



City Research Online

City, University of London Institutional Repository

Citation: Harvey-Cook, J.E. (1995). Graduate recruitment at professional entry level: clinical judgements and empirically derived methods of selection. (Unpublished Doctoral thesis, City University London)

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/7736/>

Link to published version:

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

**Graduate Recruitment at Professional Entry Level:
Clinical Judgements and Empirically Derived Methods of Selection**

Jane Elizabeth Harvey-Cook

**Thesis submitted for the degree of Doctor of Philosophy
in Occupational Psychology**

City University Business School, Department of Business Studies

September 1995

Table of Contents

1	Introduction	1
1.1	Biographical Data	5
1.2	Recruitment and Selection in the Accounting Profession	7
1.3	General Recruitment Issues	10
	1.3.1 Presentations and Literature	11
	1.3.2 Application Form Processing	14
	1.3.3 Selection Interviews	17
1.4	Other Methods of Selection	28
1.5	Summary	32
1.6	Objectives of the Study	34
 2	 Biographical Data	 39
2.1	Introduction	39
2.2	History	39
	2.2.1 Holland's Vocational Preference Theory	40
	2.2.2 Owens's Classification Of Persons	41
	2.2.3 Integration of the two Paradigms	41
2.3	Biodata Use	42
2.4	The Stability of Biodata Predictors	43
	2.4.1 Super's Life-span Life-space Theory	44
	2.4.2 The Ecology Model	45
2.5	Behavioural Consistency	46
	2.5.1 Achievement, Background and Commitment	46
2.6	Criticisms of Biodata	48
	2.6.1 Interpretability	48
	2.6.2 Organizational-Specificity	49
	2.6.3 Loss of Validity	50
	2.6.4 Faking	50
2.7	Fairness and Ethical Considerations	51
2.8	Biographical Items in Selection Instruments	57
	2.8.1 Item Selection	57
	2.8.2 'Hard' and 'Soft' Data Items	59
2.9	Conclusion	60
 3	 The Chartered Accountancy Profession	 63
3.1	Introduction	63
3.2	General Background to Recruitment in the Profession	64
3.3	Recruitment Stages	68
	3.3.1 Literature and Presentations	69

	3.3.2	Application Forms Processing	71
	3.3.3	Selection Interviews	73
3.4		Conclusion	74
Part I : The Empirical Derivation of the Biodata Models			75
4		Data and Samples	76
	4.1	Introduction	76
	4.2	Selection of the Criteria	77
	4.3	Criterion Groups	77
	4.3.1	Examination Performance Prediction Models	78
	4.3.2	Recruitment Success Prediction Models	79
	4.4	Sample Details	80
	4.5	Data	81
	4.5.1	Predictor Data	83
	4.5.2	Criterion Data	84
	4.6	Summary	88
5		Methodology	89
	5.1	Introduction	89
	5.2	The Weighted Application Blank Technique (WAB)	90
	5.3	Multivariate Statistical Methodologies	92
	5.3.1	Discriminant Analysis	92
	5.3.2	Logistic Regression (logit)	93
	5.4	Methods Adopted in This Study	96
	5.4.1	WAB Models	97
	5.4.2	Logit	100
	5.5	Summary	100
6		Results and Discussion of Models	102
	6.1	Introduction	102
	6.2	Examination Prediction Models	104
	6.3	Comparison of Logit and WAB Results for Model 1	109
	6.4	Examination Model Component Variables	110
	6.5	Recruitment Success Models	116
	6.6	Comparison of the Logit and WAB Results for Model 2	120
	6.7	Recruitment Success Model Component Variables	121
	6.8	Validation	125
	6.9	"Empirical" versus "Rational" Methodologies	127
	6.10	Conclusion	130

7	Implementation of Models	132
7.1	Introduction	132
7.2	Developing Suitable Cut-off Points	134
	7.2.1 Model 1 Cut-off Points	135
	7.2.2 Model 2 Cut-off Points	136
7.3	Assessing the Impact of the Models on Sensitive Groups . .	137
	7.3.1 Sex	137
	7.3.2 Race	139
7.4	Testing the Decision Strategy on the Validation Groups . .	140
	7.4.1 Model 1 Validation Groups	140
	7.4.2 Model 2 Validation Groups	140
7.5	Discussion of Validation Results	141
7.6	Replacing the Selection Interview	143
7.7	Tandem Use of Models	145
7.8	Scoring Software	147
7.9	Pilot Exercise	150
	7.9.1 Individual Firm Results for Model 1	152
	7.9.2 Individual Firm Results for Model 2	153
	7.9.3 Parallel Use of the Models	154
	7.9.4 Offers Made To Applicants	156
	7.9.5 Discussion With Firms	157
7.10	Implications of the Pilot	159
7.11	Summary and Conclusions	160
	Part II	164
8	Validation and Further Development of Logit Models	165
8.1	Introduction	165
8.2	Model Validation	171
	8.2.1 Sample Details	171
	8.2.2 Data	172
8.3	Model 1 Validation Results	172
8.4	Model 2 Validation Results	174
8.5	Summary of Validation Non-Exempt 1988-90 Sample . . .	175
8.6	Extension to Accounting Graduates	176
8.7	Summary of Application to Accounting Graduates	178
8.8	Development of a Pre-Sifting Strategy	179
8.9	UCCA (UCAS) Point Pre-sift Strategies	182
8.10	Summary of Pre-Sift Development	183
8.11	Analysis of Costs and Benefits	184
8.12	Summary of Costs and Revenues	190
8.13	Discussion	191

9	Conclusions and Further Work	197
9.1	Introduction	197
9.2	The Development of Selection Models	201
9.3	Model Performance versus Judgemental Methods	206
9.4	Long Term Validity	208
9.5	Validity Generalization	210
9.6	Implementation Issues	211
9.7	Limitations of this Study and Further Research	214
	Bibliography	220
	Appendix	233
	A : Original Predictor Data Items Collected	234
	B : Variables Used in Logistic Regression	238
	C : Model 1 : Examination Performance Prediction	241
	D : Model 2 Recruitment Success Prediction	246
	E : Firms Participating in Part I	251
	F : Software Output Examples	252

List of Tables

- 1.1 Non-Accounting Graduate Withdrawal Prior to PE2

- 4.1 Model 1 : Examination Prediction Sample
- 4.2 Model 2 : Recruitment Prediction Sample
- 4.3 Predictor and Criterion Variable Categories

- 5.1 Model 1 : Examination Performance Prediction (WAB)
- 5.2 Model 2 : Recruitment Success Prediction (WAB)
- 5.3 Model 1 : Examination Performance Prediction (Logit)
- 5.4 Model 2 : Recruitment Success Prediction (Logit)

- 6.1 Point Biserial Correlation Coefficients
- 6.2 Logit Model 1 Score Statistics
- 6.3 Logit Model 1 : Derivation Sample Hit-Rate
- 6.4 Logit Model 1 : Validation Sample Hit-Rate
- 6.4 WAB Model 1 Score Statistics
- 6.6 WAB Model 1 : Holdout Sample Hit-Rate
- 6.7 WAB Model 1 : Validation Sample Hit-Rate
- 6.8 Logit Model 2 Score Statistics
- 6.9 Logit Model 2 : Derivation Sample Hit-Rate
- 6.10 Logit Model 2 : Validation Sample Hit-Rate
- 6.11 WAB Model 2 Score Statistics
- 6.12 WAB Model 2 : Holdout Sample Hit-Rate
- 6.13 WAB Model 2 : Validation Sample Hit-Rate
- 6.14 Model 1 1985-7 Sample : Comparative Statistics

- 7.1 Model 1 : 1985/6 Entrants' Logit Outcome Probabilities
- 7.2 Model 2 : 1985/6 Entrants' Logit Outcome Probabilities
- 7.3 Model 1 : Item Scores by Sex of Entrant
- 7.4 Model 1 : 1985/6 Groups at Suggested Cut-off Points
- 7.5 Model 1 : 1987 Groups at Suggested Cut-off Points
- 7.6 Model 2 : 1985/6 Groups at Suggested Cut-off Points
- 7.7 Model 2 : 1987 Groups at Suggested Cut-off Points
- 7.8 Results of Replacing Interviews with Model 1
- 7.9 Results of Replacing Interviews with Model 2
- 7.10 Results of Tandem Use of Models on 1985/6
- 7.11 Model 1 Decision Strategy Applied to Piloting Firms
 - 7.11.1 Graduates
 - 7.11.2 Undergraduates (If I/IIi Obtained)
 - 7.11.3 Undergraduates (If I/IIi Not Obtained)
- 7.12 Model 2 Decision Strategy Applied to Piloting Firms
 - 7.12.1 Graduates
 - 7.12.2 Undergraduates (If I/IIi Obtained)

- 7.12.3 Undergraduates (If I/IIi Not Obtained)
 - 7.13 Combined Model Scores for Firms' Application Forms
 - 7.13.1 Graduates
 - 7.13.2 Undergraduates
 - 7.14 Stoy Hayward Offers to 1990 Applicants
 - 7.15 Kidsons Impey Offers to 1990 Applicants
 - 7.16 Total Applications Categorised by Model 1 Score
 - 7.17 Total Applications Categorised by Model 2 Score
-
- 8.1 Validation Sample Details
 - 8.2 1988-90 Non-Exempt Entrants Model 1 Validation Sample
 - 8.3 1988-90 Non-Exempt Sample Model 1 Outcome Probabilities
 - 8.4 1988-90 Non-Exempt Sample Model 1 Decision Group Outcome Probabilities
 - 8.5 1988-90 Non-Exempt Entrants Model 2 Validation Sample
 - 8.6 1988-90 Non Exempt Sample Model 2 Outcome Probabilities
 - 8.7 1988-90 Non Exempt Sample Model 2 Decision Group Outcome Probabilities
 - 8.8 1988-90 Exempt Entrants Model 1 Validation Sample
 - 8.9 1988-90 Exempt Entrants Model 1 Score Statistics
 - 8.10 1988-90 Exempt Entrants Model 1 Outcome Probabilities
 - 8.11 1988-90 Exempt Entrants Model 1 Decision Group Outcome Probabilities
 - 8.12 Presifting on ≥ 3 and ≥ 4 Grade A 'O' Level Passes
 - 8.12a Results of Adopting ≥ 3 Grade A 'O' Levels as a Cut-off
 - 8.12b Results of Adopting ≥ 4 Grade A 'O' Levels as a Cut-off
 - 8.13 Outcome of Alternative Pre-Sifting Strategies
 - 8.14 UCCA Point Scoring Strategy
 - 8.15 Total Contribution Towards Overheads Associated with Student Training

List of Figures

- 3.1 Graduate Recruitment Stages**
- 6.1 Model 1 Criterion Groups' Logit Scores**
- 6.2 Model 1 Validation Groups' Logit Scores**
- 6.3 Model 1 Hold-out Groups' WAB Scores**
- 6.4 Model 1 Validation Groups' WAB Scores**
- 6.5 Model 2 Criterion Groups' Logit Scores**
- 6.6 Model 2 Validation Groups' Logit Scores**
- 6.7 Model 2 Hold-out Groups' WAB Scores**
- 6.8 Model 2 Validation Groups' WAB Scores**
- 7.1 Model 1 1985/5 Full Sample Logit Scores**
- 7.2 Model 1 Group Outcome Probabilities**
- 7.3 Model 1 Selection Decision Strategy**
- 7.4 Model 2 1985/5 Full Sample Logit Scores**
- 7.5 Model 2 Group Outcome Probabilities**
- 7.6 Model 2 Selection Decision Strategy**
- 7.7 Stoy Hayward Decision Tree (Based on Model 1 Scores)**
- 7.8 Kidsons Impey Decision Tree (Based on Model 1 Scores)**
- 7.9 Finnies Decision Tree (Based on Model 1 Scores)**
- 8.1 Model 1 1988-90 Groups' Logit Scores**
- 8.2 Model 1 1988-90 Group Outcome Probabilities**
- 8.3 Model 2 1988-90 Groups' Logit Scores**
- 8.4 Model 2 1988-90 Group Outcome Probabilities**
- 8.5 Model 1 Exempt Graduate Groups' Logit Scores**
- 8.6 Pre-sifting Strategy Results**
 - 8.6a Reductions in Application Forms Scored Using a Minimum of Grade A 'O' Levels**
 - 8.6b Processing Outcome**
- 8.7 Estimated Contribution from Graduate Entrants by Stage of Training**

Acknowledgements

My thanks for the fact that this thesis was ever started and eventually finished must go, in the main, to two very special people.

The first is Professor Richard Taffler, my supervisor, who provided me with the stimulus to begin the thesis and the skills to complete it. He has been infinitely patient and forbearing, even though, at times, he must have despaired of my ability to provide simple facts without peripheral irrelevancies. To his everlasting credit, we still remain friends.

The second is my husband, Peter Cook, who has had to live with someone who must have seemed permanently distracted and unreasonably single-minded. When I first began this work with the accountancy profession, I started to work at home and have continued to do so until the present. I have had two children and written the thesis once already, before deciding to include the most recent validation work.

Throughout all of this, Peter has remained a very stabilizing influence, keeping me on the straight-and-narrow emotionally, particularly in times of crisis when I felt that I should never complete the thesis. Without him, I should not have had the physical or financial means to pursue the research and I am sure he will be as glad as I that this is finished and I dedicate it to him.

My children also deserve a special mention, because, although they are both still very young, they understand that there are times when I do not want to be disturbed and have learnt to gauge whether the current emergency warrants breaking my concentration. The computer desk has been witness to more tears and cuddles than any other article of furniture in our house. I hope, when they are older, they will understand the obsessional nature of research and not resent the amount of my time it took up.

I would also like to thank the ICAEW for funding the original study and Stoy Hayward, Kidsons Impey and Finnies for continued funding for the pilot and validation exercise. Within these firms, special thanks go to Chris Priestley and Sandy Boyle (Kidsons Impey), Rupert Merson (Stoy Hayward) and Cherry Birch (Finnies) for their invaluable input and support.

Lastly, I should like to thank Sue Phelps, who proof-read the draft and Toy Odiakosa, who helped me to overcome the mysteries of desk-top publishing.

Declaration

I grant powers of discretion to the University Librarian to allow this thesis to be copied in whole or in part without further reference to me. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgement.

Abstract

This research provides evidence to support the argument that selection procedures dependent upon clinical judgements, being used in the chartered accountancy profession, may well provide results not significantly different from those obtained by chance. Research has suggested that personality type, choice of vocation and performance are predictable from personal histories (Holland, 1976; Owens and Schoenfeldt, 1979; Eberhardt and Muchinsky, 1982a; Super, 1980; Wernimont and Campbell, 1968) and using a predictive model approach to scoring biographical data (biodata) is explored here as a means of improving the selection function.

Part I of this study develops predictive models for scoring the biodata of applicants to the profession. An original contribution is made by carefully comparing two empirical model-building methodologies: the generally accepted, non-parametric, Weighted Application Blank technique and the parametric, logistic regression technique. The validity of both are explicitly tested using information from a sample of 23 training offices from 22 medium size chartered accountancy firms. The sample trainees were all non-accounting graduates entrants entering between 1985 and 1987 (N=665). Evidence is provided of the superiority of the results of the parametric models, in terms of true predictive validity. Relevant theory and the important implications of the results for related biodata studies generally are discussed.

The result of applying the models to applicants, rather than recruits, is examined in a pilot study. An original approach to scoring applications is presented. Specifically developed software is provided to minimise both processing time and error margins. The biodata logit scores of the applicants and their likely success as trainees as indicated by that score, are compared with the firm's decision whether to accept or reject. Severe problems inherent in the judgemental approach to selection are revealed and the superior performance of the model-based approach demonstrated.

Part II addresses the crucial issue of long term validation of biodata models by scoring a sample of recruits from 3 representative firms' 1988-90 entrants (N=323). The evidence does not support criticism of long term validity, as the logit models demonstrate effective performance, measured in terms of the probability of correct classification, successfully predicting the criteria on those entering the profession up to 5 years after subjects used in model development. It is suggested that poor methodology may be responsible for excessive loss of validity over time in other studies and their lack of use of hard data.

In addition, original evidence is provided to support the hypothesis of the generalizability of such models (i) across organizations and (ii) across samples significantly different from the development sample. This evidence suggests that, not only may the models be used to score applicants accounting firms of different sizes (and are therefore not organization-specific) but they may be used to score accounting graduates, who differ considerably from the original development sample (indicating that they are not sample specific). The appropriateness of using these models in a manner similar to psychometric tests is considered.

An assessment of approximate net profit associated with successful, failing or partially successful trainees is made. Accounting graduate trainees are more financially viable than non-accounting graduates.

Human information processing is extremely complex. Judgemental processing, involving, as it does, the comparison of diverse information, is adversely effected by the level of complexity of the task and the number of items to consider. In addition, human judgement is notoriously susceptible to bias caused by situational determinants and personal experience. This is particularly true in the areas of recruitment and selection, where human 'clinical' judgements are used routinely to assess the potential of job applicants. The problems associated with such judgements are magnified when the applicants are not experienced in the work for which they are applying.

Techniques for evaluating the likely performance of people in various types of occupation are legion. Some are more thoroughly researched than others. Some have high face-validity which belies their low validity and reliability, others, although exhibiting good reliability and validity, may not have the face validity to make themselves popular. This might explain the continued use of the unstructured selection interview, personality tests of any kind and pseudo-scientific and fundamentally questionable methods of selection eg graphology, which are all in constant use in this country (Robertson and Makin, 1986) and the limited use of more valid measures, such as structured interviews, cognitive tests and scored biographical data (biodata).

Employers in the UK are remarkably similar to those in other European countries in their chosen methods of recruitment (Smith and Abrahamsen, 1992). There is a prevalent, if misplaced, belief that human judgemental methods, like the ubiquitous selection interview, are superior methods of evaluation. At the very least, the prospective employer and employee must meet to assess each other in social terms. However unstructured interviews, which are widely used to judge applicant suitability in terms of the interviewers' prediction of future performance, are

extensively criticised for lack of predictive validity and reliability and their basic unfairness (eg, Arvey and Campion, 1982).

Recruiters believe that good employees result from successful recruitment procedures but what of their firm's poorer employees? The recruiter obviously believed at the time of selection that they also had the right qualities for success but, barring those whose removal from the job was not performance related, this was manifestly not so. Equally, what of those not accepted for employment? What proportion of these would have been successful employees? The lowest expectation for any recruitment system is that the results are better than random selection of applicants. 'A selection procedure is useful if the organization can do a better job of hiring people with it than without it' (Guion, 1992, p330). This expectation is not always fulfilled.

Guion points out the employment decision is based on implied prediction and that such 'hypothesis-making' requires a clear understanding of the job to be filled. Thus the job must be formally analysed and the resulting description used to identify important behavioural criteria. This information should then be used to develop 'hypotheses' about potential predictors and prepare assessment strategies.

The research literature is emphatic that statistical methods of evaluation are superior to 'clinical' methods (Dawes and Corrigan, 1974; Dawes, 1977) and Herriot (1987) has argued that clinical judgement of candidates' suitability for employment should perhaps be removed from the employment interview. '*Professional* judgement, unlike "hunch" or "play-it-by-ear" decision making, is systematic, informed and based on understanding and research (Guion, 1992, p388).'

If the evaluation of the prospective employee's aptitude for the job does not take place in an interview, where then should this function be accomplished? Since it is wasteful in both manpower and cost terms to conduct interviews with unsuitable candidates, arguably it should take place before the candidate is approached.

The employer has to use what information is available to him at the time of the

assessment to judge the suitability of the applicant and this area is fraught with error. Application forms are used to make swift judgement on whether to meet the applicants and interviews and tests of various types are used to complete the assessment.

Scoring the applicant on his/her background data, supplied by the application form and cv, is often an intuitive process, not governed by any statistical methodology and yet there is sufficient data held on the humble standard application form to provide *ex ante* accurate predictions of applicant likelihood of success.

Biographical data (biodata), has been found to be a very useful source of information for building predictive models, particularly at the initial stage of recruitment. This study seeks to assess the applicability of such models for overall professional recruitment of graduates. The profession studied here is the chartered accountancy profession. Its record of graduate recruitment has not been good, with wastage and poor performance affecting over one third of entrants.

The profession recruits almost exclusively from those graduating from the established universities. Over the last decade, the wastage rate has clearly demonstrated that the profession's predominantly interview-based selection process is failing to select appropriate trainees with an acceptable survival rate from the applicant pool.

However, the profession is made up of a multiplicity of firms of differing sizes and client bases. Small local practices service local clients, medium sized firms service medium sized clients, sometimes specialising in one type of client eg, the shipping industry or the press, and large firms service public-listed clients, conglomerates and multinationals. Within firms and between firms, there may be differing codes of conduct, methods of work and idiosyncratic forms of appraisal. Nevertheless, within these frameworks, methods of recruitment across the profession are almost identical (Harvey-Cook and Taffler, 1987).

The ICAEW (Institute of Chartered Accountants of England and Wales) monitors the

training and experience of all trainees, such that at the end of a successful contract, all students should be of a similar standard. In theory, no matter what size of training firm or client base, each trainee passing the professional examinations and completing a successful three year contract within a member firm, is expected to be at a level of professional competence comparable with any other.

