0.185, 0.075]	–0.046 (0.067) [–0.178, 0.086]
Asian or Asian British	–0.011 (0.053)	–0.016 (0.056)	0.060 (0.058)	–0.032 (0.030) [–0.091, 0.028]	–0.021 (0.032) [–0.084, 0.043]	0.082* (0.033) [0.016, 0.147]
Black or Black British	–0.063 (0.066)	–0.057 (0.073)	–0.019 (0.069)	–0.181*** (0.042) [–0.263, –0.099]	–0.147** (0.045) [–0.235, –0.059]	–0.065 (0.046) [–0.156, 0.026]
Age group (Reference: 16 – 24)						
25 – 34	0.190*** (0.047)	0.153** (0.057)	0.107 (0.055)	0.168*** (0.037) [0.097, 0.241]	0.158*** (0.037) [0.085, 0.230]	0.084* (0.040) [0.006, 0.164]
35 – 49	0.319*** (0.041)	0.284*** (0.057)	0.179** (0.059)	0.274*** (0.034) [0.210, 0.340]	0.256*** (0.037) [0.182, 0.329]	0.131** (0.042) [0.050, 0.214]
50 – 64	0.285*** (0.042)	0.253*** (0.058)	0.153* (0.061)	0.238*** (0.035) [0.172, 0.306]	0.218*** (0.040) [0.140, 0.296]	0.102* (0.044) [0.016, 0.189]
65 +	0.364*** (0.090)	0.350** (0.103)	0.295** (0.104)	0.392*** (0.069) [0.256, 0.528]	0.368*** (0.074) [0.224, 0.513]	0.308*** (0.077) [0.158, 0.459]

Continued…

	MLR Estimator			Bayesian Estimator		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
	Estimates	Estimates	Estimates	Estimates	Estimates	Estimates
Relationship status (Reference: Single)						
Married or cohabiting		0.069 (0.036)	0.015 (0.036)		0.086** (0.027) [0.034, 0.139]	0.048 (0.027) [–0.005, 0.102]
Divorced or separated		–0.009 (0.049)	–0.029 (0.049)		0.035 (0.036) [–0.036, 0.106]	0.023 (0.037) [–0.048, 0.095]
Widowed		–0.017 (0.119)	–0.015 (0.115)		0.135 (0.087) [–0.036, 0.305]	0.146 (0.087) [–0.025, 0.316]
Parental status (Ref: Lone parents with school children)						
Coupled parents with school age children		–0.159* (0.077)	–0.123 (0.074)		–0.038 (0.047) [–0.131, 0.054]	–0.049 (0.050) [–0.146, 0.049]
Employees without school age children		–0.103 (0.067)	–0.068 (0.065)		–0.011 (0.039) [–0.087, 0.064]	–0.029 (0.042) [–0.111, 0.054]
Illness or Disability (Reference: Yes)						
No		0.113*** (0.031)	0.093** (0.030)		0.107*** (0.022) [0.064, 0.150]	0.086*** (0.023) [0.042, 0.131]
Region (Reference: London)						
Southern England		–0.023 (0.054)	0.014 (0.055)		0.001 (0.035) [–0.068, 0.069]	0.047 (0.036) [–0.024, 0.118]
East of England		0.002 (0.062)	0.045 (0.062)		0.040 (0.042) [–0.042, 0.122]	0.090* (0.043) [0.006, 0.175]
The Midlands		–0.010 (0.058)	0.066 (0.059)		0.021 (0.036) [–0.049, 0.090]	0.097** (0.037) [0.025, 0.169]
Northern England		0.010 (0.054)	0.091 (0.056)		0.037 (0.034) [–0.029, 0.103]	0.125*** (0.036) [0.056, 0.195]
Wales		0.022 (0.068)	0.072 (0.069)		0.029 (0.047) [–0.063, 0.122]	0.106* (0.048) [0.011, 0.201]
Scotland		0.139* (0.066)	0.212** (0.066)		0.180*** (0.043) [0.095, 0.265]	0.257*** (0.045) [0.168, 0.345]
Northern Ireland		0.021 (0.096)	0.068 (0.088)		0.089 (0.050) [–0.008, 0.186]	0.149** (0.051) [0.049, 0.250]

Continued…

	MLR Estimator			Bayesian Estimator		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
	Estimates	Estimates	Estimates	Estimates	Estimates	Estimates
Education (Reference: No qualifications)						
GCSE / O-level or lower			0.042 (0.048)			0.016 (0.033) [–0.049, 0.081]
Up to A-level			–0.104 (0.058)			–0.078 (0.041) [–0.157, 0.002]
Up to Diploma in HE			0.046 (0.056)			0.007 (0.040) [–0.071, 0.085]
University or higher degree			–0.057 (0.049)			–0.063 (0.035) [–0.131, 0.005]
No recorded data			–0.033 (0.057)			–0.049 (0.039) [–0.124, 0.028]
Occupational classification (Ref: Managers & senior officials)						
Professional occupations			–0.316*** (0.048)			–0.287*** (0.038) [–0.360, –0.212]
Associate professional & technical occupations			–0.454*** (0.046)			–0.448*** (0.035) [–0.516, –0.380]
Administrative & secretarial occupations			–0.474*** (0.055)			–0.482*** (0.039) [–0.557, –0.405]
Skilled trades occupations			–0.355*** (0.071)			–0.356*** (0.051) [–0.455, –0.256]
Personal service occupations			–0.284*** (0.072)			–0.291*** (0.047) [–0.383, –0.198]
Sales & customer service occupations			–0.523*** (0.072)			–0.549*** (0.048) [–0.642, –0.454]
Process, plant & machine operatives			–0.635*** (0.071)			–0.627*** (0.050) [–0.725, –0.529]
Elementary occupations			–0.395*** (0.073)			–0.395*** (0.048) [–0.488, –0.301]

Continued…

	MLR Estimator			Bayesian Estimator		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
	Estimates	Estimates	Estimates	Estimates	Estimates	Estimates
Full or Part-time (Reference: Part-time)						
Full-time			0.091* (0.035)			0.076** (0.026) [0.026, 0.126]
Organisational sector (Reference: Private sector)						
Public sector			0.199*** (0.043)			0.171*** (0.025) [0.123, 0.220]
Organisation size (Reference: Micro)						
Small			–0.147** (0.044)			–0.137*** (0.031) [–0.197, –0.076]
Medium			–0.220*** (0.045)			–0.217*** (0.033) [–0.280, –0.153]
Large			–0.413*** (0.043)			–0.409*** (0.031) [–0.469, –0.348]
Intercept (constrained)	0.000	0.000	0.000	0.000	0.000	0.000
R^2	0.012*** (0.003)	0.017*** (0.003)	0.080*** (0.007)	0.012*** (0.000) [0.008, 0.016]	0.019*** (0.000) [0.014, 0.024]	0.080*** (0.005) [0.070, 0.090]
Adjusted R^2	0.012	0.016	0.078			
Unweighted sample size	16,678	16,582	16,068	16,678	16,582	16,068
Model estimation time (hours)	51.40	321.66	1,017.28	3.78	8.41	9.22

Notes: Data from UKHLS, Wave 8 (2016 – 2017). Standardised coefficients and estimates in parentheses are standard errors (MLR estimator) or posterior standard deviations (Bayesian estimator) and ones in square brackets indicate 95% credible intervals. Significance tests for the MLR estimator are based on two-tailed p-values, while those for the Bayesian estimator are based on one-tailed p-values. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$. Estimates based on the MLR estimator consider the complex sample design of the UKHLS data, while estimates based on the Bayesian estimator are unweighted.

Working Conditions

Model parameter estimates for the MIMIC models predicting *working conditions* based on the MLR and Bayes estimators yielded similar results. The posterior parameter trace plots for predictors of *working conditions* for Model 3 suggested that the MCMC algorithm reached equilibria in estimating the posterior parameter distributions (Appendix 7.5). Point estimates based on the MLR estimator for Model 3 were within the Bayesian 95% CIs of the estimates based on the Bayes estimator (Table 7.4). Results from the adjusted R^2 estimates[49] suggested that demographic, socio-demographic, and socio-economic characteristics did not explain much of the variation in *working conditions*. Thus, demographic characteristics explained approximately 1% of the variation in *working conditions*, while the addition of socio-demographic characteristics resulted in the model explaining approximately 2% of the variation, and the introduction of socio-economic characteristics resulted in the model explaining approximately 8% of the variation (Table 7.4).

Considering sex, differences in *working conditions* were statistically significant in all three models. However, there was a marked increase in the magnitude of the differences when socio-economic characteristics were introduced. Expected *working conditions* were 0.171 units (*SE* = 0.032) poorer for male than female employees (Model 3). The effect of ethnic group on *working conditions* was not statistically significant in any of the three models. Differences in *working conditions* between employees aged 16 – 24 years old and those in other age groups were statistically significant in Models 1 and 2. However, when socio-economic characteristics were introduced, there was no longer a difference between employees aged 16 – 24 and 25 – 34 years old. Based on Model 3, employees aged 16 – 24 years old had poorer expected *working*

[49] Refer to Appendix 7.6 for the posterior parameter distributions and trace plots for R^2 estimates (unadjusted) for *working conditions* based on the Bayesian estimator.

conditions than those aged 35 – 49 (0.179 units, *SE* = 0.059), 50 – 64 (0.153 units, *SE* = 0.061) or 65+ (0.295 units, *SE* = 0.104) years old.

For relationship status, differences in *working conditions* not statistically significant in both models. On the other hand, for parental status, the difference in *working conditions* between lone parents and coupled parents with primary school age children was statistically significant in Model 2, while that between lone parents with and employees without primary school age children was not statistically significant. However, when socio-economic characteristics were introduced (Model 3), there was no longer a statistically significant effect of parental status on *working conditions*. There was a statistically significant difference in *working conditions* by longstanding illness or disability in both models. Expected *working conditions* were slightly better for employees without a longstanding illness or disability (0.093 units, *SE* = 0.030) than those with a longstanding illness or disability. In terms of region, the difference in *working conditions* between employees in London and Scotland was statistically significant in both models, while differences between employees in London and those in any other region were not statistically significant. On average, *working conditions* for employees in Scotland were 0.212 units (*SE* = 0.066) better than for those in London and controlling for socio-economic characteristics resulted in a larger effect (Model 3).

In terms of education, the effect on *working conditions* was not statistically significant when controlling for all other predictors. On the other hand, differences in *working conditions* between managers and senior officials and employees in any other occupational group were statistically significant. Thus, expected *working conditions* for managers and senior officials were better than for employees in any other occupational group. Controlling for all other predictors, employees in full-time employment (0.091 units, *SE* = 0.035) had slightly better expected *working conditions* than those in part-time employment, while employees working in the public sector (0.199 units, *SE* = 0.043) had better expected *working conditions* compared to those working in the private sector. The differences between the groups in both predictors were statistically significant. For organisation size, there were statistically significant differences in *working conditions* between employees in micro size organisations compared to those in organisations of other size. On average, employees in micro size organisations had better *working conditions* than those in small (–0.147 units, $SE = 0.044$), medium (–0.220 units, $SE = 0.045$) or large (–0.413 units, $SE = 0.043$) size organisations. Notably, the absolute value of the coefficients increased with increasing organisation size.

Table 7.5: MIMIC Model Results for Work-time Scheduling

	MLR Estimator			Bayesian Estimator		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
	Estimates	Estimates	Estimates	Estimates	Estimates	Estimates
Sex (Reference: Female)						
Male	–0.723*** (0.025)	–0.731*** (0.025)	–0.395*** (0.028)	–0.680*** (0.019) [–0.717, –0.643]	–0.683*** (0.019) [–0.720, –0.646]	–0.380*** (0.021) [–0.421, –0.338]
Ethnic group (Reference: White)						
Mixed	0.023 (0.089)	0.029 (0.089)	0.036 (0.098)	0.073 (0.068) [–0.061, 0.207]	0.126 (0.070) [–0.011, 0.262]	0.033 (0.066) [–0.097, 0.162]
Asian or Asian British	–0.279*** (0.053)	–0.256*** (0.055)	–0.226*** (0.056)	–0.158*** (0.032) [–0.220, –0.095]	–0.115*** (0.034) [–0.182, –0.048]	–0.107** (0.033) [–0.172, –0.042]
Black or Black British	–0.083 (0.074)	–0.061 (0.081)	–0.066 (0.070)	–0.046 (0.044) [–0.132, 0.041]	0.034 (0.048) [–0.059, 0.127]	–0.008 (0.046) [–0.098, 0.081]
Age group (Reference: 16 – 24)						
25 – 34	–0.043 (0.050)	–0.083 (0.053)	–0.040 (0.052)	0.021 (0.038) [–0.054, 0.096]	0.002 (0.040) [–0.077, 0.081]	–0.086* (0.039) [–0.162, –0.008]
35 – 49	–0.046 (0.046)	–0.102 (0.053)	–0.120* (0.052)	0.073* (0.035) [0.005, 0.142]	0.031 (0.041) [–0.049, 0.111]	–0.115** (0.040) [–0.194, –0.036]
50 – 64	–0.069 (0.045)	–0.112* (0.056)	–0.151** (0.056)	0.037 (0.036) [–0.033, 0.108]	0.021 (0.043) [–0.064, 0.106]	–0.136** (0.043) [–0.219, –0.052]
65 +	–0.067 (0.071)	–0.097 (0.080)	–0.258** (0.083)	0.042 (0.070) [–0.095, 0.180]	0.024 (0.076) [–0.125, 0.171]	–0.218** (0.074) [–0.362, –0.073]

Continued…

	MLR Estimator			Bayesian Estimator		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
	Estimates	Estimates	Estimates	Estimates	Estimates	Estimates
Relationship status (Reference: Single)						
Married or cohabiting		0.020 (0.037)	–0.043 (0.034)		0.026 (0.028) [–0.030, 0.081]	–0.006 (0.027) [–0.058, 0.047]
Divorced or separated		–0.017 (0.047)	0.003 (0.044)		–0.026 (0.038) [–0.100, 0.048]	0.007 (0.036) [–0.064, 0.077]
Widowed		0.046 (0.114)	0.088 (0.102)		0.050 (0.089) [–0.125, 0.224]	0.113 (0.084) [–0.052, 0.278]
Parental status (Ref: Lone parents with school children)						
Coupled parents with school age children		–0.054 (0.062)	0.061 (0.061)		0.082 (0.050) [–0.016, 0.181]	0.053 (0.048) [–0.040, 0.147]
Employees without school age children		–0.123* (0.051)	0.023 (0.051)		–0.024 (0.042) [–0.106, 0.058]	–0.024 (0.040) [–0.103, 0.054]
Illness or Disability (Reference: Yes)						
No		–0.109*** (0.031)	–0.056* (0.028)		–0.086*** (0.023) [–0.131, –0.040]	–0.070** (0.022) [–0.112, –0.027]
Region (Reference: London)						
Southern England		0.047 (0.054)	0.227*** (0.054)		0.154*** (0.037) [0.081, 0.226]	0.212*** (0.035) [0.143, 0.281]
East of England		0.056 (0.068)	0.181** (0.063)		0.130** (0.044) [0.044, 0.216]	0.154*** (0.042) [0.071, 0.236]
The Midlands		0.032 (0.056)	0.232*** (0.057)		0.129*** (0.038) [0.055, 0.202]	0.195*** (0.036) [0.124, 0.266]
Northern England		0.046 (0.055)	0.222*** (0.053)		0.141*** (0.036) [0.070, 0.212]	0.200*** (0.035) [0.132, 0.268]
Wales		–0.018 (0.072)	0.150* (0.068)		0.093 (0.050) [–0.005, 0.190]	0.165*** (0.048) [0.072, 0.258]
Scotland		0.124 (0.067)	0.204** (0.062)		0.231*** (0.045) [0.141, 0.320]	0.205*** (0.044) [0.119, 0.291]
Northern Ireland		–0.104 (0.083)	0.023 (0.082)		0.086 (0.053) [–0.018, 0.189]	0.084 (0.050) [–0.015, 0.183]

Continued…

	MLR Estimator			Bayesian Estimator		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
	Estimates	Estimates	Estimates	Estimates	Estimates	Estimates
Education (Reference: No qualifications)						
GCSE / O-level or lower			0.064 (0.039)			0.031 (0.033) [–0.033, 0.095]
Up to A-level			0.136** (0.052)			0.107** (0.040) [0.029, 0.184]
Up to Diploma in HE			0.014 (0.050)			0.034 (0.039) [–0.042, 0.110]
University or higher degree			0.084 (0.043)			0.085* (0.033) [0.020, 0.150]
No recorded data			0.013 (0.050)			–0.026 (0.038) [–0.101, 0.049]
Occupational classification (Ref: Managers & senior officials)						
Professional occupations			0.173*** (0.047)			0.173*** (0.036) [0.102, 0.244]
Associate professional & technical occupations			0.092* (0.043)			0.070* (0.033) [0.005, 0.136]
Administrative & secretarial occupations			0.156** (0.049)			0.146*** (0.037) [0.073, 0.219]
Skilled trades occupations			–0.213** (0.074)			–0.201*** (0.053) [–0.304, –0.097]
Personal service occupations			0.358*** (0.055)			0.309*** (0.045) [0.220, 0.397]
Sales & customer service occupations			0.331*** (0.057)			0.290*** (0.047) [0.197, 0.382]
Process, plant & machine operatives			–0.182** (0.069)			–0.171** (0.053) [–0.275, –0.068]
Elementary occupations			0.176** (0.058)			0.157*** (0.047) [0.065, 0.248]

Continued…

	MLR Estimator			Bayesian Estimator		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
	Estimates	Estimates	Estimates	Estimates	Estimates	Estimates
Full or Part-time (Reference: Part-time)						
Full-time			–0.378*** (0.032)			–0.374*** (0.025) [–0.422, –0.325]
Organisational sector (Reference: Private sector)						
Public sector			0.742*** (0.031)			0.742*** (0.022) [0.699, 0.785]
Organisation size (Reference: Micro)						
Small			0.313*** (0.036)			0.212*** (0.030) [0.153, 0.271]
Medium			0.377*** (0.040)			0.301*** (0.032) [0.238, 0.364]
Large			0.506*** (0.040)			0.422*** (0.030) [0.362, 0.482]
Intercept (constrained)	0.000	0.000	0.000	0.000	0.000	0.000
R^2	0.135*** (0.009)	0.144*** (0.009)	0.371*** (0.016)	0.120*** (0.006) [0.108, 0.133]	0.128*** (0.007) [0.115, 0.141]	0.351*** (0.010) [0.331, 0.372]
Adjusted R^2	0.135	0.143	0.369			
Unweighted sample size	16,678	16,582	16,068	16,678	16,582	16,068
Model estimation time (hours)	51.40	321.66	1,017.28	3.78	8.41	9.22

Notes: Data from UKHLS, Wave 8 (2016 – 2017). Standardised coefficients and estimates in parentheses are standard errors (MLR estimator) or posterior standard deviations (Bayesian estimator) and ones in square brackets indicate 95% credible intervals. Significance tests for the MLR estimator are based on two-tailed p-values, while those for the Bayesian estimator are based on one-tailed p-values. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$. Estimates based on the MLR estimator consider the complex sample design of the UKHLS data, while estimates based on the Bayesian estimator are unweighted.

Work-time Scheduling

The MIMIC models' parameter estimates based on the MLR and Bayes estimators for *work-time scheduling* yielded similar results. Point estimates with the MLR estimator were within the Bayesian 95% CI of the estimates based on the Bayes estimator, except for estimates for the Asian or Asian British ethnic group and organisation size (Table 7.5). For the variation in *work-time scheduling* explained by the predictors, the adjusted R^2 results suggested that socio-demographic characteristics did not explain much of the variation. Thus, demographic characteristics explained approximately 14% of the variation in *work-time scheduling*, while with the addition of socio-demographic characteristics the model still explained approximately 14% of the variation. However, the introduction of socio-economic characteristics resulted in the model explaining approximately 37% of the variation (Table 7.5).[50]

There was a statistically significant effect of sex on *work-time scheduling* in all three models, with a particularly marked decrease in the magnitude of the difference when socio-economic characteristics were introduced. On average, males had less awareness of and poorer access to other forms of *work-time scheduling* (–0.395 units, *SE* = 0.028) than female employees. In terms of ethnic groups, the effect on *work-time scheduling* was statistically significant in all three models when comparing between employees from White and Asian or Asian British ethnic backgrounds, with the latter having less awareness of and poorer access to other forms of *work-time scheduling* (–0.226 units, *SE* = 0.056) (Model 3). Otherwise, differences between employees from White and other ethnic backgrounds were not statistically significant in any of the models. From Model 1, the effect of age on *work-time scheduling* was not statistically significant, while in Model 2 only the difference between employees aged 16 – 24 and 50 – 64 years old was statistically significant. However, when socio-economic

[50] Refer to Appendix 7.8 for the posterior parameter distributions and trace plots for R^2 estimates (unadjusted) for *work-life balance* based on the Bayes estimator.

characteristics were introduced, only the difference between employees aged 16 – 24 and 25 – 34 years old was not statistically significant (Model 3). Thus, on average, employees aged 16 – 24 years old had more awareness of and better access to other forms of *work-time scheduling* than those aged 35 – 49 (–0.120 units, *SE* = 0.052), 50 – 64 (–0.151 units, *SE* = 0.056) or 65+ (–0.258 units, *SE* = 0.083) years old.

Considering relationship status, the effect on *work-time scheduling* was not statistically significant in both models. On the other hand, for parental status, only the difference in *work-time scheduling* between lone parents with and employees without primary school age children was statistically significant in Model 2. However, when socio-economic characteristics were introduced (Model 3), there was no statistically significant effect of parental status. The effect of longstanding illness or disability on *work-time scheduling* was statistically significant in both models. Employees without a longstanding illness or disability were slightly less awareness of and had poorer access to other forms of *work-time scheduling* (–0.056 units, *SE* = 0.030) than those with a longstanding illness or disability (Model 3). Regarding region, from Model 2, differences in *work-time scheduling* between employees in London and other regions were not statistically significant. However, when socio-economic characteristics were introduced, differences between employees in London and those other regions were statistically significant, except for employees in Northern Ireland (Model 3). Employees in London had less awareness of and poorer access to other forms of *work-time scheduling* than those in any other region.

Moving on to education, only the difference in *work-time scheduling* between employees with no qualifications and those with up to A-level qualifications was statistically significant. On average, employees with no qualifications had less awareness of and poorer access to other forms of *work-time scheduling* than those with up to A-level qualifications (0.136 units, *SE* = 0.052). Differences in *work-time scheduling* between managers and senior officials compared to employees in any other occupational group were statistically significant. Employees in skilled trades occupations (–0.213 units, *SE* = 0.074), or process, plant and machine operatives (–0.182 units, *SE* = 0.069) had, on average, less awareness of and poorer access to other forms of *work-time scheduling* than managers and senior officials, while this was better for employees in any other occupational group compared to managers and senior officials. The differences in *work-time scheduling* between full or part-time employment, or public or private organisational sector and *work-time scheduling* were statistically significant. On average, employees in full-time had less awareness and poorer access to other forms of *work-time scheduling* (–0.378 units, *SE* = 0.032) than those in part-time employment, while employees who worked in the public sector were more aware of and had better access to other forms of *work-time scheduling* (0.742 units, *SE* = 0.031) than those in the private sector. Lastly, considering organisation size, on average, employees in micro size organisations were less aware of and had poorer access to other forms of *work-time scheduling* than those in small (0.313 units, *SE* = 0.036), medium (0.377 units, *SE* = 0.040) or large (0.506 units, *SE* = 0.040) size companies and the differences were statistically significant.

7.3 Discussion

This chapter presented MIMIC models investigating the effects of demographic, socio-demographic, and socio-economic characteristics on *overall QWE*, *economic compensation*, *working conditions*, and *work-time scheduling* in the UK employee population. MLR and Bayes estimators were used to estimate the models, with non-informative priors used for the Bayes estimator so that parameter estimates were analogous to those of the MLR estimator. This was partly due to estimation with the MLR estimator being computationally cumbersome because of categorical observed items and the number of latent traits in the measurement model. Overall, the two estimators yielded similar results, but parameter estimates for some predictors based on the MLR estimator were outside the Bayesian 95% CI of those based on the Bayes estimator, including some age group, relationship status, and parental status categories for *overall QWE*, or organisation size for *work-time scheduling*. Importantly though, the study laid the foundation for estimating more complex models with the Bayes estimator that would otherwise not be feasible with frequentist methods, such as extending this analysis to a longitudinal analysis. Model comparison with the MLR estimator suggested that the model with demographic, socio-demographic, and socio-economic characteristics exhibited better fit to the data. On the other hand, socio-economic characteristics explained more of the variation in *overall QWE* and other *dimensions of QWE* than the demographic or socio-demographic characteristics.

In terms of findings, firstly, focusing on demographic characteristics and considering sex, results from the study suggested that females had poorer *overall QWE* and *economic compensation* than males, while males had poorer *working conditions* and less awareness of and had poorer access to other forms of *work-time scheduling* than females in the UK employee population. This supported evidence from previous literature which highlighted the disadvantages experienced by females in the labour market compared to males. Thus, according to Fredman (2004) and Piasna and Plagnol (2018), employment for females tends to be marked by poor employment security with greater impediment in accessing training, and little or non-linear career progression pathways compared to males, partly attributed to career breaks as a consequence of childrearing (Lindley 2015; Piasna and Plagnol 2018). This might explain the poorer *overall QWE* and the implications on their level of *economic compensation*. While other studies found no differences between females and males in the UK employee population in terms of job control (Gallie and Zhou 2013; Lindley 2015; Wu et al. 2021), this study found that males had poorer *working conditions* than females. The discrepancies may be partly attributed to different items of job control used in different studies (Adler 1993), but also methods of aggregation. For example, in their study Wu et al. (2021) used task order, work manner, and work pace, aggregated these by estimating their arithmetic mean which assumes equal weighting of the items and found no significant differences by sex in the UK employee population. This study, however, included job tasks and work hours in addition to the items used by Wu et al. (2021), while the conditional slopes estimated by the bifactor IRT model for each of these items on *working conditions* were not equal. This suggested that the items did not contribute an equal weight on *working conditions*. Findings from this study supported literature that indicated females were more likely to have better work-life balance (*work-time scheduling*) than males in the UK employee population. However, according to Piasna and Plagnol (2018) and Tomlinson (2007), this was attributed to job design with employers seeking low-cost and flexible labour rather than female employees' preferences or the need to accommodate family responsibilities.

In terms of ethnic background, findings from the study indicated disparities in labour market experiences by ethnic group. Compared to employees from a White ethnic background, those from Asian or Asian British, or Black or Black British ethnic backgrounds had poorer *overall QWE* and *economic compensation*, while those from an Asian or Asian British ethnic background were also less awareness and had poorer access to other forms *work-time scheduling* in the UK employee population. These findings were consistent with evidence from the study by Zwysen and Demireva (2020) who found that employees from ethnic minority backgrounds were less likely to be in jobs with high levels of economic compensation, work-life balance (*work-time scheduling*), job security, and intrinsic satisfaction than employees from a White ethnic background. This can be attributed, in part, to historical roots which defined race and ethnicity as marks of inferiority (Dillon 2020; Korpi 2018). This study also found that there were no differences in *working conditions* by ethnic group, nor between employees from White and Mixed ethnic backgrounds in terms of *overall QWE*, *economic compensation*, and *work-time scheduling*. This might be due to the relatively small proportion of employees from a Mixed ethnic background, but also studies have shown that there are variations within ethnic minority groups (Clark et al. 2022; Zwysen and Demireva 2020).

Regarding age, results from this study supported empirical evidence from Arranz et al (2019), which suggested that younger employees were more likely to be in more precarious employment and fared worse off in the labour market than older employees. In the UK employee population, employees aged between 16 – 24 years old had poorer *overall QWE* than those aged between 35 – 49 or 65 + years old, poorer *economic compensation* than those aged 25 – 34, 35 – 49 or 50 – 64 years old, and poorer *working conditions* than those aged 35 – 49, 50 – 64 or 65 + years old. While literature suggested that younger employees were more likely to participate in work-related training than older employees (Canduela et al. 2012; Dieckhoff et al. 2007), they tended to have employment contracts that offer limited employment security and economic compensation (Kim and Kurz 2001). This might explain the poorer levels of *overall QWE* and *economic compensation* among younger employees. Furthermore, Esser and Olsen (2012) found that younger employees were more likely to have less autonomy at work than older employees, partly due to limited experience when they enter the labour market, resulting in poorer *working conditions*. In terms of *work-time scheduling*, empirical evidence suggested that younger employees were more likely to have poor work-life balance than older employees partly due to working long hours to establish their careers (Sturges and Guest 2004) and perhaps fewer family commitments or responsibilities. However, this study found that employees aged between 16 – 24 years old were more aware of and had better access to other forms of *work-time scheduling* than those aged 35 – 49, 50 – 64 or 65 + years old. This could be attributed to different populations between the studies as research by Sturges and Guest (2004) focused on a population of UK graduate employees in large organisations, while this study considered the UK employee population. Furthermore, Sturges and Guest (2004) also reported unaggregated results of their indicators, such as working hours and conflict between work and non-work time. However, other studies have argued that, subjectively, there are generational differences in the centrality of work in employees' lives. Smola and Sutton (2002) and Sturges and Guest (2004) suggested that younger employees placed more importance in a 'working to live, not living to work' approach to work-life balance than older employees. This might explain results of *work-time scheduling* found in this study among younger employees.