Thus it is feasible to view the profession as a whole and develop suitable recruitment strategies for profession-wide use, providing the predictive criteria reflect professional, rather than individual firms', goals. Successful and unsuccessful trainees should form homogenous groups from which to draw sufficient information to build predictive models for classifying all graduate applicants to the profession.

Nevertheless, inter-firm and, indeed intra-firm differences in ethos, management style and training should not be reflected within such strategies as these are the areas which must be assessed at meetings not concerned with the selecting of the suitable applicant, the selection models having taken on this function. '...(A)ll organizations have a climate or culture which in turn tends to attract a certain type of person (Swinburne, 1986).

This study seeks to provide evidence that indeed it is feasible to use formal statistical predictive models to replace the selection function of recruitment interviews, proposing that such models are fairer, more valid and more reliable. In addition, taking a profession-wide approach using subjects drawn from firms of differing sizes from within a particular profession, the feasibility of developing one set of models for all firms within that profession is examined. Such a strategy would uphold arguments that reject the organization-specificity of such models (Childs and Klimoski, 1986; Rothstein *et al*, 1990) contrary to the suggestions of Hunter and Hunter (1984).

1.1 Biographical Data

The use of biographical data (biodata) for predicting future behaviour has recently increased in the USA. Drakeley (1989b) suggests that the rise in popularity of biodata has coincided with US employment legislation, concerned with the job-relatedness of information gathered by firms in their selection procedures (Hunter and Hunter 1984) and the impact of the use of cognitive tests on minority groups. Certainly Reilly and Chao (1982), in examining the validity and fairness of alternatives to using psychometric tests in selection, found that only biodata and peer evaluations provide comparable validities, although more recent research (Weisner and Cronshaw, 1988; Campion, 1988) suggests that structured interviews may provide equally high validities.

Biographical items on application forms are used routinely by firms and education establishments to form judgements as to whether, in the opinion of the processor, applicants are suitable for the position/course for which they are applying. Harvey-Cook (1984) finds that much of the information required by chartered accountancy firms on their application forms is redundant in practical terms. Decisions on the inclusion of items are based on arbitrary judgements of what constitutes useful and necessary information for selection and indicators of future success/failure.

High volumes of similar application forms received at certain times of the year, as in the case of graduate recruiters, are generally inspected by a clerk or secretary with, perhaps, instructions concerning the rejection of those applicants who are deemed to be totally unsuitable. These are conditions under which cognitive overload is most likely to occur. Processors faced with vast numbers of extremely similar forms may lose sight even of the simplest of criteria. There is simply too much information to process.

The data normally collected on an application form is usually what is termed 'hard' biographical data, ie the data is objective and verifiable. In addition, it is the kind

of information which an applicant is usually happy to divulge. Those items which might cause offence, eg sex, marital status and racial items, are still to be found on application forms and in the event may be missed out by applicants if they feel offended. However, this may not necessarily be a problem in practice, for example, only 5% of all those entering training contracts under the aegis of the ICAEW in 1993/94 had apparently chosen not to respond to the question on racial background (ICAEW, 1995). In this study only 3 of the 665 application forms were similarly deficient.

A great deal of what is termed 'biodata' and used in the development of predictive models, however, is of the 'soft' (unverifiable) type where the applicant is invited to express feelings or opinions in multiple choice format, with the inevitable acceptance of the inability of such measures to be reliably compared between subjects (eg Biographical Inventory for graduate entrant HM Inspectors of Tax, 1985; Helmreich *et al*, 1973; Matteson, 1978).

Despite the possible loss of the contribution of such more qualitative and unsubstantiable data (Asher, 1972; Stokes, Mumford and Owens, 1989), the development and validation of predictive, practical biodata models from the usual hard data found on application forms, is the basis of the first part of this study. The associated advantages stem from the continued use of firms' present accepted application style, including the SAF (standard application form), firms' individual application forms and cv's, and the ease of verification of the information and the perceived worth of such data for recruiters.

Using biodata predictively in employee selection is to employ the fruits of a combination of research paradigms : Holland's Vocational Preference theory, Super's Life-span Lifespace theory, Wernimont and Campbell's behavioural consistency approach, etc. The basis of such research is simply that prior experiences shape the way we will behave in future situations and those who have had similar patterns of past experience are likely to behave in a similar way in future. 'Such data work because behaviour tends to be consistent' (Guion, 1992,

p344).

It has been suggested that biodata instruments are not 'transportable' across organizations (Hunter and Hunter, 1984) and that empirical keys are organization-specific. This study also seeks to test Rothstein *et al*'s (1990) thesis of non-organizational specificity of such instruments in a professional environment with many separate firms of varying sizes with greatly varying client bases and criteria, albeit involved in similar work, using common predictors. Not only may professional examination performance be satisfactorily predicted by such an approach, but also successful performance within the individual organization and in client firms, as measured by performance ratings which encompass both firm and client-based behaviour.

In addition, the long term validity of biodata instruments has been questioned (Hunter and Hunter, 1984; Mitchell and Klimoski, 1982; Henry and Hulin, 1987, 1989; Ackerman, 1989; Rothstein *et al*, 1990 and Barrett and Doverspike, 1992) and this study will offer evidence to support the stability of such devices over time.

1.2 Recruitment and Selection in the Accounting Profession

In the early 1980s accounting firms generally were critically examining their recruitment procedures in order not only to improve their productivity by more successful manpower management, but to come to terms with what they were led to believe was an imminent and rapid shrinkage of the pool of young prospective employees, particularly graduates. Indeed, in 1987, Herriot spoke of the decrease in the number of graduates over the following 15 years and, as recently as 1990, the Institute of Chartered Accountants of England and Wales (ICAEW) reported a 27% drop in school leavers and a 30% increase in the demand for graduates, in the Executive Summary of their magazine 'Accountancy' (ICAEW, 1990).

A study of small firm (10 or fewer partners) recruitment completed by the ICAEW

(1989) indicates that, despite the overall increase in student numbers in the profession, there had been a noticeable drop in the numbers recruited to small firms. The reasons for this may variously stem from the staffing implications (trainees spend a lot of time on courses etc), recruitment expense (both financial and time aspects) and uncertainty of tenure associated with student accountants.

Harvey-Cook and Taffler (1987), in their study of the profession, note that only 10-15% of those submitting application forms to chartered accountancy firms were being offered an interview as literally thousands of applications may be received for very few vacancies. Indeed a recent survey carried out by the PA Consulting Group (1993) finds that an average of 95 applications are received for each general graduate vacancy, Blue Chip companies and familiar high street names attracting an average of 272 applications per vacancy. Recruitment to the profession rose from 4,683 in 1977/8 to a peak of 7,063 in 1988/9 and then fell steadily from 6,894 in 1989/90 to 3,840 in 1993/4. The intake percentage of accounting graduates has remained steady at about 20% and the non-accounting graduates at around 70% (ICAEW, 1988, 1995).

The chartered accountancy profession offers a basic training so useful that qualified accountants are much sought after in industry and commerce and therein lies the reason for the popularity of the profession with graduates, particularly those who did not choose accountancy when they took their degree. At present more than half of the ICAEW qualified members work outside accountancy (O'Kane, 1989). This loss is no longer accepted philosophically by the profession as it is, in effect, bearing industry's training costs, even though there are distinct advantages for firms in having members placed in industry.

The profession is plagued by its image and the dangers that training '...will be perceived as boring, socially barren and commercially irrelevant by students looking for intellectual challenge and an interesting life' (ICAEW, 1990).

Thus retention in the profession, and particularly to the training firm, has become

an important issue, additional to the successful passing of professional examinations and practice work performance. Changes in accounting practice have resulted in the need for fewer trainees eg, the reduced audit commitment of many larger firms, but the need for a steady stream of personnel with partner and management potential has not reduced, as firms diversify into areas such as management consultancy. Recruitment departments are under increasing pressure to generate such personnel in the face of diminishing resources and yet the sample firms had not seriously explored the characteristics conducive to long tenure or successful performance as a practising accountant.

Smaller firms are seen to be at particular risk and certainly their share of the intake had dropped from 28% in 1985 (the first sample year of this study) to 20% in 1988. Currently firms with ≥ 100 partners account for 65% of the total graduate intake (ICAEW, 1995).

The ICAEW, concerned at the levels of wastage during the training stage of graduate entrants, particularly in small and medium-sized firms, helped to fund part of the research which formed the basis of the study reported here, Harvey-Cook and Taffler (1987) having expressed the opinion that using a biodata-model selection strategy might well be an answer which could be considered on a profession-wide basis.

Methods adopted generally by companies to recruit staff are examined by an Institute of Personnel Managers (IPM) study conducted by Gill (1980) which shows that 90% of the 335 companies studied relied upon the interview to select executives, only 10% of firms used some form of psychometric testing and less than 5% used assessment centres. These findings are upheld by Harvey-Cook and Taffler (1987) and are similar to results published by Robertson and Makin (1986) which show a greater proportion of the 108 Times Top 1000 firms using graphology (7.8%) than biographical data (5.8%), while one firm uses astrology to select managers. Anderson and Shackleton (1986) conclude '...over the past 15 years, there appears to have been minimal uptake of the advances in staff resourcing technology into

organizational practices.'

They stress that it is up to personnel specialists to take advantage of the developing technology to improve their practice and call upon psychologists to '...identify and isolate the causal factors underlying the negligible transfer of research developments into wider recruitment and selection practices.'

As the accounting profession is moving through the 1990s, the effectiveness of the recruitment function has not greatly improved and drop-out and failure rates are still unacceptable. There are still many more graduates seeking employment than there are jobs available but, despite this, the PA survey is reported to find that 23% of their sample of firms are unable to fill their quota of graduates (PA Consulting Group, 1993).

The accounting profession recognises that the key to more successful recruitment and retention must be ensuring a better 'fit' between candidate attributes and performance.

'The problem of manpower planning for the 1990s is so serious that it needs to be put high on the agenda of every managing partner, no matter what the size of firm.' (ICAEW, 1990).

1.3 General Recruitment Issues

All industries which are associated with large graduate intakes are understandably anxious to take steps to ensure that their recruitment methods appear fair and efficient. The quality and quantity of those recruited must be such that firms are adequately staffed by competent employees who are likely to be successful in their work and remain with their recruiter for a reasonable length of time, in order to offset training costs, improve productivity and contribute to the overall development of the employing organization.

This study commenced late in 1988 and might be seen as answering the call made by Anderson and Shackleton (1986) above. At that time little improvement in the uptake of more technical recruitment practices had been reported in the literature, but many of the largest graduate recruiters were reported in the media as experimenting with assessment centres, outward-bound courses and proliferating personality and psychometric tests, such as the Watson-Glaser CTA (Watson and Glaser, 1991) and the Occupational Personality Questionnaire (Saville and Holdsworth, 1984) to supplement and improve their existing recruitment procedures. Nevertheless, most still see the selection interview as a fundamental part of the recruitment process and resist the notion that it is not valid.

The following sub-sections give an outline of procedures adopted generally by large recruiters of graduates at the time this research commenced.

1.3.1 Presentations and Literature

Professions and individual organizations recruiting graduates in quantity offer presentations at universities during which they explain what the organization will require from the graduates and what they can expect in return. Many also issue brochures which are distributed to university careers offices for students to peruse or which are sent to direct applicants. At presentations, there are very often current staff members, who are available to discuss the firm with prospective entrants.

Employers are keen that this publicity reflects their firm's ethos, giving a reasonable picture of life within their organization and indeed candidates' expectations of the firm are formed by such pre-selection literature and any presentation made by them at university recruiting sessions (Wanous, 1977; Rynes and Boudreau, 1986).

Herriot (1987) expresses concern about the truthfulness of pre-selection publicity material since realistic job information, vital for students making

career choices, is not always what is found in brochures. He emphasises the need for full factual information, '...not vague promises about an exciting and stimulating environment in a glossy brochure.'

Recruits, who, on the surface, seem identical in qualification and background to successful trainees, drop-out of the system. Identifying those who are not well matched to the job is a key issue, as is identifying the reasons for mismatch.

'New employees entering the accounting profession have work expectations which are not being confirmed by their actual work experience at least during the first year of employment.' (Dean *et al*, 1988).

While the latter expectations may be based on a composite of all the information applicants have received concerning their own chosen profession during the recruitment and selection processes, and their inherent needs (Ferguson and Hatherly, 1991), the problems associated with person/job mismatch should be considered to be the unsatisfactory result of all information received before joining the organization. Porter and Steers (1973) argue that the job behaviour of an employee is directly affected by the degree to which his/her prior expectations match reality.

Certainly some of those recruited by the less 'open' methods will be successful, perhaps more by chance than design, and many will not (Porter and Steers, 1973). Unfortunately the problems associated with unfulfilled expectations may not readily be removed for

' Failure to satisfy employee expectations may result in dysfunctional organizational outcomes...it may not be possible to extract higher than expected commitment levels by surpassing those expectations.' (Dean *et al*, 1988).

There is general agreement that those who have received realistic previews are able to self-select more efficiently (Premack and Wanous, 1985; Meglino and De Nisi, 1987) and thus are less likely to apply for jobs for which they

are not suited. When the applicant is fully informed, he/she is freer to make informed choices. As freedom to make choices increases, so does commitment to the selected courses of action. In addition, the fully informed candidate can, prior to commencement, develop coping strategies to deal with the more worrying aspects of the job, if and when they are encountered (Premack and Wanous, 1985).

Once the candidate has joined the organization, those who have received good information will experience an adjustment of expectations, which will be shifted downwards, and they are therefore less likely to be disappointed (Saks and Cronshaw, 1990). There is also less likely to be a mismatch between the individual's needs and the organizational climate.

'The use of realistic job previews in the recruitment of new members has shown consistent results in reducing the turnover of newcomers for a variety of organizations.' (Wanous, 1977).

The reason for this is likely to be because Organizational Reality Shock (ORS) described as '... the discrepancy between an individual's work expectations established prior to joining and the individual's perceptions after becoming a member of the organization' (Dean *et al*, 1988), is reduced when the applicant has received realistic job information based on analysis of the work involved. Thus realistic information should be used in attracting applicants to reduce the numbers of those who leave prematurely because they have suffered unacceptable ORS.

However, the IPM study (Gill, 1980) finds that 24% of those firms surveyed used no job description in their recruitment procedures. In addition, firms are unlikely to issue any information, *via* whatever channel, which they consider in any way detrimental to their corporate image, or likely to put candidates off with negative overtones. Thus it seems unlikely that much effort will be taken to reduce the gloss and improve the quality of the information, while graduate recruiters feel themselves to be in open

competition with each other for the same potential recruits, which indeed they are.

Those applicants who are, in fact, mismatched and leave, may be the victims of inadequate or less than truthful pre-entry information. Certainly, the high volume of applications received by chartered accountancy firms indicates that efficient self-selection is not taking place. Wilson (1994) finds a moderately high correlation ($r = .34$) between trainee accountants' experience of receiving a realistic preview and professional commitment and states that

'...trainee accountants are more committed to the profession and more *successful* within the host firm when they have gained prior exposure which informs them of what to expect in the job.'

1.3.2 Application Form Processing

Application forms influence decisions made at all stages of selection (Herriot and Rothwell, 1981; Dipboye *et al*, 1984; Herriot, 1984) and are according to Herriot, '...in principle a source of the most valid predictors of job success yet discovered: biodata.'

A recent unpublished IMS survey (Strebler, 1990) indicates that 90% of firms use application forms and 79% of those sampled felt that they were reliable enough to predict job performance as they stood. There has been little published research in the UK based on using such biographical data for predicting job success. What there is tends to have been carried out by the armed forces or the Civil Service. For example, graduate applicant tax inspectors are subject to a biodata inventory which has been in use since 1985, biodata has been used as a pre-selection sift for the appointments-in-administration competition for cabinet office posts (Bethel-Fox *et al*, 1988) and has been used as part of the selection procedure for Royal Naval Officer training (Drakeley *et al*, 1988).

Although, by report, other large recruiters use biodata, eg, medical schools and high street banks, there is, as yet, no published evidence from these sources. This may perhaps reflect the competitive nature of graduate recruitment as organizations using biodata may not wish it to be known, in order to keep the "edge" over others, or may fear that the components of the models will become known and test results contaminated. It should be noted that such contamination will not affect hard data items, which can and should be verified.

The application forms used by organizations recruiting graduates are usually very similar. Apart from the usual details of name, address, education, academic qualification and previous work experience, the candidate is invited to describe himself in terms of social activities and the levels of responsibility he has achieved at school and in subsequent activities.

There is, in addition, a space where the candidate is expected to give an account of his/her reasons for applying to the organization and perhaps to a particular profession. This would appear to invite the candidate to peruse the publicity material and give the recruiting organization appropriate responses gleaned from it (Keenan, 1985).

Candidates may indeed make earnest attempts to assess the match between their characteristics and those apparently required, but, as indicated above, the brochure may not give adequate or even truthful information upon which to make valid career decisions.

The majority of application forms require the names of personal and academic referees. These references are 'virtually useless as screening devices' (Anderson and Shackleton, 1986), as they contain '...no specific relevant information and such details that are included are mostly indefinite and inaccurate.' Common sense indicates that it would be highly unlikely that applicants would cite referees who might give any other than a good

reference and it is very rare for a referee to give negative responses. Evidence as to the low validity and reliability of such letters is given by Muchinsky (1979) and Reilly and Chao (1982).

Anderson and Shackleton contend that the use of referees should be limited to the factual checks of biographical information and not for eliciting opinions relating to personality characteristics or suitability for positions. However, despite their obvious inadequacy, like the selection interview, letters of reference continue to be used by recruiters, 96.3% of Robertson and Makin's (1986) sample of 108 firms used references some of the time.

Much of the information on the application form seems to be used as a prompt for suitable questions at interview and is not, in effect, used for selection purposes *per se*, eg, academic or athletic awards obtained, general health. Other information is used as an *ad hoc* predictor of future performance eg, grades of yearly undergraduate examinations, which are used to make informal predications of eventual degree class. The use of such information with no underlying statistical basis is invalid and must inevitably contribute to inappropriate selection decisions.

Finally, given that the information upon application forms is seen to be of importance to the firm (why else collect it ?), it would appear likely that the necessary measures would be taken by the recruiter to ensure that the information is appropriately gathered, coded and systematically used. This is rarely the case. Very few firms in Robertson and Makin's (1986) survey used the data on their application forms for predictive purposes, 94.2% never using biodata at all and only 1.9% of firms using it always. In addition, Wingrove *et al* (1984) report that there are factors other than those related to the actual contents of the application form, eg 'applicant wrote a lot', which were highly influential in accept/reject decisions.

The range of information demanded by an application form is wide and

varied. The items are certainly likely to be valid predictors (Herriot, 1984) but very little real use is made of them. Considering the time taken by applicants to complete them this is staggering. Nevertheless, they do so willingly in the belief that all of the information will be used to judge their suitability in an effective way. This faith appears misplaced.

1.3.3 Selection Interviews

Candidates whose application forms have been successfully vetted by firms are inevitably offered a selection interview to assess their fitness for the position being offered. There can be few selection techniques which have been so thoroughly researched as the selection interview with many of the results negative, particularly those relating to the traditional, unstructured interview.

'Judged by the acid test of psychometric efficiency - that is validity - the selection interview is a miserable failure. Review after review of the literature has indicated that the degree of validity is lower than that obtainable by means of appropriate psychological tests, and much lower than for biodata' (Herriot, 1984).

'With the possible exception of such obvious forms of quackery as graphology, phrenology and astrology, no other personnel selection technique is held in such low esteem in the research literature as the interview.'
(Dipboye *et al*, 1984).

Yet the interview remains the most widely used selection tool (Robertson and Makin, 1986; Anderson and Shackleton, 1986) despite having low reported validity and reliability. Apparently recruiters are not convinced. Orpen (1985), referring to earlier, unpublished work, finds

'...selection officers were generally unwilling to dispense with the interview as the main selection tool, even when informed that psychological tests were more valid predictors of success for that particular job.'

Campion *et al* (1988) comment 'There is no evidence that the continual warnings of researchers over the last 40 years about the limitations of the

traditional interview have decreased its prevalence.' The interview is seen as being as valid as any suitable psychometric test and just as much an indicator of future success. In the accountancy profession, since the application pre-sift is cursory, the reliability of the recruitment process is almost exclusively focused on the interview result (Harvey-Cook and Taffler, 1987), despite the warnings concerning the validity of selection interviews having been issued to the profession directly by Bevan (1989).

There is an extensive and detailed literature on the selection interview. A brief discussion only follows.

Reilly and Chao (1982) specifically compare the fairness of interviews with other methods of evaluation, including biodata and reference checks, and conclude that only biodata and peer evaluations produce validity coefficients to match those of psychometric tests, the latter having the highest predictive validity. However, Dreher *et al* (1988) argue that traditional research design issues may be responsible for the interview's unwarranted reputation for poor validity, and suggest that interviews should be behaviourally based.

Wiesner and Cronshaw's (1988) meta-analysis of employment interviews to investigate the impact of interview format and structure on interview validity, suggests the interview may, in fact, be a good selection instrument. However, structured interviews provide validity coefficients twice as high as those derived for unstructured interviews and higher validity coefficients are '...associated with more reliable interviews [ie, structured, board, situational] and the use of formal job-analytic information in developing interview questions.'

Huffcut and Arthur (1994) define interview structure as '...the reduction in procedural variability across applicants, which can translate into the degree of discretion that an interviewer is allowed in conducting the interview.'