Secondly, for socio-demographic characteristics, namely the relationship status of employees, findings from this study partly supported other literature suggesting better outcomes in the labour market for married/cohabiting employees (Bardasi and Taylor 2008; Ribar 2004). However, much of the empirical research has focused on the marriage premium when considering males, with married males more likely to have better outcomes (Bardasi and Taylor 2008; Ribar 2004; Schoeni 1995). On the other hand, evidence pertaining to females was more ambiguous with research often framed in terms of marriage penalties (Ribar 2004). Married/cohabiting employees had better levels of *overall QWE* than single employees, while there were no differences between single and divorced/separated or widowed employees. Married/cohabiting or divorced/separated employees also had better *economic compensation* compared to single employees, whereas there was no difference between single and widowed employees. However, in terms of *working conditions* and *work-time scheduling*, there were no differences by relationship status, indicating that relationship status has no influence on these aspects of the labour market. Better *overall QWE* and *economic compensation* among married/cohabiting employees may be attributed to spousal support and its stabilising influence, which can lead to accumulation of human capital and result in better outcomes in the labour market (Bardasi and Taylor 2008; Ribar 2004).

For parental status, findings from this study were not consistent with results from other literature which suggested that lone parents were particularly disadvantaged and more likely to be in precarious employment than coupled parents (Esser and Olsen 2018; Nieuwenhuis and

Maldonado 2018). Employees without primary school-age children had poorer *overall QWE* compared to lone parents with primary school-age children, while there was no difference between lone parents with and coupled parents with primary school-age children. In terms of *economic compensation*, *working conditions*, and *work-time scheduling*, there were no differences by parental status. Employees without primary school-age children may be less constrained in terms of the demands between their work and family responsibilities, and therefore less particular about the quality of their work and employment. This might explain their poorer *overall QWE*. However, the inconsistency in findings from this study and other literature might be due to the measure of parental status used in this study. This was limited to primary school-age children; thus, children aged between 5 – 11 years old but excluded very young or older children who might still be dependent on their parents.

Considering longstanding illness or disability, evidence from this study supported some previous literature that highlighted the challenges employees with a longstanding illness or disability experience in the labour market. Davidson and Kemp (2008) found that disabled employees were more likely to be in non-standard employment than non-disabled employees. Although it affords better work-life balance (Lyonette 2015), non-standard employment is characterised by job insecurity (Meager and Hill 2005), low pay, with employees often ineligible for sick pay or occupational pensions, limited pathway to promotion or career progression, as well as lower levels of job autonomy (McGovern et al. 2004). This study found that there was no difference in *overall QWE* between employees with or without a longstanding illness or disability, contrary to evidence from other literature. On the other hand, while *economic compensation* and *working conditions* were better for employees without than for those with a longstanding illness or disability, the latter were more awareness of and had better access to other forms of *work-time scheduling*. This result was consistent with evidence from other literature. Similar *overall QWE* by longstanding illness or disability could be partly attributed to government initiatives that support employees with a disability to participate in the labour market (Grover and Piggott 2015; Lewis et al. 2013). Indeed, evidence has indicated that people with disabilities are increasingly joining the workforce (Department for BEIS 2018), which may result in QWE for disabled employees being a salient social issue. This might also explain the results on *work-time scheduling* for employees with a longstanding illness or disability, although this may also be due to the higher likelihood of disabled employees being in non-standard employment (Grover and Piggott 2015; Lyonette 2015). Poorer *economic compensation* and *working conditions* among employees with a longstanding illness or disability might, perhaps, be attributed to socially embedded barriers, such as discriminatory attitudes employees with a longstanding illness or disability experience in the labour market or skills differentials (Grover and Piggott 2015).

In terms of differences across UK regions and nations, results from this study supported evidence from other studies highlighting longstanding disparities in the labour market and advantages for those residing in London and the Southern regions (Department for LUHC 2022; Jones and Green 2009; Low Pay Commission 2021). This was partly attributed to a shift from heavy industry to a knowledge economy highly centralised in these regions (Hepworth et al. 2005; Jones and Green 2009). However, the disparities did not necessarily apply to every aspect of QWE. There was no difference in *overall QWE* between employees in London and Southern England, while employees in other regions had poorer *overall QWE* than those in London. Considering *economic compensation*, while there was no difference between employees in London and Scotland, employees in London had better *economic compensation* than those in other regions. In terms of *working conditions*, employees in London had poorer levels than those in Scotland, but there were no differences between employees in London and those in other regions. For *work-time scheduling*, there was no difference between employees in London and Northern Ireland, while employees in London were less awareness of and had poorer access to other forms of *work-time scheduling* than those in other regions. Better *overall QWE* for employees in London and Southern England, and better *economic compensation* for those in London can be partly attributed to the knowledge economy which is highly centralised in these regions and require a highly skilled workforce (Hepworth et al. 2005; Jones and Green 2009). In the case of better *economic compensation* for employees in Scotland, evidence from literature suggested that there is a knowledge economy in Scotland that is dominated by three cities; thus Edinburgh, Aberdeen, and Glasgow; with average earnings in Edinburgh and Aberdeen particularly highly competitive relative to national standards (Hepworth et al. 2005). The highly skilled workforce in the knowledge economy might also explain the better *working conditions* for employees in Scotland compared to those in other regions. This, however, does not explain the poorer *working conditions* for employees in London or Southern England. This might be due to the extreme case, particularly in London, where the knowledge economy is at its most competitive, whilst also the least inclusive especially for low skilled employees (Hepworth et al. 2005; TUC 2021b), resulting in poorer *working conditions*. On the other hand, the competitive nature, along with the work demands associated with the knowledge economy in London (Hepworth et al. 2005; TUC 2021b), and the low proportion of high-quality jobs in Northern Ireland (Jones and Green 2009) might explain the results on *work-time scheduling* in these regions.

Lastly, considering socio-economic characteristics and focusing on education, results from this study were consistent with findings in other literature that highlighted education as an important investment in human capital (Okay-Somerville and Scholarios 2013; Solomon et al. 2022). However, there were variations for different aspects of the labour market in the UK employee population. Employees with no qualifications had poorer *overall QWE* than those with a university or higher degree, while there were no differences between employees with no qualifications and those with any other educational qualification. For *economic compensation*, employees with no qualifications had poorer levels than those with other educational qualification, except when compared to those with GCSE / O-level or lower qualifications, where there were no differences. On the other hand, employees with no qualifications were less awareness of and had poorer access to other forms of *work-time scheduling* than those with up to A-level qualifications, while there were no differences compared to those with any other educational qualification. However, for *working conditions*, there were no differences by educational qualifications in the UK employee population. Higher educational attainment is associated with strongly developed high-skilled workers (Gallie 2007b; Soskice 1999), who command attractive remuneration packages for their skills and this might partly explain the better *economic compensation* for employees with higher educational qualifications. Furthermore, better *overall QWE* for employees with a university or higher degree might reflect the added advantage of higher education in the labour market. On the other hand, the lack of differences in *work-time scheduling* between those with no qualifications and those with particularly high levels of educational attainment might be related to the job demands for employees with high levels of education. In terms of *working conditions*, the lack of expected differences by education could be partly attributed to graduates being employed in non-graduate occupations due to over-supply and underemployment of university graduates in the labour market (Green and Zhu 2010; Okay-Somerville and Scholarios 2013; Warhurst 2008). This may particularly be the case in a liberal market economy, like the UK, where education and training systems lack industry-specific skills post-compulsory secondary education and place more emphasis on general education (Hall and Soskice 2001; Soskice 1999, 2005).

In terms of occupational classification, findings from this study supported results from Gallie (2015) and Wheatley (2022), which suggested an occupational hierarchy in the variation of some aspects of QWE and this was partly attributable to skills differentials. In the UK employee population, managers and senior officials had better *overall QWE* and *working* *conditions* than employees in other occupational groups. While there were no differences in *economic compensation* between managers and senior officials and employees in associate professional and technical occupations, *economic compensation* for managers and senior officials was poorer compared to employees in professional occupations, but better compared to employees in other occupational groups. On the other hand, employees in skilled trades or those who worked as process, plant, and machine operatives were less awareness of and had poorer access to other forms of *work-time scheduling* compared to managers and senior officials, while this was better for employees in other occupational groups than managers and senior officials. Much of the balance of power or decision-making within organisations in the UK labour market resides with managers and senior officials (Gallie 2007b; Holman 2013; Hall and Soskice 2001; Soskice 1999) and this might partly explain their better *overall QWE* and *working conditions* compared to employees in other occupational groups. On the other hand, the nature of the work done by employees in skilled trades (e.g. agricultural, electrical, construction, or food preparation trades) or those who work as process, plant, and machine operatives (e.g. textile process, energy plant, assemblers, or transport operatives) means they have less flexibility in their working arrangements and partly explains their results on *work-time scheduling* compared to managers and senior officials. However, the results on *work-time scheduling* for managers and senior officials compared to employees in other occupational groups might be attributed to job demands associated with their roles and responsibilities (e.g. senior officials in local or national government; production, works and maintenance managers; or hospital and health service managers), such as the need to be contactable outside their standard work times. Variations in *economic compensation* can be partly explained by the skill differentials between occupational groups (Gallie 2007b; Soskice 1999), with highly skilled employees having better *economic compensation*. Thus, while there were no differences between managers and senior officials and employees in associate professional and technical occupations, and better levels for employees in professional occupations than managers and senior officials, employees in these occupational groups are highly skilled compared to those in other occupational groups.

Regarding full or part-time employment, results from this study supported findings from other studies which suggested better outcomes in the UK labour market for full-time than part-time employees, except for work-life balance (*work-time scheduling*) (Hoque and Kirkpatrick 2003; Lyonette et al. 2010; McGovern et al. 2004; Warren and Lyonette 2015). Thus, in the UK employee population, employees in part-time employment had poorer *overall QWE*, *economic compensation*, and *working conditions*, but were more awareness of and had better access to other forms of *work-time scheduling* than those in full-time employment. Evidence has shown that employees in non-standard forms of employment, such as part-time employment, tend to be marginalised in terms of training and development, and consultation at work (Hoque and Kirkpatrick 2003; Lyonette et al. 2010; Warren and Lyonette 2015). Other studies also argued that, by design, part-time jobs required fewer skills and lower levels of training than full-time jobs and this is associated with skills differentials and marginal productivity (Gallie 2007b; Hall and Soskice 2001; Holman 2013b). Consequently, compared to part-time jobs, full-time jobs are more likely to have better economic compensation, better prospects for promotion, greater job security (Warren and Lyonette 2015), as well as greater autonomy (McGovern et al. 2004), but poorer work-life balance (Lyonette 2015).

For organisational sector, results from this study were partly consistent with findings in other studies. Cribb et al (2014), Murphy et al (2020) and Rubery (2013) supported evidence of a public sector pay premium and a more skilled workforce in the public than private sector. Rubery (2013) also highlighted better outcomes in the public than private sector in terms of provisions for work-life balance. This study found poorer *overall QWE* among employees in public sector than private sector organisations, while employees in public sector organisations had better *economic compensation*, *working conditions*, and more awareness of and better access to other forms of *work-time scheduling* than for those in private sector organisations. While, as a liberal market economy, there is minimal state or government involvement in the regulation of the UK labour market (Gallie 2007b; Hall and Soskice 2001; Holman 2013b; Soskice 1999), the government as a public sector employer, was more likely to adhere to the regulations it has set, such as paying the NMW or NLW, providing pension schemes, including availability of flexible working arrangements than private sector organisations. Furthermore, evidence suggested that pay was more uniformly distributed within the public sector and less so in the private sector, with pay at the top of the distribution higher in the private than public sector (Cribb et al. 2014; Lucifora and Meurs 2006). This, as well as evidence suggesting a more skilled workforce in the public than private sector (Cribb et al. 2014; Murphy et al. 2020; Rubery 2013), might explain the better *economic compensation*, *working conditions*, and *work-time scheduling* among public sector employees compared to private sector employees. However, the poorer *overall QWE* among public sector employees than private sector employees might be due to the greater variety of jobs available in the private sector than public sector, as a share of the UK labour market. This may offer employees in the private sector more opportunities to work in areas of their interest compared to the public sector and result in better *overall QWE*. Furthermore, an underutilisation of skills among the workforce in the public than private sector as evidence suggests that, on average, the public sector workforce is more highly skilled than the private sector workforce (Cribb et al. 2014) might partly explain poorer *overall QWE* for public sector employees.

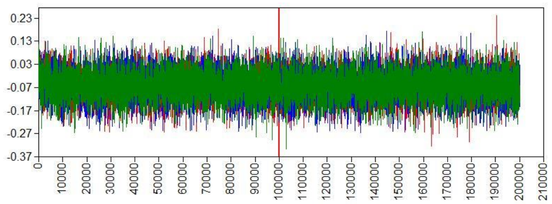
Findings from this study supported, in part, evidence from other literature about the effect of organisational size on different aspects of QWE, while also highlighting some of the ambiguity in the literature. Bryson et al. (2021), Forth et al. (2006) and Storey et al. (2010) found that employees in larger firms had poorer QWE than those in smaller firms. This was partly due to formally centralised systems in large firms designed to increase efficiency and productivity but can be detrimental to skills application and development for employees. However, larger firms also have resources, including human resource management systems, to consciously design jobs of high quality to attract employees with appropriate skills compared to smaller firms (Bryson et al. 2021). This argument was supported by the results from this study, which indicated better *overall QWE* among employees in micro size organisations than those in small or medium size organisations, while there was no difference between employees in micro or large size organisations. On the other hand, employees in micro size organisations had poorer *economic compensation*, and were less aware of and had poorer access to other forms of *work-time scheduling*, but better *working conditions* compared to those in small, medium, or large size organisations. This was consistent with findings from Forth et al. (2006), who found better levels of pay in medium or large size organisations compared to small size organisations due to their greater resources and flexibility in setting remuneration packages to attract employees with appropriate skills. Smaller organisations were also less likely to have formal practices that supported work-life balance than larger organisations (Forth et al. 2006), which might explain level of awareness and access to other forms of *work-time scheduling*. However, informal systems in smaller organisations meant their employees were more likely to report having access to various flexible working arrangements if needed, as well as greater autonomy (Forth et al. 2006). This may partly explain the better *working conditions* for employees in micro size organisations compared to those in small, medium, or large size organisations found in this study.

Lastly, while the measure of QWE did not capture any aspect of the *social dialogue* dimension, the bivariate analyses in Chapter 4 indicated statistically significant associations between predictors of QWE and collective bargaining. For example, males, employees from Asian or Asian British ethnic backgrounds, or employees aged 16 – 24 years old were more likely to report not having recognised trade unions or staff associations at their workplace. On the other hand, single employees, lone parents with primary school age children, those with no longstanding illness or disability, or employees in the south of England were more likely to report not having recognised trade unions or staff associations at their workplace. Furthermore, employees in part-time employment, private sector organisations, micro size organisations, with no educational qualifications, or in routine and manual occupations were also more likely to report not having recognised trade unions or staff associations at their workplace.

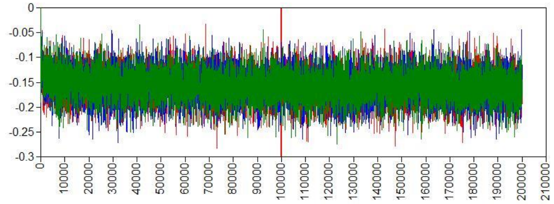
The next chapter will provide a summary of the research and highlights contributions the study has made to the topic of measuring QWE. This will consider the theoretical, methodological and substantive contributions. It will also consider the limitations of the research and how this could be developed further.

7.4 Appendices

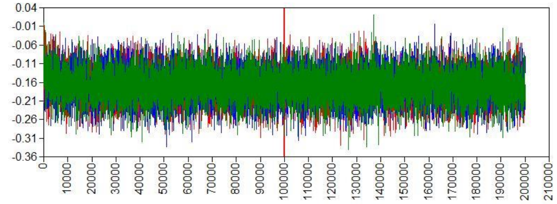
7.4.1 Appendix 7.1: Posterior Parameter Trace Plots for MIMIC Model 3 for *Overall QWE* based on the Bayesian Estimator



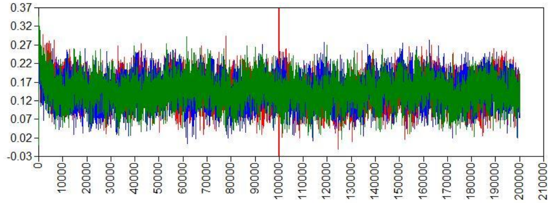
Ethnic group: Mixed



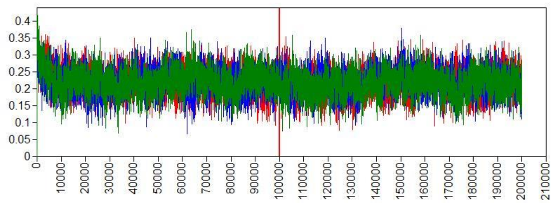
Ethnic group: Asian or Asian British



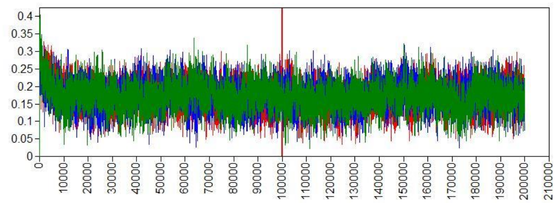
Ethnic group: Black or Black British



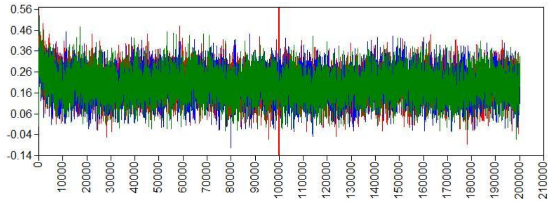
Age group: 25 - 34 years old



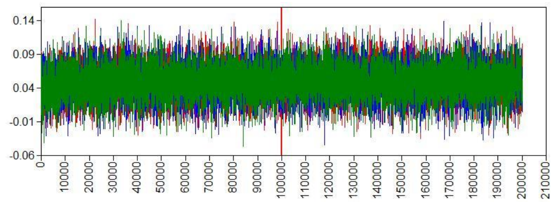
Age group: 35 - 49 years old



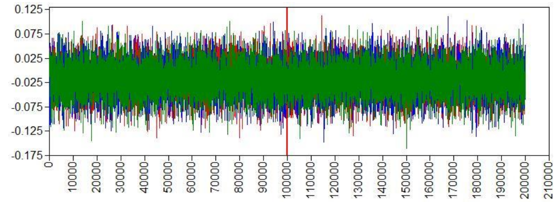
Age group: 50 - 64 years old



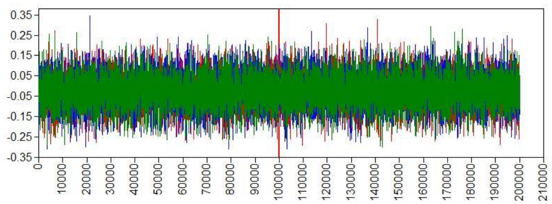
Age group: 65 + years old



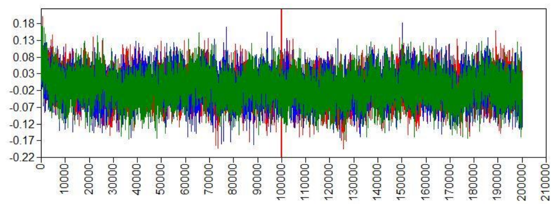
Relationship status: Married or cohabiting



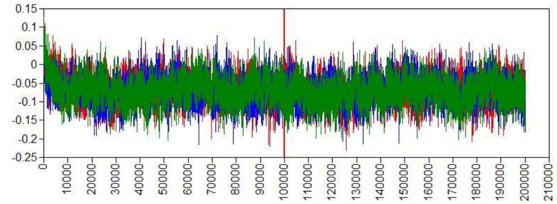
Relationship status: Divorced or separated



Relationship status: Widowed



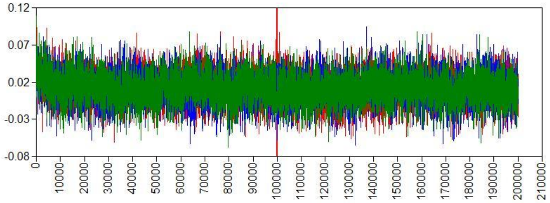
Parental status: Coupled parents with school age children



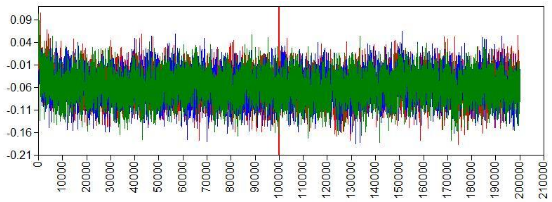
Parental status: Employees without school age children

Continued…

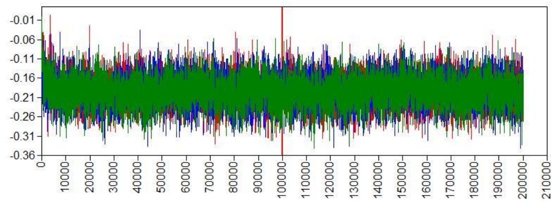
Continued…



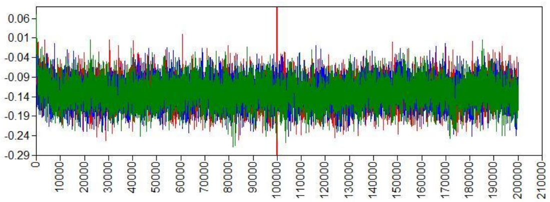
Illness or disability: No



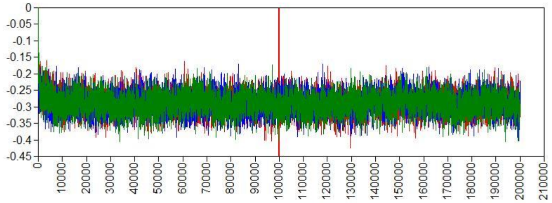
Region: Southern England



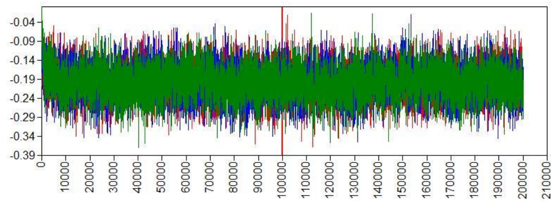
Region: East of England



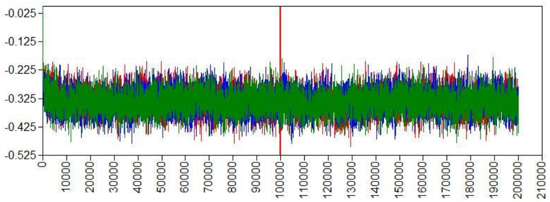
Region: The Midlands



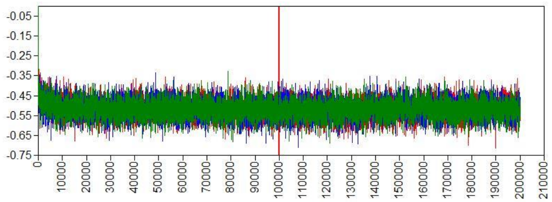
Region: Northern England



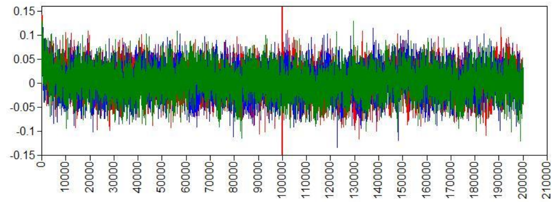
Region: Wales



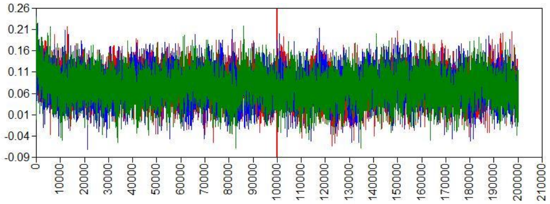
Region: Scotland



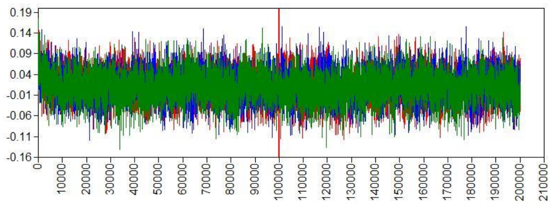
Region: Northern Ireland



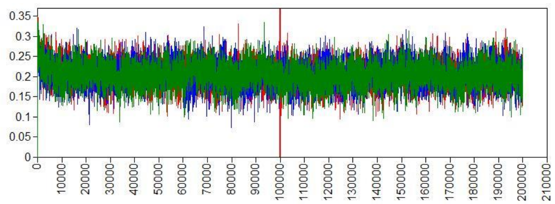
Education: GCSE / O-level or lower



Education: Up to A-level



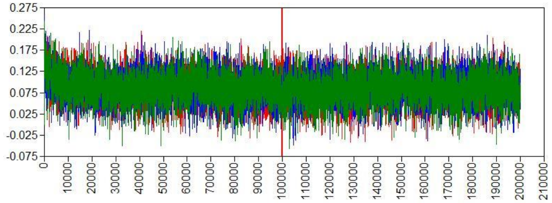
Education: Up to diploma in higher education



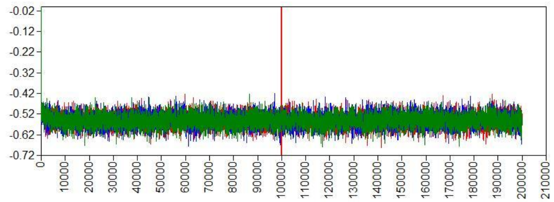
Education: University or higher degree

Continued…

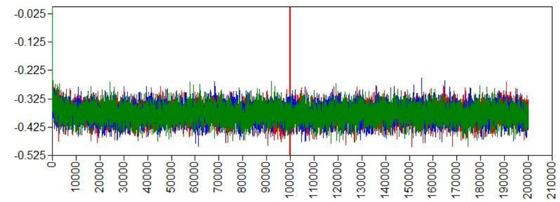
Continued…



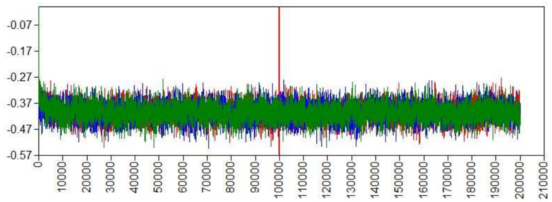
Education: No recorded data



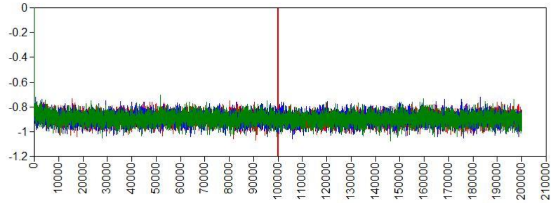
Occupational classification: Professional occupations



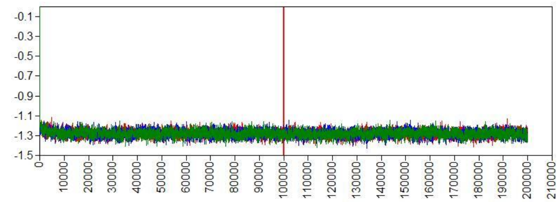
Occupational classification: Associate prof. & technical occ.



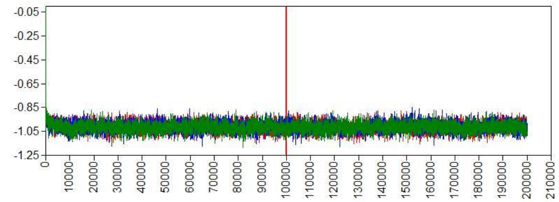
Occupational classification: Admin. & secretarial occ.



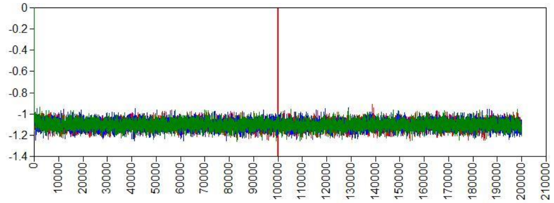
Occupational classification: Skilled trades occupations



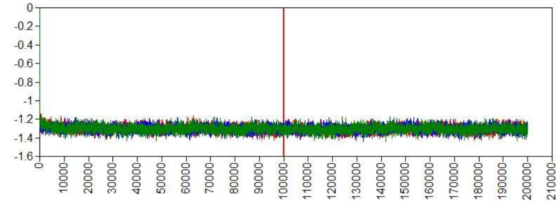
Occupational classification: Personal service occupations



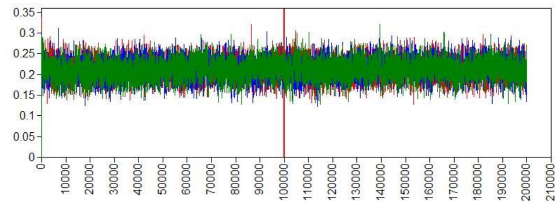
Occupational classification: Sales & customer service occ.



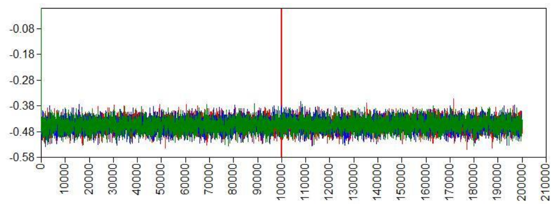
Occupational classification: Process, plant & machine op.



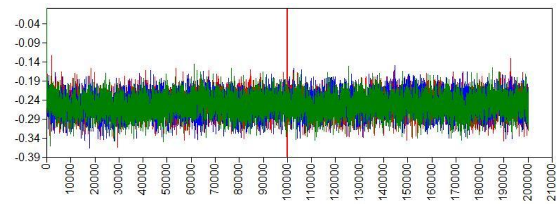
Occupational classification: Elementary occupations



Full or part time employment: Full-time



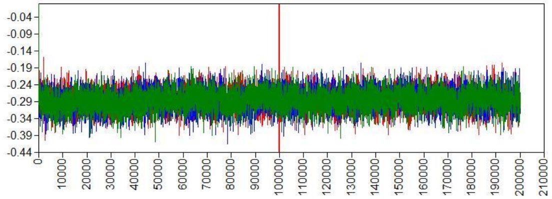
Organisational sector: Public sector



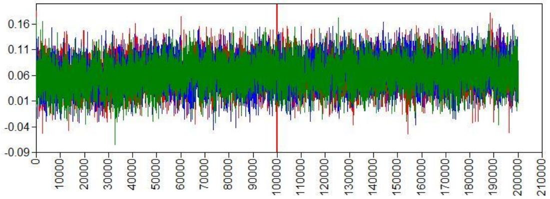
Organisation size: Small

Continued…

Continued…



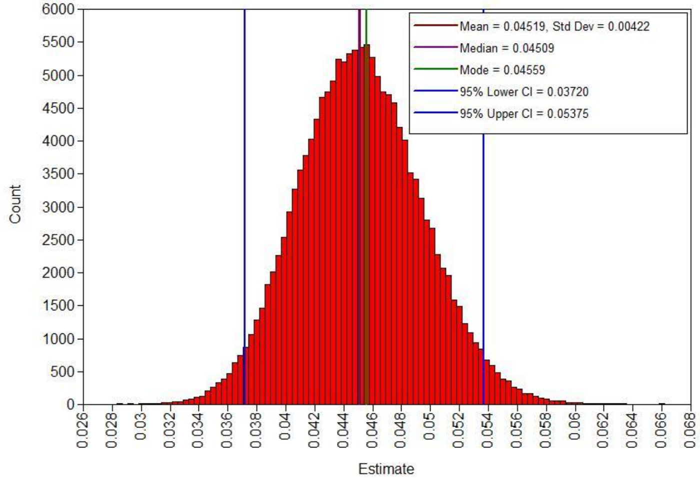
Organisation size: Medium



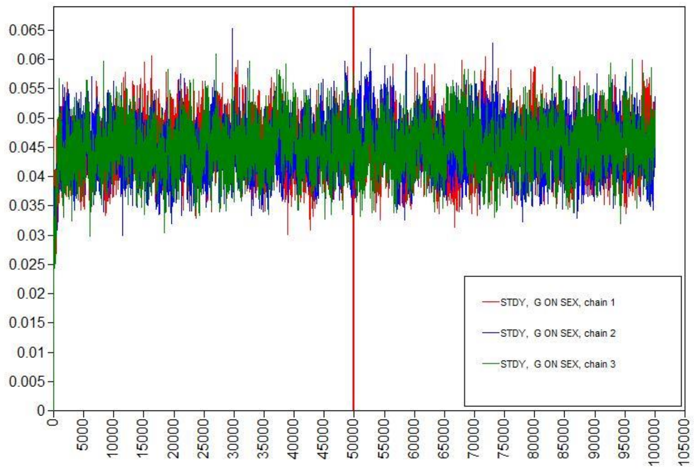
Organisation size: Large

Notes: The estimates are standardised parameter estimates. The three chains in the trace plots for all predictors mixed well, indicating that the MCMC algorithm reached equilibria in estimating the posterior parameter distributions. See notes from Figure 7.x for additional information.

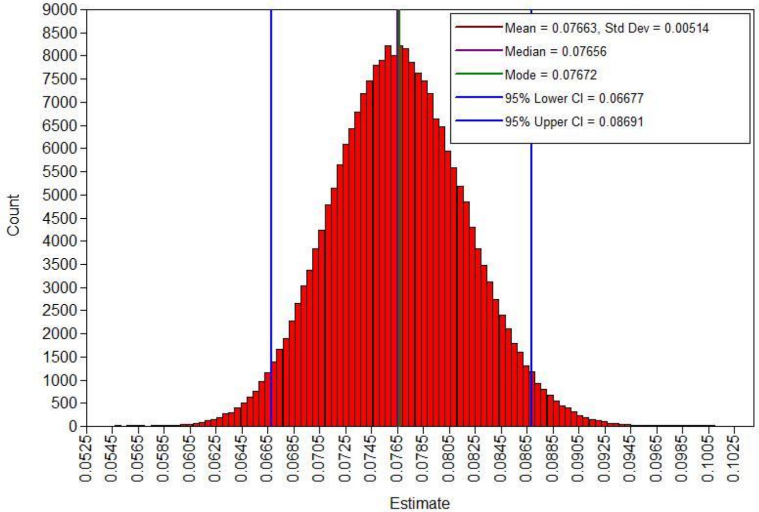
7.4.2 Appendix 7.2: Posterior Parameter Distributions and Trace Plots for R^2 Estimates for *Overall QWE* based on the Bayesian Estimator



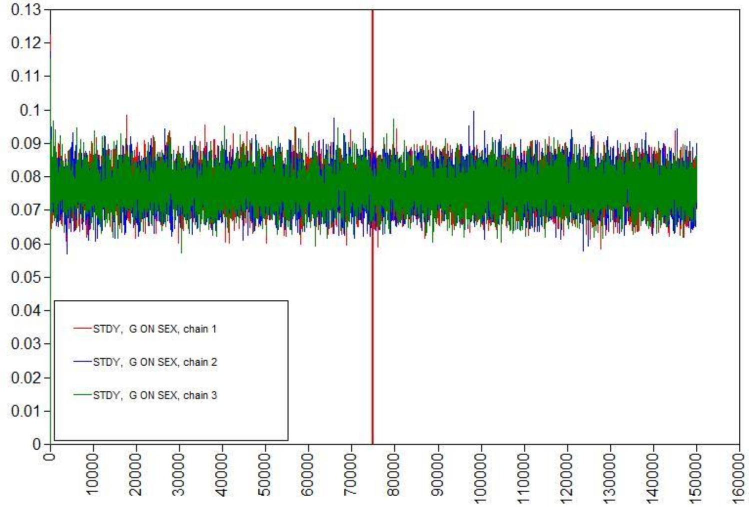
a) Model 1 Posterior Parameter Distribution



a) Model 1 Posterior Parameter Trace Plot

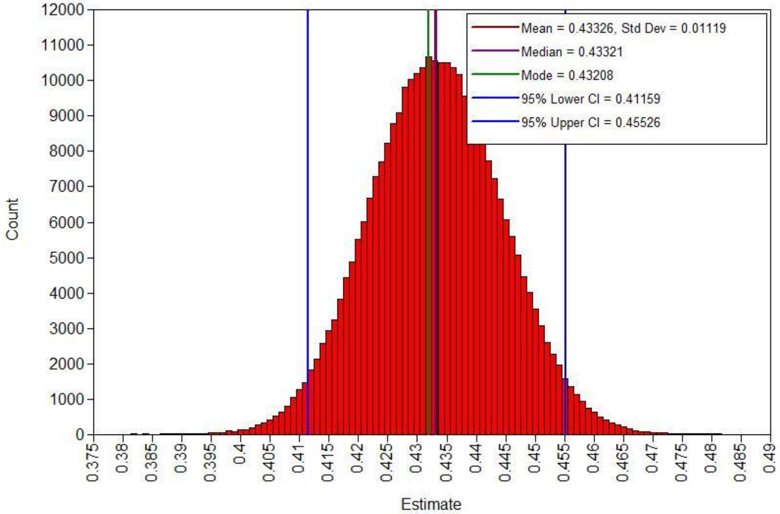


b) Model 2 Posterior Parameter Distribution

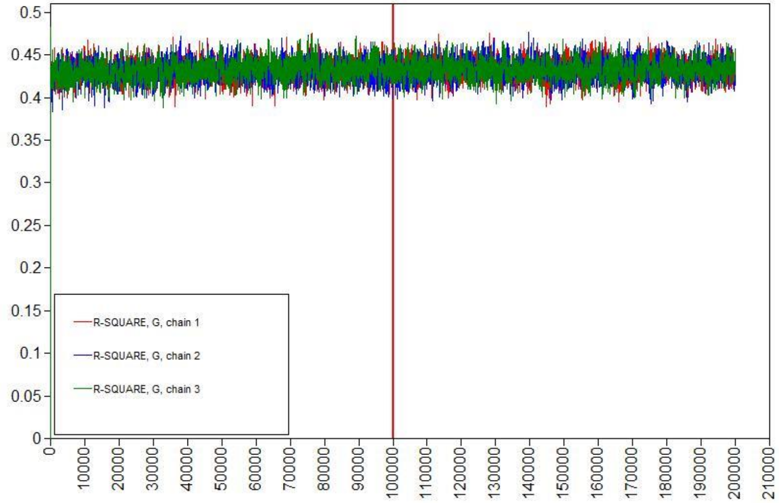


b) Model 2 Posterior Parameter Trace Plot

Continued…



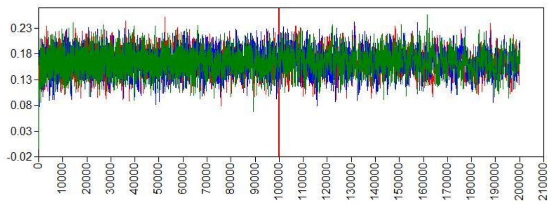
c) Model 3 Posterior Parameter Distribution



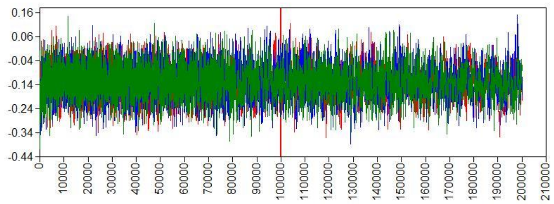
c) Model 3 Posterior Parameter Trace Plot

Notes: The estimates are standardised parameter estimates. The histograms show normal posterior parameter distributions estimated by the models and the posterior means, posterior standard deviations, and 95% credible intervals printed within the charts correspond to the unadjusted R^2 estimates for the MIMIC Models for *overall QWE* (Table 7.x). The trace plots show relatively well mixed chains, indicating that the MCMC algorithm reached equilibria in estimating the posterior parameter distributions for these estimates in all three models.

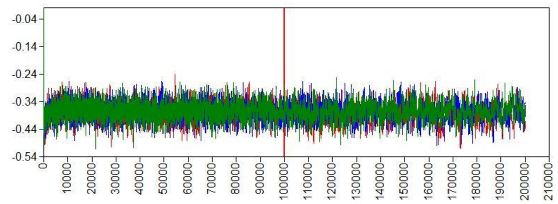
7.4.3 Appendix 7.3: Posterior Parameter Trace Plots for MIMIC Model 3 for *Economic Compensation* based on the Bayesian Estimator



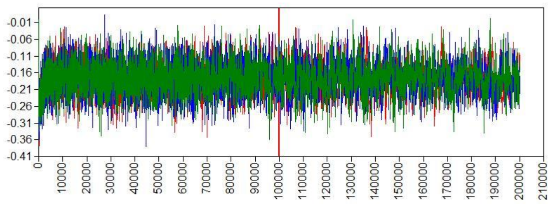
Sex: Male



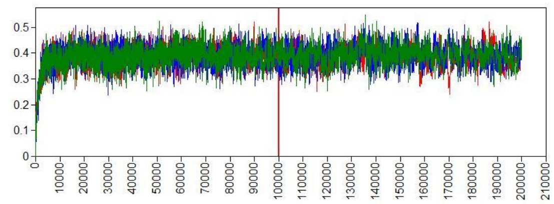
Ethnic group: Mixed



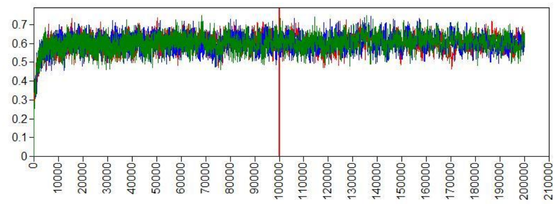
Ethnic group: Asian or Asian British



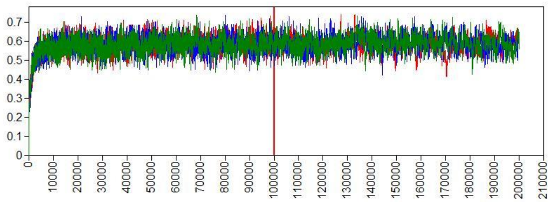
Ethnic group: Black or Black British



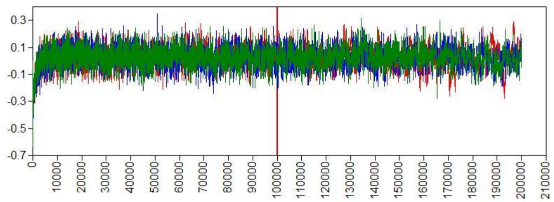
Age group: 25 - 34 years old



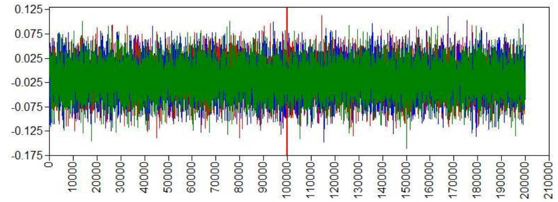
Age group: 35 - 49 years old



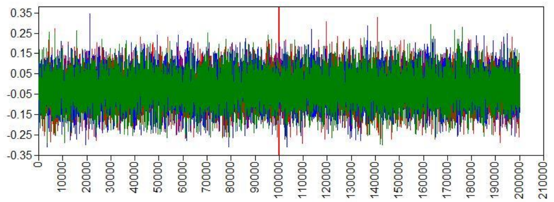
Age group: 50 - 64 years old



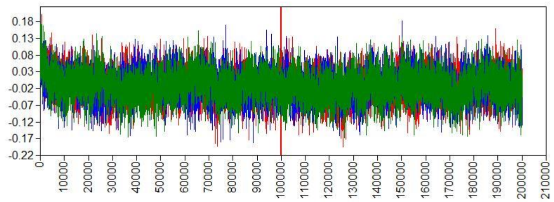
Age group: 65 + years old



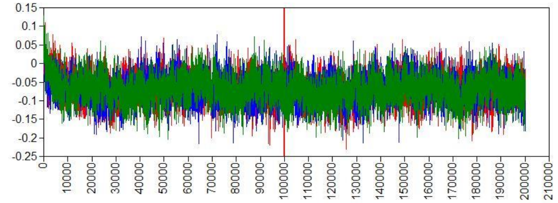
Relationship status: Married or cohabiting



Relationship status: Divorced or separated



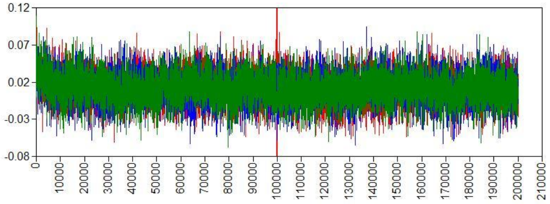
Relationship status: Widowed



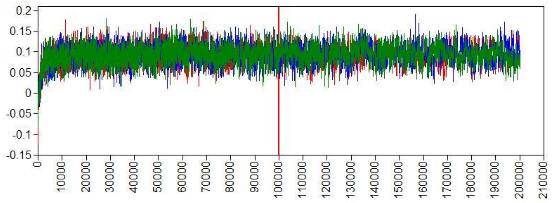
Parental status: Coupled parents with school age children

Continued…

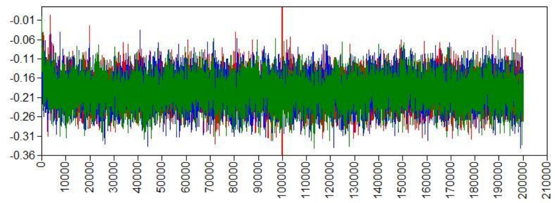
Continued…



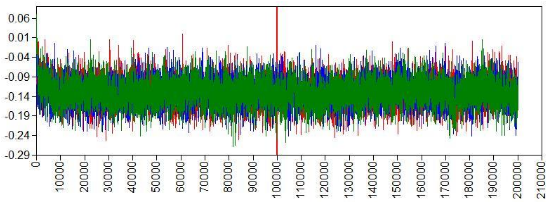
Parental status: Employees without school age children



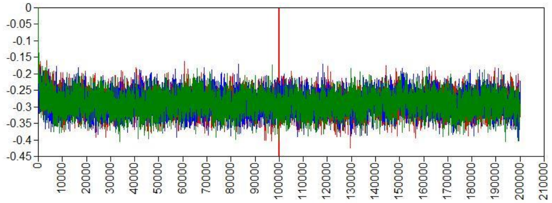
Illness or disability: No



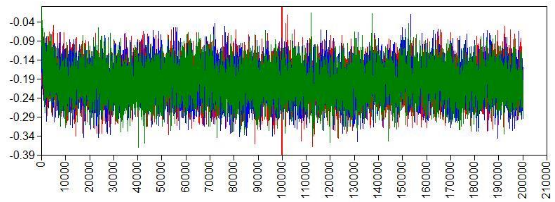
Region: Southern England



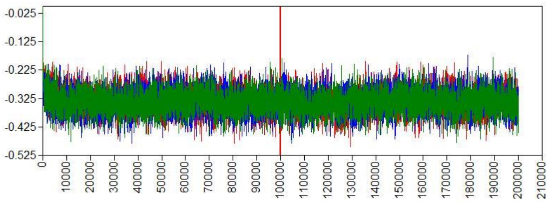
Region: East of England



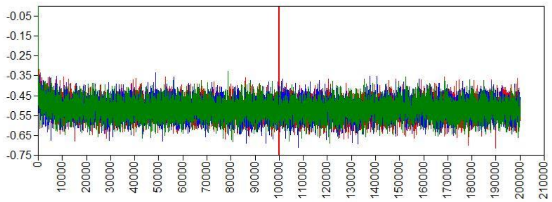
Region: The Midlands



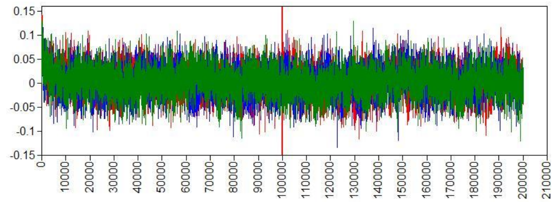
Region: Northern England



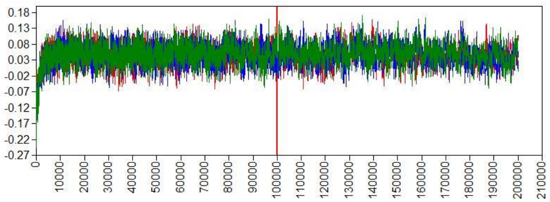
Region: Wales



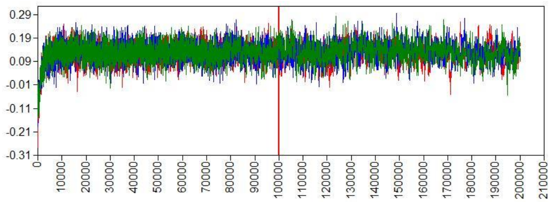
Region: Scotland



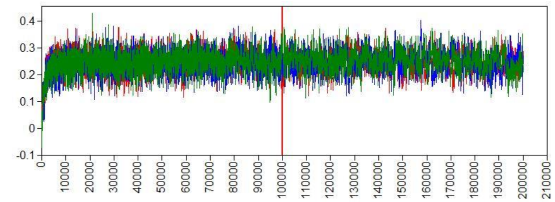
Region: Northern Ireland



Education: GCSE / O-level or lower



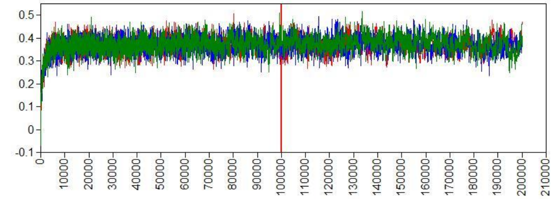
Education: Up to A-level



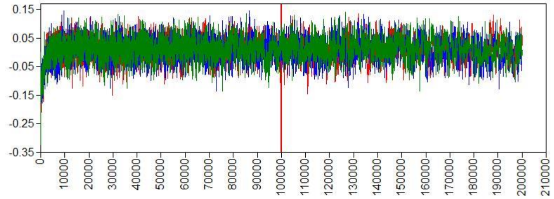
Education: Up to diploma in higher education

Continued…

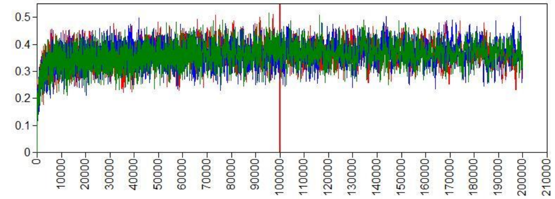
Continued…



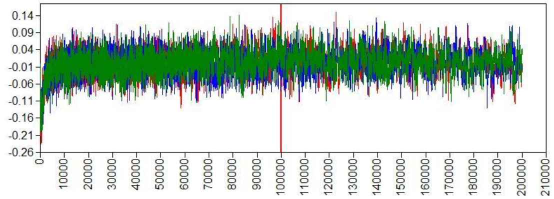
Education: University or higher degree



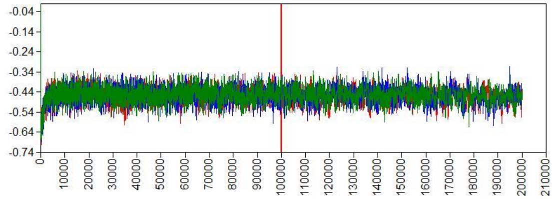
Education: No recorded data



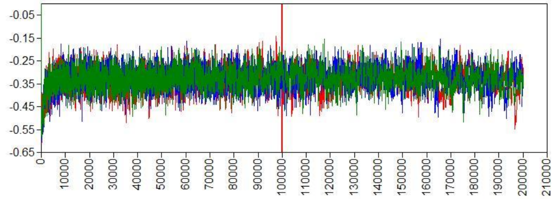
Occupational classification: Professional occupations



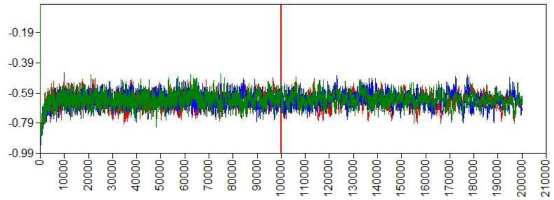
Occupational classification: Associate prof. & technical occ.



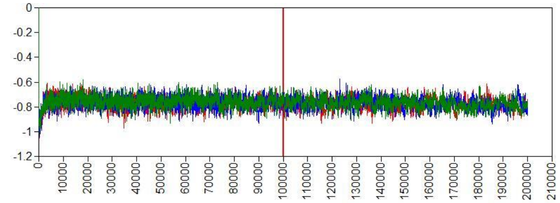
Occupational classification: Admin. & secretarial occ.



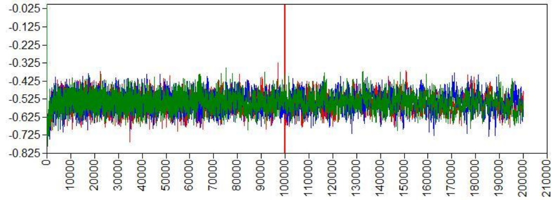
Occupational classification: Skilled trades occupations



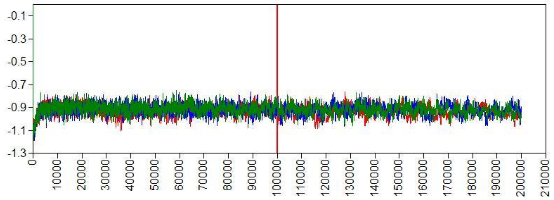
Occupational classification: Personal service occupations



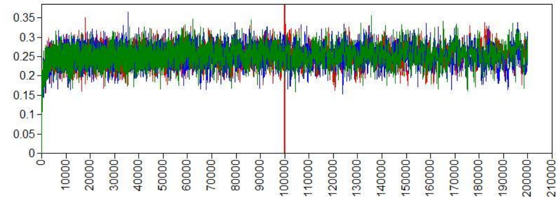
Occupational classification: Sales & customer service occ.



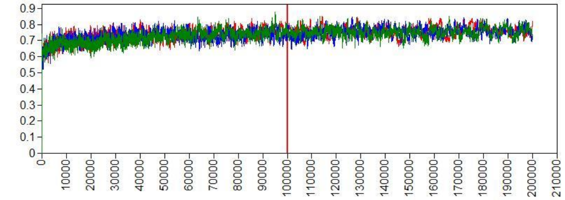
Occupational classification: Process, plant & machine op.



Occupational classification: Elementary occupations



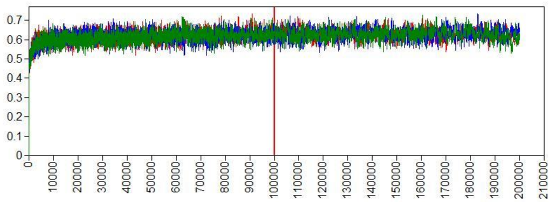
Full or part time employment: Full-time



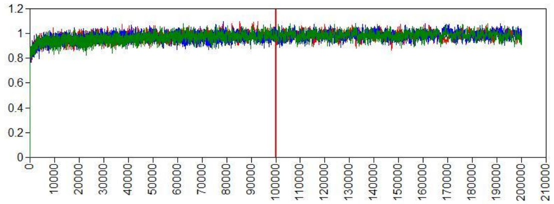
Organisational sector: Public sector

Continued…

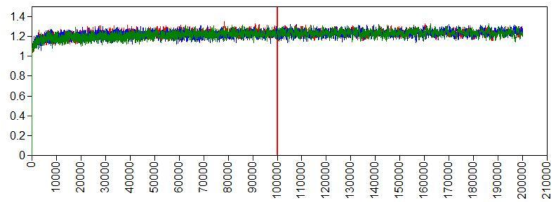
Continued…



Organisation size: Small



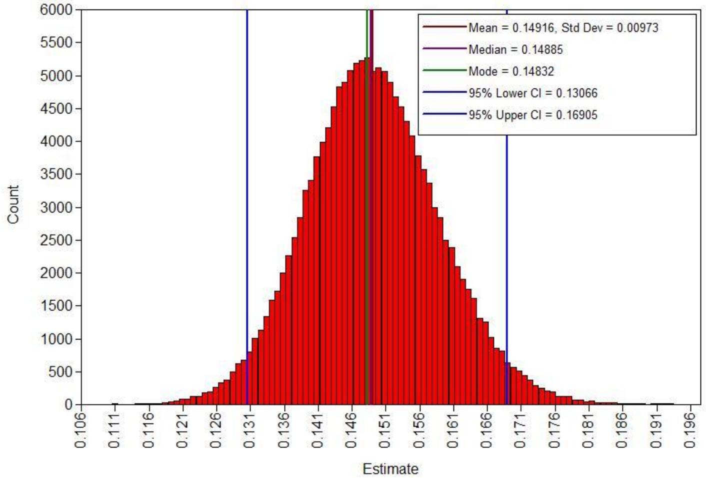
Organisation size: Medium



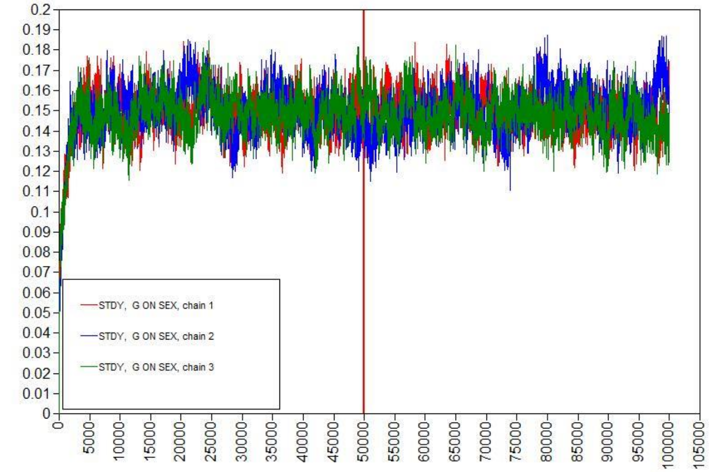
Organisation size: Large

Notes: The estimates are standardised parameter estimates. The three chains in the trace plots for all predictors mixed well, indicating that the MCMC algorithm reached equilibria in estimating the posterior parameter distributions. See notes from Figure 7.x for additional information.