Harris (1990) discusses the recent development of three types of interview structure: Behaviour Description Interviews [BDI] (Janz, 1982), Situational Interviews [SI] (Latham *et al*, 1980) and Comprehensive Structured Interviews [CDI] (Purcell *et al*, 1980).

The first concentrates on past employment experience and uses responses to predict future performance. Orpen (1985) compares BDIs with unstructured interviews and finds BDI validity for predicting supervisory ratings after one years ($r = 0.48$) to be significantly higher than for unstructured interviews ($r = 0.08$), and, for predicting sales performance, the difference is even more marked, $r = 0.61$ v 0.05 .

The second type of structuring, situational interviews, focuses on what the interviewee would do in a given situation (his/her intentions). This technique, pioneered by Latham *et al* (1980), is grounded in the theory of goal setting, which states that intentions, or goals, are the immediate precursors of a person's behaviour.

SIs provide the interviewer with examples of actual work situations, rated on anchored scales, which measure applicant intentions. They tap 'meaningful' areas of predictor behaviour for signs and/or samples of future job-related behaviour (Wernimont and Campbell, 1968). Questions are of the 'what would you do if...' variety and are based on a critical incidents job-analysis.

Latham *et al* (1980) report high inter-rater reliabilities for SIs and Wright *et al* (1989) obtained a reasonable mean correlation of 0.39 for their meta-analysis of 13 studies. Latham (1989) cites 8 studies using SIs with predictive validities ranging from 0.14 to 0.45 and compares them favourably with cognitive test results. He also provides evidence from 10 studies to support the valid, bias-free, reliable and practical nature of SIs.

Janz and Purcell *et al* liken BDIs and CSIs, the latter as yet the subject of

limited research, to cognitive tests, although there is no evidence so far that significantly better results are achieved.

Why should structuring the interview increase its predictive validity? Dawes (1977) states that any structured form of decision making is superior to unstructured forms. Maurer and Faye (1988) suggest that structuring interviews may simply prevent the interviewer from collecting too much information and the employment of mechanical scoring systems regulates their ratings.

Unstructured interviews yield more information than structured interviews, but it is disorganized and 'noisy' and therefore makes inter-candidate evaluation more or less impossible. 'Given this state of affairs, it is not at all clear how interviewers manage to discriminate between candidates on any rational basis.' (Keenan and Wedderburn, 1980).

Hakel (1989) believes that inter-rater agreement is *the* key issue in assessing the reliability of the selection interview, although he warns that '...vigorous pursuit of interrater agreement might produce an interview...that is nothing more than a self-reported work sample' (p291). Nevertheless, there may be disadvantages in curtailing the range of information gathered. Hakel defines this as a problem of construct validity, '...specifically the trade-off between bandwidth and fidelity. The interview is...a wideband low fidelity technique (p291).'

Carlson *et al* (1971) examine the consistency with which people interviewing the same candidate actually agree. Their results indicate that only structured interviews yield information which has reliable inter-rater agreement. This is also supported by the work of Schmitt (1976).

Campion *et al* (1988) argue that structuring raises the psychometric properties of the interview and report a corrected validity of 0.56 in their

study. They, and Wright *et al* (1989), suggest that highly structured interviews are the equivalent of verbal ability tests and more recently, Huffcut and Arthur (1994) provide evidence that such interviews may provide overall validity comparable with that of ability tests and that validity increases with increasing structure.

The theory of validity of predictors in selection (Smith, 1994) states that the characteristics of work performance fall within three domains: the universal (required by all work), the occupational (required by certain jobs) and the relational (required to relate to others in specific settings). The validity of the predictor measure is a function of the extent to which it reflects the dependence of work performance upon these three domains. He suggests that situational interviews focus less on the 'irrelevant domain' and more on the 'occupational' (and possibly the universal) domain(s).

As interviewers ask applicants about the past in order to make predictions of likely future behaviour, interview validity must depend on the memory of the applicant and pre-supposes that the responses will be truthful. This is not to suggest that applicants deliberately falsify their responses, although they may do so, but they are playing a role of which they are aware certain characteristics are expected and will modify their behaviour and responses accordingly (Super, 1980). In addition, not all aspects of the past behaviour of the applicant will be good predictors of future success and some variables will be better predictors than others.

Relatively little emphasis is placed on the behaviour of the interviewer, although there is ample evidence that there is a strong relationship between candidate responses and interviewer behaviour (Keenan, 1976,1977,1978; Keenan and Wedderburn, 1980). Research demonstrates that positive non-verbal interviewer behaviour can increase the desire of the interviewee to accept an offer (Liden and Parsons, 1986) and applicants take this behaviour as an indication of their likely success (Rynes and Miller, 1983).

Interviewer behaviour may also be manipulated by applicants' impression management techniques (Gilmore and Ferris, 1989).

Even though structuring interviews restricts information gathering to that set out by the structure, interviewers are, unfortunately, subject to many biases, eg primacy, halo, leniency and severity effects (Shuh, 1978; Arvey and Campion, 1982; Reilly and Chao, 1982; Harris, 1990) which are concerned with the order of information presentation and the tendency of raters to raise or lower their overall ratings on the basis of the candidate's ratings on particular aspects.

The physical characteristics of applicants have significant impact on interview outcome and most recruiters are happy to admit that they are not impressed by candidates who, by their careless dress, have indicated that they place little importance on the interview. Although this is understandable in a conventional professional atmosphere, if unscientific, there are far more insidious biases at work in the interview situation. For example, those whose physical appearance does not impress or indeed creates a negative impression on the interviewer, will be discriminated against, as will those who do not match the personal stereotype the interviewer is seeking.

Physically unattractive people are thus at a considerable disadvantage as they are perceived less favourably than attractive people with identical qualifications (Haefner, 1977; Rynes and Gerhardt, 1990; Morrow, 1990).

Applicant sex is also an important discriminatory variable. Heilman and Saruwatari (1979) found that when females represented 25% or less of the pool of candidates, they were evaluated less favourably than when the proportion of female candidates was higher. Females are persistently discriminated against at interviews (Dipboye *et al*, 1975; Dipboye *et al*, 1977). Dipboye *et al* conclude that even though "attractive" women are more highly rated and recommended for higher salaries than "unattractive" ones

generally, in jobs which are masculine stereotyped (as is accountancy), they are consistently evaluated lower than "unattractive" women and both are consistently rated lower than men, regardless of job. Indeed, in this case, it is a disadvantage to be attractive : this has been termed the "beauty is beastly" effect (Heilman and Saruwatari, 1979).

There has been little work in the UK on how racial minorities are treated at interview. What work there has been tends to have been carried out in the USA where equal opportunities legislation is further advanced than in the UK. The indications are that black candidates are not treated less favourably than white candidates in the USA (Rothstein *et al*, 1990) and may even be treated more favourably.

However, in the UK, there is a difference between the proportion of various minority groups in the population and the proportion of the same groups in professional work (Brennon and McGeevor, 1990). The disproportionality may be more a function of the education system in the UK, which produces fewer university graduates from minority groups to enter the employment pool, rather than restrictive recruitment practices.

The CNAA survey of ethnic minorities in the graduate labour market in 1985 (Brennan and McGeevor, 1990) indicates that 0.55% of the university graduate sample was Afro-Caribbean, 0.89% was African and 1% was Asian. Of those recruited to the chartered accountancy firms in 1990, the percentages were 0.32%, 0.87% and 7.38% respectively. There is an observable overweighting in the Asian category and a very small difference between the other two categories. On this evidence it seems unlikely that the profession is treating minority applicants differently from non-minorities.

The minor differences in the accounting profession may have arisen from the practice of recruiting mainly from the university graduate population, in which minority groups are under-represented. Polytechnics or 'new'

universities traditionally recruit from those who have lower academic achievement than the established universities accept, and those with vocational qualifications. Since ethnic minority communities are commonly found in inner-city areas, where standards of educational attainment are lower than the national average, they are more likely to gain their degrees at polytechnics.

While both the Commission for Racial Equality and the Equal Opportunities Commission monitor employment procedures for racial minorities and different genders, the physically and mentally handicapped workforce does not enjoy such protection. Campion and Arvey (1989) note the paucity of recent research concerning interview impact on handicapped applicants making it impossible to assess the presence or absence of unfair treatment.

Arvey and Campion (1982) ask why the interview is still central to virtually all recruitment processes (Robertson and Makin, 1986; Anderson & Shackleton, 1986) 'given its relatively low validity, reliability and susceptibility to bias and distortion' and discuss some of the reasons for this prevalence. The most likely seems to be that suggested by Kahneman and Tversky (1973):

'... people are prone to experience much confidence in highly fallible judgements, a phenomenon that may be termed the **illusion of validity**. Like other perceptual and judgement errors, the illusion of validity often persists even when its illusory character is recognised. When interviewing a candidate, for example, many of us will have experienced great confidence in our prediction of his future performance, despite our knowledge that interviews are notoriously fallible.'

Since we rely in everyday life upon our judgements of people and situations, it is inevitable that we should adhere to those judgements, even in the face of conflicting evidence, for acceptance of such evidence would undermine our faith in our ability to cope with the environment (Festinger, 1957).

Certainly, at some stage, both parties need to meet to assess applicant fitness for the job. Recruiters in organizations probably rightly believe that they cannot assess such essential attributes as value priorities, motivation and life-style expectations, other than in face-to-face meetings and the need for such assessment is emphasised by Smith (1994).

- ' At an intuitive level it is easy to see that a brilliant engineer at odds with his or her organization is likely to produce little. A mediocre person who fits the organization like a glove is likely to do better than many would expect.'

He implies that the use of 'relationship variables' as predictors is largely ignored by researchers.

It is, unfortunately, very easy for candidates to make a good impression on the interviewer by simply agreeing with him/her and complimenting him/her (Keenan, 1977). This behaviour was found to enhance the chances of a successful selection outcome, even when the candidate's qualifications were inadequate (Gilmore and Ferris, 1989).

If we like people like ourselves and tend to recruit them at interview, this would not, logically, seem a bad thing, since those interviewing are likely to be successful job holders. However, those similar to the incumbents of previous years may not be what the firm is seeking, even though the recruits will have to work in the same organizational structure. Collins (1987) reports that the national director of human resources for Price Waterhouse in the USA stated that, 'It is clear that the public accountancy profession is not that same as it was 20 years ago.' The same article quotes the dean of a prestigious American graduate business school as saying, 'It becomes abundantly clear that the next generation of accountants cannot succeed by replicating the last.' In addition, such approaches perpetuate any social or racial bias within professions.

The assessment of applicant fit, however, does not have to be undertaken at

the same time as the *selection* decision is made, since an applicant may be competent to do the job, but be not suited to the work environment of the firm.

The problems concerned with the lack of reliability and validity of interviews may be overcome by removing the selection function from the interview and changing its status (Herriot, 1987). The interview then becomes a negotiation, rather than an assessment session with one-way information flow. The two parties may concentrate on the necessary interaction to assess candidate 'fit' within the organization (Rynes and Gerhardt, 1990) and the formation of a 'social contract' (Argyris, 1960, referred to in Herriot, 1987).

Barber *et al* (1994) find that a 'recruitment only' interview can result in substantial increases in the information acquired by applicants, enhancing the organization's ability to communicate with them. They suggest that such an interview format is ideal for organizations who want to counteract a low profile or poor image.

Evidence from Rynes and Gerhart (1990) suggests that, when general employment skills ('universal characteristics' (Smith, 1994)) are held constant, interpersonal skills, goal orientation and physical attractiveness contribute to the assessment of 'fit' (these items are not available on the application form) and academic qualities, social involvement and work experience (all available on the application form) do not contribute. It would appear that interviews may have a unique contribution in tapping areas which may not be available at the application stage, although Rynes and Gerhardt call for further research into the whole question of 'fit'.

Reilly and Chao (1982) suggest that the low reported validity of the selection interview might be caused by confusion about its rules and objectives. In the case of the chartered accountancy profession, where on-campus, and indeed in-firm interviews, are conducted by many different people from the same

firm, there is an obvious need to standardise the procedures to ensure that candidates are assessed on an equal footing. The same groups should be rejected or accepted, independent of interviewer, and those rejected are those who would ultimately have proved to be unsuccessful. In short, the system should be valid and reliable.

Despite this, only 44% of firms in Harvey-Cook and Taffler's (1987) survey had a written interview structure, and these were not necessarily based on job-analysis. Structuring the interview is a process dependent on job-analysis which should be undertaken by suitably qualified professionals, although there are 'off-the-peg' systems available which allow firms to analyse their work and prepare job descriptions, eg the Common-Metric Questionnaire, (Harvey, 1994).

The problem of inter-rater reliability can certainly be addressed by using BARS [Behaviourally Anchored Rating Scales] (Smith and Kendall, 1963) or trait assessments (Latham *et al*, 1980). Latham *et al* find that their BARS benchmarks show striking similarity to the responses of interviewees. They suggest that the situational interview might be improved by combining it with methods that focus on episodes of past behaviour and such an approach might be considered by the accounting profession (see Orpen above).

However, Carlson *et al* (1971) ask the fundamental question, 'Why spend the additional effort and expense of using a structured interview when commercially available and inexpensive paper and pencil ability tests predict just as well?' They admit that interviews have greater 'face-validity' than the psychometric tests with which they are frequently compared and employers *want* to use interviews to meet prospective employees. In addition, applicants whether from social majorities or minorities, may believe such 'content-orientated' methods of selection (Schmitt and Ostroff, 1986) to be fairer than those not so obviously connected with the job, such as cognitive and personality tests.

While Hakel (1989) emphasises the need to take into account practical considerations in changing selection procedures, even refining the selection interview may be a costly and time consuming business.

Huffcut and Arthur (1994) draw attention to the costs of structuring interviews (p188): '... results are particularly relevant to practitioners, who must weigh cost and development time against anticipated results when deciding how much to structure an interview.' Maurer and Fay (1988) make similar comments concerning the development of situational interviews: 'The use and characteristics of an interview method are reflected in its expected costs and practical viability...advantages must be weighed against the cost and feasibility of performing the critical incidents analysis upon which this form of interview is based.'

1.4 Other Methods of Selection

Firms recruiting graduates may choose to use other methods for assessment. Among these are personality tests, cognitive tests and job-related tests eg, in-tray exercises (individual situational decision making or job-related information tests), group discussion/decision exercises, situational inventories and assessment centres (which are combinations of many different types of test).

Downs (1970) reports that candidates feel that work-related tests give them an opportunity to show what an interview cannot. One of the most positive aspects of such tests is that candidates can see the relevance of the tests to the job (high face-validity) and so respond in a positive manner to them. Smither *et al* (1993) find

'...newly hired entry-level managers judged simulations, interviews and cognitive tests, with relatively concrete item types, to be significantly more job-related than personality, biodata and cognitive tests with relatively abstract item types.'

However, the evidence from Becker and Colquitt (1992) would seem to indicate that

the areas most likely to be faked on a biodata form are exactly those which are more job-relevant. Their obvious job-relevance opens them to abuse.

The high 'face-validity' of the ubiquitous personality test, as poorly rated by many as the interview (Blinkhorn and Johnson, 1990), may account for the recent rise in the use and popularity of such tests at the selection stage and the regularity of the spirited correspondence concerning the nature and valid use of such tests in the psychological press (*The Psychologist* January, March, May, November, 1990; June, July, September, November, 1991; February, March, May, 1993).

Herriot (1987) suggests the popularity of personality tests can be explained as firms use them to assess how applicants will fit in and how far they will rise in the firm's hierarchy. The underlying belief is that work being a group activity, relationships with colleagues are important. Senior management should be 'home-grown' and therefore 'steeped in the organizational culture'. The personality test gives the manager a 'profile' of a likely candidate which is identifiably similar to his self-perception and so it is employed to identify others like him.

Jackson and Rothstein (1993) refute Blinkhorn's and Johnson's allegations of poor validity, commenting that their evidence is too restrictive on many measures and a biased sampling of the research. They stress that the contribution of personality tests should be assessed 'in terms of the degree to which they contribute to improved criterion predictability in the context of other predictors.'

Nonetheless, Guion (1992) concurs with Blinkhorn's and Johnson's critique, criticising the quality of research into the practical utility of personality tests on the basis of '...too many concurrent rather than predictive designs and too little replication (p343)'. He quotes earlier work (Guion and Gottier 1965) concluding '...it is difficult...to advocate...the use of personality measures in most situations...as instruments of decision.' Although Guion agrees that personality tests can be useful in a multiple assessment context, he is clear that '...the evidence does not exist to justify the use of personality measures, without specific research for

specific purposes, as a basis *for employment decisions* (p343).’ (emphasis added)

This would seem to indicate that it is unwise to use such devices as stand alone measures but they may be considered as useful additions to the selection process, a response echoed by Herriot (1985). Personality tests might prove valuable for identifying personality congruence with the ‘conventional’ type of Holland (1976) or indeed with the present day ‘successful’ chartered accountant personality profile. However, Crawley *et al* (1990) find interpreting the results of personality tests (the Occupational Personality Questionnaire and Myers-Briggs Type Inventory) in their assessment centres far from easy. Needless to say, the personality attributes assessed by such tests must be demonstrably job-related for their use in selection procedures.

Assessment centres (ACs), which combine several assessment techniques, have a very good reputation for validity, though a less good one for reliability, ie, they predict the successful outcome at an acceptable level but may not maintain their performance over time (Turnage and Muchinsky, 1984). A disadvantage of ACs is that they are situation-specific methods and must be developed for each given selection problem.

Again, the emphasis is on thorough and systematic job-analysis to ensure the validity of the centre (Blanksby and Iles, 1990). Only through such analysis can the ‘...qualities, skills, characteristics, or abilities or dimensions that are required for effective job performance’ be assessed. Abstract ‘dimensions’ of performance are identified, used to form the component tests and combined to produce the overall assessment rating (OAR).

However, Robertson and Kandola (1982) note that work-sample tests in assessment centres provide consistently higher validities than other tests which are not so closely related to the job; in fact ‘...it is exercises (work-samples) *not* dimensions that best represent the underlying structure of assessor ratings...’ Robertson *et al* (1987). Crawley *et al* (1990) further emphasise the need for exercises based on the actual job in assessment centres, in order to provide content validity, as their research

suggests that the *sign* rather than *sample* rationale of Wernimont and Campbell (1968) is of dubious worth.

There is still uncertainty over what the Overall Assessment Rating (OAR) measures and why ACs work (Klimoski and Brickner, 1987; Robertson *et al*, 1987) and research is still continuing in this area. Unfortunately, ACs are very expensive to run, due to the number of component exercises and the need for many assessors. Where large numbers of candidates are assessed, as with graduate recruitment procedures in accounting firms, the process would be impractical, time-consuming and prohibitively expensive. ACs are more usually associated with assessment of attributes such as the management potential of the individual already within the organization.

Situational inventories (Motowidlo and Tippins, 1993) which are low-fidelity work simulations, are the paper and pencil equivalents of situational interviews, with multiple-choice format questions. Motowidlo and Tippins find relatively low predictive validity for such devices ($r = 0.3$ for job performance) although their study provided only 36 subjects with sufficient criterion information for analysis. However, such inventories are relative newcomers to selection and will, no doubt, be subject to further research and Motowidlo and Tippins note that, whereas the job-related questions are so specific it might be expected that those with no job experience would not score well, this is not the case in practice. Such an approach might well, therefore, be considered by graduate recruiters, since, like biodata approaches, they use few staff resources and reduce interview commitment and costs.

Cognitive psychometric tests are designed to score such attributes as clerical aptitude (DAT Battery, General Clerical Test, Clerical Ability Battery) language ability (MLAT), verbal reasoning (WAIS), spatial awareness (DAT), and critical thinking (Watson-Glaser Critical Thinking Appraisal). Furthermore, most of the more complex tests, such as the WAIS and the Watson Glaser CTA, target specific areas of competence, viz (in the case of the CTA) : drawing inferences, recognising

assumptions, deductive reasoning, logical interpretation and argument evaluation.

Even were specific tests to be identified as useful tools for specific professions, applicants making multiple applications within the profession might find themselves repeatedly completing them with the inevitable consequences of practice effects, boredom and frustration. However, use on a profession-wide basis might be considered, where any applicant completes the test once and a professional body, such as the ICAEW, provides the result to member firms.

Guion (1992) reports a reduction in the use of highly specific tests and a trend towards general measures. He draws attention to the *Journal of Vocational Behaviour* which, in 1986, devoted an issue to the importance of general intelligence, *g*, in employment settings and concludes that 'a general factor of intelligence is quite often a better predictor of success, in training and on a job, that is an "optimally weighted" set of more specific scores (p341).'

Such tests have provided the best accuracy as predictors for employment purposes (Ghiselli, 1973; Owens, 1976; Hunter and Hunter, 1984) but their use, too, necessitates the establishment of predictor-criteria relationships after job analysis has been carried out. That is, specific levels of the targeted test areas must be identified as significant predictors of performance in job-related areas, providing point-to-point correspondence (Asher, 1972). Unless these clear relationships are demonstrated, the use of such tests may alienate applicants (Campion *et al*, 1988) and, in addition, may technically contravene employment and equal opportunity legislation (Owens, 1976; Pace and Schoenfeldt, 1977).

1.5 Summary

The selection mechanisms in general use for graduate recruitment ie, promotional material, pre-selection based upon application form vetting and interviews, are not viewed favourably in the research literature and face criticism for their poor validity

and reliability. There is mounting concern that such methods are too fallible and may well not be adequate in the face of increased competition for suitable prospective recruits.

Recruiters are understandably anxious to refine their current procedures while being very reluctant to relinquish any of their traditional methods, even in the face of pressing evidence as to their inappropriate nature (Arvey and Campion, 1982; Orpen, 1985; Campion *et al*, 1988; Bevan, 1989; Herriot, 1987).

Promotional material and the use of application forms can be improved, the former by greater attention to providing pertinent, reliable information and the latter by structured appropriately developed scoring systems. The selection interview may perhaps be most resistant to improvement.

Herriot (1984) argues that managers are certainly aware that the selection interview has severe failings and they should recognise that, as both parties are making decisions, the situation should be recognised as one of information exchange and perhaps of negotiation and selling.

It has been suggested that structuring the interview raises its validity by removing some of the 'noise' and opportunities for bias. In addition, structured situational interviews, with standardised scoring systems to control inter-rater reliability, can provide predictive validities comparable with cognitive tests.

In order to raise the validity of the selection process, either the validity of the selection interview must be increased by the methods described above, or other more valid methods of selection, such as the use of scored biodata, adopted. Model-driven biodata scoring strategies can be designed to remove those with little chance of success from the applicant pool before other recruitment strategies are applied.

Adopting such an approach, candidates primed by realistic job information, and thus in a secure position from which to make self-selection decisions (Williams, 1984;

Premack and Wanous, 1985; Meglino and De Nisi, 1987), are selected by the firm at the application-form stage when only those whose biodata score indicates a high likelihood of success in the job are approached.

The candidate is free to assess what the organization can offer, and how well they will fit in, at the interview (Herriot, 1987). The organization, with similar knowledge, will be in a much better position to negotiate with the candidate. Thus the interview may be used by both parties to assess candidate 'fit' (Herriot, 1987; Rynes and Gerhart, 1990). The usefulness of such 'recruitment only' interviews is emphasised by Barber *et al* (1994).

Biodata models which are based on factual data may well produce results as valid as those achieved by psychometric tests (Reilly and Chao, 1982). Using them to determine selection criteria may, in effect, be using a biodata model as a proxy for a standardized test and, it will be argued, remove the need for a 'selection' interview, by reliably predicting the candidates' likely performance in the job before the employer makes the approach to the applicant.

Biodata models are developed from criteria related to job performance, eg, professional performance and examination results, are obviously job-related and their use would obviate the need for specific research into which particular standardized tests would yield satisfactory predictive indicators for that job. In addition, the predictive power of strategies based such standardized measures is less likely to decrease over time than those based on more flexible items.

1.6 Objectives of the Study

The chartered accountancy profession wishes to reduce the numbers of those wrongly recruited, whether the losses are due to mis-match, examination failure, poor performance within the firm, or any combination of these. The non-accounting graduate attrition rates for the entry years 1980-1986 were 47%, 42%, 37%, 31%

32%, 36% and 36% (ICAEW, 1992, 1993). The Education and Training Board of the ICAEW was concerned with such wastage and its implications for the profession as a whole, its member firms and, not least, failing students.

This concern is justified. Examination of the thesis study sample of trainees indicates that only 34% of recruits are successful in terms of passing both their Foundation and Intermediate examinations at the first attempt. Thirty percent are outright failures, having failed or referred in both, or failed the first and then left. Thirty-six percent are intermediate cases who have failed or referred either of these examinations. These statistics are not better than would be expected from chance recruitment.

At the request of the ICAEW this research is primarily directed at non-accounting graduates, since they constitute by far the larger part of the trainee accountant graduate entry cohorts (78% : 22%) [ICAEW 1992].

The development sample is drawn from non-accounting graduates entering a sample of small and medium size accounting firms during 1985, 1986 and 1987.

The focus of the research is the first two years of the training contract because it is within this period that the recruitment failures become manifest. Table 1.1 below demonstrates the attrition of the non-accounting degree graduate trainees prior to PE2.

Table 1.1					
Non-Accounting Graduate Trainee Withdrawal Prior to PE2*					
Entry	1985	1986	1987	1988	1989
% Lost	31.3	32.2	31.2	32.4	32.4

(* ICAEW 1991/2)

This thesis compares empirically derived methods of personnel selection (biodata models) with clinical judgemental methods (selection interviews), within the context of graduate recruitment in the UK chartered accountancy profession, i.e., in an industrial setting.

An original comparison is made between two model building methodologies, viz: the non-parametric standard recommended procedure, the informal WAB technique of England (1971) and a formal parametric statistical technique, logistic regression (logit).

Siegel (1956) stresses that the advantages of non-parametric tests stem from their ability to accommodate ranked or categorical data, to waive the assumption of normal distribution of scores, to be simply computed and to be applied to small samples. However, Norusis (1990) points out that such tests are generally less powerful than parametric tests and are most useful in situations where parametric procedures are not appropriate.

While Mitchell and Klimoski (1982) use a WAB-derived model to compare with their formally derived model, a true comparison is not made between methodologies since each is developed using different sample data. This thesis will provide a proper comparison between formal and informal methodologies, based on analyses using identical data sets.

Biodata models have been associated with loss of validity over time which may have been variously due to poor methodology and/or poor development (Shuh, 1967; Owens and Schoenfeldt, 1979; Eberhardt and Muchinsky, 1982b; Mitchell and Klimoski, 1982; Schmitt *et al*, 1984) but more recent work has challenged the assumption of diminishing validity (Rothstein *et al*, 1990; Barrett and Doverspike, 1992).

This thesis examines the out-of-sample performance of predictive selection models over 4 years after development and seeks to provide evidence supporting their

continuing use as predictors.

Biodata models have also been presumed to be organization-specific (Hunter and Hunter, 1984) and therefore restricted in use to groups reflected by the derivation sample.

This thesis seeks to provide original evidence that this is not the case and that development samples drawn from organizations across a profession may provide models applicable to the sampled profession as a whole.

In addition, evidence of the robust nature of such models will be provided, demonstrating that where the model criteria apply, models may be 'transferred' to samples markedly different from the derivation sample (the accounting graduate entrants).

Finally, there is a paucity of the reported use of such models once developed. Since predictive models are developed for practical applications, this is not helpful. Some of the issues of diminishing long-term validation may well result from practical use and/or abuse and lessons may be learnt from the experience of users. This thesis therefore reports a pilot exercise and issues arising from 'live' implementation of a model-based selection process in an industrial setting.

Specifically the objectives are:-

- 1) To develop predictive selection models using an established non-parametric technique, the Weighted Application Blank, and an appropriate parametric technique, logit.
- 2) To test the predictive-validity of both techniques and to compare the results achieved by each.
- 3) To compare the results of the selected technique with the results of the current judgementally-based procedures.

- 4) To examine the crucial area of long term validity, providing positive evidence of the reliability of predictive models.
- 5) To explicitly test the theory of organizational specificity in the light of professional, rather than organizational, entry.
- 6) To provide original evidence to support the robust nature of predictive biodata models indicating that they may, like cognitive tests, be used on cases differing substantially from the intended group but to whom the model criteria are applicable.
- 7) To explore implementation issues not addressed in the literature, using data from a representative sample of firms from within the target profession.

Chapter 2 provides a review and discussion of the use of biographical data as a source of predictive data and Chapter 3 provides an account of recruitment issues and the current methods of selection employed in the chartered accountancy profession.

The empirical research in this study is presented in two sections. Part I describes the empirical derivation of predictive models and Part II reports their long-term validation and practical implementation in an industrial setting and, in addition, also provides the evidence of their robust nature and organizational non-specificity.

The development of the selection models, the comparison between the two methodologies and the implementation of a model-based selection procedure in a representative sample of firms are reported in Chapters 5 to 7. These are followed in Chapter 8 by a description and discussion of Part II of this study, relating to the assessment of the long-term validation of the models and their further development.

The final section, Chapter 9, draws conclusions from the study and suggests areas which would benefit from further research.

2.1 Introduction

The theme central to any employment selection strategy is the match between the skills and personality of the candidate and those required by the job. This involves exploring the characteristics of the job, then predicting the extent to which any given candidate will fit the job's requirements. There are many selection techniques in current use which have varying degrees of success and most involve some kind of assessment of the life history of the applicant. Methods used to treat this biographical data (biodata) also vary to a great extent in terms of content validity and reliability.

This chapter will examine the nature of biodata and some of the advantages and disadvantages of its use for predicting applicants' job behaviour.

2.2 History

Biographical data, or biodata, has had a long history of use for predictive purposes, dating back to the early part of the century (Owens, 1976 and Monahan and Muchinsky, 1983) and this reflects its

'...intuitive and intrinsic validity based on the fact that it speaks directly to a central measurement axiom, namely that what a man will do in the future is best predicted by what he has done in the past.' (Owens, 1976).

Owens reports heavy and valid military use of biodata during the Second World War, however, according to Reilly and Chao (1982), it did not achieve wider recognition or use until the 1950s. Reilly and Chao believe that the adverse impact

of equal opportunities legislation in the USA may well have contributed to the adoption of 'fairer' methods of selection to replace the standardised tests which, although highly successful, had been found to have adverse impact on minority groups (Hunter and Hunter, 1984). Biodata is currently a well used tool, both in selection and staff-development and the volume of research recently attributable to it bears witness to this.

The development of the use of scored biodata for predictive purposes is the result of a combination of several theories of personality and development. Two distinct paradigms which reflect the common-sense view that past patterns of experience will moderate our approach to our future lives and consequently our choice of career:- Holland's Vocational Preference Theory (Holland, 1976) and Owens attempt at Classification of Persons using biodata from American college students (Owens and Schoenfeldt, 1979), are integrated in the work of Eberhardt and Muchinsky (1982a). They conclude that significantly different personal history experiences differentiate the lives of members of different vocational types.

2.2.1 Holland's Vocational Preference Theory

Holland's theory suggests that our prior experiences shape the way we approach decisions about our future and members of a particular vocation will have similar histories of personal development. Thus personality type and vocational groups are congruent and individuals with similar personalities choose similar vocations. People can be characterised by their degree of similarity to one or more of six vocational types, viz: Realistic, Investigative, Artistic, Social, Enterprising and Conventional. Membership of these groups is seen to be the result of an interaction between a variety of cultural and personal forces and accountants are found in the Conventional group.

2.2.2 Owens's Classification Of Persons

Owens and Schoenfeldt (1979) derived a 389-item Autobiographical Inventory (AI) from a pool of 2000 items reflecting significant aspects of early life experience, including both those within and beyond the control of the individual. The inventory consists of questions relating to items such as family life and extrafamilial relationships, school-related activities, interests and attitudes derived from life experiences and sports participation. The results of applying this inventory to college freshmen, in conjunction with a combination of psychometric tests, provide evidence of adolescent subgroups identified by similar patterns of past experience.

Subsequently the Biographical Questionnaire (BQ), a shortened form of the AI, was developed and another sample of freshmen scored. Owens's and Schoenfeldt's subjects, and the subgroups to which their scores assigned them, have provided strong longitudinal evidence to support the predictive validity of biodata for a wide range of criterion performances, over many years.

2.2.3 Integration of the two Paradigms

Eberhardt and Muchinsky (1982a) report the administration of the BQ and Holland's Vocational Preference Inventory to 816 students exposing the relationships between past experience and vocational preference. They, too, conclude that significantly different life history experiences do differentiate the lives of the various vocational types. Further support is supplied by the work of Neiner and Owens (1985) in a longitudinal study conducted over 7 years, examining early life experiences and early post-college experiences.

Owens (1976) provides a succinct overview of the use of biodata at that time which not only discusses history, prevalence of use, methods of combining and scoring items and statistical approaches, but suggests possible future

industrial applications for biodata.

2.3 Biodata Use

More recently Owens's review of biodata approaches has been revisited by Stokes and Reddy (1992) who find that many of his ideas are currently under positive scrutiny. In particular, biodata may not only be useful for predicting performance but also for 'motivation-saturated' criteria such as accident and safety behaviours.

Owens notes that the fact that 'types' of people with similar backgrounds will find themselves in similar occupations could create, *inter alia*, problems concerning group dynamics, creativity, development of the individual and development of work teams. Some of these issues are already subject to biodata research. Stokes and Reddy (1992) report that the subgrouping approach lends itself to a wide range of behavioural prediction, enhancing selection, promotion assessment, employee counselling etc. In addition, the approach enables employers continually to update the data which refines its predictive ability.

Current use of biodata in the United States is widespread and most available literature concerns American studies. There has, however, been an upsurge of use in the UK, although this seems to have been mostly confined to the armed forces (Drakeley, 1984; Drakeley *et al*, 1988; Winter *et al*, 1987) and the Civil Service (Bethel-Fox *et al* 1988; Oates 1985), with the exception of a few very large employers, such as (by report) ICI, high street banks, British Telecom and The Post Office. This might, perhaps, account for the fact that Robertson and Makin in their 1986 survey, find that only 6 of the 108 firms sampled used biodata for management selection purposes and, also, for the lack of literature concerning the use of biodata techniques in the UK.

Recent studies show that successful biodata models have indeed been developed for a wide variety of predictive purposes in employment, eg, career success (Childs and

Klimoski, 1986), CPA (Certificate of Public Accountancy) examination performance of accountants (Dunn and Hall, 1984), success in professional training (Helmreich *et al*, 1973; Drakeley, 1984; Drakeley *et al*, 1988), leadership effectiveness (Kuhnert and Russell, 1990), management potential (Ritchie and Boehm, 1977), potential for faking (Cohen and Lefkowitz, 1974), performance (Schmitt *et al*, 1984; Hunter and Hunter, 1984; Gable *et al*, 1989), turnover (Gable *et al*, 1989), voluntary withdrawal from training (Drakeley *et al*, 1988) and assessment centre ratings (Bethel-Fox *et al*, 1984; Drakeley *et al*, 1988) amongst others.

2.4 The Stability of Biodata Predictors

While longitudinal studies (eg Rothstein *et al*, 1990) provide long term evidence of the effectiveness of biographical data, concern is still frequently expressed that effects may not be stable and predictive validity, therefore, not reliable.

An example may serve to underline this concern. The factors upon which Owens's original 1968 questionnaire were based differed for men and women. Owens and Schoenfeldt, in 1979, report refactoring and finding the factors almost precisely duplicated, yet Eberhardt and Muchinsky (1982b) find only partial congruence in their replication, all male factors having female factor analogues. They argue that the changing life experiences of females are affecting the factor structure. Indeed, in the light of the changes brought about by Equal Opportunities legislation in the USA, it would be remarkable if this had not been so.

However, Reilly and Chao (1982) suggest that different biodata keys may than be needed for men and women but Rothstein *et al* (1990) found that biodata validities are not moderated by sex. Stokes and Reddy (1992) indicate there may be differences in mean factor scores for males and females.

Davis (1984), in another longitudinal study examining college graduates, finds that the similarity of homogeneous biodata subgroups does decline over time. This

supports the view expressed by Owens and Schoenfeldt (1979) and Eberhardt and Muchinsky (1982) that, as biodata subgroup members are exposed to a variety of new situations and environments, their experiential paths diverge necessitating reclassification of biographical types.

Although reliability coefficients for biodata forms obtained by test-retest methods have been found to be in the 0.8 to 0.9 range over considerable time periods (Mumford and Stokes, 1992), Shaffer *et al* (1986), examining Owens's BQ, report average test-retest reliability coefficients to be only 0.56 for males and 0.58 for females when evaluating item reliability. They also find factor stability to be very varied, ranging from 0.49 to 0.91 (mean 0.78) for men and from 0.5 to 0.88 (mean 0.76) for women, with the more objective factors maintaining significantly greater stability in their five year study.

The likelihood that a biodata measure might provide dwindling reliability, for whatever reason, necessitates monitoring. However, where the predictive criteria remain stable and the predictor items are objective, such a problem is likely to be minimized.

2.4.1 Super's Life-span Life-space Theory

Super (1980) provides some explanation of the changing validity of different aspects of biodata. He states that the theatres and roles of a person's life all interact and influence each other. He suggests that non-occupational positions, occupied before the commencement of adult career, influence the type of occupation taken up and the ways in which role expectations are met.

'Thus amount and type of schooling is one determinant of occupation entered, and the first occupational position, both its type and job performance, is one determinant of later occupational positions open to the individual...The more adequately, in self-perception and that of others, the adolescent plays pre-occupational roles, especially those of student and part-time worker, the more likely are success and satisfaction in occupational roles.' (Super, 1980)

The likelihood that changing personal experience affects the predictive performance of biodata is self-evident. The importance of such variables as pastimes, interests and levels of commitment etc, as predictor measures, must fluctuate according to the level of individual development, as recent experiences supersede earlier ones.

2.4.2 The Ecology Model

Mumford *et al* (1990), having tracked individuals 6 to 8 years after graduating from college, find that while, for some subgroups, continuity in personality, interests and goals is evident, for others there is discontinuity brought about by intervening life experiences. Their research led them to propose a general theoretical framework which they term the *ecology* model and which rationalizes recent research findings in this area.

This model takes into account the fact that interaction between the individual and the environment is a continuous process and, since time and energy are finite, the individual will be forced to choose between courses of action, always seeking to maximise the likelihood of achieving long term success, in terms of adaptation to the environment. Each choice refines the individual's characteristics and a discernable 'development trajectory' emerges, which may be predicted.

There would thus appear to be a 'shelf-life' for different biodata categories as the individual is 'refined' by his/her interaction with the environment, elements of past experience and intellect waxing and waning in influence. This is a matter for concern to those who use such information in predictive models and may be a key to the reported loss of validity associated with such models over time. However, predictive models developed from objective biodata and focusing on time-specific criteria, eg, training success, should not lose validity over time, provided that the relevance of the criteria remains unchanged.

2.5 Behavioural Consistency

The key to understanding why biodata works as a predictor, is that behaviour tends to be consistent (Guion, 1992). The behavioural consistency approach (Wernimont and Campbell, 1968) suggests that predictors are in fact samples of future job-related behaviours, rather than signs of them, indicating that future behaviour will be an adaptation of past or present behaviour. Thus lack of relevant experience will result in diminished performance of the task and, conversely, relevant experience will enhance performance and, according to Super (1980), success and satisfaction.

For example, past examination success may be an indicator of future examination success because the individual has developed good preparation skills and examination technique and has had practice in sitting examinations, not merely because he/she has facility in the specific subject examined. Asher (1972) suggests that accurate prediction may be a function of a point-to-point correspondence between predictor space and criterion space: the more points the predictor and criterion have in common, the more accurate the prediction will be.

2.5.1 Achievement, Background and Commitment

However, individual differences may render some types of biodata more effective than others. Drakeley (1989a) attempts to integrate existing accounts of the criterion-related validity of biodata within the context of social opportunity structure and quotes Rodger (1965) who states that success in any enterprise can be said to depend on three things:- capacity, inclination and opportunity. Drakeley argues that sociological constraints work with the situational constraints, reported by Super (1980), to shape aspirations and attitudes.

He proposes that if capacity, inclination and opportunity are necessary for success, then three classes of item :- achievement, background and

commitment, are necessary to account for the criterion-related validity of biodata.

Achievement data are those such as measures of educational achievement or awarded positions of responsibility, background data are those such as parents' occupation, number of siblings and type of school attended and commitment data are those such as affiliations to societies and leisure pursuits. He finds significant evidence of the effects of achievement and commitment, but not of background, in the success of Royal Navy officer recruits in the initial stages of training. It would, however, be foolish to suggest that background does not affect both achievement and commitment but whereas achievement (as measured in the above terms) and commitment are within the control of the individual, background is not.

The use of some background items for selection purposes, therefore, is seen to be controversial (section 2.7). Certain areas are controlled by legislation, eg, use of racial and sexual information, but others, arguably having as great an unfair impact, eg, size of family and position in it, residential area and physical characteristics (including disability) are not subject to such control.

Measures of achievement are arguably easiest to assess, being readily available in terms of examination results etc and are probably representative of the general measure of ability, *g*, which fluctuates little over time. However, measures of commitment are not so easy to develop because they may be derived from unquantifiable data and subject to continual change (Super, 1980) as experience shapes our attitudes and expectations. Whereas general ability is likely to have significant impact on performance throughout a career, the usefulness of commitment measures is likely to be related to key stages in individual development (Super, 1980; Stokes and Reddy, 1992).

2.6 Criticisms of Biodata

2.6.1 Interpretability

The fact that criterion-related validity is often identified but is uninterpretable, has led to biodata models being criticised for 'blind empiricism' (Baehr and Williams, 1967). Personnel psychology as a science demands explanations of the relationships between criterion and predictor and biodata models frequently provide instances where the relationship is far from obvious. Most quoted examples are the importance of primogeniture in successful navy diver training (Helmreich *et al*, 1973) and attendance at a circus for successful door-to-door salesmen (Appel and Feinberg, 1969, quoted by Drakeley, 1989b).

As Drakeley points out, it is quite possible to research every item to find sufficient evidence to support the predictor-criterion relationship, but this would be an almost impossible task and for all practical purposes, unrealistic; personnel psychology should, above all else, be an applied science. Nevertheless, researchers may be expected to make efforts to 'explain' component variables of even integrative models, in which the score achieved is dependent upon the interaction between as many variables as there are in the model and where, in reality, such explanations are probably futile.