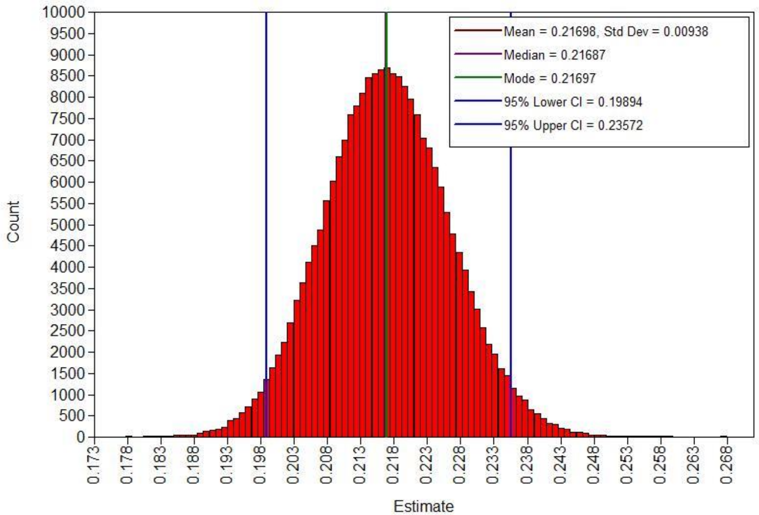
7.4.4 Appendix 7.4: Posterior Parameter Distributions and Trace Plots for R^2 Estimates for *Economic Compensation* based on the Bayesian Estimator



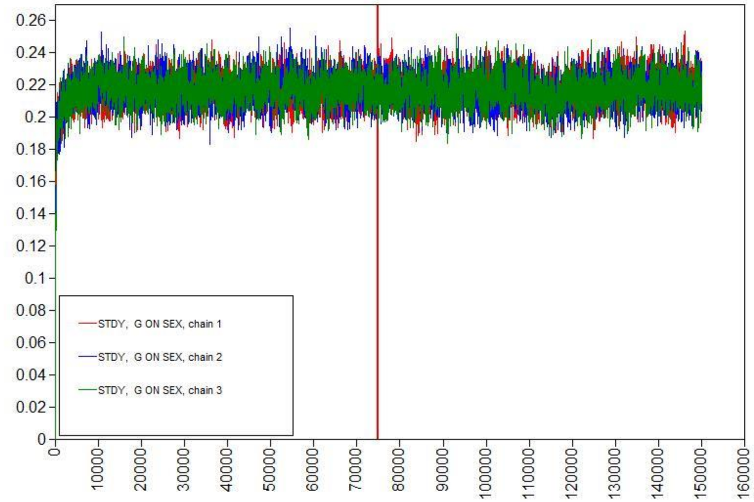
a) Model 1 Posterior Parameter Distribution



a) Model 1 Posterior Parameter Trace Plot

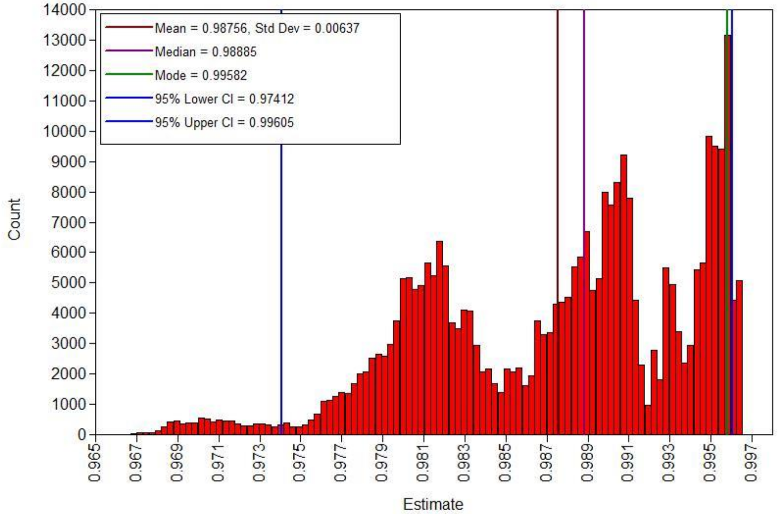


b) Model 2 Posterior Parameter Distribution

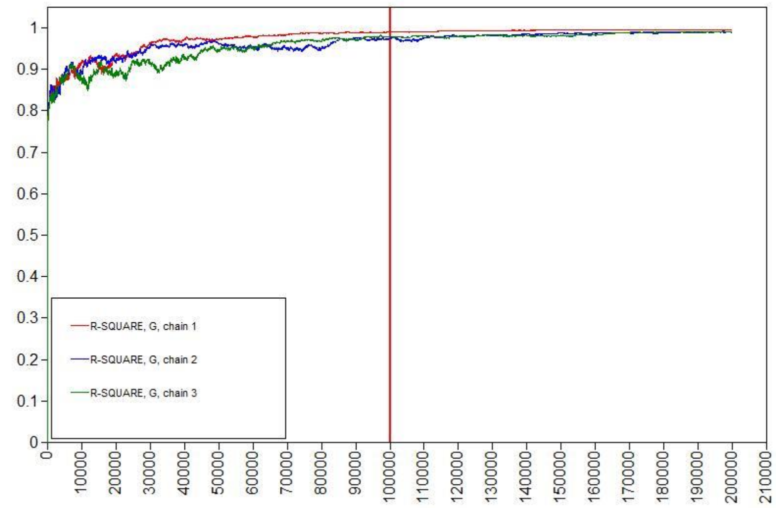


b) Model 2 Posterior Parameter Trace Plot

Continued…



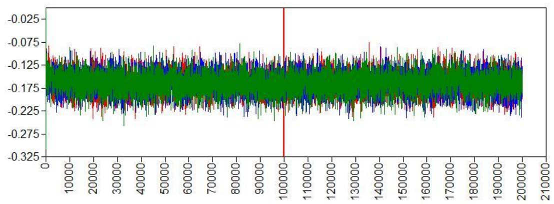
c) Model 3 Posterior Parameter Distribution



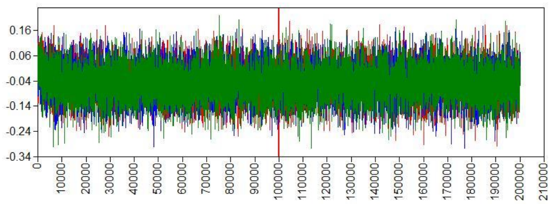
c) Model 3 Posterior Parameter Trace Plot

Notes: The estimates are standardised parameter estimates. The histograms show posterior parameter distributions estimated by the models and the posterior means, posterior standard deviations, and 95% credible intervals printed within the charts correspond to the unadjusted R^2 estimates for the MIMIC Models for *economic compensation* (Table 7.x). From the trace plots, the MCMC algorithm reached equilibrium in estimating the posterior parameter distribution for R^2 for Model 2; however, chains for Model 1 and particularly Model 3 did not mix well, suggesting that they did not reach equilibria in estimating posterior parameter distributions for R^2 for these models and the posterior parameter distribution for R^2 for Model 3 did not follow a normal distribution. This might be due non-informative priors.

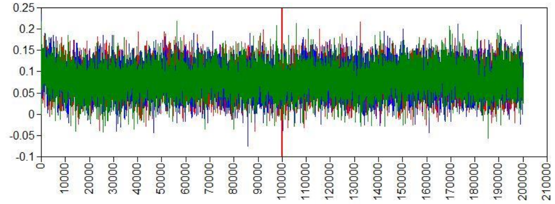
7.4.5 Appendix 7.5: Posterior Parameter Trace Plots for MIMIC Model 3 for *Working Conditions* based on the Bayesian Estimator



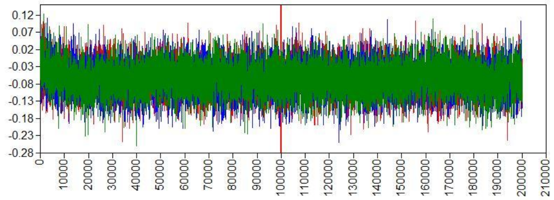
Sex: Male



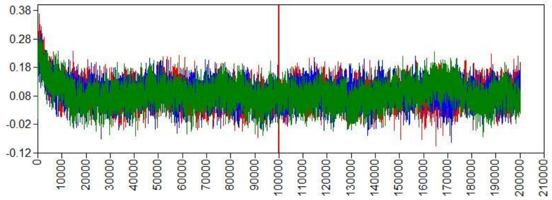
Ethnic group: Mixed



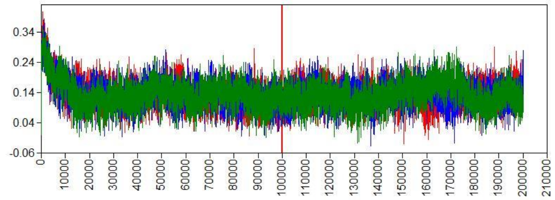
Ethnic group: Asian or Asian British



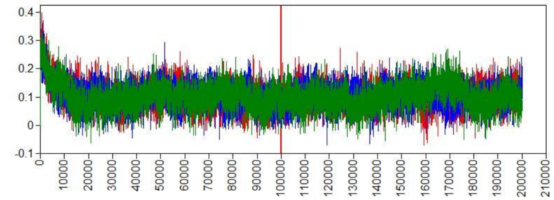
Ethnic group: Black or Black British



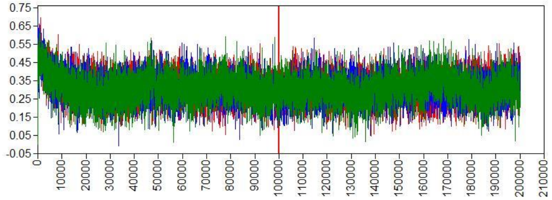
Age group: 25 - 34 years old



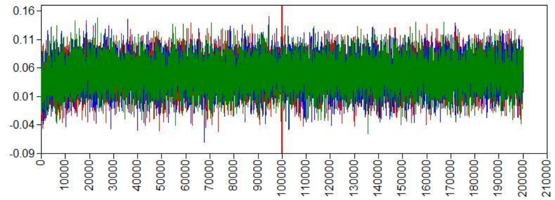
Age group: 35 - 49 years old



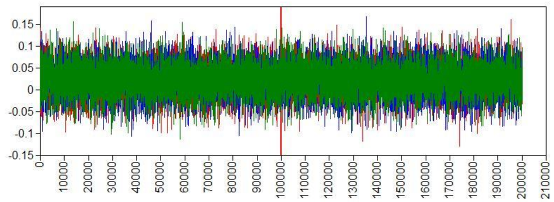
Age group: 50 - 64 years old



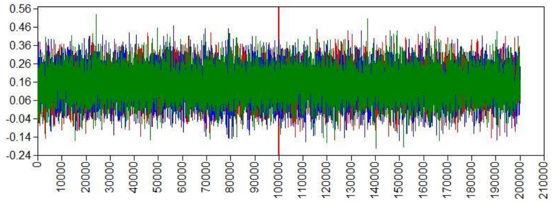
Age group: 65 + years old



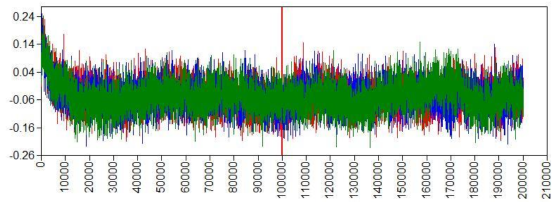
Relationship status: Married or cohabiting



Relationship status: Divorced or separated



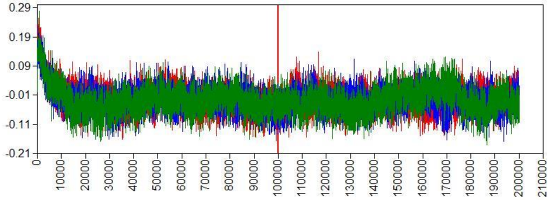
Relationship status: Widowed



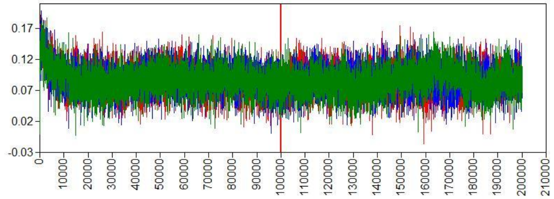
Parental status: Coupled parents with school age children

Continued…

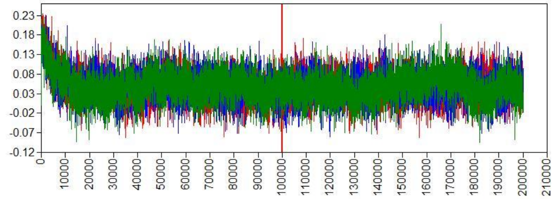
Continued…



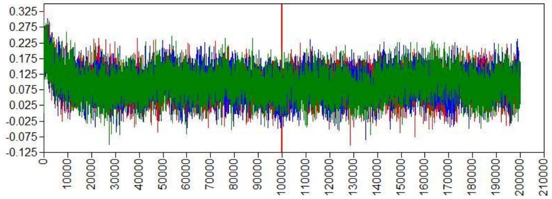
Parental status: Employees without school age children



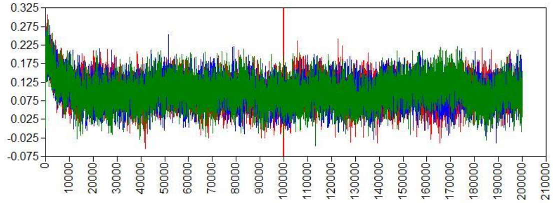
Illness or disability: No



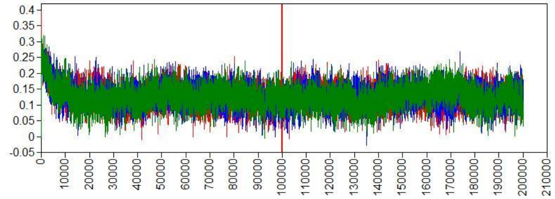
Region: Southern England



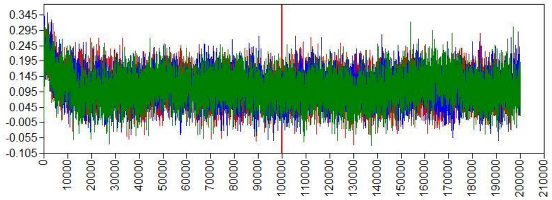
Region: East of England



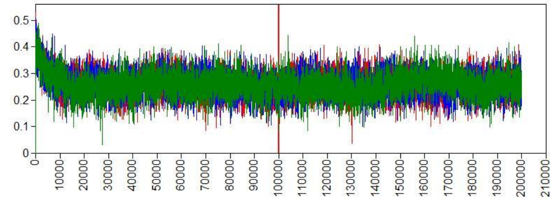
Region: The Midlands



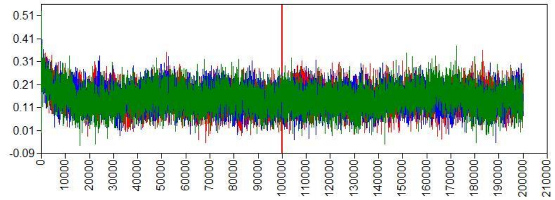
Region: Northern England



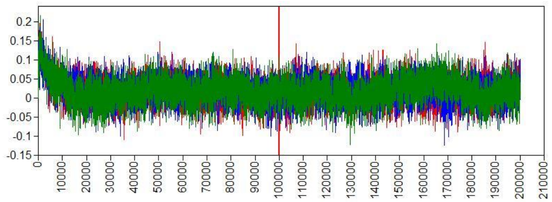
Region: Wales



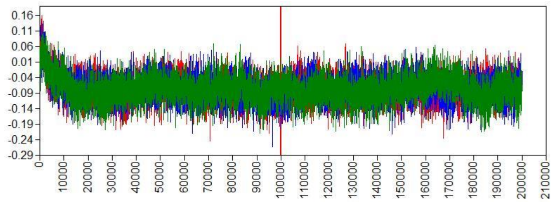
Region: Scotland



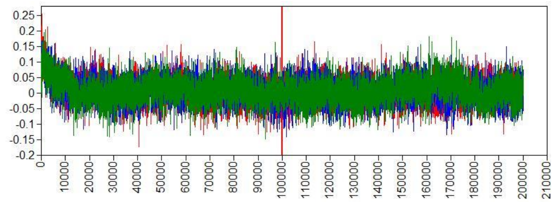
Region: Northern Ireland



Education: GCSE / O-level or lower



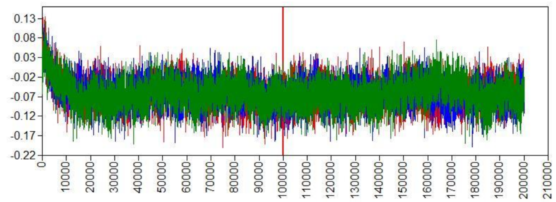
Education: Up to A-level



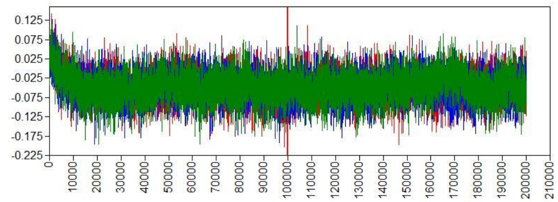
Education: Up to diploma in higher education

Continued…

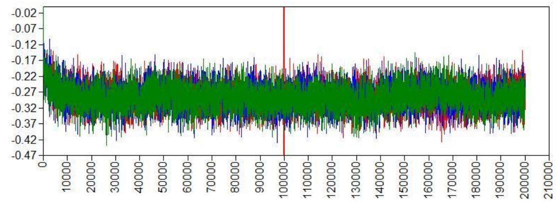
Continued…



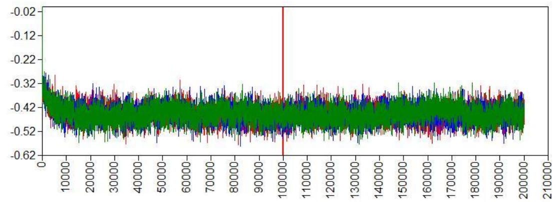
Education: University or higher degree



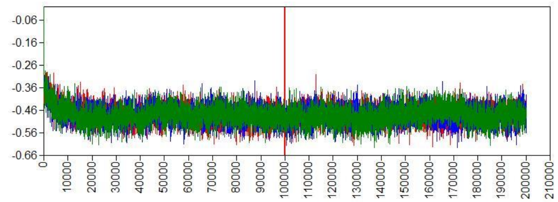
Education: No recorded data



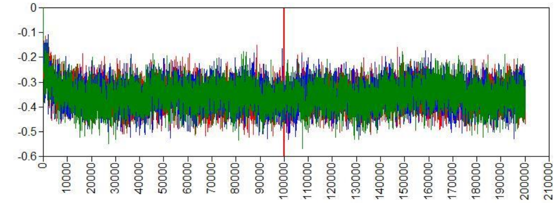
Occupational classification: Professional occupations



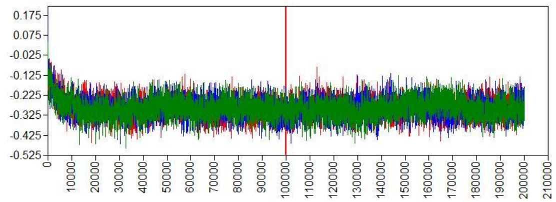
Occupational classification: Associate prof. & technical occ.



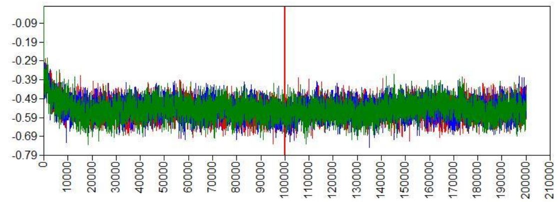
Occupational classification: Admin. & secretarial occ.



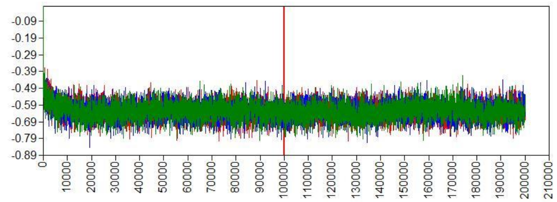
Occupational classification: Skilled trades occupations



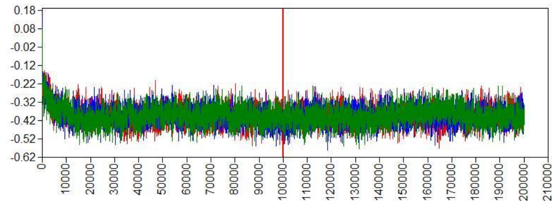
Occupational classification: Personal service occupations



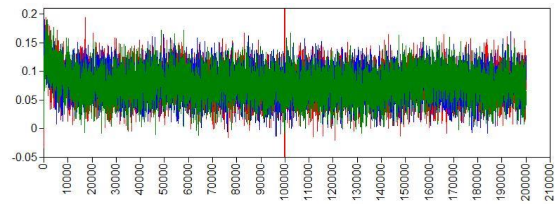
Occupational classification: Sales & customer service occ.



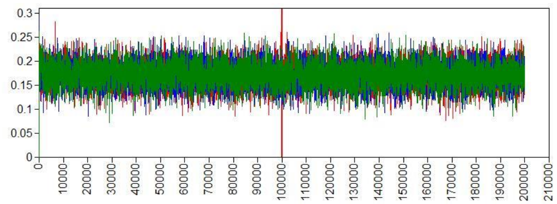
Occupational classification: Process, plant & machine op.



Occupational classification: Elementary occupations



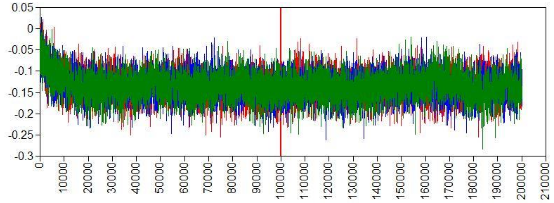
Full or part time employment: Full-time



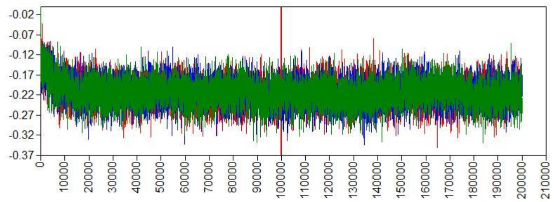
Organisational sector: Public sector

Continued…

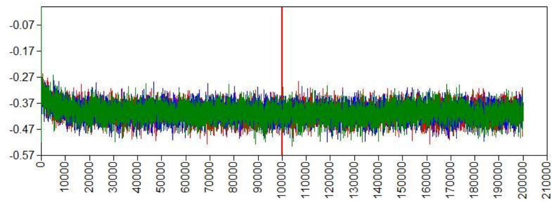
Continued…



Organisation size: Small



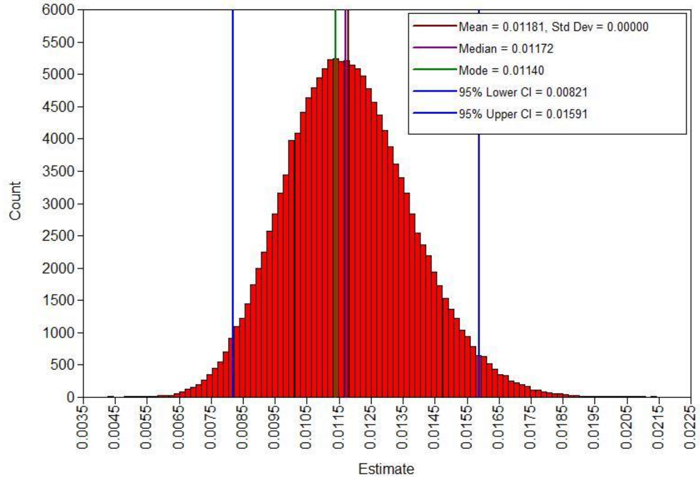
Organisation size: Medium



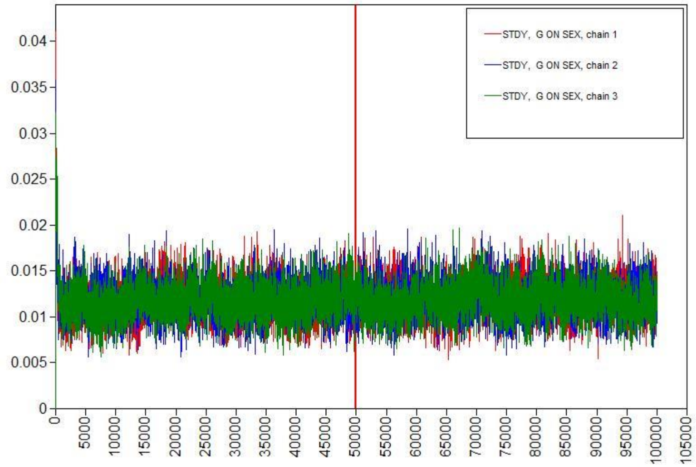
Organisation size: Large

Notes: The estimates are standardised parameter estimates. The three chains in the trace plots for all predictors mixed well, indicating that the MCMC algorithm reached equilibria in estimating the posterior parameter distributions. See notes from Figure 7.x for additional information.

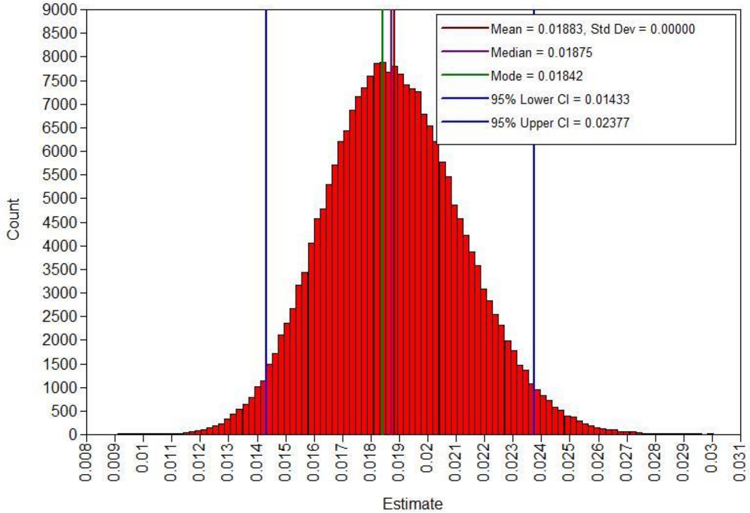
7.4.6 Appendix 7.6: Posterior Parameter Distributions and Trace Plots for R^2 Estimates for *Working Conditions* based on the Bayesian Estimator



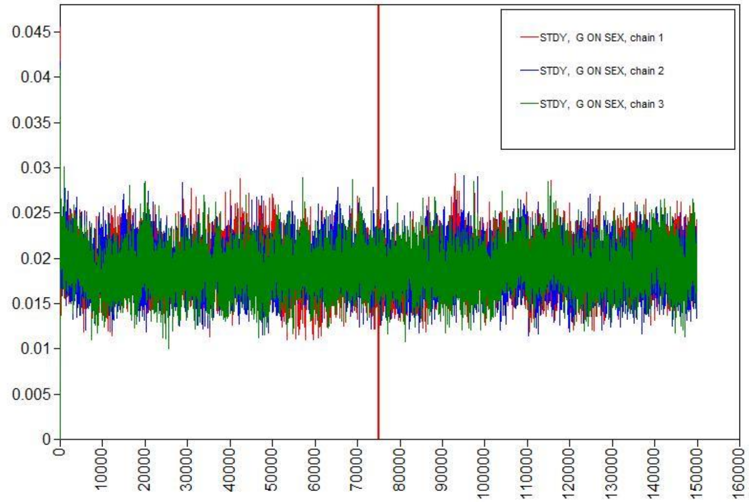
a) Model 1 Posterior Parameter Distribution



a) Model 1 Posterior Parameter Trace Plot

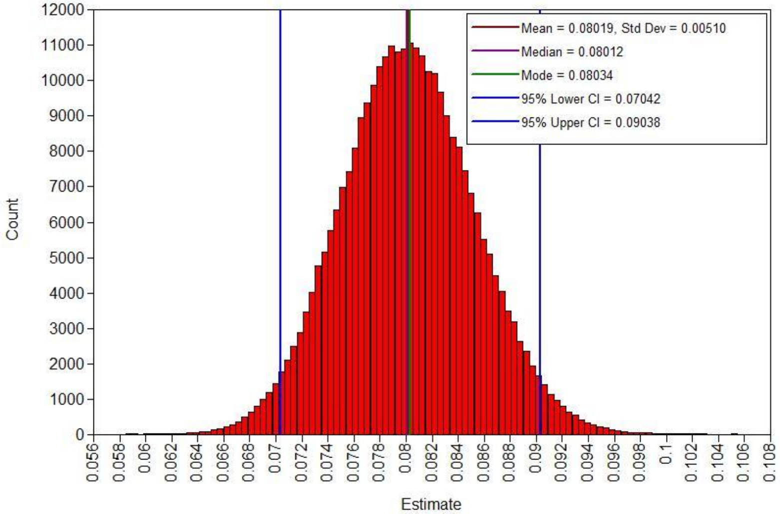


b) Model 2 Posterior Parameter Distribution

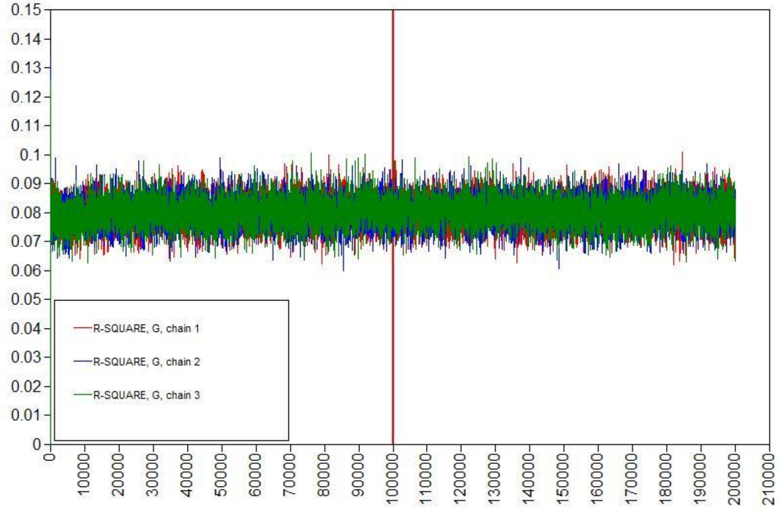


b) Model 2 Posterior Parameter Trace Plot

Continued…



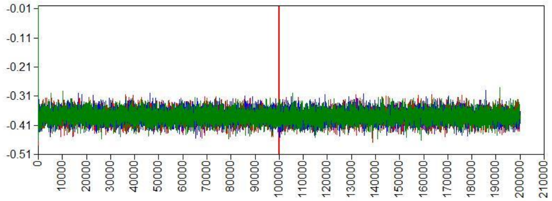
c) Model 3 Posterior Parameter Distribution



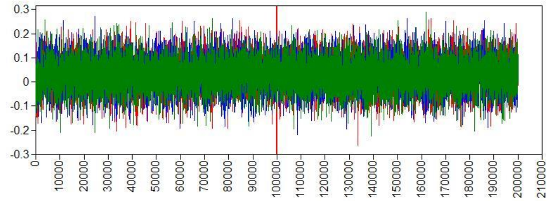
c) Model 3 Posterior Parameter Trace Plot

Notes: The estimates are standardised parameter estimates. The histograms show normal posterior parameter distributions estimated by the models and the posterior means, posterior standard deviations, and 95% credible intervals printed within the charts correspond to the unadjusted R^2 estimates for the MIMIC Models for *working conditions* (Table 7.x). The trace plots show relatively well mixed chains, indicating that the MCMC algorithm reached equilibria in estimating the posterior parameter distributions for these estimates in all three models.