'Rational' methods of biodata item keying are used to construct models with the deliberate intention of being able to identify the nature of the relationships between predictor and criterion. Thus such methods derive keys by using expert judgement to decide, *a priori*, which predictive items should be included in the measure. These items are then subject to statistical analysis to derive the scoring keys.

'Empirical' approaches derive their keys by experimental methods directed

primarily at criterion prediction. Individual items which do not differentiate significantly in their ability to predict are discarded. While the models derived in this way are undoubtedly accurate, the predictor-criterion relationships may appear inexplicable.

Comparisons between the two methods have not provided any conclusive evidence of the superiority of either and Stokes and Reddy (1992) suggest that a combination of the two may provide 'maximum predictability and psychological interpretability'.

Mitchell and Klimoski (1982), however, are firmly in favour of using successful empirically derived methods even when the nature of predictor/criterion relationships is not clear, because 'It may not be fair to ask that every application of personnel theory must contribute towards psychological theory, particularly when this theoretical contribution comes at the price of higher research expenditure as well as a loss of predictive power.'

2.6.2 Organizational-Specificity

It has been suggested that a stumbling block in the path of general use of biodata models is that the keys, unlike cognitive test items, are not transportable between jobs and organizations (Hunter and Hunter, 1984), ie, they are organization-specific. However, Rothstein *et al* (1990) provide evidence that biodata predictors may be constructed to be valid across organizations and stable over time.

They further argue that high levels of situational specificity are not necessarily an inherent property of non-cognitive measures such as biodata scales and emphasise the advantages of a consortium-based, multiple organization approach, such as is undertaken in this study. Certainly, an instrument which is successful in recruitment for one high street bank will

almost certainly be equally successful in another.

2.6.3 Loss of Validity

Section 2.4 above outlines the reason why certain items of biodata may not remain good predictors over time. Theyer (1977) argues that biodata model validity is moderated by 'age, organizational practices and procedures, criterion and temporal changes in the nature of the job among other factors', all of which is undeniably true and fits well with the ecology model proposed by Mumford *et al* (1990).

However, Rothstein *et al* (1990) have recently found that biodata predictors yield substantial validities up to 11 years after development, despite many societal changes taking place in the interim. They suggest that

'...methods of biodata scale construction and validation based on large samples and successive replications produce both validities that generalize across organizations and across other potential moderator variables *and* validities that tend to be stable over fairly long periods of time....That is, generalizability and temporal stability of biodata validities may both depend on the same processes of scale construction.'

2.6.4 Faking

Finally, biodata instruments have been criticised for their susceptibility to 'faking', whether deliberate or not. This problem is obviously more associated with biodata inventories than application form scoring systems. Such 'response distortion' can be restricted by using verifiable data and ensuring that applicants are aware that checking of the information is routine (Anderson and Shackleton, 1986).

Verifiable items have been found to be less likely to be falsified than soft

items (Asher, 1972; Cascio, 1975; Shaffer *et al* 1986; Becker and Colquitt, 1992), however, Goldstein (1971) found considerable discrepancies between information given on application forms and previous employers' records. This is in direct contrast to two studies quoted by Owens (1976) - (Mozel and Coza, 1952 and Keating *et al*, 1950) - which found a high correlation between hard item responses and previous employer records.

Becker and Colquitt (1992), in a study concerned with which items are amenable to faking and which are faked in practice, find that subjects can fake a biodata instrument when instructed to do so, but that, in practice, the faked items are 'less historical, objective, discrete, verifiable and external than other items, and...*more* job relevant'. Perhaps applicants attempt to 'second guess' the reasons for inclusion of soft data items and respond accordingly. Goldstein (1971) suggests that such discrepancies may well be unintentional and reflect the applicant's perception of himself and his/her worth.

Asher (1972) argues that accountability is necessary when using biodata and that it will only be present if the applicant's permission is sought for the verifying of the information with past employers, teachers etc, thus informing the candidate that his responses will be checked. This need is further emphasized by the work of Hough *et al* (1990) and Schrader and Osburn (1977). However, responses to soft data items can rarely be checked and, in addition, may fluctuate over time (Mumford *et al*, 1990).

2.7 Fairness and Ethical Considerations

Pace and Schoenfeldt (1977) stress that for compliance with the equal opportunities legislation in the USA (and presumably this is also relevant to the UK), all items in a biodata model should be demonstrably job-related. Owens (1976) argues that since the items used in biodata models and the scoring keys are empirically derived, only

job-relevant items will be included and responses evaluated in terms of their relationship to subsequent job success. He states that this means there can be no justification for complaints of wilful discrimination.

While the contention rages over whether there are in reality inherent genetic differences in ability between races or social classes (Flynn, 1989 and Rushton, 1990), there are observable cultural differences in outlook, aspiration and *developed* ability (Hunter and Hunter, 1984). Cognitive tests, even where defensibly job-related and exhibiting impressive predictive validities, adversely affect minority groups. Hunter and Hunter (1984), having reviewed the literature of the preceding 15 years concerning the attempts to reverse the adverse impact of tests on such groups, conclude that

'The evidence is clear: the difference in ability test scores is mirrored by a corresponding difference in academic achievement and performance on the job. Thus, the difference in mean test scores reflects a real difference in mean *developed* ability. If the difference is poverty and hardship, then it will vanish as poverty and hardship are eliminated.'

Biodata measures have shown themselves to be at least as good predictors of success in both academic and industrial settings as cognitive tests, while not being so highly correlated with race. Owens (1976) suggests that biodata may therefore provide an alternative predictor for selection because, 'All in all, the available evidence would seem to suggest that the major dimensions of biodata response are quite stable across cultures, age, race, and sex groups, and companies.' (p623)

Although legislation ensures that recruitment practices are not obviously discriminatory, new strategies should be critically appraised, not only in the light of employment legislation, but in terms of what is socially acceptable. Biodata models may contain variables which appear to favour certain social groups, if the sample from which they are developed is biased towards certain social sectors. In the UK there is a powerful sociological argument that, as there has been little likelihood of those from a low social class background entering professions like chartered accountancy, models based on successful job holders may be biased against those

from such groups.

Drakeley (1989a) also suggests that the lack of professional incumbents from lower social classes reflects the aspirations and academic attainments of the various class groups in the UK. Children are certainly influenced from the earliest age by the attitudes of parents, family and peers to higher education and delayed earning capacity. There are reasons to suspect that low parent and teacher expectations are responsible for the low numbers of graduates in racial minority groups, particularly Afro-Caribbeans (Brennon and McGeevor, 1991).

In addition, as discussed in section 1.3.3 above, socially disadvantaged students graduating from polytechnics have reduced likelihood of being selected for professions such as chartered accountancy for, while it may be argued that the CNAA (Council for National Academic Awards) ensures that a degree from a polytechnic is comparable in theory with a degree from a university, they are not commonly perceived to be equal. If the new universities (former polytechnics) continue to exhibit differences in entry qualifications and quality of degree status, the professions will continue their arguably unfair recruitment practices. This is unlikely to be affected by simply calling polytechnics universities.

A further consideration of social inequality may arise from the ramifications of student funding. Those whose parents are unable to contribute to their costs will be deterred from taking up higher education as they will not have the means to live or provide for study equipment. This problem will be exacerbated in the near future when students will be expected to pay towards their tuition fees.

The implication that items which appear to perpetuate bias or prejudice should not be part of the data pool is clear. Nevertheless, the work of Holland, Owens and Super, *inter alia*, indicates that experience is important in determining job choice and success and that those who do not have the right experience base will not have such a good chance of success, or achieve similar job satisfaction, as those who do. While it is inevitable that sociological influences affect future behaviour and are

beyond the control of the individual, there is a case for examining apparently contentious data for evidence of obvious unfairness, rather than dismissing it out-of-hand. Drakeley (1989a), for example, does not find significant background effects in his study while commenting that such effects are certainly present.

For example, by repute, those from independent school backgrounds attain better levels of qualification at school. Collier and Mayer (1985) report that the type of school attended is significantly related to applicants' likelihood of entering the prestigious Politics, Philosophy and Economics course at Oxford University. However, it does not necessarily follow that independent schooling is a better preparation for higher education, or a professional career, than state schooling. It may well be that those who have attended independent schools tend to recruit those from a similar background and the whole system becomes cyclic.

Nevertheless such candidates are likely to have had role models which will lead them to perceive professions like politics, chartered accountancy or the law as everyday careers, whereas those from different backgrounds may not even consider them or know what they involve. In short, aspirations must reflect experience and the need to provide adequate recruitment information is once again emphasized.

During this research, personnel staff, raising objections to biodata use, have argued that those educated in the public sector should have some kind of compensatory weight applied to their academic qualifications. This argument is fraught with problems, since the different types of education offered in the public sector and the effects of extraneous factors, such as the area of the school, would all have to be taken into account.

How, for example, should a candidate who has attained poor qualifications at a recognised public (independent) school be compared with one who had achieved similar qualifications at a state comprehensive, or a grammar school? The fact that they have each achieved a similar level in standardised tests is the only reasonable method of comparison, for background effects - positive or negative - over which

they have no control, are not tested in such examinations, even though they have affected the outcome. State examinations measure developed ability, not inherent ability. The advantages or disadvantages of different types of schooling are not simply academic; social aspects may be of equal or greater importance.

Practically, since the 'damage' has been done by the applicant's past experience, it is unrealistic to expect that employers will make efforts to deliberately accept candidates whose backgrounds indicate that they will not be successful. This would be unacceptable and irresponsible. The employer is in the position of maximizing the benefits achieved from the applicants available. If there is good reason to rate one group of individuals more highly than another, he/she will inevitably do so, as the costs of recruiting unsuitable candidates penalise both firm and recruit.

The 'bone of contention' is whether the employer has indeed good reason to favour one type of candidate over another. Recruitment methods which rely heavily on the basis of social background may be held to blame for socially biasing the professions. The Harvey-Cook (1984) survey of recruitment methods employed in the chartered accountancy profession finds no obvious evidence of social bias at the pre-selection stage, where the achievement of minimum academic skills and extracurricular activities were of primary consideration. However, the most likely area for this type of discrimination to operate is the selection interview, where accent, physical characteristics, sex, race or indeed social background, all affect interviewer judgement.

The Commission for Racial Equality and the Equal Opportunities Commission monitor the employment practices of recruiters in the UK and attempt to ensure that they are fair in terms of race and sex, however, the issue of physical disability, overseen by the Employment Service, is not adequately monitored. Although there are recommendations concerning the desirable proportion of disabled people within organisations, these are not yet backed by legislation.

An example of this anomaly is found in the chartered accountancy profession. None

of the firm offices sampled by Harvey-Cook (1987), nor those sampled between 1985-90 for this study had any trainees with serious physical disabilities. Since the application forms of unsuccessful candidates are not kept, there was no way of ascertaining whether such applicants were not being offered interviews, not accepted at interview nor, indeed, whether there were any such applicants. Firms are certainly aware of such personal data as there is generally a section on application forms relating to applicants' state of health and whether they have a particular disability.

Employers are hesitant, perhaps for many good reasons, in adopting positive recruitment strategies with set targets related to their proportion of such employees. Questioned about their lack of disabled employees, the firms studied indicated the lack of facilities on site or at clients' establishments and cite problems concerning employing those with impaired mobility. While such popularly presented problems certainly exist, there is no reason why they should not be surmounted and the Employment Service provides support and advice for employers who wish to employ physically handicapped people.

However, there may actually be very few physically disabled accounting trainees for other reasons. For example, there is a shortage of role-models for disabled people in the professions and those who have a considerable physical handicap are unlikely to have attended local state schools and may well complete their full-time education at special residential colleges, rather than polytechnics and universities. Most professional graduate recruitment is aimed at the latter establishments.

The impact of the component variables of predictive models and overall scores can only be checked to ensure that they do not favour one particular group over another if members of such groups are available within the derivation and validation samples. There are very few disabled members of the chartered accountancy profession and there were none in the samples used in this study. Clearly general biodata models should not be applied to such candidates.

2.8 Biographical Items in Selection Instruments

Biographical items may be collected from application forms and then weighted (Weighted Application Blank) or may be taken from a self-report inventory (Biographical Inventory Blank).

Biographical data has been described as 'historical and verifiable pieces of information about an individual' (Asher, 1972). However much of the information currently wrongly labelled biodata is neither historical nor verifiable (Mael, 1991) but relies on the subject giving his/her opinions or impressions in a series of multiple choice questions. 'It is not uncommon to find items about internal states, opinions, and reactions to hypothetical situations in biodata measures.' (Mael, 1991).

Mael (1991) attempts to provide a conceptual rationale for biodata. He argues that Owens's often quoted axiom concerning past behaviour, strictly interpreted, could limit biodata to actions of the same type, 'such as previous job or activities that required similar skills.' How, then, are the items used in biodata instruments selected by occupational psychologists ?

2.8.1 Item Selection

While the development of biodata inventories has received a good deal of research attention (Owens, 1976; Mumford and Stokes, 1992; Asher, 1972; Owens et al, 1962), little attention has been directed towards the identification of specific component items. Mumford and Owens (1987) investigate this area and advise that the first step is to 'define the domain', by defining the job and the criteria of interest, then to assess the behaviours which distinguish successful and unsuccessful performers.

Smith (1994) provides research which defines the types of domain of the various predictors in selection and groups them under the general headings of 'occupational', 'general' and 'relationship', the latter, he believes, being

frequently ignored by recruiters.

Once the domain has been identified, the biodata items which are likely to be good predictors of the criterion can be generated. The task may be approached in a number of ways, including using life-history essays, biographical information and interviews with supervisors and incumbents (Stokes and Reddy, 1992).

In the case of recruitment to a profession, where the majority of those who apply have no experience of the career on which they will embark, the choice of suitable biodata predictor variables depends, to a great extent, on expert judgement. Those who are already in the profession can provide invaluable information on what types of skill are required by the job and those who develop the biodata measure may provide further input from reviews of the relevant literature.

However, the pool of items can be restricted by the nature of the instrument being developed. For example, a biodata inventory can capture a great deal of personal information since it is completed by the subject to be scored, although there are disadvantages associated with inventories and these are outlined in sections 2.6.4 and 2.8.2 below. Whereas scoring life histories, or application forms, may restrict the scorer to the data provided in the normal course of application.

Indeed, this study, attempting to provide interim measures to increase the validity of the recruitment process in the accountancy profession, is restricted to the biodata vouchsafed in the normal fashion by graduate entrants on their application forms. While it has been noted that this approach is limited, perhaps losing much of the evidence of prior behaviour and experience which determines future job behaviour, the advantages of the speed and ease of model construction outweighed the latter cost.

2.8.2 'Hard' and 'Soft' Data Items

Items of biographical data may be termed 'hard' or 'soft' with reference to their amenability to verification. Items which are verifiable are such things as academic record, place of birth, parental occupation etc and are those used in this study. These are described as 'hard' items and are those which are associated with application forms or blanks. Hard data items tend to be more reliable than the soft items (Owens,*et al* 1962; Asher, 1972) generally found on self-report inventories and are amenable to coding since the response categories are usually discrete and have interval or ratio scales.

Items which are not considered to be verifiable are expressed opinions, abstract value judgements and reported attitudes, particularly those which rely on memory of past events, or choice of a particular behaviour in a given situation. These items, in multiple-choice response format, are difficult to code for scoring purposes since they are almost impossible to scale appropriately for statistical analysis, having at the most an arbitrary rank order. Consider appropriately coding the following example from Asher (1972) while bearing in mind individual differences in perceptions of 'strictness'.

'Were your parents:

- a. Very strict with you*
- b. Usually strict with you*
- c. Seldom strict with you*
- d. Never strict with you'*

Nevertheless, despite the coding and statistical difficulties associated with such data, there is evidence of their being powerful predictors of future behaviour (Mumford *et al*, 1989).

2.9 Conclusion

The use of biodata for predictive purposes has been a feature of recruitment for almost a century. Many studies show the effectiveness (Owens, 1976 and Monahan and Muchinsky, 1983) and robustness of biodata's predictive powers (Hunter and Hunter, 1984; Schmitt *et al*, 1984; Hough *et al*, 1983; Rothstein *et al*, 1990).

Biodata approaches are criticised on several grounds, notably:

- (i) the lack of predictor-criterion relationship explanation
- (ii) adverse impact on socially sensitive groups
- (iii) susceptibility to faking; and
- (iv) loss of validity.

Since the models are developed from job-related criteria, the component variables must be job-related and the need for precise explanations of the relationships is questionable. However, plausible explanations of predictor/criterion relationships may be made more possible by combining 'rational' and 'empirical' methodologies in biodata model construction.

Background moderates the whole process of recruitment and selection and minimises the probabilities of applicants from certain socio-economic groups applying to the higher professions. However, bias in the selection process is inevitable until those background aspects which differentiate unfairly between social groups no longer exist (Hunter and Hunter, 1984). Needless to say, biodata models, used in the selection process, cannot control for the adverse impact of background effects in any subsequent interview or other forms of assessment. Nonetheless, a carefully developed biodata instrument may, in fact, be less biased than more traditional approaches (Owens, 1976).

Research has indicated that hard types of biodata item are less susceptible to faking than soft items and are, in effect, the only items which can be verified. However, the wisdom of checking the biodata of applicants who are selected using biodata models, is emphasised, since applicants can, either inadvertently or deliberately, falsify their responses (Goldstein, 1971; Becker and Colquitt, 1992)).

In addition, hard data items are usually interpretable on interval or ratio scales of measurement and are thus more amenable to statistical analysis, since they do not lead to questionable assumptions or arbitrary categorization.

The problem of diminishing validity may be reduced by the use of large sample sizes and proper cross-validation (Hunter and Hunter, 1984), although models developed from hard data items do not necessarily suffer from such shrinkage. Evidence from Barrett and Doverspike (1992) indicates that predictive validities may remain stable or even increase. The stability of the criterion measure may be the key to reliability (Rothstein *et al*, 1990). Nevertheless, long term validity of biodata models should be assessed (Shuh, 1967; Owens and Schoenfeldt, 1979; Eberhardt and Muchinsky, 1982) since behavioural choices, which refine the individual's goals, cause fluctuations in the influence of biodata items over time (Mumford *et al*, 1990).

Biodata models which successfully predict the selection criterion, not only on a separate cross-validation sample, but on successive independent samples, can be considered as viable alternatives to selection interviewing. Employers must surely be considered justified in the use of such models as they have been found to equal, in terms of predictive validity, or even better, the types of psychometric test which are in common use for selection purposes (Reilly and Chao, 1982; Drakeley *et al*, 1988).

Such models offer the promise of lower recruitment costs and greater productivity resulting from more accurate recruitment decisions. From the applicant's point of view, they reduce the likelihood of being recruited to a profession to which they are not suited, a mistake which is arguably more costly for them than the recruiter. In

addition, application scoring systems have little impact on the applicant, who has willingly provided the information in the knowledge that he/she will be judged upon it.

In the next chapter, recruitment and selection strategies in use in the chartered accountancy profession at the time of the original research are discussed, in the light of their effectiveness in terms of the relevant literature.

3.1 Introduction

The chartered accountancy profession focuses almost exclusively on the selection interview as its prime recruitment tool. Chapter One deals briefly with this form of recruitment and points out more shortcomings than positive indicators. It is suggested that selection methods which rely so heavily upon such a poor tool are ineffective and lead to the unacceptable Type I and Type II error rates in recruitment terms which have caused mounting concern. Further, it is suggested here that the use of biodata models in place of the selection interview would dramatically reduce these rates and, as a direct result, lower recruitment costs.

The profession is the largest single recruiter of graduates in the UK, accounting for over 10% of all graduates taking up employment (University Statistics Record, 1992). Sharing anxieties common to other graduate recruiters, the profession is concerned with the drop-out rates suffered by entrants during their training period, particularly in the early stages, when a loss of up one third of the intake has not been unusual, although the most recent figures are more optimistic. In the sample studied here, the loss of approximately 4% occurring before the first examination (Graduate Conversion or Foundation) is deemed to be commitment-related and those after, generally ability-related.

Another source of concern is the examination first time pass rate which is arguably responsible for the majority of later withdrawals. The sample statistics discussed in section 1.6 above suggest that the current recruitment system is extremely wasteful in terms of recruitment and training costs. Such wasteful strategies, allowing as they do for the levels of drop-out associated with previous intakes, reflect unfavourably on the profession in recruitment rounds, as well as having a highly negative effect on failing students.

As a result the profession is seeking acceptable changes in methods which might help to reduce this wastage and improve the overall quality of those selected. 'Traditional ways of doing things may no longer be acceptable or profitable...Recruitment and retention of staff are going to require much more effort, imagination and money.' (Accountancy, January 1990, p143)

The methods used by the profession do not differ greatly from those adopted by most other large professional recruiters (Robertson and Makin, 1986; Smith and Abrahamson, 1992) who place heavy reliance on the selection interview preceded by the usual application form vetting. The criteria used for this vetting are not usually anything other than subjective, nor are they based on evidence as to their relative importance as predictors of future performance (Wingrove *et al*, 1988). This vetting has been described as a 'paper-sift' (Drakeley, 1988).

This chapter will discuss the extant recruitment and selection methods in the Chartered Accountancy profession in England and Wales and ways in which improvements might conceivably be achieved, for

'Failure to give these problems the attention they deserve is likely to leave the firms in a weak competitive position as they struggle to provide client services with a dwindling or second-rate workforce.'
(Accountancy, January 1990, p143).

3.2 General Background to Recruitment in the Profession

The profession in the UK relies almost entirely upon graduates for its trainees. Over the last 15 years approximately 90% of those taken by firms for training have been graduates (ICAEW, 1995). Most large and medium sized firms recruit solely from the 'old' university graduate population, taking those of any discipline. Those with accounting degrees make up around 20% of the total intake.

Those holding such a degree, or a degree with a very large component of

mathematics, accountancy, economics and law, are exempt from the Graduate Conversion (Foundation) Course (GCC) and examination, which takes place after approximately 6 months of the training contract, and which all other graduate entrants must take. The training contract for both accounting and non-accounting graduates is for three years.

Holders of degrees which cover subjects in the Graduate Conversion (Foundation) examination may obtain individual exemptions from these subjects at the discretion of the ICAEW, which monitors and steers education and training within the profession.

Firms may require an exempt graduate to take a Test of Accounting Competence (TAC) to satisfy the firm and the ICAEW that both accounting and non-accounting graduates are at a similar level of accounting competence before they take the two Professional Examinations. The first, PE1 (Intermediate), is taken by non-exempt graduates roughly 21 months into the three year contract, unless there have been examination referrals or failures at GCC. Exempt students take PE1 at approximately 18 months into their contract. The second examination, PE2 (Final), is taken near to the end of the three year training contract for both types of trainee.

The graduate conversion course (GCC) examination was changed in 1987 to incorporate Data Processing (information technology). In 1990, the whole examination structure was changed. Data Processing was removed to reflect the fact that most students are now computer literate when they commence training. In addition, Quantitative Techniques, previously a major stumbling block, was also removed and incorporated in the later levels of examinations where its application is essential to understanding the concepts examined. The GCC examination is now known as the Foundation Examination and, apart from the removal of Data Processing and Quantitative Techniques, remains virtually unchanged, although the Law syllabus now concentrates on Contract and Company Law, dropping previous components of Trust and Insolvency Law.

The first professional examination, PE1, is now called the Intermediate Examination and also remains virtually unchanged, examining elements of Financial Reporting, Auditing, Financial Planning and Control, Business Finance Decisions and Taxation. PE2, now called the Final Examination, again remains relatively unchanged, although certain items have been shuffled from one syllabus to another. Subjects examined are Financial Reporting (auditing), Business Planning and Evaluation and Advanced Taxation. The paper is complemented by a multi-disciplinary case study.

The examinations are founded upon professional requirements and represent standardized measures of accounting performance, albeit theoretical.

There has been an effort to ensure that the examinations reflect the current nature of the work, although accounting firms still express the opinion that the examinations, even with the revised syllabuses, do not do so. However, the ICAEW is convinced that they are still effective and those who wish to be considered as qualified members must have successfully passed these examinations, each within a given length of time, though not necessarily within the three year training contract.

Nevertheless, the ICAEW has the current syllabus and examination structure under review. It seems likely that a core of essential subjects will remain mandatory but other topics will form a range of options. In addition, the system may be 'modulized' so that students do not all sit their examinations at the same time. This would obviously benefit firms who suffer chronic manpower shortages at certain times of the year.

While graduates are under training contract and subject to the various examinations, they work in client firms, as well as their own, gaining experience in the practical aspects of accountancy. Firms charge their clients for the trainees' services according to their level of attainment. Thus those who have suffered examination referrals or failures or simply deferred taking their examinations, will not be as profitable for their firm since their fees will not match those of their contemporaries who have been successful.

However, Benveniste *et al* (1986) find that differences in terms of revenue associated with examination failure and withdrawal are not so great as are commonly expected, but report that the unquantifiable costs, such as disruption to staff schedules, are much greater. Unsuccessful trainees will inevitably be associated with higher examination fees and more lost hours for study, ie lower productivity. In addition, those who do not perform well in their practice work are also costly to the firm in that they do not impress clients in the right way and their inefficiency or inability to work well with others is counter-productive.

When the contracted term is over, the more desirable students will be encouraged to stay with the firm, gain their practising certificates from the ICAEW, without which they cannot undertake independent practice as chartered accountants, and remain until they are elected to partnership. Firms frequently find that those who they had hoped would remain with the firm after qualifying choose to leave and conversely, many of those they had hoped might seek new pastures choose to remain within the firm. More than half of the qualified chartered accountant members of the ICAEW work outside professional practice (O'Kane, 1989).

Although the profession generally may be seen to commit generous resources to the selection and recruitment functions, and to go about them in an apparently systematic manner, there is little evidence that any other than a very few firms of the size sampled are using any measures which are well regarded by the extant literature on good recruitment and selection practice. Harvey-Cook and Taffler (1987) find that only 16% of the firms they studied had carried out a job analysis, which is considered essential for identifying necessary candidate attributes for successful job performance (Herriot, 1987). In addition, although firms seemed to be in general agreement as to what makes a good candidate, only 8 (32%) of the firms studied indicated that they actually had a candidate profile, used as a reference item during selection.

Firms believe that candidates who at university have exhibited examination and general academic success will perform well in the professional examinations and that

those who have been team members, in particular, those who have sought and/or obtained positions of responsibility at school and university, will have the necessary characteristics to perform well in practice work. Although the research suggests that these heuristics are related to important predictors (Baehr and Williams, 1967; Glennon *et al*, 1966; Herriot, 1984), their use is not statistically governed. They are simply the traditional 'rules of thumb' for ability-related prediction.

Vacation work is examined for evidence of motivation, commitment and experience which will contribute to the overall skills of the candidate. Again, there appears to be no statistical basis for such judgements even though the literature suggests that they are important predictors. The individual is assessed in terms of specific traits. Evidence that candidates possess these traits is sought in their biographical data, provided by the application forms. Thus judgements of the applicant's social and practical skills are made upon simplistic analyses of information supplied by the application form, supported by unstructured assessment of their behaviour at interview and in group exercises (Macan and Dipboye, 1990).

Research cited above has shown that there are indeed very good predictors of future performance in the life histories of applicants (Holland, 1976; Owens and Schoenfeldt, 1979; Super, 1980, etc) but their use as predictors should be governed by unbiased statistical methods, rather than 'gut-feelings'.

3.3 Recruitment Stages

In general the recruitment process in accounting firms follows three stages and the progression through these is illustrated in Figure 3.1. During the first stage the undergraduate has access to information from his/her careers advisory service which will recommend the various lists of graduate opportunities (GO, GET etc) and may well advise perusal of the literature concerning certain professions such as chartered accountancy. The on-campus promotions provided by different firms and professions will also be available to reinforce the information in such literature.

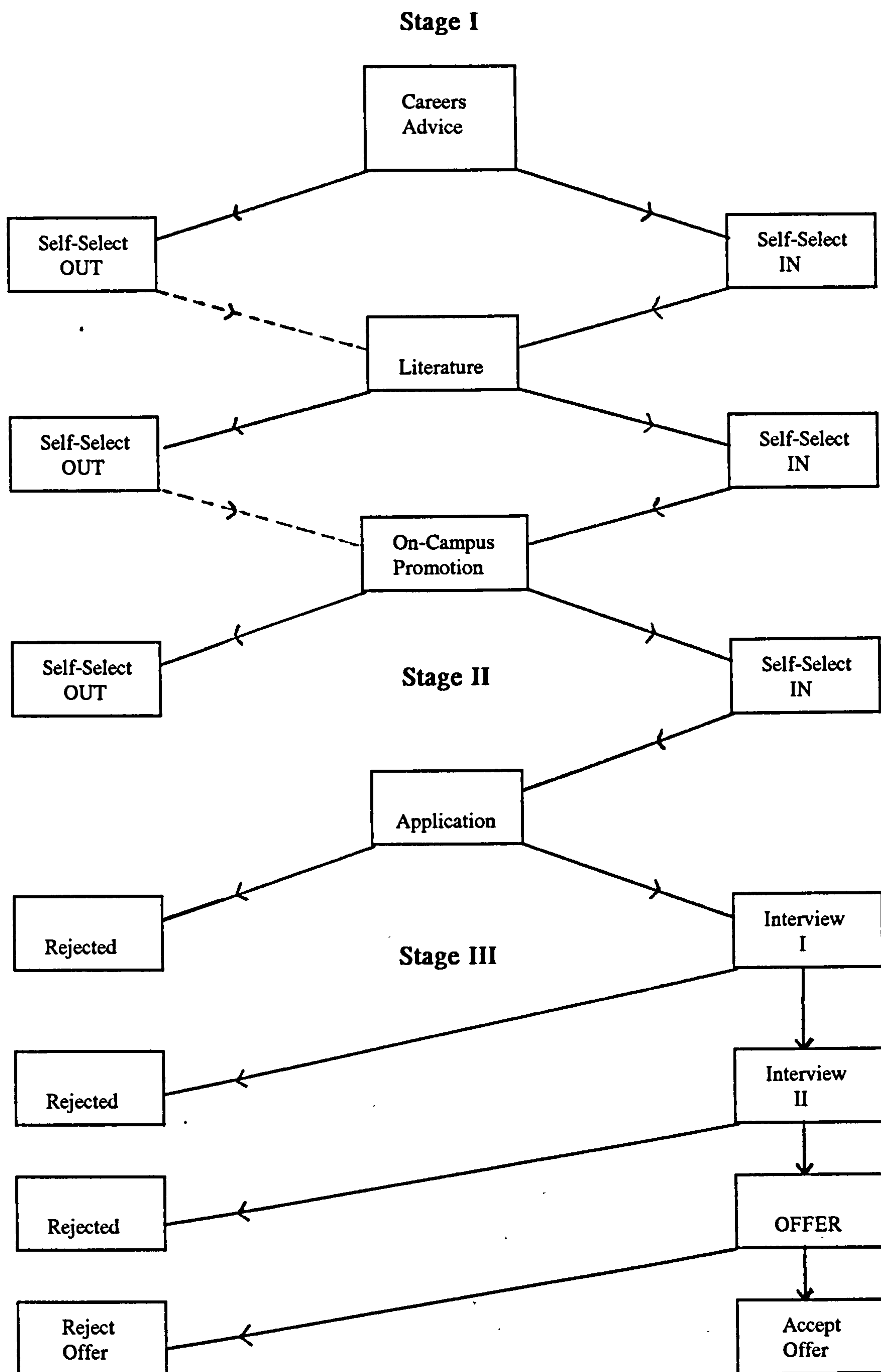


Figure 3.1 Graduate Recruitment Stages

In the light of the literature and the presentations, students make self-selection decisions and may seek confirmatory evidence of the fitness of their choices at each of the stages, particularly at meetings with firms' representatives.

Having used the information to 'select in', the student submits an application form. The second stage of the process depends on what happens to that application. If an interview results, the applicant's correctness of selection is confirmed. If no interview is secured or interviews are not successful, his/her selection decision is undermined. Obviously, the information upon which the decision was made has been misleading, or (s)he must consider himself unable to make good job choices.

Herriot's (1987) opinion of this progression is '...a strategy where very large numbers of applicants are reduced to manageable proportions in an almost entirely random but very expensive procedure.'

3.3.1 Literature and Presentations

The brochure, which is the firm's first contact with the student, is seen as a very important marketing device but, since it introduces the student to the profession, it also constitutes a job preview and thus should give an honest and clear representation of the true nature of the job (see section 1.2.1 above). Harvey-Cook and Taffler (1987) find that only 16% of their sample of 25 accounting firms of all sizes, had carried out job analyses, putting them in a position to be able to give accurate information.

Herriot (1987) emphasises the need for realistic job information as do Wanous (1977) and Keenan (1985). The ICAEW and its member firms are understandably fairly defensive about the content of their recruitment literature and feel it is a true representation of the facts (Wilson, 1989).

Firms are, however, faced with a dilemma: should they offer clear, and perhaps apparently negative information, so that recruits are really aware of

the realities of life as a trainee, and thus perhaps 'put off' potentially good candidates, or should they attempt to gloss over the more negative aspects, perhaps making the point that those with the 'right' outlook etc, will make the grade? They are understandably fearful that if their brochure is not positive enough, another firm will attract the better candidates, although Wanous (1977) concludes that giving realistic job information does not effect the firm's ability to attract candidates. Unfortunately, if incorrect or simply circumspect information has been given, the likelihood of the right candidates applying may be reduced. 'Painting too rosy a picture will build up unrealistic expectations in recruits and this in turn will lead to higher turnover and lower job satisfaction.' (Wilson, 1989)

Wilson suggests that realistic job information is not what is available to prospective trainee accountants and is directly responsible for the disillusionment which contributes to commitment-related withdrawal. The student is not free to make valid choices because he is not faced with full, accurate information and will therefore not be able to self-select 'in' or 'out' accurately (Meglino and De Nisi, 1987).

Behavioural consistency theories suggest that people are unlikely to aspire to positions or to seek entry to a profession to which they are, by dint of their background and experience, unsuited. Nevertheless, many applicants who apply to the accountancy profession are unsuitable and must have received incorrect signals from the profession. Thus recruitment errors are 'built in' to the current system before either party has approached the other.

In a study conducted by Rhode *et al* (1977) the most negative quality of day to day accountancy is seen to be dull work, other aspects commonly mentioned being work hours, budget pressures and office politics. However this study may well not reflect the current scene where sampling methods and improved technology have removed some of the more repetitive tasks from a trainee's experience, although O'Kane (1989) reports boredom as

being one of the principle factors in trained accountants' decisions to leave accountancy. It would nevertheless seem appropriate to find an acceptable way to introduce candidates to these and other negative aspects, so that those who have accepted them from the outset as being integral to the training period, but a necessary evil, can develop the necessary coping strategies.

It would also seem appropriate that such information comes on a one-to-one basis from those who are already under training contract (Herriot, 1987) if it is unlikely to form part of a brochure designed to create an attractive picture of the job. Unfortunately, those students taken on milk-round to aid the presentation are likely to be the better students who may be unable to give the candidates an unbiased view of the less attractive aspects of the work, since they have already come to terms with them. These students have a vested interest in the success of their chosen profession and must promote it positively in order to maintain personal integrity. Students will not be taken if they are exhibiting negative behaviour or failing in either examination or practice work terms.

Such changes that are made in accountancy firms' promotional information are typically the result of a marketing exercise designed to make the firm appear more attractive to applicants and may even, in fact, be responsible for attracting more unsuitable applicants.

3.3.2 Application Forms Processing

The majority of applicants to chartered accountancy firms are those in their final year of a degree course and who have been exposed to the pre-application material discussed in section 3.3.1 above. The completed application forms arrive at firms in large numbers in the autumn and gradually tail off through the winter and spring. In the firm, the applications are examined by a clerk or secretary for the ICAEW's minimum academic requirements of passes at GCSE in Mathematics and English and for the

firms' usual requirements of 9 UCCA points indicating average 'A' level performance.

When this pre-sift is completed, the forms of those meeting these requirements are passed to a more senior member of staff who decides which applicants from the pool will be interviewed. There is general agreement about the items on the application form which are important predictors of future success and there is evidence to support their use as predictors (Herriot, 1984). However, the methods adopted by firms to sift through forms vary, and, the larger the firm, the more likely this process is undertaken by more than one person, resulting in inconsistent decisions even within the same firm.

Wingrove *et al* (1984) find that applications can be rejected on the way in which they have been completed, regardless of suitability for the job and suggest that this '...resulted in inferences on the part of recruiters regarding motivation to work for the organization.'

Inevitably, when large amounts of information are presented for perusal, without specific focus, the sifter loses sight of the real task and candidates' applications may be rejected upon criteria not perhaps related to their contents (Drakeley, 1989, Wingrove *et al*, 1984). Wingrove *et al* point out that the number of variables taken into account by processors is 'greater than decision makers are normally thought capable of consciously processing'. However only two out of the 25 of the accounting firms in Harvey-Cook and Taffler's (1987) survey indicate that they had any systematic method of scoring their application forms. Thus, in the majority of firms having no scoring system, and where applicants are supplying so much information, inaccuracies are inevitable.

The academic information is treated as most important as an ability indicator and the social information is secondary, providing predictions of motivation

and commitment. Cues from other areas of the form are taken to make personal judgements of very questionable validity, eg, neatness of completion cited by Wingrove *et al* and used by 79% of Harvey-Cook and Taffler's (1987) sample.

3.3.3 Selection Interviews

Large firms, and coalition groups of smaller firms, who make on-campus presentations at those universities which have reputations for producing good candidates, will inform the careers advisers of those they wish to interview during their visit. Small firms who do not make on-campus presentations usually interview at their own offices. In both cases the candidate is usually given the opportunity to discuss aspects of the firm and accountancy generally with the firms' existing trainees.

Accounting firms, as a rule, operate a two interview system whereby those who are successful after their first interview will be invited for a second interview, with a more senior manager or partner, and, upon the basis of this, will be either offered a training contract or rejected. Some firms, however, only have one interview, accepting or rejecting at that time.

The recommended ways of increasing the reliability and validity of the selection interview are to structure the interview and perhaps make it situational (Carlson *et al*, 1970; Schmitt, 1976; Dreher *et al*, 1978; Latham *et al*, 1984). This means that the same job-related questions, based on job analysis, are asked of all candidates and the interviewer should have examples of good, average and poor answers to compare with those of the candidate. Only 16% of firms sampled by Harvey-Cook and Taffler (1987) had carried out a job analysis but 44% used an interview schedule. None had 'anchors' to guide response rating. None of the firms reports using situational interviews.

3.4 Conclusion

The chartered accountancy profession operates a three tier system of recruitment : promotion and literature, application form processing and interview(s). The system, however is very wasteful with more that one third of students leaving pre-qualification. Although all three areas might be improved by various means, there is, in reality, considerable resistance to real change as far as promotion, literature and the use of interviews are concerned. However, a useful and acceptable contribution could be made by structuring the use of the application form using appropriately developed biodata models, thereby removing the selection function from the interview. Nevertheless the mismatch problems associated with the promotional literature will remain.

Firms using biodata models to select interviewees would be assured that the high likelihood of the subsequent success of those candidates on the chosen criteria, means that the interview can be used more as an information gathering session for both parties, rather than a selection tool *per se* and, in the case of a highly rated candidate, as a selling exercise.

Chapter 4 describes the data and samples used in Part I of this study, concerned with the development and piloting of suitable biodata models for selecting applicants to the accountancy profession and the remaining chapters in the section describe the empirical derivation of biodata models, their validation and implementation.

Part I

The Empirical Derivation of the Biodata Models

4.1 Introduction

It is incumbent on the chartered accountancy profession to reduce the numbers of those unsuccessfully recruited, whether the losses are due to mis-match, examination failure, poor performance or any combination of these. A continuing high wastage rate has implications for the profession as a whole, its individual member firms and, not least, failing students. Hoskin and Steele (1991) note that failure and withdrawal rates are not changing greatly over time and that expensive recruitment mistakes are prevalent.

Part I of this research study is concerned primarily with addressing the issue of improved selection to the profession by assessing the implications of using scored biodata to select suitable **non-accounting** graduate applicants to the accounting profession and compares the results of using such methods with those of the obtained by current interview-based selection procedures.

Chapter 2 indicated that the recruitment and selection methods used by accounting firms when this study began are not highly rated in terms of validity and reliability and that the high wastage rates may be attributable to this fact. At the request of the ICAEW this research was initially directed at the selection of non-accounting degree graduates, since they constituted by far the larger part of the trainee accountant graduate entries (78% : 22% in 1988) [ICAEW 1992]. The model development sample is therefore derived from the non-accounting graduates entering medium-size professional accounting firms during 1985 and 1986 and the derived models are validated on 1987 entrants.

The size of the sampled firms reflects the ICAEW's consideration of those seen to be in most need of improved selection procedures. The firm offices providing the

trainee biodata were chosen to reflect the need to provide an adequate number of successful and unsuccessful subjects, in terms of the chosen criteria, to form sufficiently large sample groups for model derivation and validation.

This research concentrates on predicting behaviour in the first two years of the training contract, ie the run-up to and sitting of the second professional examination. ICAEW statistics (1992) reveal that the majority of wastage occurs in this period, with minimal loss in the third year. Both wastage due to examination failure and overall performance are used as criteria in developing the models.

The verifiable biographical data collected from the trainees' records, to provide the predictor data, reflects those areas which the literature has revealed to be valuable in predicting performance and tenure. However, the data pool was reduced by the multiple formats of application encountered, which led to unacceptable levels of missing data for certain items.

4.2 Selection of the Criteria

By general consensus, the firms taking part in the study identified their areas of concern, ie those which they would wish to predict when assessing likely trainees, as being

- (i) success in professional examinations,
- (ii) successful performance in practice work; and
- (iii) likelihood of leaving before termination of contract.

4.3 Criterion Groups

Two dichotomous groups are generally necessary for the development of biodata

models. Criteria for group membership are identified here to differentiate between desirable and undesirable applicants. Two separate pairs of models are developed to reflect the profession's concerns but it should be noted that the criteria definitions do not take trainee-centred criteria like job satisfaction or any post-training criteria into account. Those who do not perform well in their examinations and/or who do not gain satisfactory marks for their practice work, may well be those who are not well-matched to their firm, not necessarily the profession. In another firm, they might be successful. On the other hand there are others who are unsuccessful in one or other examination but who may well be excellent accountants in practice and be well satisfied in their work.