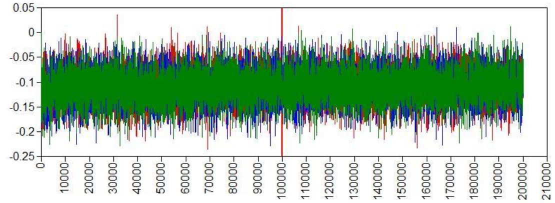
7.4.7 Appendix 7.7: Posterior Parameter Trace Plots for MIMIC Model 3 for *Work-time Scheduling* based on the Bayesian Estimator



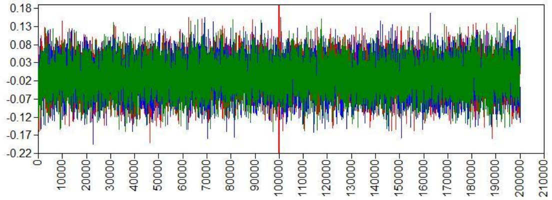
Sex: Male



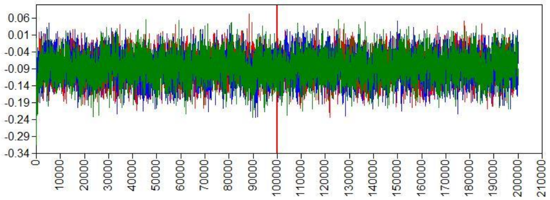
Ethnic group: Mixed



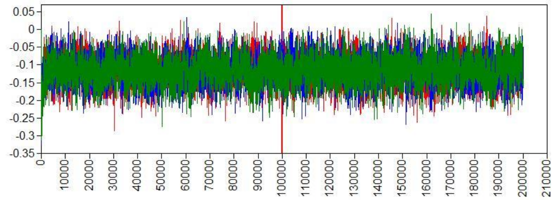
Ethnic group: Asian or Asian British



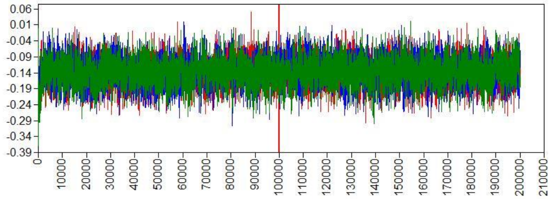
Ethnic group: Black or Black British



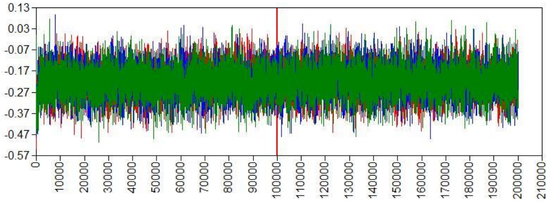
Age group: 25 - 34 years old



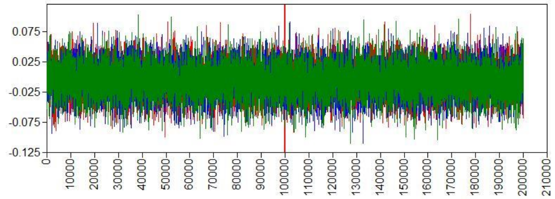
Age group: 35 - 49 years old



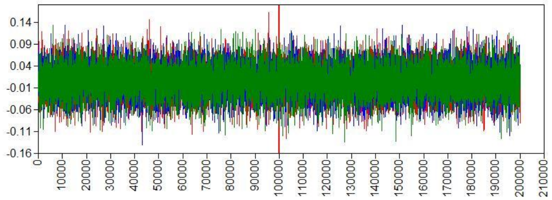
Age group: 50 - 64 years old



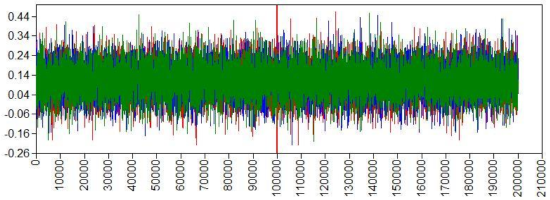
Age group: 65 + years old



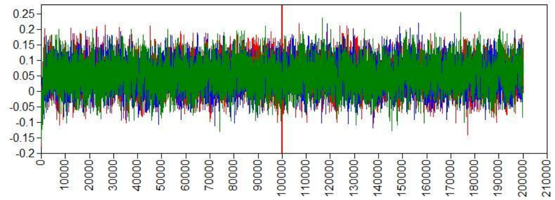
Relationship status: Married or cohabiting



Relationship status: Divorced or separated



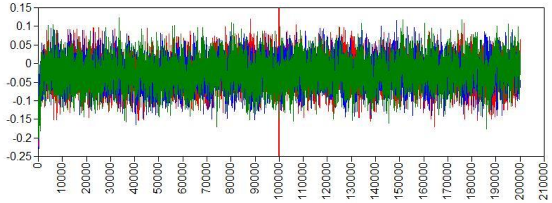
Relationship status: Widowed



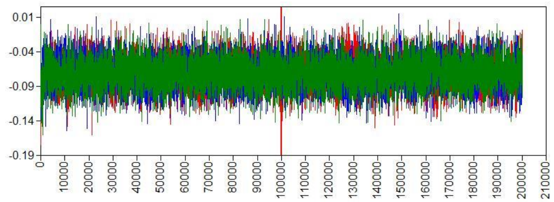
Parental status: Coupled parents with school age children

Continued…

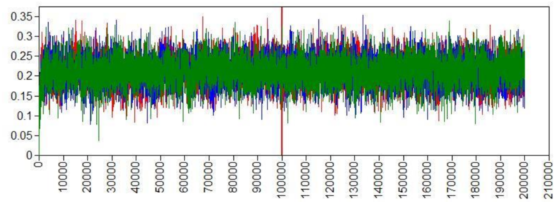
Continued…



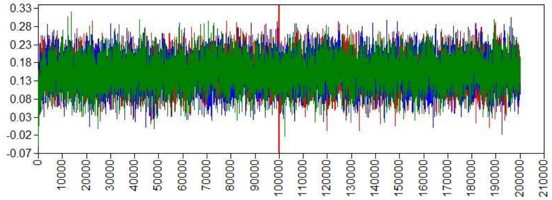
Parental status: Employees without school age children



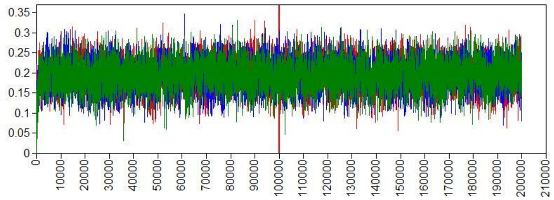
Illness or disability: No



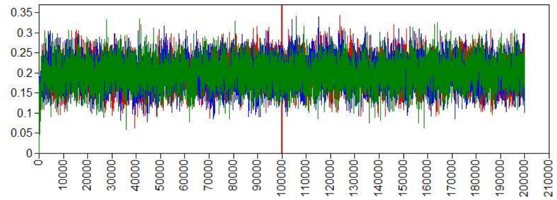
Region: Southern England



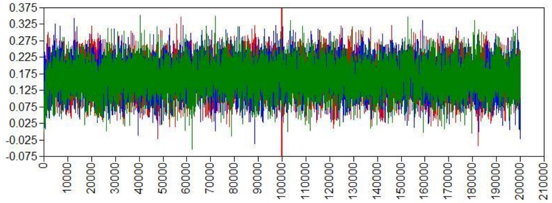
Region: East of England



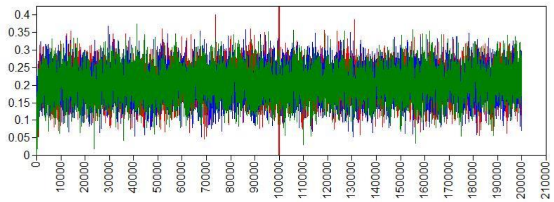
Region: The Midlands



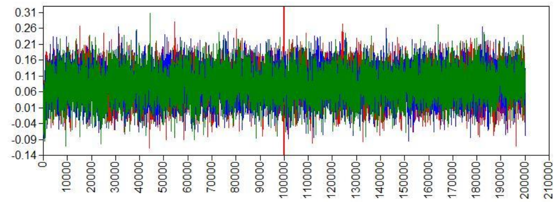
Region: Northern England



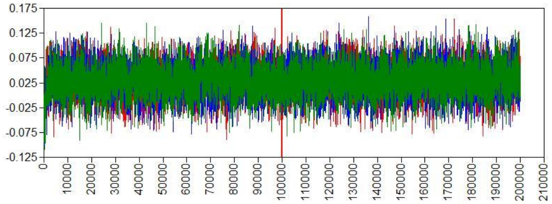
Region: Wales



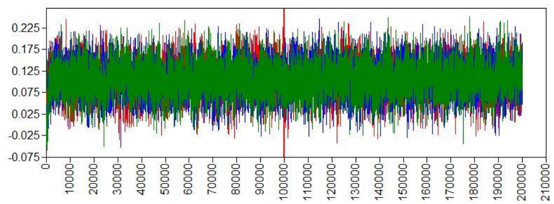
Region: Scotland



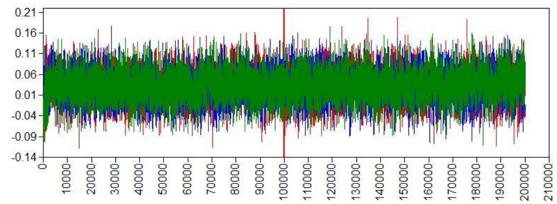
Region: Northern Ireland



Education: GCSE / O-level or lower



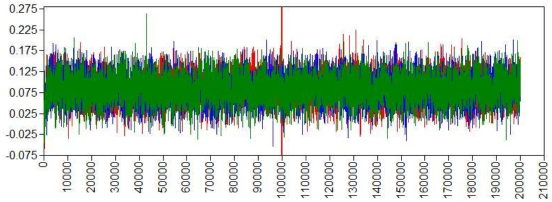
Education: Up to A-level



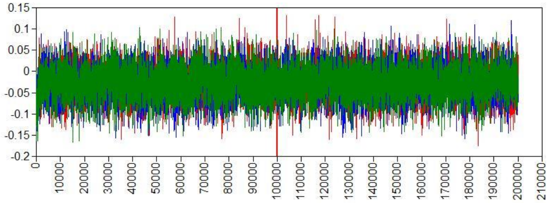
Education: Up to diploma in higher education

Continued…

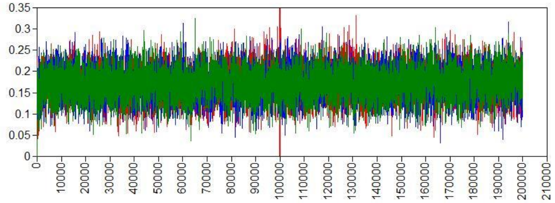
Continued…



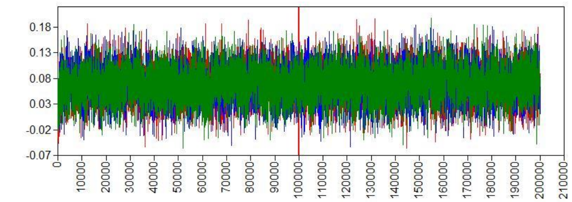
Education: University or higher degree



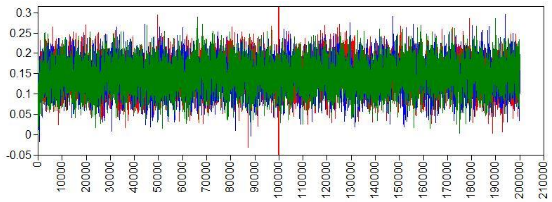
Education: No recorded data



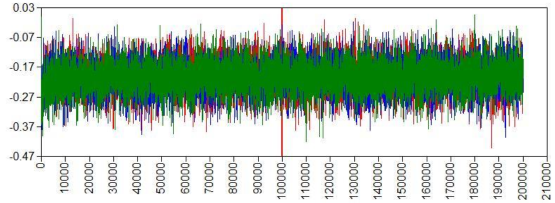
Occupational classification: Professional occupations



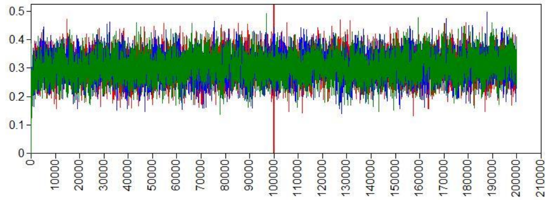
Occupational classification: Associate prof. & technical occ.



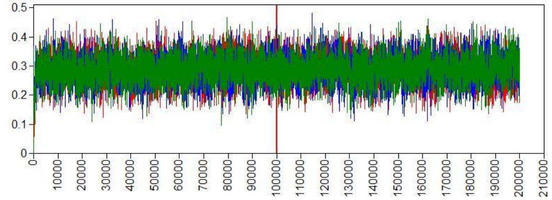
Occupational classification: Admin. & secretarial occ.



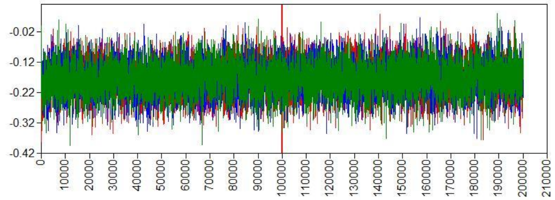
Occupational classification: Skilled trades occupations



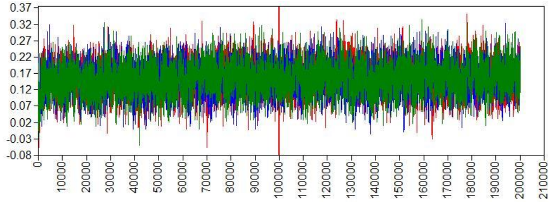
Occupational classification: Personal service occupations



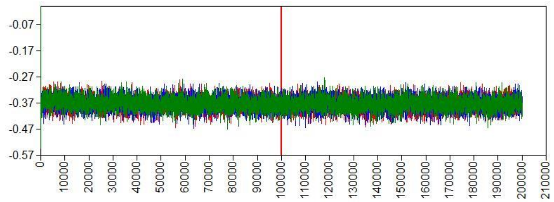
Occupational classification: Sales & customer service occ.



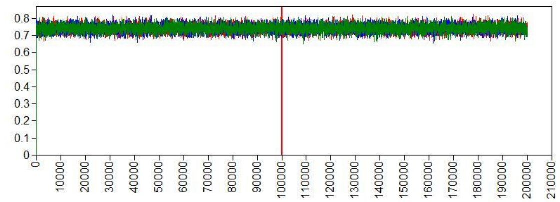
Occupational classification: Process, plant & machine op.



Occupational classification: Elementary occupations



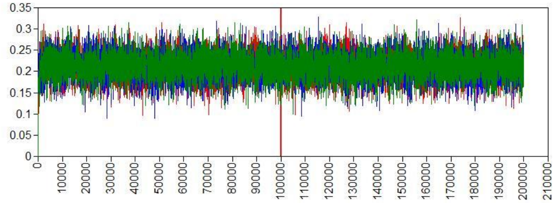
Full or part time employment: Full-time



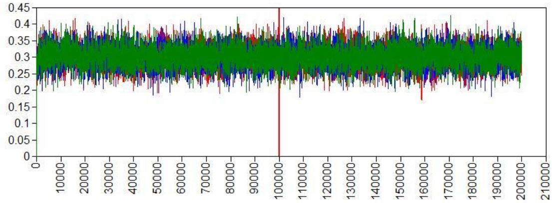
Organisational sector: Public sector

Continued…

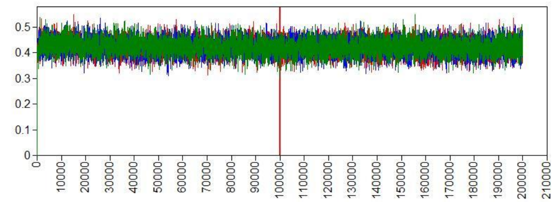
Continued…



Organisation size: Small



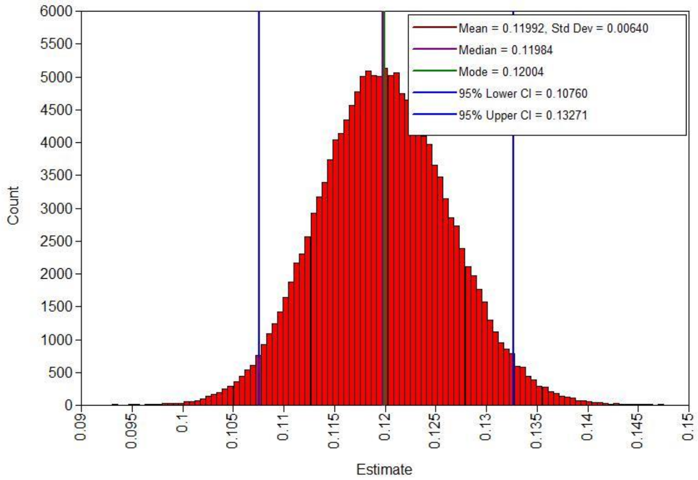
Organisation size: Medium



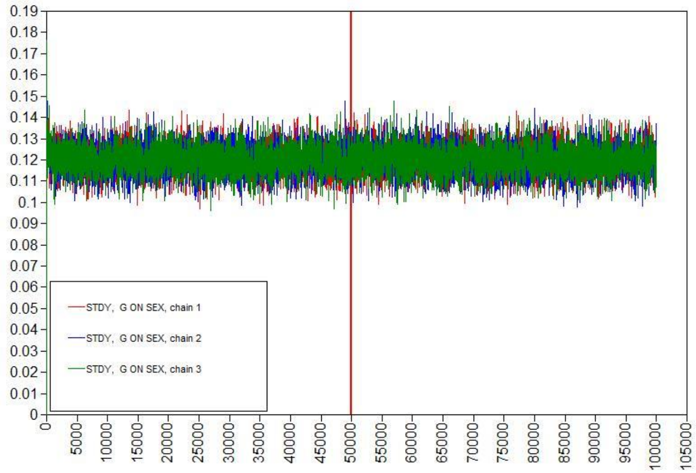
Organisation size: Large

Notes: The estimates are standardised parameter estimates. The three chains in the trace plots for all predictors mixed well, indicating that the MCMC algorithm reached equilibria in estimating the posterior parameter distributions. See notes from Figure 7.x for additional information.

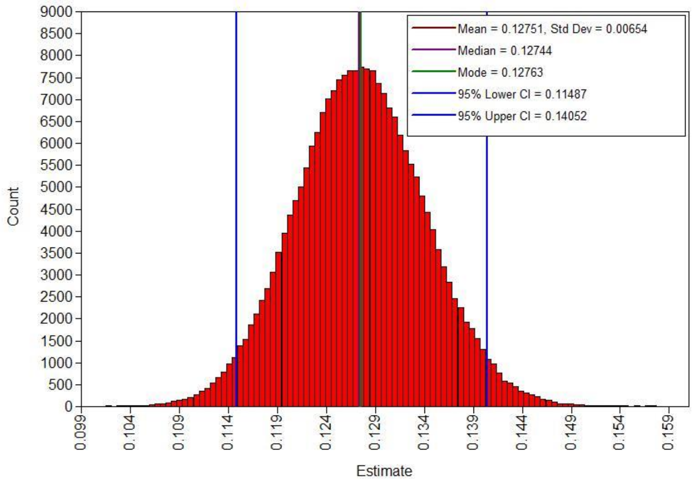
7.4.8 Appendix 7.8: Posterior Parameter Distributions and Trace Plots for R^2 Estimates for *Work-time Scheduling* based on the Bayesian Estimator



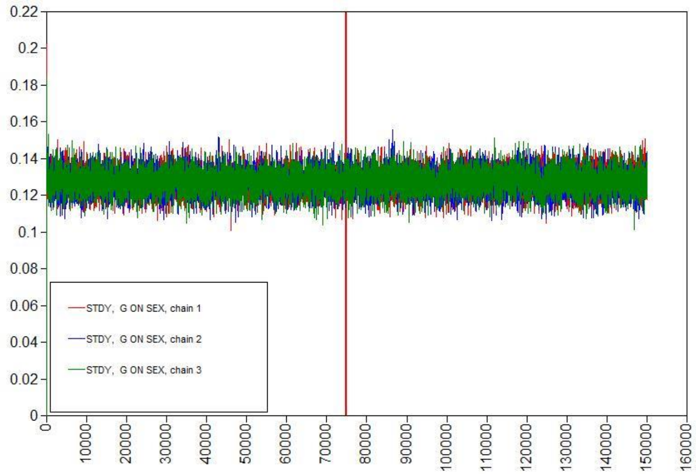
a) Model 1 Posterior Parameter Distribution



a) Model 1 Posterior Parameter Trace Plot

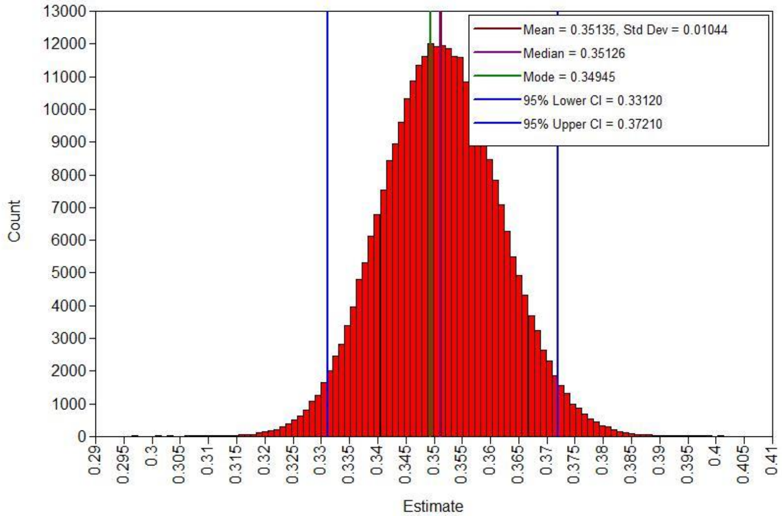


b) Model 2 Posterior Parameter Distribution

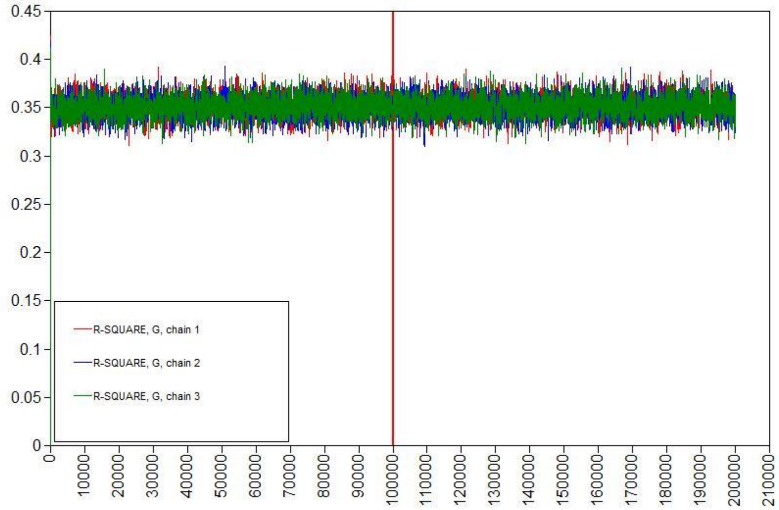


b) Model 2 Posterior Parameter Trace Plot

Continued…



c) Model 3 Posterior Parameter Distribution



c) Model 3 Posterior Parameter Trace Plot

Notes: The estimates are standardised parameter estimates. The histograms show normal posterior parameter distributions estimated by the models and the posterior means, posterior standard deviations, and 95% credible intervals printed within the charts correspond to the unadjusted R^2 estimates for the MIMIC Models for *work-time scheduling* (Table 7.x). The trace plots show well mixed chains, indicating that the MCMC algorithm reached equilibria in estimating the posterior parameter distributions for these estimates in all three models.

Chapter 8 Summary and Conclusions

This chapter provides a summary of the research and highlights contributions the study has made to the topic of measuring quality of work and employment (QWE). The study aimed to apply advanced statistical methods, more specifically item response theory (IRT) modelling, to address some of the limitations with existing measurement instruments of QWE. The first section considers the theoretical contributions of the study to the conceptualisation of QWE and the development of a framework for measuring QWE. The second section focuses on the methodological contributions of the study. This highlights limitations with the data and how it could be improved, as well as the application of IRT modelling and how it addressed limitations of existing measures of QWE. This also considers the application of frequentist and Bayesian approaches to model parameter estimation. The third section focuses on the substantive contributions, specifically highlighting findings from this study that were inconsistent with previous literature. The fourth section considers the study's limitations, while the fifth section outlines how this research could be further developed, and the last section draws conclusions from the research.

8.1 Theoretical Contributions

This study has made some theoretical contributions to the conceptualisation of QWE by developing a framework for measuring QWE. The concept of QWE is complex and elusive as it is difficult to define precisely and cannot be directly measured (Muñoz de Bustillo et al. 2009, 2011a, 2011b). Proposed definitions recognise its diverse attributes (Kalleberg 2011) while also being worker-centred and associated with well-being. Thus, in general, QWE can be defined as the degree to which a job has work and employment attributes that enhance or diminish the well-being of workers (Burchell et al. 2014; Felstead et al. 2019; Green 2006; Holman 2013b; Muñoz de Bustillo et al. 2011a).

The theoretical framework developed for this study was based on the frameworks from the QuInnE project (Erhel and Guergoat-Larivière 2016; European Commission 2018) and the Department of BEIS (Taylor et al. 2017). These frameworks presented more comprehensive dimensions of a measure of QWE compared to frameworks in other literature (see Table 2.3). For example, they included dimensions highlighted as important in the social sciences literature on QWE, such as wages, employment quality, education and training, working conditions, work-life balance, and participation and representation (Erhel and Guergoat-Larivière 2016; European Commission 2018; Taylor et al. 2017). Other frameworks included dimensions that were outcomes rather than inputs of QWE, such as well-being (Gifford 2018; Irvine, White, and Diffley 2018) or were unrelated to QWE, such as ethics (UNECE 2010). On the other hand, other frameworks excluded important dimensions; for instance, Muñoz de Bustillo et al. (2011b) excluded collective representation.

Furthermore, some frameworks aggregated dimensions that should be disaggregated, such as the OECD framework, which aggregated working conditions, skills development, and work-life balance into one dimension (Cazes et al. 2015). Conversely, other frameworks disaggregated dimensions that should be aggregated; for example, the EWCS framework disaggregated work intensity and physical environment (Eurofound 2012, 2017b). Other frameworks also assigned job attributes to dimensions that did not fit or cohere theoretically. For example, the ETUI framework included job security as part of the working conditions dimension when this fitted better with the non-standard forms of employment dimension (Leschke et al. 2008, 2012; Piasna 2017).