4.3.1 Examination Performance Prediction Models

The choice of model development groups for these models reflects the most common cause of withdrawal from training recorded in firms' records: examination failure. Whereas the examination performance models are developed to predict successful professional examination performance versus failure, firms' concerns are more directed towards preventing loss of revenue. Table 4.1 details the nature of the two groups and their numbers and intermediate groups.

The two dichotomous groups are:

- (i) those who pass both the GCC (Foundation) and PE1 (Intermediate) examinations at the first attempt within two years; and
- (ii) those who fail or are referred in both, or fail or refer GCC at the first attempt and then leave the firm, or do not catch up with their contemporaries and pass PE1 by the end of the two year period.

The first Professional Examination is chosen as the criteria, rather than both PE1 and PE2 (Final) for two reasons:

Table 4.1

Model 1 : Examination Prediction Samples

Criterion Group	Development Sample (1985/6)	Validation Sample (1987)	
Pass	134	48	
Fail	106	90	
Intermediate	<u>202</u>	<u>85</u>	
	<u>442</u>	<u>223</u>	
Group Details			
Group	Details	1985/6	1987
Pass	Pass both GCC and PE1 at first at first attempt within two years	134	48
Fail	Fail or refer in both GCC and PE1	38	24
	Fail GCC and leave	39	27
	Fail GCC first attempt and PE1 not taken within 2 years	<u>29</u> 106	<u>39</u> 90
Intermediate	Pass GCC at first attempt and PE1 not taken	6	4
	Pass GCC at first attempt and leave	17	10
	Fail GCC first attempt but pass PE1 within 2 years	16	22
	Leave before GCC	14	15
	Pass GCC first attempt but fail PE1	126	34
	Pass GCC and PE1 at first attempt but after 2 years	<u>23</u> <u>202</u>	<u>0</u> <u>85</u>
		<u>442</u>	<u>223</u>

- (i) Once PE1 is passed, there is little subsequent wastage. ICAEW statistics for the loss, prior to PE1, of the non-accounting graduate entries 1985-1989 are 31%, 31%, 31%, 32%, 32%, but for post PE1 the same entries suffer 5%, 4%, 0.4%, 0.5%, 0.9% loss.
- (ii) Firms are more relaxed about their students' performance on PE2 than on PE1.

The latter relates to the lack of material differences in hourly charge-out rates for successful and unsuccessful third year students, compared with the more significant differences at earlier stages of training.

4.3.2 Recruitment Success Prediction Models

With adequate quality and quantity of practice work performance data, it should have been possible to develop parallel models to predict performance alone, with the added advantage of having a continuous criterion variable. This is not the case, as for reasons given below in section 4.5.2, pure work-related performance data is not sufficiently reliable to warrant separate treatment. However, those who drop-out, for whatever reason, represent a group which it is very desirable to identify as they are clear recruitment failures. Thus the second pair of models developed focuses on predicting:

- (i) those who will drop-out (those who leave their firm for any other than uncontrollable reasons within 2 years) and who thus represent clear **recruitment failures** and
- (ii) those who combine examination success, as defined above, with good practice-work ratings for their first six months in the firm (there were inadequate cases with a 12 month rating for this purpose). These trainees may be considered **recruitment successes**.

The details of the two groups of subjects and the intermediate cases are given in Table 4.2.

Table 4.2

Model 2 : Recruitment Prediction Sample

Criterion Group	Development Sample (1985/6)	Validation Sample (1987)
Recruitment Success	85	29
Leaver	83	42
Intermediate	<u>193</u>	<u>99</u>
	<u>361</u>	<u>170</u>

Group Details

Group	Details	1985/6	1987
Recruitment Success	Pass both GCC and PE1 trainees at first attempt with average or above appraisal rating at six months	85	29
Leaver	Leavers within the first 2 years of training contract	83	42
Intermediate	Fail both GCC and PE1 but still with the firm after 2 years	26	1
	Fail GCC but pass PE1 at first attempt	11	1
	Fail GCC first attempt and PE1 not taken	9	56
	Pass GCC first attempt and PE1 not taken	5	7
	Pass GCC first attempt and fail or refer PE1	98	25
	Pass GCC and PE1 first attempt but poor appraisal rating	21	9
	Pass GCC and PE1 first attempt but after 2 years	23	0
		<u>193</u>	<u>99</u>
		<u>361</u>	<u>170</u>

4.4 Sample Details

This study is primarily concerned with the medium-sized offices of professional chartered accountancy firms who may be perceived to be at a competitive disadvantage, compared with the large firms (≥ 100 partners), when attracting recruits. The latter are usually able to commit large resources to the personnel function and are thus able to initiate research on an individual basis. However, smaller firms will not be able to commit such resources and are unlikely to have such well-established personnel functions. They are also recruiting, in many cases, from those who have not been accepted by larger firms, or who are aware that they do not possess the level of qualification required by larger firms.

The criteria used for the selection of participating firms were

- (i) size of firm and
- (ii) total number of students under contract by firm office.

The 1987 ranking by fee income (The Accountant, 1988) highlights a clear scree, separating the then largest 12 firms from the others, in terms of size, with the remaining firms all with fee income below £40m. For the purposes of this research only firms with 1987 turnover below this figure were considered.

Because of the need to ensure an adequate pool of subjects within each of the three sample years from a manageable number of firms' offices, only those with a minimum of 20 students under training contract were considered, since at any time at least three entry years of trainees will be represented.

The resulting group consisted of 27 offices from 26 different firms and were identified with the collaboration of the Education and Training and Statistics departments of the ICAEW. Twenty-two of these firms agreed to co-operate with the research, one of which had two offices which were included in the sample. The participating firm offices chosen were mainly centred in or around London, although 4 firm offices were out of London in Hull, Basingstoke, Canterbury and Norwich.

A list of participating firms may be found in Appendix E.

In order to be able to examine as recent a sample as possible and yet be able to follow through to two complete years of the training contract, the entrants to the firms for the years 1985, 1986 and 1987 are studied in Part I.

Since the full examination results were not available for the 1987 entrants at the beginning of the study, the first two years (1985/6) form the development sample for the predictive models and the last year (1987) is used to validate the using the examination results when they became available. The final sample of 665 subjects consists of 442 1985 and 1986 entrants and 223 1987 entrants.

4.5 Data

Biographical data models, although robust and extremely useful, have suffered criticisms of pure empiricism and the 'failure to relate results to meaningful dimensions of behaviour' (Baehr and Williams, 1967). The Baehr and Williams study develops a 15 factor explanatory model accounting for 66.6% of the variance, in an attempt to expose the underlying relationships between biodata and occupational classification, which may be used as 'meaningful and significant background parameters for the general, male, employed population.'

Their factors 1, 2, 9 and 11 are concerned with school and higher education (school achievement, higher education achievement, school activities and educational vocational consistency). Factors 6, 7, 8 and 10 reflect family background (early family responsibilities, parental family adjustment, situational stability and professional, successful parents). Factors 3 and 4 are 'drive' and 'leadership' and group participation and factors 12, 13 and 14 are work-orientated (vocational decisiveness, vocational satisfaction and selling experience). Finally, factor 15 is good health.

Baehr and Williams suggest that, although the first 8 factors are those which make a satisfactory model, attention should also be drawn to the 5 uncorrelated second order factors from their analysis (educational background and achievement, upward mobility and drive, personal-social leadership, financial achievement and background and stability and status quo orientation). They suggest that these may represent broad behaviour patterns which present potential for the operational use of biodata.

Indeed, these factors may be seen to represent the areas that most recruiters feel their traditional procedures are identifying and are reflected in the twelve categories of Glennon *et al*'s (1966) Catalog of Life History Items and the choice of item category in Helmreich *et al* (1973) and many other studies (Matteson, 1978; Herriot, 1984; Childs and Klimoski, 1966; Bethel-Fox *et al*, 1988). However, many of the areas covered in the Glennon *et al*'s catalogue of items relate to unverifiable items used to provide questions for developing self-report Biographical Inventories. Baehr and Williams, on the other hand, stress that they use 'quantifiable personal background data items.'

Dunn and Hall (1984) in a US study specifically concerned with CPA examination performance, choose their items from various studies that are all similarly focused. A candidate's Scholastic Aptitude Test (SAT) score is used to indicate aptitude, grade point average in academic study to reflect academic accounting ability topics, public accounting knowledge to reflect practical experience and hours of self-study to assess motivation.

Examination performance is clearly related to academic ability and aptitude and measures are available in the form of results from previous examinations and such information as academic prizes awarded etc. Therefore, in this study, all data relating to academic performance was collected from the trainee's records.

However, motivational aspects which interact with academic variables to predict acceptable performance, versus dropping-out, may be indicated accurately, but perhaps less obviously, by responses to items concerning levels of social activity and

responsibility, available on application forms and verifiable. The seemingly more obvious, but unsubstantiable responses to item categories, such as those found in Glennon *et al*'s (1966) list, may only apply to the respondent at the time of response and may change through experiential effects (Owens and Schoenfeldt, 1979 and Davis, 1984; Mumford *et al*, 1990). In addition, as the nature of day-to-day experience of employees within an organization changes, the response categories may become redundant or inaccurate as predictors. This will have an adverse impact on the reliability of models based on such data, accounting for much of the shrinkage experienced.

As this study is concerned with constructing models from which applicants' scores may be verified, the data list is only concerned with items which are hard and verifiable and are available on the generality of standard graduate application forms. Nevertheless, all the items collected reflect the first four factors of Baehr and Williams's analysis and all of the second order factors.

The data were collected on templates (Appendix A) and the 23 firm offices were all visited between January and April 1989. All available firm records were analysed. As a result, data from 442 subjects from 1985/1986 and 223 from 1987 subjects was collected. Firms were contacted again in January 1990 to provide PE1 examination results for the 1987 entrants which had by that time become available.

4.5.1 Predictor Data

The template of original data items collected, given in Appendix A, was moderated by considerations as to what impact certain variables might have on minority groups (Cascio *et al*, 1988) and by the desire to restrict the information to verifiable data for reasons given in section 2.8 and above.

Unfortunately, since Part I of the study involves 22 firms with their own individual application forms and accepting standard application forms and cvs, there were missing items in several cases, eg parental occupation,

school based social activities and driving licence.

Table 4.3 summarises the predictor and criterion variables by category type. Of the original list of 67 predictor variables, 40 had to be removed from the analysis, due mainly to data availability problems, a few being those which might be considered to be racially or sexually biased. The reduced lists of variables may be found in Appendix B.

4.5.2 Criterion Data

England (1971) emphasises that criterion variables should be measured as carefully as possible since '...the entire study is dependent on the adequacy and accuracy of the criterion. The most the WAB can do is to predict the criterion used in its development.' He draws attention to the recommendations of Blum and Naylor (1968) who list characteristics which are necessary and/or desirable in any criterion:

1. Realistic
2. Understandable
3. Relevant
4. Representative
5. Related to other criteria
6. Acceptable to job analyst
7. Acceptable to management
8. Inexpensive
9. Constant from one situation to another
10. Reliable
11. Predictable
12. Measurable
13. Uncontaminated and bias-free
14. Discriminating

Professional examination results represent a fairly standard method of categorising trainees, since all trainees have been subject to the same syllabus. In the case of the GCC examination, which is set and marked by

Table 4.3

Predictor and Criterion Variable Categories

Predictor Variables	Number of Items
Academic (eg. number of 'O' level GCEs)28	
Employment (eg. number of jobs while at school)	6
Home Life (eg. area of family residence)	6
Social Involvement/ Responsibility (eg. school prefect, team captain)	16
Personal (eg.marital status, driving licence)	11
<hr/>	
Criterion Variables	
Professional Examination Progress	9
Withdrawal Record	9
Summary Practice Work Record (overall categories)	2

private training firms, the results are monitored by the ICAEW. Professional Examinations 1 and 2 are set and marked by the ICAEW. As such, they can be seen to conform with Blum and Naylor's list.

Although all the recruits used in Part I of this study were non-accounting graduates, some degree disciplines do offer exemptions from the various elements of the GCC examination and the amount of exemption was collected as a predictor variable.

Subjects' performance in the GCC examination and both Professional Examinations were recorded. The results of a possible three attempts at each examination were collected (pass, refer, fail) and this involved the firms being contacted at regular intervals after the original data collection was completed, to update the data in line with the most recent examination sitting results.

The problems arising from the use of so many different firm offices became apparent when taking performance related information from different record sources. The many and diverse rating systems characteristics, in effect, certainly preclude items 9 and 10 and perhaps item 13 in Blum and Naylor's list.

In general, firms make the students responsible for their own files and for collecting their performance ratings. This results in a wide disparity of quality and quantity of information. Where the performance is rated on a job-by-job basis, the student may have little chance to chase up the relevant senior or manager when the assignment is complete, since both parties will then move on to another. Inevitably, there will be missing ratings and even where it is mandatory for the student to collect them, they may not have been completed immediately after the job and their accuracy will reflect the length of the intervening period. In addition, since none of the firms studied had any form of behavioural anchors to guide the assessor, inter-rater

reliability is not a feasible expectation.

When data collection was completed, it became obvious that, although for some firms in-depth categorical performance ratings are available and up-to-date, this is by no means sample wide. Although the ICAEW do monitor the progress of students by means of a six monthly progress appraisal form, there were no guidelines covering the completion of the various categories of practice-related behaviour on it and again inter-rater reliability is likely to be poor. However, since this form at least offered some standard means of comparison between students in different firms, being used by all firms for their trainees, it is used here in preference to the firms' own assessment forms, where it can be found.

This unfortunately smacks of 'throwing the baby out with the bathwater' but is unavoidable since it is more-or-less impossible to generalize between the different items and rating categories in use in the different firms. Where the ICAEW form is not available, the average overall rating for the first six months is used, where that is available.

Although there is general agreement among the staff partners concerning the importance of monitoring performance, this is belied by the way it is actually recorded and monitored. A certain amount of shrinkage in the number of performance ratings available is to be expected, due to leavers records being incomplete, but it is surprising that there were not sufficient ratings available for use at the 12 month stage, nor for any 6 month interval thereafter. This reflects the casual way in which performance-related data is treated in the size of firm studied, and is mystifying, since salary in many firms is performance related.

The result of this casual attitude is that the criteria for the prediction of overall performance model relies upon the only rating that is adequately represented for the majority of subjects, ie that made 6 months in to the

contract.

Overall assessment ratings are used, rather than the various categorical ratings, since this is felt to represent a general picture of the subjects' performance, within the given period. In the event, subjects at 6 months into their training are unlikely to have had a great variety of practical experience, having spent much time in examination preparation and on training courses. However, it is during this period that ORS will have maximum effect and Guion (1992) refers to it as the 'honeymoon period'.

Furthermore, the rating scale used on the ICAEW form is a four point scale (presumably to avoid central tendency) with no obvious average. However, after discussion with firms' recruitment partners, it was concluded that the upper two categories would be considered to be above average and average, while the lower categories would be considered to be below average and very poor. Even so, some partners had completed these forms by creating a box between the second and third categories to make an average!

Nevertheless it is felt that this study can justifiably use an overall rating, from whatever source, since the criterion is average or above average performance, coupled with a successful examination record. Although there are likely to be differences in opinion between ratings of performance of those rated average or above, there is likely to be more-or-less total agreement concerning behaviour which falls short of 'acceptable', ie is below average or in need of improvement.

An ICAEW recommendation concerning performance rating and associated record keeping during the training contract, particularly with regard to scales for performance rating is very desirable. It is not clear what use if any is made of the performance rating by the ICAEW.

It is clear that, with such margins for error incorporated within the

performance ratings, the likelihood of models based on such unreliable data maintaining validity is low.

4.6 Summary

This chapter describes the samples of graduate trainees selected and the data collected to develop and validate biodata recruitment models for medium size accounting firms. Problems associated with missing data have been discussed.

The final sample consists of 665 non-accounting graduate trainees entering 23 offices of 22 different firms during the years 1985-7. The first two years are used for model building and the last for true ex-ante predictive tests.

The predictor data are all collected from the application forms or cv's of the entrants and the criterion variables, reflecting firms recruitment concerns, relate to professional examination records, tenure and overall success within the firm. There was a surprising paucity of data relating to the latter available in firms, with no consistent approach to scoring performance within the profession apparent. The poor quality of such data in terms of inter- and intra-firm rater agreement was also noted.

Chapter 5 discusses the two methodologies used to derive the models, viz, the simple, non-parametric WAB approach of England (1971) and the more complex, parametric technique: logistic regression (logit). The derived models and their predictive validity are discussed in Chapter 6 and the results of validating the developed models on the 1987 validation sample are reviewed. A comparison is then drawn between the validation results of the two methodologies and conclusions drawn about the relative efficiency and operational utility of the two modelling approaches.

5.1 Introduction

There has been some debate concerning the appropriate treatment of biographical data in the literature over the last decade or so. The most commonly cited empirical method is that of England (1971) described in section 5.2 below.

Mitchell and Klimoski (1982) argue that 'empirical methods' such as the WAB are probably able to produce models of both greater predictive power and validity than more 'rational' approaches and that the former should be used, especially since, at the time of their report, they were easier and cheaper to develop and could be produced by any personnel manager, without recourse to an independent consultant (England, 1971). Their work has had a great deal of influence on recent thinking.

However, there are serious methodological problems associated with their own study. The regression function of their 'rational' method was derived using their subjects' scores on six components, drawn from factor analysis, as independent variables. This may be criticised on the grounds that factors are not real but are experimental artifacts. In addition, the items initially submitted to the factor analysis were derived from 4 *a priori* life history *constructs* categorised upon **hypothetical continua** which cannot be said to directly equate to the items and categories used in the contrasting empirical method, the WAB (England, 1971).

Moreover, 69% of the information in their sample was removed by the decision to use only 6 of the components (presumably with eigenvalues > 1), although they cite the 31% of variance explained as being comparable with two other biodata studies : Morrison (1962) - 5 components explaining 23% of the variance - and Baehr and Williams (1967) - 8 components explaining 43.3% of the variance.

Given the problems with their statistical methodology, the complexity of the 'rational' methods employed by Mitchell and Klimoski in their efforts to develop Life History Dimensions based on a *theoretically meaningful set of constructs*, and the subsequent changes in data scaling, it is hardly surprising that they achieve such different derivation and cross-validities.

This research makes direct comparison between the non-parametric WAB approach and the parametric logit procedure and this forms an important part of this thesis. The methods are comparable in cost terms since both may be easily adopted using personal computers, although the multivariate methodology does require some specialist knowledge to modify and interpret component variables. Since the data are all factual, a viable comparison can be made between the two methodologies tested.

5.2 The Weighted Application Blank Technique (WAB)

In general terms, the methods which are described as empirical are those which make direct comparison between the two chosen dichotomous groups, without undue concern *a priori* over whether the derived model variables can be adequately explained in terms of their relationship to the criterion. The standard approach is the non-parametric Weighted Application Blank (WAB) technique of England (1971).

This technique involves the development and use of a scored biodata inventory. Predictor variables are polychotomous and the criterion variable dichotomous. The methodology does not require the assumption of linear relationships in the data as do the conventional statistical approaches. The score derived from the model is a linear composite of weights which represents the degree of difference between the criterion groups' members on all predictor variables which show significant differences between criterion group responses.

Typically, a count of the percentages of the two model groups falling within the response categories of a group of suitably chosen biographical items is made. The differences between the model groups' percentages is calculated for each variable and the differences weighted according to tables (Stead & Shartle, 1940).¹ England is clear upon the point of missing information :

'...in cases where there are persons who have given no answer, or inadequate answers on a particular item, the usual procedure is to eliminate these persons from the calculation of weights for that item and thus to work with a reduced number of subjects.' (p28)

Missing cases are thus not included in the percentage calculations given here, however, it should be noted that Mitchell and Klimoski (1982) replaced their missing values with means. The 'net weights' may then be converted into 'assigned weights' for analysis, which is recommended by England, as this is essentially a practical methodology, designed for ease of use.

The assigned weights range from 0 to 2. The 0 assigned weight reflects a significant negative net weight of -4 or less, an assigned weight of 2 indicates a significant positive net weight of +4 or more, while an assigned weight of 1 indicates an intermediate, and therefore less strong, net weight of +/- 1, 2 or 3. The table of assigned weights created by Strong and printed in England (p.28) has been modified to weight fewer chance differences between the rating groups, but information on the establishment and correction of these tables is not available.

Those items which fail to discriminate between the two groups are discarded. The weights are used to construct a linear additive model for scoring subjects and the accept/reject decision is based on the score distributions of the groups, interpreted in the light of the firm's recruitment needs and their varying costs of misclassification. Examples of the scoring tabulation may be found in appendices C and D.

¹ Although these tables are quoted in England (1961,1971) and Guion (1965), no record of the original source could be found anywhere.

England advocates applying the derived model to the hold-out sample to set the appropriate cut-off point, since model performance will be biased upwards when applied to data from which they were derived, due to this 'sample-bias'.

Weiss (1976) argues that, since approaches as the WAB tend to make fewer statistical assumptions and require fewer parameter estimates, they capitalise less on chance, compared with multivariate techniques, and are more likely to be stable in cross-validation than other approaches.