Notwithstanding the comprehensive nature of the frameworks from the QuInnE project (Erhel and Guergoat-Larivière 2016; European Commission 2018) and the Department of BEIS (Taylor et al. 2017), this study improved the conceptualisation of their dimensions. For example, the ‘wages’ dimension was framed as *economic compensation* to capture non-wage pecuniary rewards, such as access to a pension scheme, or pay progression in the form of annual increments. The 'education and training' dimension was framed as *training and progression* to associate skills development and progression at the workplace. On the other hand, the 'employment quality' dimension was framed as *employment security* as 'employment quality' is used to describe a broad component of QWE capturing employment relations (Cazes et al. 2015; Muñoz de Bustillo et al. 2011a). This study's theoretical framework of QWE consisted of six dimensions: *economic compensation*, *training and progression*, *employment security*, *working conditions*, *work-life balance* (or *work-time scheduling*), and *social dialogue* (see Figure 2.1).

8.2 Methodological Contributions

Data

The study considered different sources of data and conducted a detailed review of the European Working Conditions Survey (EWCS), the Labour Force Survey (LFS), and the United Kingdom Household Longitudinal Study (UKHLS) (see Table 3.1). The UKHLS was selected because it had more appropriate survey items that captured different dimensions of QWE based on the theoretical framework in Figure 2.1 than the LFS, while it had a larger UK sample than the EWCS. Furthermore, whereas this study conducted a cross-sectional analysis of the UKHLS, the UKHLS also has a panel design which would enable the investigation of changes in QWE over time, an analysis that this research has laid the foundations for.

Methodologically, the review highlighted the longstanding challenges related to survey instruments measuring attributes of QWE and presented suggestions for improving the survey instruments. For example, while the LFS had items related to *training and progression* and *working conditions*, such as training opportunities, accidents and ill health, these captured either outcomes of work or referred to both previous and current employment (see Appendix 3.3). To measure QWE, the survey questions should be framed to capture attributes of the work that expose employees to hazards, such as handling hazardous substances or exposure to loud noise, but also relate to current employment for clarity of the job being evaluated. The LFS also had *work-time scheduling* items that captured formal flexible working arrangements (see Appendix 3.3); however, these referred to respondents' actual working arrangements. An appropriate measure of QWE should capture the formal flexible working arrangements available at a workplace rather than individual respondents' preferences.

Regarding the EWCS, the earnings indicator was based on net pay (see Appendix 3.2). However, this is a measure of disposable income rather than QWE, as respondents may have different deductions from their pay unrelated to QWE, for example, student loan payments. Therefore, an appropriate measure of QWE should be based on gross pay.

Limitations associated with the UKHLS are highlighted in Section 8.4; however, methodological issues related to a potential processing error for the 'progression prospects' item. This item had three valid response options, '*no*', '*yes*', and '*does not apply*' (see Table 3.2), but the last response lacked clarity on who these respondents were. The UKHLS Support Team suggested these were respondents who may have reached the top tier of their roles with no further progression prospects, and in this study, this category was recoded as '*no*'. To reduce the potential for processing error in the measurement of this item, it would be appropriate for the response options to be '*no*' or '*yes*'.

Measuring Quality of Work and Employment

There have been numerous efforts to develop a measure of QWE, as evidenced by the substantial number of instruments in the literature (see Table 2.2). However, existing measures of QWE have some shortcomings. These related to their composition, with some measures including components unrelated to QWE, while others omitted important components of QWE. There were also issues related to the aggregation of indicators, with no consensus on whether to report unaggregated results of the indicators of QWE, aggregate these within dimensions, and/or aggregate the dimensions into an overall measure. Intertwined with the aggregation issues was the weighting of indicators on the aggregate measure with no consensus on how the weights should be assigned. Furthermore, the evaluation of measurement equivalence of the instruments for different groups was seldom considered in the literature on measuring QWE, even though it is a prerequisite for between-group comparisons. This study proposed IRT modelling as a method to develop a measurement instrument of QWE that addresses some of these shortcomings.

First, while the composition of the measure of QWE was rooted in different traditions of the social sciences literature on QWE (Section 2.1.3) and the theoretical framework of QWE (Section 2.1.5), IRT modelling enabled the testing of the hypothesised latent structure of the observed items and evaluated how well the model fitted the data. Initially, 25 items across six dimensions were considered in developing the measure of QWE, but the '*weekend working*' and '*collective bargaining*' items were excluded in the subsequent IRT modelling due to a violation of the local independence assumption. That is, given the hypothesised latent structure of the observed items, the model did not sufficiently explain responses to these items. This approach contrasts with other approaches to developing measures of QWE, such as estimating averages of the observed items, which put together items within dimensions and/or an overall measure of QWE without evaluating whether the items cohere within the dimensions and/or the overall measure.

Second, the study compared different competing graded-response IRT models, that is, unidimensional, correlated factors, second-order factor, and bifactor models, and presented new knowledge that suggested the measurement of QWE was better modelled by a bifactor model. The bifactor model postulated that responses to the observed items were explained by *overall QWE* given other *dimensions of QWE* in the measurement model and other *dimensions of QWE* over and above *overall QWE*. Properties of the measurement instrument indicated that a unidimensional solution did not sufficiently account for the common variance among the items measuring QWE (see Table 5.7). This provided empirical evidence supporting the consensus within labour market research that QWE is a multidimensional concept. The structure of the bifactor model was such that the overall factor accounted for the common variance shared by all the observed indicators, given the specific factors in the measurement model. In contrast, specific factors accounted for the common variance among the observed indicators within that specific factor, over and above the overall factor (Brown and Croudace 2015; Chen et al. 2012, 2006; Reise 2012; Reise et al. 2018). Furthermore, all the factors in the measurement model were assumed to be mutually orthogonal (Cai et al. 2011; Reise 2012), which meant that *overall QWE* and other *dimensions of QWE* could be investigated and reported independently. Importantly, this addressed some of the discourse in the literature on whether to aggregate items within *dimensions of QWE* and/or into a measure of *overall QWE*. On the other hand, reporting unaggregated items, as suggested by Irvine et al. (2018) and the ONS (2019, 2022) (see Section 2.1.4), does not account for the complexity of the relationships that exist between the items measuring QWE.

Third, IRT modelling addressed the lack of consensus on assigning weights of the observed items on the aggregated measures of QWE. The bifactor model models a within-item multidimensional latent structure (Adams et al. 1997; Desjardins and Bulut 2018; Paek and Cole 2020) with each item directly associated with *overall QWE* and another specific *dimension of QWE* (see Figure 5.1 (d)). Item parameters estimated by the bifactor IRT model were used to determine the item's weight on *overall QWE* and the specific *dimension of QWE*. This contrasts other approaches that arbitrarily assign weights or assume equal weighting without a theoretical explanation. Criticisms of model-based approaches to determining weights are that the results may be counterintuitive and not necessarily align with the purposes of public policy (Sehnbruch et al. 2020). However, patterns in the data may be more informative in estimating weights, while it can also be argued that data should drive public policy.

Lastly, the study applied differential item functioning (DIF) to evaluate measurement equivalence of the measure of QWE for each of the demographic, socio-demographic, and socio-economic characteristics. Much of the literature on the measurement of QWE implicitly assumes measurement equivalence of the instruments without this being evaluated. For example, of the existing or proposed measurement instruments of QWE reviewed in this study (see Table 2.2), only the European Intrinsic Job Quality Index (Cascales Mira 2021) evaluated measurement equivalence. This study used the iterative hybrid ordinal logistic regression/IRT approach by Choi et al. (2011) to detect DIF, and this method evaluated uniform and non-uniform DIF but also provided estimates of the magnitude of DIF. However, its limitation is that it assumes the observed items measure a unidimensional latent structure (Choi et al. 2011), which may lead to a misidentification of DIF for observed items with a multidimensional latent structure. An alternative approach extended multiple indicators multiple causes (MIMIC) modelling to evaluate DIF by investigating whether a predictor directly affected the observed items given the model, with a statistically significant direct effect indicating the presence of DIF (Wang and Wang 2020). While this method accounted for the multidimensional latent structure of the observed items, it only evaluated uniform DIF and did not provide estimates of the magnitude of DIF. A method of detecting uniform and non-uniform DIF for within-item multidimensional latent structures that also provides estimates of the magnitude of DIF would be helpful in this research area.

Estimation Methods

This study also contributed to the discourse on the application of frequentist and Bayesian approaches to model parameter estimation. For the frequentist approach, the robust maximum likelihood (MLR) estimator was used to estimate model parameters with standard errors robust to non-normality and non-independence of observations associated with data obtained from complex sampling designs (Muthén and Muthén 2017; Wang and Wang 2020), such as the UKHLS. However, the MIMIC models were computationally cumbersome to estimate with the MLR estimator due to categorical observed items and the number of latent traits in the measurement model (Muthén and Asparouhov 2012). This was evidenced by the time the models took to reach convergence (see Tables 7.2 – 7.5).

The Bayesian approach was also used to estimate the MIMIC models using non-informative priors to obtain parameter estimates analogous to those based on the MLR estimator (Johnson and Sinharay 2016; Muthén and Asparouhov 2012). Bayes estimators are also robust to data non-normality, while the estimation of complex models with a high number of latent traits was computationally less cumbersome than with the MLR estimator (Muthén and Asparouhov 2012; Wang and Wang 2020). Both estimation methods yielded similar results, although, for some predictors, parameter estimates based on the MLR estimator were outside the Bayesian 95% CI based on the Bayes estimator. The more important implication is that the study established a foundation for conducting more complex analyses with the Bayesian methods that would otherwise not be feasible with the frequentist methods, such as extending this analysis to a longitudinal study investigating changes in QWE over time.

8.3 Substantive Contributions

The study compared and predicted levels of *overall QWE* and other *dimensions of QWE* by demographic (sex, ethnic group and age group), socio-demographic (relationship status, parental status, illness or disability and region), and socio-economic (education, occupational classification, full or part-time, organisational sector and organisation size) characteristics. There were statistically significant associations between these characteristics and *overall* or other *dimensions of QWE* in the UK employee population. Considered as individual predictors, this study suggested that demographic or socio-demographic characteristics did not explain much of the variation in *overall* or other *dimensions of QWE*, with the effects either small or negligible (see Tables 6.2 – 6.3). In contrast, socio-economic characteristics explained more of the variation in the latent traits (see Table 6.4). Occupational classification had a large effect on *economic compensation*, while full or part-time, organisational sector and organisation size had moderate effects on *economic compensation*. Occupational classification also had a moderate effect on *overall QWE*, while organisational sector had a moderate effect on *work-time scheduling*. Notably, whereas education is an important investment in human capital (Okay-Somerville and Scholarios 2013; Solomon et al. 2022), it had small effects on *overall QWE* and other *dimensions of QWE*. This could be due to a lack of industry-specific skills in post-compulsory secondary education in liberal market economies, such as the UK, whose education and training systems place more emphasis on general education (Hall and Soskice 2001; Soskice 1999, 2005).

Substantively, findings from this study broadly supported evidence from other literature about levels of QWE for different groups of UK employees. While this suggested that IRT modelling largely replicated other methods of measuring QWE, some results were inconsistent with previous literature. These discrepancies were attributed to different reasons, including different study populations, differences in the observed items used for measuring QWE, and the methodology of developing the measure of QWE. Furthermore, the study highlighted that IRT modelling and the bifactor model provided a more nuanced understanding of differences in QWE between some groups of employees that would otherwise not be feasible with other methods.

First, in contrast to other studies that found no differences in work autonomy (*working conditions*) by sex (Gallie and Zhou 2013; Lindley 2015; Wu et al. 2021), this study found that males had poorer *working conditions* than females. This was partly attributed to different items of work autonomy used in different studies and the method of aggregating the items. For example, in their study, Wu et al. (2021) used work pace, work manner, and task order as measures of work autonomy (*working conditions*) and aggregated these by estimating the arithmetic mean of the items, which assumes equal weighting. This study, however, included job tasks and work hours in addition to the items used by Wu et al. (2021), while the conditional slopes estimated by the bifactor IRT model for each of these items on *working conditions* were not equal (see Table 5.3 and Appendix 5.1). This suggested that the items did not contribute an equal weight on *working conditions*, as implied by the method of estimating the arithmetic mean of the items.

Second, this study also found that younger employees were more aware of and had better access to other forms of *work-time scheduling*, in contrast to evidence suggesting that they were more likely to have poorer outcomes than older employees (Sturges and Guest 2004). This discrepancy could be attributed to different populations between the studies. Sturges and Guest (2004) focused on a population of UK graduate employees in large organisations at different stages of their careers, while this study focused on a general UK employee population. Furthermore, Sturges and Guest (2004) reported unaggregated results of the items measuring different attributes of work-life balance, such as working hours and conflict between work and non-work time. This approach does not capture the complexity of the relationships between the items measuring work-life balance (or *work-time scheduling*), as opposed to IRT modelling which aggregated the items within a dimension of *work-time scheduling*.

Third, findings from this study supported previous literature that highlighted the challenges experienced by employees with a longstanding illness or disability in the labour market (Davidson and Kemp 2008; Meager and Hill 2005). Thus, employees with a longstanding illness or disability had poorer *economic compensation* and *working conditions* but more awareness of and better access to other forms of *work-time scheduling* than those without. However, this study found no differences in *overall QWE* between employees with longstanding illness or disability and those without, in contrast to other literature. This could be attributed to different methods used in the studies. For example, Meager and Hill (2005) reported unaggregated results of the items measuring different attributes of QWE, such as pay, work-related training, employment type, and working patterns. However, this does not capture the complexity of the relationships between the items. The better-than-expected *overall QWE* among employees with a longstanding illness or disability could be partly attributed to government initiatives supporting their participation in the labour market (Grover and Piggott 2015; Lewis et al. 2013). Evidence has indicated that people with disabilities are increasingly joining the workforce (Department for BEIS 2018), which may result in QWE for disabled employees being a salient social issue in the labour market.

Fourth, this study's results supported literature highlighting longstanding disparities in the labour market by region with advantages for those residing in the London and Southern England regions (Department for LUHC 2022; Jones and Green 2009; Low Pay Commission 2021). However, the study also highlighted better outcomes for employees in Scotland and suggested that they had better *working conditions* and were more aware of and had better access to other forms of *work-time scheduling*, along with similar levels of *economic compensation* compared to those in London, while employees in London had better *overall QWE* than those in Scotland. Substantively, the better outcomes for employees in Scotland could be partly because of a knowledge economy centralised in major cities such as Edinburgh, Aberdeen, and Glasgow, which require a highly skilled workforce (Hepworth et al. 2005), similar to London. Methodologically, IRT modelling and the bifactor model provided a more nuanced understanding of QWE that would otherwise not be feasible with other approaches to measuring QWE. Thus, ordinarily, *overall QWE* would be an average of other *dimensions of QWE,* and this would have resulted in better *overall QWE* for employees in Scotland compared to those in London. However, the bifactor model made it plausible that despite better outcomes on other *dimensions of QWE* for employees in Scotland, they had poorer *overall QWE* than those in London. Substantively, *overall QWE* captured aspects of effective gross pay, work autonomy items, awareness of formal flexible working arrangements available at the workplace (flexi-time, compressed hours, annualised hours, home working and other flexibility), and being able to vary working hours on an informal basis as these had moderate to high loadings. On the other hand, effective gross pay, pension provision, and pay progression had moderate to high loadings on *economic compensation*, while job tasks, work pace, work manner, and task order had high loadings *working conditions*, and part-time, term-time, job sharing, flexi-time, compressed hours and annualised hours had moderate to high loadings on *work-time scheduling* (Appendix 5.1). Considering, specifically employees in London and Scotland, more of the common variance among indicators loading moderate to high on *overall QWE* was captured for employees in London than Scotland, given other *dimensions of QWE* in the measurement model. In contrast, for *working conditions*, or *work-time scheduling*, more of the common variance among indicators loading moderate to high or high within these dimensions was captured for employees in Scotland than London.

Fifth, results from this study were consistent with findings in other literature that highlighted disadvantages for employees with no educational qualifications in the labour market (Okay-Somerville and Scholarios 2013; Solomon et al. 2022), although this varied for different aspects of the labour market. Thus, employees with no qualifications had poorer *overall QWE*, *economic compensation*, and were less aware of and had poorer access to other forms of *work-time scheduling* than those with some form of education, but there were no differences in *working conditions* by educational qualifications. The lack of expected differences in *working conditions* could be attributed to graduates working in areas unrelated to their area of study (Green and Zhu 2010; Okay-Somerville and Scholarios 2013; Warhurst 2008). This may particularly be the case in a liberal market economy, like the UK, where education and training systems lack industry-specific skills post-compulsory secondary education and place more emphasis on general education (Hall and Soskice 2001; Soskice 1999, 2005).

Lastly, as with the example relating to comparison by region, IRT modelling and the bifactor model also provided a more nuanced understanding of QWE in the comparison by organisational sector. Results from this study supported evidence from other literature that indicated better outcomes for employees in public sector organisations than those in the private sector (Cribb et al. 2014; Murphy et al. 2020; Rubery 2013). Employees in public sector organisations had better *economic compensation*, *working conditions*, and were more awareness of and had better access to other forms *work-time scheduling* than those in private sector organisations. This could be attributed to the role of the government as a public sector employer. Whilst there is minimal government involvement in the regulation of the UK labour market as a liberal market economy (Gallie 2007b; Hall and Soskice 2001; Holman 2013b; Soskice 1999), the government is more likely to adhere to the regulations it has set, such as paying the NMW or NLW, providing pension schemes, and flexible working arrangements, than private sector organisations. However, despite this, the study found that public sector employees had poorer *overall QWE* than private sector employees. This result would not be feasible with other approaches to measuring QWE where *overall QWE* would be the average of the other *dimensions of QWE*. Thus, in the bifactor model, more of the common variance among indicators loading moderate to high on *overall QWE* was captured in private than public sector employees, given other *dimensions of QWE* in the measurement model. In the cases of *economic compensation*, *working conditions*, or *work-time scheduling*, more of the common variance among indicators loading moderate to high or high within these dimensions was captured in public than private sector employees, over and above *overall QWE*. Substantively, the better *overall QWE* for private sector employees could be due to a greater variety of jobs available in the private sector than the public sector as a share of the labour market, with more opportunities to work in areas of their interest.

8.4 Limitations

There were several limitations in conducting this research, and these were mainly related to a lack of availability of appropriate items measuring different attributes of QWE in social survey data. First, there were measurement issues relating to the construct validity of the *working conditions* dimension using the UKHLS. This dimension was exclusively measured by different attributes of work autonomy and did not consider other aspects such as work intensity, job variety, or measures of health and safety such as exposure to hazards at work.

Second, model identification in IRT modelling, and indeed latent variable modelling in general, requires a minimum of three items per latent trait or factor. The UKHLS had two items measuring each of the *training and progression* and *employment security* dimensions, while the *social dialogue* dimension was measured by a single item. The implications for this in estimating the bifactor model were that the slopes for the items measuring the *training and progression* and *employment security* dimensions were constrained to be equal for model identification. This contributed to estimated scores for these dimensions not being a good representation of the latent traits and, therefore, not suitable for subsequent analysis (see Section 5.2.6); although the items still contributed to *overall QWE*. However, the measure of QWE did not capture any aspect of the *social dialogue* dimension, as responses to the item were not sufficiently explained by the model.

These limitations can be addressed by including additional survey items in the UKHLS on *training and progression*, *employment security*, and *social dialogue* dimensions, as well as items measuring *working conditions* beyond work autonomy. Possible items can be adapted from other social surveys with appropriate QWE indicators. For example, the EWCS has survey items such as training participation and training days provided by the employer for the *training and progression* dimension, predictable hours for the *employment security* dimension, direct participation and support at the workplace for the *social dialogue* dimension, and a range of health and safety, work intensity and job variety items for the *working conditions* dimension (see Appendix 3.2).

More broadly, the longstanding challenges related to the availability of appropriate survey instruments for measuring QWE could be addressed by developing item banks of attributes capturing different dimensions of QWE. This is the case with other social science disciplines that experienced practical and technical issues associated with measurement and applied IRT modelling. For example, the Programme for International Student Assessment (PISA) was launched in educational measurement to develop and conduct a large-scale international assessment for monitoring educational system outcomes related to student achievement (OECD 1999, 2019). On the other hand, in health measurement, the Patient-Reported Outcomes Measurement Information System (PROMIS) project was launched to develop item banks of different health domains that provided efficient, flexible, and publicly available measurements of patient-reported health outcomes (Cella et al. 2010; Gershon et al. 2016).

8.5 Further Research

This study has contributed to the measurement of QWE with the application of IRT modelling and presented new knowledge that suggested the measurement of QWE was better captured by a bifactor model. Further research should investigate whether the measurement of QWE using a bifactor IRT model can be replicated with data from a different place, time and target population.

First, in terms of place, theoretical debates suggest institutional structures have important implications for the functioning of labour markets, including QWE, with empirical evidence from comparative analysis studies suggesting cross-national variations based on the institutional regimes (Gallie 2007a, 2009; Green et al. 2013; Holman 2013a). Further research should seek to extend the application of the bifactor IRT model to develop a measure of QWE at the country level and by the institutional regime. This could be based on the theoretical framework of QWE developed in this study and use data from the EWCS. The study could investigate how *overall QWE* and other *dimensions of QWE* compare between European countries and by institutional regime based on the *varieties of capitalism* approach. This approach places firms/organisations at the centre of understanding the capitalist economies, which differ in how organisations resolve their coordination problems (Gallie 2007b; Korpi 2006). The economies gravitate towards polar ends of the forms of coordination, from *coordinated market economies* (such as Denmark, Germany and Switzerland), *mixed* (or *Mediterranean*) *market economies* (such as France, Portugal and Spain) to *liberal market economies* (such as Ireland and the UK) (Hall and Soskice 2001; Hancké, Rhodes, and Thatcher 2007; Lallement 2011; Soskice 1999). In addition, DIF should be used to evaluate the measurement equivalence of the measurement instrument by country and institutional regime.

Second, further research should extend the bifactor IRT model to investigate changes in *overall QWE* and other *dimensions of QWE* over time. This could be a trend analysis using cross-sectional data to investigate whether there has been a decline, improvement or no change in the QWE of the workforce over time and inform labour market policy. Data from the EWCS, a cross-sectional survey conducted every five years, can be used to investigate the trends in European countries and institutional regimes. Notably, the EWCS has small sample sizes for individual countries. This may have implications for the feasibility of disaggregating the analysis by different groups within countries and IRT models with a high number of latent traits. However, the Bayes estimator can be used to conduct the analysis as this has superior performance when working with small samples of observed data than frequentist approaches (Muthén and Asparouhov 2012; Wang and Wang 2020). Furthermore, the UKHLS can also be used to conduct a trend analysis for the UK employee population using the cross-sectional samples of its data. While it is an annual panel survey, the UKHLS has two-year rational modules from Wave 2 (2010 – 2011) with items measuring QWE and trend analysis can be conducted at more frequent intervals than with the EWCS.

Third, a longitudinal analysis using the panel sample of the UKHLS could also be conducted to investigate changes in *overall QWE* and other *dimensions of QWE* over time among individual UK employees. This can help investigate how QWE changes for individual employees, for example, over the course of life or career events such as changes in health circumstances, returning to work from maternity leave, gaining new educational qualifications, a change in occupational classification, or the impact of economic shocks on the labour market such as the COVID-19 pandemic. The model structure can be replicated across different time points to extend the bifactor model for longitudinal analysis. Regarding model estimation, *overall QWE* and/or other *dimensions of QWE* would be specified to be uncorrelated within time points, while between time points, corresponding latent traits can be specified to be correlated. Furthermore, slope and intercept item parameters for corresponding items can be constrained to be equal between time points. On the other hand, *overall QWE* and other *dimensions of QWE* at the first time point would be assumed to have a mean of zero and a unit variance with the means and variances freely estimated at other time points to capture changes over time. The bifactor model with one time point in this study was computationally cumbersome to estimate with the MLR estimator, and a longitudinal analysis may not be feasible with this estimator. However, estimation with the Bayes estimator may be more feasible.

Lastly, further research should seek to apply the bifactor IRT model to measure QWE for a different target population, such as self-employed workers. These were excluded from this current study because of a lack of comparable attributes of QWE between employees and the self-employed in social survey data. For example, in the UKHLS, self-employed respondents were not asked survey questions related to pension provision, pay bonuses or annual increments, job security, formal or informal flexible working arrangements, or collective bargaining. Furthermore, for some survey questions asked of employees and the self-employed, the attributes may have different implications in measuring QWE. For example, while training prospects for employees captured the expectation of work-related training provided by the employer, for the self-employed, the responsibility rested with the respondent, as is the nature of self-employment. The scope of this research could repeat the research that has been proposed for employees using the UKHLS and the EWCS.

Viva Reflections

This study assumed that the measurement model for QWE was reflective in nature, that is, the observed indicators were dependent on QWE, as opposed to a formative measurement model, where QWE would be assumed to be dependent on the observed indicators (Coltman et al. 2008; Rose et al. 2023). Further research should consider whether QWE is better represented by a formative measurement model. Theoretically, a formative measurement model is also characterised by the latent trait being dependent on a constructivist interpretation, with a change in the observed indicators likely to result in a change in the latent trait score (Borsboom et al. 2003). Furthermore, since the observed indicators define the latent trait, the latent trait is sensitive to the number and type of indicators included in the measurement model (Coltman et al. 2008; Rose et al. 2023). On the other hand, the latent trait in a reflective measurement model exists independently of the observed indicators (Borsboom et al. 2004; Rossiter 2002). Additionally, a change in the latent trait must precede variation in the observed indicators and the inclusion or exclusion of an indicator does not affect the content validity of the latent trait (Coltman et al. 2008; Rose et al. 2023). Empirical considerations contrasting between reflective and formative measurement models include the intercorrelation between observed indicators. While high positive intercorrelations between observed indicators are desirable in reflective models, observed indicators can have any pattern in formative models, although they should have the same directional relationship (Coltman et al. 2008; Rose et al. 2023).

Alongside considering a formative measurement model for QWE, further analysis can explore the validity of the latent traits. This can be achieved by splitting the sample into two equal random samples and conducting an exploratory factor analysis (EFA) with different numbers of latent traits using the first sample to explore the dimensionality of the data. The second sample can then be used to estimate a confirmatory model based on the dimensionality suggested by the EFA and evaluate model fit.

Due to the computational challenges in model estimation, partly attributable to the number of latent variables in the measurement model, and particularly with the MLR estimator, interaction terms between predictors of QWE were not considered. Including interaction terms in the structural model can help facilitate the investigation of intersectionality in the labour market between different groups in society. For example, the intersectionality between gender and marital status, with research on this topic often framed as a marriage premium for males (Bardasi and Taylor 2008; Ribar 2004; Schoeni 1995) and a marriage penalty for females (Ribar 2004), as well as gender and the presence of children with lone parents being predominantly female (Esser and Olsen 2018; Klett-Davies 2016; Nieuwenhuis and Maldonado 2018). Future research can consider simplifying the model estimation to enable the introduction of interaction terms to the model less cumbersome. This can be done, for example, by saving the measurement model and then adding the structural model, or using the factor scores as the dependent variables, although this would introduce some measurement error.

This current research defined parental status based on having primary school age children in the household. However, this did not distinguish between parents with pre-primary school age children, who require more parental care, as well as those with post-primary school age children who were still dependent. Future research should seek derive a parental status variable that distinguishes between parents with pre-primary school age, primary school age, post-primary school age children, and those with no dependent children.

Future research can consider whether education should be classified as a socio-economic or socio-demographic characteristic. Other socio-economic characteristics examined in this study, such as full or part-time employment, organisational sector, organisation size, and occupational classification, were directly linked to work and employment, compared to education being a characteristic of an individual. While empirical evidence suggested that education was associated with greater job resources (Solomon et al. 2022) with specific educational attainment required for jobs, evidence also indicated that some jobs are not linked to education. For example, evidence suggested that the expansion of access to higher education resulted in the over-supply and underemployment of university graduates in the labour market, leading to graduates being employed in non-graduate occupations (Green and Zhu 2010; Okay-Somerville and Scholarios 2013; Warhurst 2008).

In terms of pay, this study used respondents' effective gross pay which was their hourly pay relative to the NMW or NLW. However, by definition, the NMW and NLW are age dependent measures, and an important criticism of these is that younger workers get lower pay than older workers for the same work. Future research can consider relative comparison of respondents' absolute pay categorised, for example, as quartile distributions or relative to a percentage of the median pay

The study used a sequential regression approach to introduce a set of predictors into the model. Demographic characteristics were introduced first, followed by the set of socio-demographic, and socio-economic characteristics. This provided information to evaluate, for example, what socio-economic characteristics added to the prediction of overall and other dimensions of QWE over and above demographic, and socio-demographic characteristics. Future research should consider, in more detail, the effects of additional predictors on the relationship between overall and other dimensions of QWE, and other predictors. For example, while the introduction of socio-economic characteristics resulted in differences in overall QWE between 16 – 24 and 25 – 34 year-olds being statistically insignificant, this also resulted in the difference between 16 – 24 and 65 + years old being statistically significant.

Lastly, for the *progression prospects* item, the valid response options (1 'yes', 2 'no', and 3 'does not apply') could not be presented in an ordinal level of measurement (Table 3.2). Respondents who selected 'does not apply' were those that were at the top tier of their jobs and in this study were recoded as having no further progression prospects as they could not progress any further in their roles. However, it can also be argued that they had exceeded lower ranking job roles and therefore could be recoded as having had progression prospects in their job. As this response option was a valid response, recoding it as either 'yes' or 'or' introduces a processing error and future research can consider treating the 'does not apply' response as missing.

8.6 Conclusion

As highlighted by Marx, work is fundamental to our humanity (Warren 2016), and the fact that workers spend substantially more of their adult lives on work activities than any other activity, except for sleep (Sinclair et al. 2020), elevates the QWE of workers in the social agenda. However, in contrast to job quantity, which forms part of the standard analysis of the labour market, the lack of consensus on a definition and inherent challenges in its conceptualisation and operationalisation have hindered QWE from being a salient social and labour market policy issue. This is highlighted by the substantial number of measurement instruments in the literature. The study contributed to the conceptualisation of QWE by developing a theoretical framework for measuring QWE. It also made methodological contributions which highlighted longstanding issues related to survey instruments measuring attributes of QWE and presented suggestions for improving the survey instruments. The study also suggested developing item banks of attributes capturing different dimensions of QWE, as in other disciplines in social sciences that experienced practical and technical issues associated with measurement. Furthermore, the study proposed IRT modelling as a method to develop a measurement instrument of QWE that addresses the shortcomings of existing measurement instruments. These shortcomings related to how items measuring QWE should be aggregated and weighted, or whether to report *overall* and/or other *dimensions of QWE*, as well as evaluating measurement equivalence of the instrument, and the study presented new knowledge that suggested the measurement of QWE was better modelled by a bifactor IRT model. The study also made some substantive contributions which suggested that socio-economic characteristics explained more of the variation in *overall* or other *dimensions of QWE*. In contrast, demographic or socio-demographic characteristics did not explain much of the variation in the latent traits. Additionally, findings suggested that IRT modelling largely replicated other methods of measuring QWE. However, some results were inconsistent with previous literature, and the bifactor IRT model provided a more nuanced understanding of differences in QWE between some groups. This study was limited to a UK employee population, and further research should seek to replicate the measurement of QWE using a bifactor IRT model with data from a different time, place and target population. This would provide robust methodology in the measurement of QWE and enable standard analyses of the labour market to look beyond job quantity but also focus on QWE and inform policy for the betterment of the labour market experiences of workers.

References

Adams, Raymond J., Mark Wilson, and Wen-chung Wang. 1997. 'The Multidimensional Random Coefficients Multinomial Logit Model'. *Applied Psychological Measurement* Vol.21(1):1–23.

Adler, Marina A. 1993. 'Gender Differences in Job Autonomy: The Consequences of Occupational Segregation and Authority Position'. *The Sociological Quarterly* Vol. 34(3):449–65.

Agresti, Alan, and Barbara Finlay. 2014. *Statistical Methods for the Social Sciences*. Fourth Edition. Harlow, Essex: Pearson Education Limited.

Angoff, William H. 2009. 'Perspectives of Differential Item Functioning Methodology'. Pp. 3–23 in *Differential Item Functioning*, edited by P. W. Holland and H. Wainer. New York, NY: Routledge.

Anker, Richard, Igor Chernyshev, Philippe Egger, Farhad Mehran, and Joseph A. Ritter. 2003. 'Measuring Decent Work with Statistical Indicators'. *International Labour Review* Vol. 142(2):147–78.

Arends, Iris, Christopher Prinz, and Femke Abma. 2017. *Job Quality, Health and at-Work Productivity*. *OECD Social, Employment and Migration Working Papers*. No. 195. Paris: OECD Publishing.

Arranz, José M., Carlos García-Serrano, and Virginia Hernanz. 2019. 'Job Quality Differences among Younger and Older Workers in Europe: The Role of Institutions'. *Social Science Research* Vol. 84:1–17.

Asparouhov, Tihomir, and Bengt O. Muthén. 2021. *Bayesian Analysis of Latent Variable Models Using Mplus*. *Technical Report*. Version 5. Los Angeles, CA.

Baldwin, Scott A. 2019. *Psychological Statistics and Psychometrics Using Stata*. College Station, TX: Stata Press.

Bardasi, Elena, and Mark Taylor. 2008. 'Marriage and Wages: A Test of the Specialisation Hypothesis'. *Economica* Vol. 75:569–91.

Bartholomew, David J., Fiona Steele, Irini Moustaki, and Jane I. Galbraith. 2008. *Analysis of Multivariate Social Science Data*. Second Edition. Boca Raton, FL: Chapman & Hall / CRC Press.

Bartholomew, David, Martin Knott, and Irini Moustaki. 2011. *Latent Variable Models and Factor Analysis: A Unified Approach*. Third Edition. Chichester, West Sussex: John Wiley & Sons Ltd.

Bartolucci, Francesco, Silvia Bacci, and Michela Gnaldi. 2016. *Statistical Analysis of Questionnaires: A Unified Approach Based on R and Stata*. Boca Raton, FL: CRC Press.

Bell, David N. F., Robert A. Hart, Olaf Hübler, and Wolfgang Schwerdt. 2000. *Paid and Unpaid Overtime Working in Germany and the UK*. *IZA Discussion Papers*. No. 133. Bonn, Germany: Institute for the Study of Labour (IZA).

Bescond, David, Anne Chataignier, and Farhad Mehran. 2003. 'Seven Indicators to Measure Decent Work: An International Comparison'. *International Labour Review* Vol. 142(2):179–211.

Birnbaum, Allan. 2008. 'Some Latent Trait Models and Their Use in Inferring an Examinee's Ability'. Pp. 395–479 in *Statistical Theories of Mental Test Scores*, edited by F. M. Lord and M. R. Novick. Charlotte, NC: Information Age Publishing Inc.

Bock, R. Darrell. 2010. 'The Nominal Categories Model'. Pp. 33–49 in *Handbook of Modern Item Response Theory*, edited by W. J. van der Linden and R. K. Hambleton. New York, NY: Springer.

Bock, R. Darrell, and Robert D. Gibbons. 2021. *Item Response Theory*. Hoboken, NJ: John Wiley & Sons, Inc.

Bock, R. Darrell, and Michele F. Zimowski. 2010. 'Multiple Group IRT'. Pp. 433–48 in *Handbook of Modern Item Response Theory*, edited by W. J. van der Linden and R. K. Hambleton. New York, NY: Springer.

Bonifay, Wes. 2020. *Multidimensional Item Response Theory*. Thousand Oaks, CA: SAGE Publications, Inc.

Bonifay, Wes, and Li Cai. 2017. 'On the Complexity of Item Response Theory Models'. *Multivariate Behavioral Research* Vol. 52(4):465–84.

Bonnet, Florence, José B. Figueiredo, and Guy Standing. 2003. 'A Family of Decent Work Indexes'. *International Labour Review* Vol. 142(2):213–38.

Bosch, Gerhard. 2009. 'Low-Wage Work in Five European Countries and the United States'. *International Labour Review* Vol. 148(4):337–56.

Brown, Andrew, Andy Charlwood, and David A. Spencer. 2012. 'Not All That It Might Seem: Why Job Satisfaction Is Worth Studying despite It Being a Poor Summary Measure of Job Quality'. *Work, Employment and Society* Vol. 26(6):1007–18.

Brown, Anna, and Tim J. Croudace. 2015. 'Scoring and Estimating Score Precision Using Multidimensional IRT Models'. Pp. 307–33 in *Handbook of Item Response Theory Modelling: Applications to Typical Performance Assessment*, edited by S. P. Reise and D. A. Revicki. New York, NY: Routledge.

Bryan, Mark L., and Almudena Sevilla. 2017. 'Flexible Working in the UK and Its Impact on Couples' Time Coordination'. *Review of Economics of the Household* Vol. 15(4):1415–37.

Bryant, Fred B., and Albert Satorra. 2012. 'Principles and Practice of Scaled Difference Chi-Square Testing'. *Structural Equation Modelling* 19:372–98.

Bryson, Alex, Christine Erhel, and Zinaida Salibekyan. 2021. 'Perceptions of Non-Pecuniary Job Quality Using Linked Employer-Employee Data'. *European Journal of Industrial Relations* Vol. 27(2):113–29.

Buchy, Lisa, Yasser Ad-Dab'bagh, Ashok Malla, Claude Lepage, Michael Bodnar, Ridha Joober, Karine Sergerie, Alan Evans, and Martin Lepage. 2011. 'Cortical Thickness Is Associated with Poor Insight in First-Episode Psychosis'. *Journal of Psychiatric Research* Vol. 45(6):781–87.

Burchell, Brendan, Kirsten Sehnbruch, Agnieszka Piasna, and Nurjk Agloni. 2014. 'The Quality of Employment and Decent Work: Definitions, Methodologies, and Ongoing Debates'. *Cambridge Journal of Economics* Vol. 38:459–77.

Cai, Li. 2016. 'Two-Tier Item Factor Analysis Modelling'. Pp. 421–35 in *Handbook of Item Response Theory, Volume One: Models*, edited by W. J. van der Linden. Boca Raton, FL: CRC Press.

Cai, Li, Albert Maydeu-Olivares, Donna L. Coffman, and David Thissen. 2006. 'Limited-Information Goodness-of-Fit Testing of Item Response Theory Models for Sparse 2p Tables'. *British Journal of Mathematical and Statistical Psychology* Vol. 59:173–94.

Cai, Li, Ji Seung Yang, and Mark Hansen. 2011. 'Generalised Full-Information Item Bifactor Analysis'. *Psychological Methods* Vol. 16(3):221–48.

Canduela, Jesus, Matthew Dutton, Steve Johnson, Colin Lindsay, Ronald W. McQuaid, and Robert Raeside. 2012. 'Ageing, Skills and Participation in Work-Related Training in Britain: Assessing the Position of Older Workers'. *Work, Employment and Society* Vol. 26(1):42–60.

Carpenter, Hannah. 2018. *UK Household Longitudinal Study: Wave 8 Technical Report*. JN 260131869. Colchester, Essex: The Institute for Social and Economic Research, University of Essex.

Carrié, Françoise, Patricia Findlay, Chris Tilly, and Chris Warhurst. 2012. 'Job Quality: Scenarios, Analysis and Interventions'. Pp. 1–22 in *Are Bad Jobs Inevitable? Trends, Determinants and Responses to Job Quality in the Twenty-First Century*, edited by C. Warhurst, F. Carré, P. Findlay, and C. Tilly. Basingstoke, Hampshire: Palgrave Macmillan.

Cascales Mira, María. 2021. 'New Model for Measuring Job Quality: Developing an European Intrinsic Job Quality Index (EIJQI)'. *Social Indicators Research* Vol. 155:625–45.

Cazes, Sandrine, Alexander Hijzen, and Anne Saint-Martin. 2015. *Measuring and Assessing Job Quality: The OECD Job Quality Framework*. *OECD Social, Employment and Migration Working Papers*. No. 174. Paris: OECD Publishing.

Cella, David, William Riley, Arthur Stone, Nan Rothrock, Bryce Reeve, Susan Yount, Dagmar Amtmann, Rita Bode, Daniel Buysse, Seung Choi, Karon Cook, Robert DeVellis, Darren DeWalt, James F. Fries, Richard C. Gershon, Elizabeth A. Hahn, Jin-Shei Lai, Paul Pilkonis, Dennis A. Revicki, Matthias Rose, Kevin Weinfurt, and Ron D. Hays. 2010. 'The Patient-Reported Outcomes Measurement Information System (PROMIS) Developed and Tested Its First Wave of Adult Self-Reported Health Outcome Item Banks: 2005 - 2008'. *Journal of Clinical Epidemiology* Vol. 63(11):1179–94.

Chalmers, Philip R. 2012. 'Mirt: A Multidimensional Item Response Theory Package for the R Environment'. *Journal of Statistical Software* Vol. 48(6):1–29.

Charlwood, Andy, and Mike Terry. 2007. '21st-Century Models of Employee Representation: Structures, Processes and Outcomes'. *Industrial Relations Journal* Vol. 38(4):320–37.

Chen, Fang Fang, Adele Hayes, Charles S. Carver, Jean-Philippe Laurenceau, and Zugui Zhang. 2012. 'Modelling General and Specific Variance in Multifaceted Constructs: A Comparison of the Bifactor Model to Other Approaches'. *Journal of Personality* Vol. 80(1):219–51.

Chen, Fang Fang, Stephen G. West, and Karen H. Sousa. 2006. 'A Comparison of Bifactor and Second-Order Models of Quality of Life'. *Multivariate Behavioral Research* Vol. 14(2):189–225.

Chen, Fang Fang, and Zugui Zhang. 2018. 'Bifactor Models in Psychometric Test Development'. Pp. 325–45 in *The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development: Volume One*, edited by P. Irwing, T. Booth, and D. J. Hughes. Chichester, West Sussex: John Wiley & Sons Ltd.

Chen, Wen-Hung, and David Thissen. 1997. ‘Local Dependence Indexes for Item Pairs Using Item Response Theory’. *Journal of Educational and Behavioural Statistics* Vol. 22(3):265–89.

Choi, Seung W. 2021. ‘Differential Item Functioning for Multidimensional Constructs (E-Mail Correspondence)’.

Choi, Seung W., Laura E. Gibbons, and Paul K. Crane. 2011. ‘Lordif: An R Package for Detecting Differential Item Functioning Using Iterative Hybrid Ordinal Logistic Regression / Item Response Theory and Monte Carlo Simulations’. *Journal of Statistical Software* Vol. 39(8):1–30.

Chung, Heejung, and Mariska van der Horst. 2020. ‘Flexible Working and Unpaid Overtime in the UK: The Role of Gender, Parental and Occupational Status’. *Social Indicators Research* Vol. 151(2):495–520.

Clark, Ken, Nico Ochmann, David N. F. Bell, and Wolfgang Schwerdt. 2022. *Good Job, Bad Job, No Job? Ethnicity and Employment Quality for Men in the UK*. *IZA Discussion Papers*. No. 15099. Bonn, Germany: Institute of Labour Economics (IZA).

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioural Sciences*. Second Edition. Hillsdale, NJ: Lawrence Erlbaum Associates.

Collins, Linda M., and Stephanie T. Lanza. 2010. *Latent Class and Latent Transition Analysis with Applications in the Social, Behavioural, and Health Sciences*. Hoboken, NJ: John Wiley & Sons, Inc.

Cribb, Jonathan, and Carl Emmerson. 2014. *Workplace Pensions and Remuneration in the Public and Private Sectors in the UK*. *IFS Report*. Briefing Note 151. London: Institute for Fiscal Studies.

Cribb, Jonathan, Carl Emmerson, and Luke Sibieta. 2014. *Public Sector Pay in the UK*. *IFS Report*. No. R97. London: Institute for Fiscal Studies.

Davidson, Jacqueline, and Peter A. Kemp. 2008. ‘Sickness Benefits in a Polarised Labour Market’. *Benefits* Vol. 16(3):225–33.

von Davier, Matthias. 2016. ‘Rasch Model’. Pp. 31–48 in *Handbook of Item Response Theory, Volume One: Models*, edited by W. J. van der Linden. Boca Raton, FL: CRC Press.

De Boeck, Paul. 2016. ‘Psychological Testing’. Pp. 313–28 in *Handbook of Item Response Theory, Volume Three: Applications*, edited by W. J. van der Linden. Boca Raton, FL: CRC Press.

De Boeck, Paul, and Mark R. Wilson. 2016. 'Explanatory Response Models'. Pp. 565–80 in *Handbook of Item Response Theory, Volume One: Models*, edited by W. J. van der Linden. Boca Raton, FL: CRC Press.

DeMars, Christine E. 2013. 'A Tutorial on Interpreting Bifactor Model Scores'. *International Journal of Testing* Vol. 13(4):354–78.

Department for BEIS. 2018. *Good Work: A Response to the Taylor Review of Modern Working Practices*. London: The Stationery Office.

Department for BEIS. 2021. 'National Minimum Wage and National Living Wage Rates'. *GOV.UK*. Retrieved 29 December 2021 (https://www.gov.uk/national-minimum-wage-rates).

Department for BEIS. 2022. *Companies House Accounts Guidance*. London: The Stationery Office.

Department for BEIS. n.d. 'Part-Time Workers' Rights'. *GOV.UK*. Retrieved 29 December 2021 (https://www.gov.uk/part-time-worker-rights).

Department for LUHC. 2022. *Levelling Up the United Kingdom*. Command Paper 604. London: Her Majesty's Stationery Office.

Desjardins, Christopher D., and Okan Bulut. 2018. *Handbook of Educational Measurement and Psychometrics Using R*. Boca Raton, FL: CRC Press.

Dieckhoff, Martina, Jean-Marie Jungblut, and Philip J. O'Connell. 2007. 'Job-Related Training in Europe: Do Institutions Matter?' Pp. 77–103 in *Employment Regimes and the Quality of Work*, edited by D. Gallie. Oxford: Oxford University Press.

Dillon, Michele. 2020. *Introduction to Sociological Theory: Theorists, Concepts, and Their Applicability to the Twenty-First Century*. Third Edition. Hoboken, New Jersey: John Wiley & Sons Ltd.

Drasgow, Fritz, Michael V. Levine, and Esther A. Williams. 1985. 'Appropriateness Measurement with Polychotomous Item Response Models and Standardised Indices'. *British Journal of Mathematical and Statistical Psychology* Vol. 38(1):67–86.

Dueber, David. 2021. 'BifactorIndicesCalculator: Bifactor Indices Calculator'. Retrieved 29 December 2021 (https://CRAN.R-project.org/package=BifactorIndicesCalculator).

Edgell, Stephen, and Edward Granter. 2020. *The Sociology of Work: Continuity and Change in Paid and Unpaid Work*. Third Edition. London: SAGE Publications Ltd.

Edwards, Michael C., R. J. Wirth, Carrie R. Houts, and Andrew J. Bodine. 2015. 'Three (or Four) Factors, Four (or Three) Models'. Pp. 159–82 in *Handbook of Item Response Theory Modelling: Applications to Typical Performance Assessment*, edited by S. P. Reise and D. A. Revicki. New York, NY: Routledge.

Erhel, Christine, and Mathilde Guergoat-Larivière. 2016. *Innovations and Job Quality Regimes: A Joint Typology for the EU. QuInnE Working Paper*. No. 3. Luxembourg: Office for Official Publications of the European Communities.

Esping-Andersen, Gøsta. 1990. *The Three Worlds of Welfare Capitalism*. Cambridge: Polity Press.

Esser, Ingrid, and Karen M. Olsen. 2012. 'Perceived Job Quality: Autonomy and Job Security within a Multi-Level Framework'. *European Sociological Review* Vol. 28(4):443–54.

Esser, Ingrid, and Karen M. Olsen. 2018. 'Matched on Job Qualities? Single and Coupled Parents in the European Comparison'. Pp. 285–310 in *The Triple Bind of Single-Parent Families: Resources, Employment and Policies to Improve Well-being*, edited by R. Nieuwenhuis and L. C. Maldonado. Bristol: Policy Press.

Eurofound. 2007. *Fifteen Years of Working Conditions in the EU: Charting the Trends*. Dublin, Ireland: European Foundation for the Improvement of Living and Working Conditions.

Eurofound. 2012. *Trends in Job Quality in Europe*. Luxembourg: Publications Office of the European Union.

Eurofound. 2016a. *6th European Working Conditions Survey – Sampling Implementation Report*. Luxembourg: Publications Office of the European Union.

Eurofound. 2016b. *6th European Working Conditions Survey – Technical Report*. Luxembourg: Publications Office of the European Union.

Eurofound. 2017a. *European Working Conditions Survey, 2015*. 4th Edition. Colchester, Essex: UK Data Service.

Eurofound. 2017b. *Sixth European Working Conditions Survey – Overview Report (2017 Update)*. Luxembourg: Publications Office of the European Union.

European Commission. 2008. *Employment in Europe 2008*. Luxembourg: Office for Official Publications of the European Communities.

European Commission. 2018. *Quality of Jobs and Innovation Generated Employment Outcomes (QuInnE) Project: Periodic Technical Report*. Luxembourg: Office for Official Publications of the European Communities.

European Parliament. 2010. *The Lisbon Strategy 2000 - 2010*.

European Parliament. 2014. *Briefing Note for the Meeting of The...*

Felstead, Alan, Duncan Gallie, Francis Green, and Golo Henseke. 2019. 'Conceiving, Designing and Trailing a Short-Form Measure of Job Quality: A Proof-of-Concept Study'. *Industrial Relations Journal* Vol. 50(1):2–19.

Field, Andy, Jeremy Miles, and Zoë Field. 2012. *Discovering Statistics Using R*. London: SAGE Publications Ltd.

Finch, W. Holmes, and Jocelyn E. Bolin. 2017. *Multilevel Modelling Using Mplus*. Boca Raton, FL: CRC Press.

Finch, W. Holmes, Jr., and Brian F. French. 2015. *Latent Variable Modelling with R*. New York, NY: Routledge.

Findlay, Patricia, Arne L. Kalleberg, and Chris Warhurst. 2013. 'The Challenge of Job Quality'. *Human Relations* Vol. 66(4):441–51.

Fontaine, Idriss, Ismael Gálvez-Iniesta, Pedro Gomes, and Diego Vila-Martin. 2020. 'Labour Market Flows: Accounting for the Public Sector'. *Labour Economics* Vol. 62.

Forth, John, Helen Bewley, and Alex Bryson. 2006. *Small and Medium-Sized Enterprises: Findings from the 2004 Workplace Employment Relations Survey*. London: Department of Trade and Industry.

Fox, Jean-Paul. 2010. *Bayesian Item Response Modelling: Theory and Applications*. New York, NY: Springer.

Fredman, Sandra. 2004. 'Women at Work: The Broken Promise of Flexicurity'. *Industrial Law Journal* Vol. 33(4):299–319.

Fumagalli, Laura, Gundi Knies, and Nick Buck. 2017. *Understanding Society: The UK Household Longitudinal Study Harmonised British Household Panel Survey (BHPS) User Guide*. Colchester, Essex: University of Essex.

Gallie, Duncan. 2007a. 'Production Regimes and the Quality of Employment in Europe'. *Annual Review of Sociology* Vol. 33:85–104.

Gallie, Duncan. 2007b. 'Production Regimes, Employment Regimes, and the Quality of Work'. Pp. 1–34 in *Employment Regimes and the Quality of Work*, edited by D. Gallie. Oxford: Oxford University Press.

Gallie, Duncan. 2009. 'Institutional Regimes and Employee Influence at Work: A European Comparison'. *Cambridge Journal of Regions, Economy and Society* Vol. 2:379–93.

Gallie, Duncan. 2015. 'Class Inequality at Work: Trends to Polarisation?' Pp. 22–41 in *Unequal Britain at Work*, edited by A. Felstead, D. Gallie, and F. Green. Oxford: Oxford University Press.

Gallie, Duncan, and Ying Zhou. 2013. 'Job Control, Work Intensity, and Work Stress'. Pp. 115–41 in *Economic Crisis, Quality of Work, and Social Integration: The European Experience*, edited by D. Gallie. Oxford, UK: Oxford University Press.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. 2013. *Bayesian Data Analysis*. Third Edition. Boca Raton, FL: CRC Press.

Gershon, Richard C., Ron D. Hays, and Michael A. Kallen. 2016. 'Health Measurement'. Pp. 349–63 in *Handbook of Item Response Theory, Volume Three: Applications*, edited by W. J. van der Linden. Boca Raton, FL: CRC Press.

Ghai, Dharam. 2003. 'Decent Work: Concept and Indicators'. *International Labour Review* Vol. 142(2):113–45.

Gibbons, Robert D., R. Darrell Bock, Donald Hedeker, David J. Weiss, Eisuke Segawa, Dulal K. Bhaumik, David J. Kupfer, Ellen Frank, Victoria J. Grochocinski, and Angela Stover. 2007. 'Full-Information Item Bifactor Analysis of Graded Response Data'. *Applied Psychological Measurement* Vol. 31(1):4–19.

Gifford, Jonny. 2018. *UK Working Lives: The CIPD Job Quality Index - Survey Report 2018*. *Research Report*. London: Chartered Institute of Personnel and Development.

Glas, Cees A. W. 2016. 'Maximum-Likelihood Estimation'. Pp. 197–216 in *Handbook of Item Response Theory, Volume Two: Statistical Tools*, edited by W. J. van der Linden. Boca Raton, FL: CRC Press.

Gorsuch, Richard L. 1983. *Factor Analysis*. Second Edition. Hillsdale, NJ: Erlbaum.

Green, Francis. 2006. *Demanding Work: The Paradox of Job Quality in the Affluent Economy*. Princeton, NJ: Princeton University Press.

Green, Francis, Tarek Mostafa, Agnès Parent-Thirion, Greet Vermeylen, Gijs van Houten, Isabella Biletta, and Maija Lyly-Yrjanainen. 2013. 'Is Job Quality Becoming More Unequal?' *Industrial and Labour Relations Review* Vol. 66(4):753–84.

Green, Francis, and Yu Zhu. 2010. 'Overqualification, Job Dissatisfaction, and Increasing Dispersion in the Returns of Graduate Education'. *Oxford Economic Papers* Vol. 62(4):740–63.

Grint, Keith, and Darren Nixon. 2015. *The Sociology of Work*. Fourth Edition. Cambridge: Polity Press.

Grover, Chris, and Linda Piggott. 2015. 'Disabled People, Work and Welfare'. Pp. 1–21 in *Disabled People, Work and Welfare: Is Employment Really the Answer?*, edited by C. Grover and L. Piggott. Bristol: Policy Press.

Gutiérrez-Galve, Leticia, Claudia A. M. Wheeler-Kingshott, Daniel R. Altmann, Gary Price, Elvina M. Chu, Verity C. Leeson, Antonio Lobo, Gareth J. Barker, Thomas R. E. Barnes, Eileen M. Joyce, and María A. Ron. 2010. 'Changes in the Frontotemporal Cortex and Cognitive Correlates in First-Episode Psychosis'. *Biological Psychiatry* Vol. 68(1):51–60.

Haberman, Shelby J. 2016. 'Models with Nuisance and Incidental Parameters'. Pp. 151–70 in *Handbook of Item Response Theory, Volume Two: Statistical Tools*, edited by W. J. van der Linden. Boca Raton, FL: CRC Press.

Hall, Peter A., and David Soskice. 2001. 'An Introduction to Varieties of Capitalism'. Pp. 1–68 in *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage*, edited by P. A. Hall and D. Soskice. Oxford: Oxford University Press.

Hambleton, Ronald K., Frederic Robin, and Dehui Xing. 2000. 'Item Response Models for the Analysis of Education and Psychological Test Data'. Pp. 553–81 in *Handbook of Applied Multivariate Statistics and Mathematical Modelling*, edited by H. E. A. Tinsley and S. D. Brown. San Diego, CA: Academic Press.

Hancké, Bob, Martin Rhodes, and Mark Thatcher. 2007. 'Introduction: Beyond Varieties of Capitalism'. Pp. 3–38 in *Beyond Varieties of Capitalism: Conflict, Contradictions, and Complementaries in the European Economy*, edited by B. Hancké, M. Rhodes, and M. Thatcher. Oxford: Oxford University Press.

van Haren, Neeltje E. M., Hugo G. Schnack, Wiepke Cahn, Martijn P. van den Heuvel, Claude Lepage, Louis Collins, Alan C. Evans, Hilleke E. Hulshoff Pol, and René S. Kahn. 2011. ‘Changes in Cortical Thickness During the Course of Illness in Schizophrenia’. *Archives of General Psychiatry* Vol. 68(9):871–80.

Heeringa, Steven G., Brady T. West, and Patricia A. Berglund. 2017. *Applied Survey Data Analysis*. Second Edition. Boca Raton, FL: CRC Press.

Heinen, Ton. 1996. *Latent Class and Discrete Latent Trait Models: Similarities and Differences*. Thousand Oaks, CA: SAGE Publications.

Hepworth, Mark, James Binks, and Britta Ziemann. 2005. *Regional Employment and Skills in the Knowledge Economy: A Report for the Department of Trade and Industry*. London: The Stationery Office.

HM Treasury. 2008. *Annual Report 2007 - 2008*. Cm 7408. London: The Stationery Office Limited.

Holland, Paul W., and Howard Wainer. 2009. ‘Preface’. Pp. xiii–xv in *Differential Item Functioning*, edited by P. W. Holland and H. Wainer. New York, NY: Routledge.

Holler, Markus. 2013. *Methodenbericht Zur Weiterentwicklung Des DGB-Index Gute Arbeit in Der Erhebungsperiode 2011/2012. [Methods Report on the Development of the “Good Work“ Index by the German Trade Union Confederation in the Field Period 2011/2012]*. Stadtbergen, Germany: INIFES, Internationales Institut für Empirische Sozialökonomie gGmbH.

Holman, David. 2013a. ‘An Explanation of Cross-National Variation in Call Centre Job Quality Using Institutional Theory’. *Work, Employment and Society* Vol. 27(1):21–38.

Holman, David. 2013b. ‘Job Types and Job Quality in Europe’. *Human Relations* Vol. 66(4):475–502.

Hoque, Kim, and Ian Kirkpatrick. 2003. ‘Non-Standard Employment in the Management and Professional Workforce: Training, Consultation and Gender Implications’. *Work, Employment and Society* Vol. 17(4):667–89.

Howell, Chris. 2003. ‘Varieties of Capitalism: And Then There Was One?’ *Comparative Politics* Vol. 36(1):103–24.

Hu, Li-tze, and Peter M. Bentler. 1999. ‘Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives’. *Structural Equation Modelling: A Multidisciplinary Journal* Vol. 6(1):1–55.

ILO. 1999. *Decent Work*.

Irvine, Gail, Douglas White, and Mark Diffley. 2018. *Measuring Good Work: The Final Report of the Measuring Job Quality Working Group*. *Report*. Dunfermline, Scotland: Carnegie UK Trust.

Johnson, Matthew S., and Sandip Sinharay. 2016. 'Bayesian Estimation'. Pp. 237–57 in *Handbook of Item Response Theory, Volume Two: Statistical Tools*, edited by W. J. van der Linden. Boca Raton, FL: CRC Press.

Jones, Paul S., and Anne E. Green. 2009. 'The Quantity and Quality of Jobs: Changes in UK Regions, 1997 - 2007'. *Environment and Planning A: Economy and Space* Vol. 41(10):2474–95.

Jung, Wi Hoon, June Sic Kim, Joon Hwan Jang, Jung-Seok Choi, Myung Hun Jung, Ji-Young Park, Ji Yeon Han, Chi-Hoon Choi, Do-Hyung Kang, Chun Kee Chung, and Jun Soo Kwon. 2011. 'Cortical Thickness Reduction in Individuals at Ultra-High-Risk for Psychosis'. *Schizophrenia Bulletin* Vol. 37(4):839–49.

Junker, Brian W., Richard J. Patz, and Nathan VanHoudnos. 2016. 'Markov Chain Monte Carlo for Item Response Models'. Pp. 271–312 in *Handbook of Item Response Theory, Volume Two: Statistical Tools*, edited by W. J. van der Linden. Boca Raton, FL: CRC Press.

Kalleberg, Arne L. 2011. *Good Jobs, Bad Jobs: The Rise of Polarised and Precarious Employment Systems in the United States, 1970s to 2000s*. New York, NY: Russell Sage Foundation.

Kenway, P. 2008. *Eradicating Child Poverty: The Role of Key Policy Areas. Addressing in-Work Poverty*. York: Joseph Rowntree Foundation.

Kim, Anna, and Karin Kurz. 2001. *Precarious Employment, Education and Gender: A Comparison of Germany and the United Kingdom*. *IDEAS Working Paper*. No. 39. Mannheim, Germany: Mannheim Centre for European Social Research (MZES).

Kim, Seock-Ho, Allan S. Cohen, Cigdem Alagoz, and Sukwoo Kim. 2007. 'DIF Detection and Effect Size Measures for Polytomously Scored Items'. *Journal of Educational Measurement* Vol. 44(2):93–116.

King's College London, and Working Families. 2021. *Working Parents, Flexibility and Job Quality: What Are the Trade-Offs?* London: Working Families and Family-friendly Working Scotland.

Klett-Davies, Martina. 2016. *Under Pressure? Single Parents in the UK*. Gütersloh: Bertelsmann Stiftung.

Kline, Rex B. 2016. *Principles and Practice of Structural Equation Modelling*. Fourth Edition. New York, NY: The Guilford Press.

Knöchel, Christian, Johanna Reuter, Britta Reinke, Michael Stäblein, Katharina Marbach, Richard Feddern, Kristina Kuhlmann, Gilberto Alves, David Prvulovic, Sofia Wenzler, David E. J. Linden, and Viola Oertel-Knöchel. 2016. 'Cortical Thinning in Bipolar Disorder and Schizophrenia'. *Schizophrenia Research* Vol. 172(1):78–85.

Korpi, Walter. 1985. 'Power Resources Approach vs. Action and Conflict: On Causal and Intentional Explanations in the Study of Power'. *Sociological Theory* Vol. 3(2):31–45.

Korpi, Walter. 2006. 'Power Resources and Employer-Centred Approaches in Explanations of Welfare States and Varieties of Capitalism: Protagonists, Consenters, and Antagonists'. *World Politics* Vol. 58(2):167–206.

Korpi, Walter. 2018. 'The Iceberg of Power below the Surface: A Preface to Power Resources Theory'. Pp. vii–xiv in *Power Resource Theory and the Welfare State: A Critical Approach*, edited by J. S. O'Connor and G. M. Olsen. Toronto: University of Toronto Press.

Lakens, Daniël. 2013. 'Calculating and Reporting Effect Sizes to Facilitate Cumulative Science: A Practical Primer for t-Tests and ANOVAs'. *Frontiers in Psychology* Vol. 4(863):1–12.

Lallement, Michel. 2011. 'Europe and the Economic Crisis: Forms of Labour Adjustment and Varieties of Capitalism'. *Work, Employment and Society* Vol. 25(4):627–41.

Lawton, K. 2009. *Nice Work If You Can Get It: Achieving a Sustainable Solution to Low Pay and in-Work Poverty*. London: IPPR.

Leschke, Janine, and Maarten Keune. 2008. 'Precarious Employment in the Public and Private Sectors: Comparing the UK and Germany'. Pp. 197–231 in *Privatisation and Liberalisation of Public Services in Europe: An Analysis of Economic and Labour Market Impacts*, edited by M. Keune, J. Leschke, and A. Watt. Brussels: European Trade Union Institute.

Leschke, Janine, and Andrew Watt. 2014. 'Challenges in Constructing a Multi-Dimensional European Job Quality Index'. *Social Indicators Research* Vol. 118(1):1–31.

Leschke, Janine, Andrew Watt, and Mairéad Finn. 2008. *Putting a Number on Job Quality? Constructing a European Job Quality Index*. *Working Paper*. 2008.03. Brussels: European Trade Union Institute.

Leschke, Janine, Andrew Watt, and Mairéad Finn. 2012. *Job Quality in the Crisis - An Update of the Job Quality Index (JQI)*. *Working Paper*. 2012.07. Brussels: European Trade Union Institute.

Levy, Roy, and Robert J. Mislevy. 2016. *Bayesian Psychometric Modelling*. Boca Raton, FL: CRC Press.

Lewis, Ruth, Lynn Dobbs, and Paul Biddle. 2013. '"If This Wasn't Here I Probably Wouldn't Be": Disabled Workers' Views of Employment Support'. *Disability and Society* Vol. 28(8):1089–1103.

van der Linden, Wim J. 2016a. 'Introduction'. Pp. 1–10 in *Handbook of Item Response Theory, Volume One: Models*, edited by W. J. van der Linden. Boca Raton, FL: CRC Press.

van der Linden, Wim J. 2016b. 'Unidimensional Logistic Response Models'. Pp. 13–30 in *Handbook of Item Response Theory, Volume One: Models*, edited by W. J. van der Linden. Boca Raton, FL: CRC Press.

van der Linden, Wim J., and Ronald K. Hambleton. 2010. 'Item Response Theory: Brief History, Common Models, and Extensions'. Pp. 1–28 in *Handbook of Modern Item Response Theory*, edited by W. J. van der Linden and R. K. Hambleton. New York, NY: Springer.

Lindley, Joanne. 2015. 'Gender Differences in Job Quality'. Pp. 42–61 in *Unequal Britain at Work*, edited by A. Felstead, D. Gallie, and F. Green. Oxford: Oxford University Press.

Lindley, Joanne, and Stephen Machin. 2012. 'The Quest for More and More Education: Implications for Social Mobility'. *Fiscal Studies* Vol. 33(2):265–86.

Lord, Frederic M., and Melvin R. Novick. 2008. *Statistical Theories of Mental Test Scores*. Charlotte, NC: Information Age Publishing Inc.

Low Pay Commission. 2021. *National Minimum Wage: Low Pay Commission Report 2021*. London: The Stationery Office.

Lucifora, Claudio, and Dominique Meurs. 2006. 'The Public Sector Pay Gap in France, Great Britain and Italy'. *Review of Income and Wealth* Vol. 52(1):43–59.

Lynn, Peter. 2009. *Sample Design for Understanding Society*. *Working Paper*. 2009–01. Colchester, Essex: The Institute for Social and Economic Research, University of Essex.

Lyonette, Clare. 2015. 'Part-Time Work, Work-Life Balance and Gender Equality'. *Journal of Social Welfare and Family Law* Vol. 37(3):321–33.

Lyonette, Clare, Beate Baldauf, and Heike Behle. 2010. *'Quality' Part-Time Work: A Review of the Evidence*. London: Government Equalities Office.

Mair, Patrick. 2018. *Modern Psychometrics with R*. Cham, Switzerland: Springer.

Martinaitis, Žilvinas, Aleksandr Christenko, and Jonas Antanavičius. 2021. 'Upskilling, Deskilling or Polarisation? Evidence on Change in Skills in Europe'. *Work, Employment and Society* Vol. 35(3):451–69.

Marx, Karl. 2000. 'Economic and Philosophical Manuscripts'. Pp. 83–121 in *Karl Marx: Selected Writings*, edited by D. McLellan. Oxford: Oxford University Press.

Masters, Geofferey N., and Benjamin D. Wright. 2010. 'The Partial Credit Model'. Pp. 101–21 in *Handbook of Modern Item Response Theory*, edited by W. J. van der Linden and R. K. Hambleton. New York, NY: Springer.

Maydeu-Olivares, Albert, and Harry Joe. 2005. 'Limited- and Full-Information Estimation and Goodness-of-Fit Testing in 2n Contingency Tables: A Unified Framework'. *Journal of the American Statistical Association* Vol. 100(471):1009–20.

Maydeu-Olivares, Albert, and Harry Joe. 2006. 'Limited Information Goodness-of-Fit Testing in Multidimensional Contingency Tables'. *Psychometrika* Vol. 71(4):713–32.

Maydeu-Olivares, Alberto, and Harry Joe. 2014. 'Assessing Approximate Fit in Categorical Data Analysis'. *Multivariate Behavioural Research* Vol. 49(4):305–28.

McGovern, Patrick, Deborah Smeaton, and Stephen Hill. 2004. 'Bad Jobs in Britain: Nonstandard Employment and Job Quality'. *Work and Occupations* Vol. 31(2):225–49.

Meager, Nigel, and Darcy Hill. 2005. *The Labour Market Participation and Employment of Disabled People in the UK. IES Working Paper*. WP1. Brighton: Institute for Employment Studies.

Monnakgotla, Jabulile, and Morné Oosthuizen. 2021. *Job Quality in South Africa: A Proposed Index for Ongoing Monitoring of Job Quality. Working Paper*. WP202103. Cape Town, SA: Development Policy Research Unit, University of Cape Town.

Morizot, Julien, Andrew T. Ainsworth, and Steven P. Reise. 2007. 'Toward Modern Psychometrics: Application of Item Response Theory Models in Personality Research'. Pp. 407–23 in *Handbook of Research Methods in Personality Psychology*, edited by R. W. Robins, R. C. Fraley, and R. F. Krueger. New York, NY: The Guilford Press.

Muñoz de Bustillo, Rafael, Enrique Fernández-Macías, José-Ignacio Antón, and Fernando Esteve. 2009. *Indicators of Job Quality in the European Union*. Brussels: European Parliament, Department of Employment and Social Affairs.

Muñoz de Bustillo, Rafael, Enrique Fernández-Macías, José-Ignacio Antón, and Fernando Esteve. 2011a. 'E Pluribus Unum? A Critical Survey of Job Quality Indicators'. *Socio-Economic Review* Vol. 9(3):447–75.

Muñoz de Bustillo, Rafael, Enrique Fernández-Macías, José-Ignacio Antón, and Fernando Esteve. 2011b. *Measuring More than Money: The Social Economics of Job Quality*. Cheltenham, UK: Edward Elgar Publishing Limited.

Muraki, Eiji. 2010. 'A Generalised Partial Credit Model'. Pp. 153–64 in *Handbook of Modern Item Response Theory*, edited by W. J. van der Linden and R. K. Hambleton. New York, NY: Springer.

Murphy, Philip, David Blackaby, O'Leary, and Anita Staneva. 2020. 'Understanding What Has Been Happening to the Public-Sector Pay Premium in Great Britain: A Distributional Approach Based on the Labour Force Survey'. *British Journal of Industrial Relations* Vol. 58(2):273–300.

Mussmann, Frank. 2009.'The German "Good-Work" Index (DGB-Index Gute Arbeit)'. Presented at the European Trade Union Institute (ETUI) Conference 2009, 18-19 March, Brussels, Belgium.

Muthén, Bengt O., and Tihomir Asparouhov. 2012. 'Bayesian Structural Equation Modelling: A More Flexible Representation of Substantive Theory'. *Psychological Methods* Vol. 17(3):313–35.

Muthén, Linda K., and Bengt O. Muthén. 2017. 'Mplus: Statistical Analysis with Latent Variables: User's Guide'.

Nieuwenhuis, Rense, and Laurie C. Maldonado. 2018. 'The Triple Bind of Single-Parent Families: Resources, Employment and Policies'. Pp. 1–27 in *The Triple Bind of Single-Parent Families: Resources, Employment and Policies to Improve Well-being*, edited by R. Nieuwenhuis and L. C. Maldonado. Bristol: Policy Press.

OECD. 1999. *Measuring Student Knowledge and Skills: A New Framework for Assessment*. Paris: PISA, OECD Publishing.

OECD. 2013. *How Is Life? Measuring Well-Being*.

OECD. 2019. *PISA 2018 Assessment and Analytical Framework*. Paris: PISA, OECD Publishing.

OECD. 2022. *Part-Time Employment Rate (Indicator)*. Paris: OECD Publishing.

Office for National Statistics. 2022. *Labour Market Overview, UK: November 2022*. *Statistical Bulletin*. London: Office for National Statistics.

Okada, Kensuke. 2017. 'Negative Estimate of Variance-Accounted-for Effect Size: How Often It Is Obtained, and What Happens If It Is Treated as Zero'. *Behaviour Research Methods* Vol. 49:979–87.

Okay-Somerville, Belgin, and Dora Scholarios. 2013. 'Shades of Grey: Understanding Job Quality in Emerging Graduate Occupations'. *Human Relations* Vol. 66(4):555–85.

Olejnik, Stephen, and James Algina. 2003. 'Generalised Eta and Omega Squared Statistics: Measures of Effect Size for Some Common Research Designs'. *Psychological Methods* Vol. 8(4):434–47.

Olsen, Gregg M., and Julia S. O'Connor. 2018. 'Introduction Understanding the Welfare State: Power Resources Theory and Its Critics'. Pp. 3–33 in *Power Resource Theory and the Welfare State: A Critical Approach*, edited by J. S. O'Connor and G. M. Olsen. Toronto: University of Toronto Press.

ONS. 2016. *Labour Force Survey User Guide Volume 1 - LFS Background and Methodology 2016*. *User Guide*. London: Office for National Statistics.

ONS. 2019. *Job Quality Indicators in the UK - Hours, Pay and Contracts: 2018*. *Article*. London: Office for National Statistics.

ONS. 2020. *Public Sector Employment, UK: December 2019*. *Statistical Bulletin*. London: Office for National Statistics.

ONS. 2022. *Job Quality in the UK - Analysis of Job Quality Indicators: 2021*. *Article*. London: Office for National Statistics.

Orlando, Maria, and David Thissen. 2000. 'Likelihood-Based Item-Fit Indices for Dichotomous Item Response Theory Models'. *Applied Psychological Measurement* Vol. 24(1):50–64.

Orlando, Maria, and David Thissen. 2003. 'Further Investigation of the Performance of S − X2: An Item Fit Index for Use with Dichotomous Item Response Theory Models'. *Applied Psychological Measurement* Vol. 27(4):289–98.

Paek, Insu, and Ki Cole. 2020. *Using R for Item Response Theory Model Applications*. Abingdon, Oxon: Routledge.

Piasna, Agnieszka. 2017. *'Bad Jobs' Recovery? European Job Quality Index 2005-2015*. *Working Paper*. 2017.06. Brussels: European Trade Union Institute.

Piasna, Agnieszka, and Anke Plagnol. 2018. 'Women's Job Quality Across Family Life Stages: An Analysis of Female Employees Across 27 European Countries'. *Social Indicators Research* Vol. 139:1065–84.

Pollert, Anna, and Andy Charlwood. 2009. 'The Vulnerable Worker in Britain and Problems at Work'. *Work, Employment and Society* Vol. 23(2):343–62.

Powell, Andrew. 2021. *Disabled People in Employment*. *Briefing Paper*. No. 7540. London: House of Commons Library.

R Core Team. 2020. 'R: A Language and Environment for Statistical Computing'.

Raykov, Tenko, and George A. Marcoulides. 2011. *Introduction to Psychometric Theory*. New York, NY: Routledge.

Raykov, Tenko, and George A. Marcoulides. 2018. *A Course in Item Response Theory and Modelling with Stata*. College Station, TX: Stata Press.

Reckase, Mark D. 2009. *Multidimensional Item Response Theory*. East Lansing, MI: Springer.

Reckase, Mark D. 2016. 'Logistic Multidimensional Models'. Pp. 189–209 in *Handbook of Item Response Theory, Volume One: Models*, edited by W. J. van der Linden. Boca Raton, FL: CRC Press.

Reise, Steven P. 2012. 'The Rediscovery of Bifactor Measurement Models'. *Multivariate Behavioral Research* Vol. 47(5):667–96. doi: 10.1080/00273171.2012.715555.

Reise, Steven P., Wes E. Bonifay, and Mark G. Haviland. 2013. 'Scoring and Modelling Psychological Measures in the Presence of Multidimensionality'. *Journal of Personality Assessment* Vol. 95(2):129–40.

Reise, Steven P., Wes Bonifay, and Mark D. Haviland. 2018. 'Bifactor Modelling and the Evaluation of Scale Scores'. Pp. 677–707 in *The Wiley Handbook of Psychometric Testing: A Multidisciplinary Reference on Survey, Scale and Test Development: Volume Two*, edited by P. Irwing, T. Booth, and D. J. Hughes. Chichester, West Sussex: John Wiley & Sons Ltd.

Reise, Steven P., Tyler M. Moore, and Mark G. Haviland. 2010. 'Bifactor Model and Rotations: Exploring the Extent to Which Multidimensional Data Yield Univocal Scale Scores'. *Journal of Personality Assessment* Vol. 92(6):544–59.

Reise, Steven P., and Dennis A. Revicki. 2015. 'Introduction: Age-Old Problems and Modern Solutions'. Pp. 3–12 in *Handbook of Item Response Theory Modelling: Applications to Typical Performance Assessment*, edited by S. P. Reise and D. A. Revicki. New York, NY: Routledge.

Revicki, Dennis A., Wen-Hung Chen, and Carole Tucker. 2015. 'Developing Item Banks for Patient-Reported Health Outcomes'. Pp. 334–63 in *Handbook of Item Response Theory Modelling: Applications to Typical Performance Assessment*, edited by S. P. Reise and D. A. Revicki. New York, NY: Routledge.

Ribar, David C. 2004. *What Do Social Scientists Know about the Benefits of Marriage? A Review of Quantitative Methodologies*. *IZA Discussion Papers*. No. 998. Bonn, Germany: Institute for the Study of Labour (IZA).

Rijmen, Frank, Minjeong Jeon, and Sophia Rabe-Hesketh. 2016. 'Variational Approximation Methods'. Pp. 259–70 in *Handbook of Item Response Theory, Volume Two: Statistical Tools*, edited by W. J. van der Linden. Boca Raton, FL: CRC Press.

Rodriguez, Anthony, Steven P. Reise, and Mark G. Haviland. 2016a. 'Applying Bifactor Statistical Indices in the Evaluation of Psychological Measures'. *Journal of Personality Assessment* Vol. 98(3):223–37.

Rodriguez, Anthony, Steven P. Reise, and Mark G. Haviland. 2016b. 'Evaluating Bifactor Models: Calculating and Interpreting Statistical Indices'. *Psychological Methods* Vol. 21(2):137–50.

Rubery, Jill. 2013. 'Public Sector Adjustment and the Threat to Gender Equality'. Pp. 43–83 in *Public Sector Shock: The Impact of Policy Retrenchment in Europe*, edited by D. Vaughan-Whitehead. Geneva: International Labour Office.

Samejima, Fumiko. 2010. 'Graded Response Model'. Pp. 85–100 in *Handbook of Modern Item Response Theory*, edited by W. J. van der Linden and R. K. Hambleton. New York, NY: Springer.

Samejima, Fumiko. 2016. 'Graded Response Models'. Pp. 95–107 in *Handbook of Item Response Theory, Volume One: Models*, edited by W. J. van der Linden. Boca Raton, FL: CRC Press.

Sarkar, Sudipa, and Jonny Gifford. 2018. *UK Working Lives: The CIPD Job Quality Index - Appendix 2: Methodology 2018*. *Research Report*. London: Chartered Institute of Personnel and Development.

Satorra, Albert, and Peter M. Bentler. 2010. 'Ensuring Positiveness of the Scaled Difference Chi-Square Test Statistic'. *Psychometrika* Vol. 75(2):243–48.

Sawilowsky, Shlomo S. 2009. 'New Effect Size Rules of Thumb'. *Journal of Modern Applied Statistical Methods* Vol. 8(2):597–99.

Scherer, Stefani, and Nadia Steiber. 2007. 'Work and Family in Conflict? The Impact of Work Demands on Family Life'. Pp. 137–78 in *Employment Regimes and the Quality of Work*, edited by D. Gallie. Oxford: Oxford University Press.

Schoeni, Robert F. 1995. 'Marital Status and Earnings in Developed Countries'. *Journal of Population Economics* Vol. 8(4):351–59.

Seeman, Melvin. 1959. 'On the Meaning of Alienation'. *American Sociological Review* Vol. 24(6):783–91.

Sehnbruch, Kirsten, Pablo González, Mauricio Apablaza, Rocío Méndez, and Verónica Arriagada. 2020. 'The Quality of Employment (QoE) in Nine Latin American Countries: A Multidimensional Perspective'. *World Development* Vol. 127:1–20.

Sinclair, Robert R., John Morgan, and Elyssa Johnson. 2020. 'Implications of the Changing Nature of Work for Employee Health and Safety'. Pp. 489–508 in *The Cambridge Handbook of the Changing Nature of Work*, edited by B. J. Hoffman, M. K. Shoss, and L. A. Wegman. Cambridge: Cambridge University Press.

Skrondal, Anders, and Sophia Rabe-Hesketh. 2004. *Generalised Latent Variable Modelling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: Chapman & Hall / CRC Press.

Smola, Karen Wey, and Charlotte D. Sutton. 2002. 'Generational Differences: Revisiting Generational Work Values for the New Millennium'. *Journal of Organizational Behaviour* 23:363–82.

Solomon, Brittany C., Boris N. Nikolaev, and Dean A. Shepherd. 2022. 'Does Educational Attainment Promote Job Satisfaction? The Bittersweet Trade-Offs Between Job Resources, Demands, and Stress'. *Journal of Applied Psychology* Vol. 107(7):1227–41.

Soskice, David. 1999. 'Divergent Production Regimes: Coordinated and Uncoordinated Market Economies in the 1980s and 1990s'. Pp. 101–34 in *Continuity and Change in Contemporary*

Capitalism, edited by H. Kitschelt, P. Lange, G. Marks, and J. D. Stephens. Cambridge: Cambridge University Press.

Soskice, David. 2005. 'Varieties of Capitalism and Cross-National Gender Differences'. *Social Politics: International Studies in Gender, State and Society* Vol. 12(2):170–79.

Sribney, William M. 1998. 'Two-Way Contingency Tables for Survey or Clustered Data'. *Stata Technical Bulletin* Vol. 8(45):33–49.

Stan, Ana D., Carol A. Tamminga, Kihwan Han, Jong Bae Kim, Jaya Padmanabhan, Neeraj Tandon, Matthew E. Hudgens-Haney, Matcheri S. Keshavan, Brett A. Clementz, Godfrey D. Pearlson, John A. Sweeney, and Robert D. Gibbons. 2020. 'Associating Psychotic Symptoms with Altered Brain Anatomy in Psychotic Disorders Using Multidimensional Item Response Theory Models'. *Cerebral Cortex* Vol. 30(5):2939–47.

StataCorp. 2021. 'Stata Statistical Software: Release 17'.

Steffgen, Georges, Philipp E. Sischka, and Martha Fernandez de Henestrosa. 2020. 'The Quality of Work Index and the Quality of Employment Index: A Multidimensional Approach of Job Quality and Its Links to Well-Being at Work'. *International Journal of Environmental Research and Public Health* Vol. 17(21):1–31.

Storey, David J., George Saridakis, Sukanya Sen-Gupta, Paul K. Edwards, and Robert A. Blackburn. 2010. 'Linking HR Formality with Employee Job Quality: The Role of Firm and Workplace Size'. *Human Resource Management* Vol. 49(2):305–29.

Stucky, Brian D., and Maria Orlando Edelen. 2015. 'Using Hierarchical IRT Models to Create Unidimensional Measures from Multidimensional Data'. Pp. 183–206 in *Handbook of Item Response Theory Modelling: Applications to Typical Performance Assessment*, edited by S. P. Reise and D. A. Revicki. New York, NY: Routledge.

Sturges, Jane, and David Guest. 2004. 'Working to Live or Living to Work? Work / Life Balance Early in the Career'. *Human Resources Management Journal* Vol. 14(4):5–20.

Tabachnick, Barbara G., and Linda S. Fidell. 2014. *Using Multivariate Statistics*. Sixth Edition. Harlow, Essex: Pearson Education Limited.

Tangian, Andranik. 2007. 'Analysis of the Third European Survey on Working Conditions with Composite Indicators'. *European Journal of Operational Research* Vol. 181:468–99.

Taylor, Matthew, Greg Marsh, Diane Nicol, and Paul Broadbent. 2017. *Good Work: The Taylor Review of Modern Working Practices*. London: The Stationery Office.

Thissen, David, and Lynne Steinberg. 2010. 'A Response Model for Multiple-Choice Items'. Pp. 51–65 in *Handbook of Modern Item Response Theory*, edited by W. J. van der Linden and R. K. Hambleton. New York, NY: Springer.

Toland, Michael D. 2014. 'Practical Guide to Conducting an Item Response Theory Analysis'. *Journal of Early Adolescence* Vol. 34(1):120–51.

Toland, Michael D., Isabella Sulis, Francesca Giambona, Mariano Porcu, and Jonathan M. Campbell. 2017. 'Introduction to Bifactor Polytomous Item Response Theory Analysis'. *Journal of School Psychology* Vol. 60:41–63.

Tomlinson, Jennifer. 2007. 'Employment Regulation, Welfare and Gender Regimes: A Comparative Analysis of Women's Working-Time Patterns and Work-Life Balance in the UK and the US'. *International Journal of Human Resource Management* Vol. 18(3):401–15.

TUC. 2021a. *Jobs and Recovery Monitor - Insecure Work*. Research and Reports. London: Trades Union Congress.

TUC. 2021b. *Levelling up at Work: Fixing Work to Level up Across the UK*. London: Trades Union Congress.

UNECE. 2010. *Measuring Quality of Employment: Country Pilot Reports*. Geneva: United Nations Economic Commission for Europe.

UNECE. 2015. *Handbook on Measuring Quality of Employment: A Statistical Framework*. Geneva: United Nations Economic Commission for Europe.

University of Essex, Institute for Social and Economic Research. 2018. *Understanding Society: Waves 1-8, 2009-2017 and Harmonised BHPS: Waves 1-18, 1991-2009. [Data Collection]*. 11th Edition. UK Data Service. SN: 6614.

Vidal, Matt. 2016. 'Fordism and the Golden Age of Atlantic Capitalism'. Pp. 283–305 in *The Sage Handbook of the Sociology of Work and Employment*, edited by S. Edgell, H. Gottfried, and E. Granter. London: SAGE Publications Ltd.

Wang, Jichuan, and Xiaoqian Wang. 2020. *Structural Equation Modelling: Applications Using Mplus*. Second Edition. Chichester, West Sussex: John Wiley & Sons Ltd.

Warhurst, Chris. 2008. ‘The Knowledge Economy, Skills and Government Labour Market Intervention’. *Policy Studies* Vol. 29(1):71–86.

Warren, Tracey. 2016. ‘Work and Social Theory’. Pp. 34–51 in *The Sage Handbook of the Sociology of Work and Employment*, edited by S. Edgell, H. Gottfried, and E. Granter. London: SAGE Publications Ltd.

Warren, Tracey, and Clare Lyonette. 2015. ‘The Quality of Part-Time Work’. Pp. 62–82 in *Unequal Britain at Work*, edited by A. Felstead, D. Gallie, and F. Green. Oxford: Oxford University Press.

Wheatley, Dan. 2022. *CIPD Good Work Index 2022: Survey Report*. London: Chartered Institute of Personnel and Development.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Second Edition. Houston, TX: Springer.

Wright, Sally, Chris Warhurst, Clare Lyonette, and Sudipa Sarkar. 2018. *Understanding and Measuring Job Quality: Part 2 - Indicators of Job Quality*. *Research Report*. London: Chartered Institute of Personnel and Development.

Wu, Yi-Jung, Xiaojie Xu, and Jingying He. 2021. ‘Gender, Educational Attainment, and Job Quality in Germany, Sweden, and the UK: Evidence from the 2015 European Working Conditions Survey’. *Sustainability* Vol. 13(23):1–16.

Yung, Yiu-Fai, David Thissen, and Lori D. McLeod. 1999. ‘On the Relationship between the Higher-Order Factor Model and the Hierarchical Factor Model’. *Psychometrika* Vol. 64(2):113–28.

Zwysen, Wouter, and Neli Demireva. 2020. ‘Ethnic and Migrant Penalties in Job Quality in the UK: The Role of Residential Concentration and Occupational Clustering’. *Journal of Ethnic and Migration Studies* Vol. 46(1):200–221.