Models derived by this method may appear to work well (Lee and Booth, 1974; Dawes and Corrigan, 1974) but the weighting system is unsophisticated and offers little in the way of explanation of the contribution made by each of the component variables, indeed, this is one of the most important criticisms of such approaches.

5.3 Multivariate Statistical Methodologies

Although the study quoted (Mitchell and Klimoski, 1982) favoured factor analysis followed by multiple regression of the component scores, this was the result of the need to reduce the data generated to interpretable ('meaningful') constructs, ie, to take a rational approach. However, choosing the variables from the data pool which are known to be good predictors of future success in criterion terms may also be considered as taking a rational approach.

There has been discussion in the literature concerning the most appropriate multivariate methodology for models such as those used in biodata studies, which use dichotomous dependent variables, and the most likely procedures are linear discriminant analysis (LDA) and logistic regression (logit).

5.3.1 Discriminant Analysis

LDA is probably the multivariate technique which most closely resembles the

WAB. It is, in essence, a weighted average technique. 'Linear combinations of independent variables are formed and serve as the basis for classifying cases into one of the groups' (Norusis, 1990).

The LDA equation takes the form :

$$D = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots$$

where β_0 is the constant, $\beta_1, \beta_2 \dots$ are the coefficients estimated from the data and $x_1, x_2 \dots$ are the values of the independent variables. The coefficient weights are estimated so that they result in the "best" separation between the training sample groups (Norusis, 1990) ie, the ratio *between groups sum of squares : within groups sum of squares* is maximized.

Unlike the WAB model, an estimation of the contribution of a single component variable to the model may be derived by the use of the Mosteller-Wallace technique.

5.3.2 Logistic Regression (logit)

Logistic regression is used to directly estimate the probability of an event occurring. The technique requires fewer assumptions of normality and equality of variance and covariance than discriminant analysis and the resulting score (the logit) falls within the range 0-1.

The logit model takes the form :

$$p = \frac{1}{1 + e^{-z}}$$

Where p is the logit score, providing a measure of the probability of the event occurring in the range 0 to 1 and z is given by

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \dots$$

where β_0 is a constant, $\beta_1, \beta_2 \dots$ are the coefficients or weights and $x_1, x_2 \dots$ are the predictor variables.

The odds of an event occurring (*probability of the event occurring / probability of the event not occurring* or $p/(1-p)$) may be written as

$$e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots}$$

The *logit* is the log of the odds ratio, ie:

$$\begin{aligned} \log \frac{\text{Prob (event)}}{\text{Prob (no event)}} \\ = \log \left(\frac{p}{1-p} \right) \\ = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots \end{aligned}$$

Unlike multiple regression, in logistic regression the relationship between the dependent and independent variables is non-linear. In multiple regression, the regression co-efficient represents the amount the dependent variable changes in relation to a one unit change in the independent variable. In logistic regression the relative importance of variable x_j is measured by $\exp(\beta_j)$ which denotes the factor by which the odds ratio $(p/(1-p))$ increases for a one unit change in that variable.

The statistical significance of a variable in the model may be tested using the Wald statistic (Maddala, 1988), which has a χ^2 distribution, and is the square of the ratio of the β to its standard error.

The R statistic may be used to examine the partial correlation between a predictor variable and the dependent variable. Norusis (1990, pB42) gives the equation for R , which falls within the range -1 to +1, as

$$R = \sqrt{\left(\frac{\text{Wald statistic} - 2}{-2LL} \right)}$$

Where -2LL is -2 x the log likelihood of the model, described below.

Variable Selection

As the relationship between predictor and criterion is non-linear, an iterative approach is required for parameter estimation. Using the Logit procedure in SPSSPC+ with a forward stepwise approach and the likelihood ratio (LR) [*likelihood for the reduced model/likelihood for the complete model*] to determine which variables should be deleted from the model, all the variables are entered at the first stage.

Each variable is then eliminated in turn. The change in the log likelihood is then examined and if the variable's contribution to the model is not significant, ie the coefficient for that variable is not significantly greater than 0, the variable will not remain in the model and the next variable will be considered. The process ceases when all variables have been examined and no further variables are eligible for entry or removal,

Assessing Model Fit

The statistic representing the difference between the log likelihood statistic for the model with a constant term only and that for the model containing the significant independent variables, and known as the model χ^2 , provides the equivalent of the overall F test in multiple regression (Norusis, 1990, pB46). This follows the χ^2 distribution with m degrees of freedom, where m is the number of parameters estimated.

Due to the binary nature of the criterion variables, in this study problems arise in comparing model performance with studies using more conventional modelling methods which report *validity coefficients* in terms of the Pearsonian product moment correlation coefficient calculated between predictor score and criterion. This statistic cannot be validly computed where either of these variables is dichotomous.

A biserial correlation coefficient, r_b , provides comparison with Pearson's r , but its use depends on there being a true dichotomy in the data. There is some question as to whether the variables in this study are fundamentally dichotomous, even though they are treated as such in the analysis. Guilford and Fruchter (1978, p311) advise 'If there is a little doubt that the distribution is a genuine dichotomy, r_{pbi} should be computed and interpreted. When in doubt, the point biserial r is probably the safer choice.' Therefore, the latter, more stringent statistic is provided here to allow more direct comparison with other studies.

The formula for the point biserial correlation statistic is

$$r_{pbi} = \frac{\overline{X}_p - \overline{X}_q}{\sigma} \sqrt{pq} \quad (p309)$$

Where \overline{X}_p = mean of X values for the higher group
 \overline{X}_q = mean of X values for the lower group
 p = proportion of cases in the higher group
 q = proportion of cases in the lower group
 σ = standard deviation of the total sample in the continuous variable X

Note that Guilford and Fruchter emphasise that the point biserial correlation, r_{pbi} , is far more stringent than the biserial correlation, r_b , and may lead to a conservative estimate of Pearson's r .

5.4 Methods Adopted in This Study

Efron (1975) shows that where the underlying assumptions of LDA hold, ie the data is distributed multivariate normal with equal covariance matrices, LDA performs better than logit. However, when these assumptions are violated, particularly as was the case with this research, where one or more of the independent variables are binary, problems may result, at least in theory (Press and Wilson, 1978). However Gilbert (1968) compared LDA and logit, specifically in the case of binary variables, and demonstrates little actual difference in the resulting models.

Such models are obviously prone to search bias, resulting from searching in a subset of variables to gain the best fit for the model, and sample bias, which results in the model fitting the data sample rather than the underlying population. The former problem is inherent in the development of all predictive models but the latter may be overcome by applying the model to a separate validation sample or by using a hold-out test (jack-knife technique), eg, the Lachenbruch Test (Lachenbruch, 1967). Such techniques remove each case in turn, fitting the model to the remainder. The examination of the misclassification rate for the held-out cases gives an indication of the degree of sample bias and adjustments may be made where necessary.

Two different functions were fitted to the data described above :

- (i) the linear additive non-parametric (WAB) and
- (iii) the logistic regression (logit).

The resulting models, the results of their application and the validation results are reviewed in Chapter 6. The following sections describe and discuss the two approaches in detail as they apply to the data and samples described in Chapter 4.

5.4.1 WAB Models

England (1971) is specific in his instructions and these were followed precisely. Two suitably dichotomous groups of subjects are selected to

represent the desirable and undesirable employees, these are the criterion groups. Within each of the two criterion groups the subjects are split randomly into two sub-groups, in this case, a model development group of 66.6% and a hold-out group of 33.4%.

The numbers in the two criterion subgroups (desirable and undesirable employees) and their corresponding hold-out groups should be above his indication of the lowest acceptable size for these groups (50 and 25 respectively) and this study conforms with his recommendations.

A random sample of 66.6% of the 1985 and 1986 subjects is used to develop the linear additive models. The remaining 33.4% are used to form a hold-out sample and the 1987 subjects are used as full validation sample, for all developed models.

Unfortunately, the nature of the GCC examination changed for the 1987 groups and this resulted in a lower pass-rate for that examination. The implications of this are discussed later.

The resulting groups are an Examination Pass Group of 89 and a Fail Group of 67 (total 156), comprising the model development groups for the examination success prediction model, and a Recruitment Success Group of 85 and a Recruitment Failure (Leaver) Group of 83 (total 168), for the recruitment success prediction model.

The frequencies in the variable categories for the Pass/Fail and Success/Leaver groups of the 67% model derivation sample were examined and, where appropriate, ie, where extremely small numbers of cases fell within certain item categories, truncations were made. The percentages of the desirable and undesirable groups' responses in the various predictor categories are tabled in Appendices C and D with the associated assigned weights.

Assigned Weights (England, p.28) are used, as England suggests, rather than the net weights used by Mitchell and Klimoski (1982), who suggest that the net weights may be more sensitive than the assigned weights although do not expand upon this.

The resulting 'significant' variables and assigned weights for the two models are given in Tables 5.1 and 5.2. All items which successfully differentiate between the two criterion subgroups, according to the tables, are retained. However, there is no indication of whether these items are statistically significant.

The next stage of the WAB procedure is to score the hold-out sample on the weighted variables to produce a total score and then examine the distribution of scores for these groups in order to select appropriate cut-off points. The point at which the maximum difference between groups is found is used in conjunction with histograms of the two groups' scores, to select the cutting score. In one aspect England is deficient since there is no indication of how to treat cases with missing data items when scoring them.

There is the possibility that missing values should be replaced with the mean of other cases or discarded. The advantage of replacing with the mean is the maintenance of the group size, but this may be offset by the disadvantage of score distribution distortion, particularly where there is an almost binary score structure in the variables which have a score of either 2 or 0. In the logit model-building process, subjects with missing data are not used and thus it was decided to score only those with full data on the WAB items. Naturally this resulted in a reduction in group sizes. These problems are discussed further in Chapter 6.

The final step in the model development process is to use a separate sample to assess the validity of the derived models and, if possible, make an assessment of their reliability. The validation sample is drawn from the

Table 5.1

***Model 1 : Examination Performance Prediction (WAB)**

Item	Category	Assigned Weight
Number of jobs at university	<2	2
	>=2	0
Accountancy related jobs at university	No	0
	Yes	2
Information Technology skills	Below average	0
	Average or above	2
Number of secondary schools attended	1	2
	> 1	0
Number of 'O' levels passed	<=7	0
	>7	2
Number of grade A 'O' levels	<4	0
	>=4	2
Number of 'O' level Maths/Science/Tech	<4	0
	>=4	2
UCCA Points	0-9	0
	>9	2
Number of 'A' level Maths/Science/Tech	<2	0
	>=2	2
Number of 'A' level Arts/Languages	None	2
	Any	0
Degree discipline	Maths/Science/Tech	2
	Other	0
Responsibility club/soc at university	None	2
	Any	1
Small group active pastimes at university	None	2
	Any	0
Small gp non-active pastimes at university	<2	0
	>=2	2

** Model developed on 67% sample of 1985/86 entrants*

Table 5.2

Model 2 : Recruitment Success Prediction (WAB)

Item	Category	Assigned Weight
Area of residence	London and S.E.	2
	Other	0
Number of jobs at school	<2	2
	>=2	0
Information Technology skills	Below average	0
	Average or above	2
Number of 'O' level Maths/Science/Tech	<4	0
	>=4	2
Studied humanities at 'O' level	Yes	2
	No	0
Number of 'A' level Maths/Science/Tech	<2	0
	>=2	2
Number of 'A' level Arts/Languages	None	2
	Any	1
Degree discipline	Maths/Science/Tech	2
	Arts & Combined	1
	Social Sci/Humanities	0
Degree class	I/IIi	2
	< IIi	0
Number of sports teams at school	<2	1
	>=2	2
Number of sports teams at university	0	1
	1	0
	>1	1
Level of responsibility at school	None or Prefect	1
	Headboy/girl/Monitor	2

** Model developed on 67% sample of 1985/86 entrants*

1987 entrant population.

5.4.2 Logit

The full criterion groups and subgroups of 1985 and 1986 entrants and the same variable set were used to develop the logit models. The logit models also differ in the number of subjects within the model groups, since those cases with missing data items have been removed from the analysis.

Using SPSS.PC+ 3.1, all variables with less than 10% missing cases were submitted in a forward stepwise logistic regression procedure, using the likelihood ratio statistic as the test for inclusion or removal from the model (section 5.3.2). The component variables resulting from this process were then submitted to analysis again and forced into the final model. The model χ^2 is used to assess the fit of the models.

The two logit models resulting from this process are outlined in Tables 5.3 and 5.4, which give the β , Wald statistic, partial correlation statistic (R) and the $\exp(\beta)$ for each component variable, in addition to the model χ^2 statistic. The models are discussed and interpreted in Chapter 6. Both exhibit levels of significance sufficiently high to reject the null hypothesis that the models differ significantly from the perfect model.

5.5 Summary

This thesis examines the likely approaches to biodata model development and specifically compares the recommended and commonly used Weighted Application Blank (WAB) technique (England, 1971) with more recent multivariate statistical procedures available on personal computer packages.

Both the non-parametric WAB procedure and two of the most common multivariate

Table 5.3

Model 1 : Examination Performance Prediction (Logit)
Significant Variables and Associated Statistics

Variable	β	Standard Error	Wald Stat.	Significance	<i>R</i>	exp(β)
Number of grade A 'O'levels	0.37	0.08	21.62	0.00	0.25	1.45
First/II.1 degree (0/1)	1.19	0.32	13.50	0.00	0.19	3.29
Number of Art/ Language 'A' levels	- 0.69	0.19	12.88	0.00	- 0.19	0.50
Science/ Technology degree	0.86	0.33	6.75	0.01	0.12	2.36
Headboy/ headgirl (0/1)	1.87	0.79	5.64	0.02	0.12	6.50
Independent School (0/1)	0.82	0.36	5.25	0.02	0.10	2.27
Constant	-1.64	0.39	17.32	0.00		

Model $\chi^2 = 71.93$, with 6 degrees of freedom, $p < 0.0001$

Table 5.4

Model 2 : Recruitment Success Prediction (Logit)
Significant Variables and Associated Statistics

Variable	β	Standard Error	Wald Stat.	Significance	R	$\exp(\beta)$
Number of 'A' level Sci/Tech subjects	0.61	0.16	14.81	0.00	0.25	1.84
Headboy/girl (0/1)	2.36	0.87	7.31	0.01	0.16	10.54
Number of teams and societies at school	0.43	0.18	5.56	0.02	0.13	1.53
Percentage exempt from GCC Examination	2.53	1.10	5.30	0.02	0.12	12.51
First/ II _i Degree	0.87	0.40	4.77	0.03	0.11	2.39
Number of solo active pastimes at university	-0.46	0.23	4.03	0.05	-0.10	0.63
Constant	-2.14	0.53	16.61	0.00		

Model $\chi^2 = 44.08$, with 6 degrees of freedom, $p < 0.0001$

parametric model formulations, LDA and logit, are described and discussed in this chapter. For reasons relating to the binary nature of many of the component variables, discriminant analysis is rejected in favour of logit, which provides the probability of the successful outcome for the scored cases.

The dichotomous criterion variables necessitate the use of an estimated validity coefficient to provide comparison with other studies. The point biserial correlation coefficient for each model is calculated for this purpose.

Thus the WAB and logit are the comparative methods used to develop the two predictive models of concern in this thesis, viz:- Model 1 to predict accounting trainee examination success v failure and Model 2 to predict recruitment success v. failure (leaving) in the early stages of training.

The component variables of the developed models are provided in Tables 5.1 to 5.4. The large number of variables in the 'overfitted' WAB models contrasts severely with the few in the logit models and the problems of uncertainty over the statistical significance of the WAB component variables is highlighted.

The following chapter examines both pairs of models in detail and assesses their criterion-related validity, with particular reference to the arguments of Mitchell and Klimoski (1982) concerning the predictive power and validity of "empirical" model building techniques. The issue of out-of-sample validation will be discussed.

6.1 Introduction

The two methods of model building employed in this study have been discussed in Chapter 5 and a general comparison of the simple WAB methodology made with the more statistically acceptable logit method. The WAB is seen to be essentially a 'rule of thumb' method for the practitioner, who may develop predictive models without particular statistical knowledge; all that is needed is adequate data and sample sizes. Nevertheless, such simple models are crude, giving no indication of the significance of the components, nor of strength of their relationships with each other or the criterion measure.

Baehr and Williams (1967) are often quoted for their criticism of such 'blind empiricism' in biodata modelling which results in the inability to explain predictor-criterion relationships. Such criticism led Mitchell and Klimoski to compare a more 'rational' approach, in which statistical measures are used to develop the actual predictor variables from the criterion measure, with the standard simple scoring technique, the WAB.

The development of the 'rational' approach stems from the knowledge that personal histories do significantly contribute to future work behaviour: *inter alia*, career choice, tenure and performance (Holland, 1966; Owens, 1976; Super, 1980; Eberhardt and Muchinsky, 1982a). Thus the relationships between the predictor items and the criterion may be established before the models are developed, by initially identifying a 'theoretically meaningful' set of constructs.

In the case of Mitchell and Klimoski (1982), the model building procedure following this initial stage was extremely complex. Continuous variables were arbitrarily ranked to turn them into categorical variables and, as in the case of this research,

item score distributions were examined and modified to provide more normal distributions. The resulting 86 items were then factor analysed to provide the components which are actually used as predictor items in a multiple regression analysis.

A 'rational' approach is adopted to item compilation in this empirical study and, although the variables chosen reflect areas known to be good predictors of academic success and work performance, they are only those commonly available on standard application forms and are all verifiable. This relates to the fact that the models developed here are for accounting firms to use as scoring devices, rather than the Mitchell and Klimoski models which were developed for use in a self-assessment inventory and include soft and hard data items.

The multivariate statistical empirical approach adopted here is able to offer some explanation of the relative power of predictor/criterion relationships of the component variables and the significance of their contribution to the model, while using real items of biodata rather than components derived from factor analysis.

Biodata models have been generally criticised for their lack of reliability, but the multivariate statistical approaches have been found to be less subject to shrinkage problems (Mitchell and Klimoski, 1982). This issue will be partially addressed here, but problems with non-stationarity of the original validation sample preclude firm conclusions.

Mitchell and Klimoski (1982), however, argue that the simple empirical WAB approach is the superior method, particular for practical purposes, and is more valid. This study will argue that this is not necessarily correct and that the methodology used here offers not only superior long term performance but, in contrast to the WAB, provides the evidence of the relative strength of the component variables without recourse to their over-complex 'rational' procedures.

The models derived in this study and their component variables are discussed in this

chapter, with their relative strengths and weaknesses, with a view to selecting the most appropriate strategy for providing practical recruitment tools. The problem of non-stationarity of the environment is also discussed as it directly effects the issue of validation.

6.2 Examination Prediction Models

Logit Model 1

Initially, a forward stepwise logit procedure developed a logit model with six significant variables which are shown in Table 5.3. The final model results from the resubmission of the significant variables to the logistic regression procedure, but with forced entry of the variables.

Table 5.3, in addition to the β 's and the standard error, gives the Wald Statistic and its level of significance. Three of the variables are achievement-related and three commitment-related. Only one of the variables has a negative β : number of 'A' level arts and languages.

The model χ^2 (71.93, $p < 0.0001$) allows the null hypothesis that the coefficients for all the variables in the model are 0, and that the model differs significantly from the perfect model, to be rejected.

Table 6.1 provides the conservative point biserial correlation coefficient for all four developed models compared across development and validation samples. The logit Model 1 value for the derivation sample of 0.53 compares well with the 0.36 reported by Mitchell and Klimoski (1982) for their rational biodata method and Reilly's and Chao's (1982) average validities of 0.39 for biodata predicting training performance and 0.35 overall for predictive biodata models. The results are equally impressive compared with those for the highly rated situational interview, for example the predictive validities across 15 occupations, summarised in Latham (1989), which range from 0.14 to 0.45.

Table 6.1

Point Biserial Correlation Coefficients
(Estimated Validity Coefficients)

Scoring Method	Derivation Sample	Validation Sample*
Logit Model 1	0.532	0.318
WAB Model 1	0.513	0.223
Logit Model 2	0.540	0.272
WAB Model 2	0.381	0.190

* Note, however, that the formula for the point biserial correlation coefficient is dependent on the proportions of the two groups within the sample. Where the proportions change, as in this case, the coefficient may not be reliable.

Huffcut and Arthur (1994), in a meta-analysis restricted to studies concerning interviews for entry level jobs, report an estimated r for all such interviews to be 0.37, with highly structured interviews providing an estimated $r = 0.57$. However, their figures are corrected to counteract the effects of range restriction, etc, and therefore are biased upwards. Our reported point biserial correlation coefficients, on the other hand, are likely to be biased downwards (see section 5.3.3 above) as estimates of Pearson's r and are based on a single study.

The performance of logit Model 1 also compares well with other studies concerned with entry level selection methods, eg the reported validities of 0.3 for in-tray exercises (Robertson and Kandola, 1982), and 0.53 (obtained by meta-analysis) for cognitive test batteries (Hunter and Hunter, 1984).

The frequency distributions of scores for the model building groups (Pass Group $n = 131$ and Fail Group $n = 98$) are shown in Figure 6.1 and for the validation sample in Figure 6.2. Group mean scores and standard deviations for the model derivation sample and validation sample are found in Table 6.2.

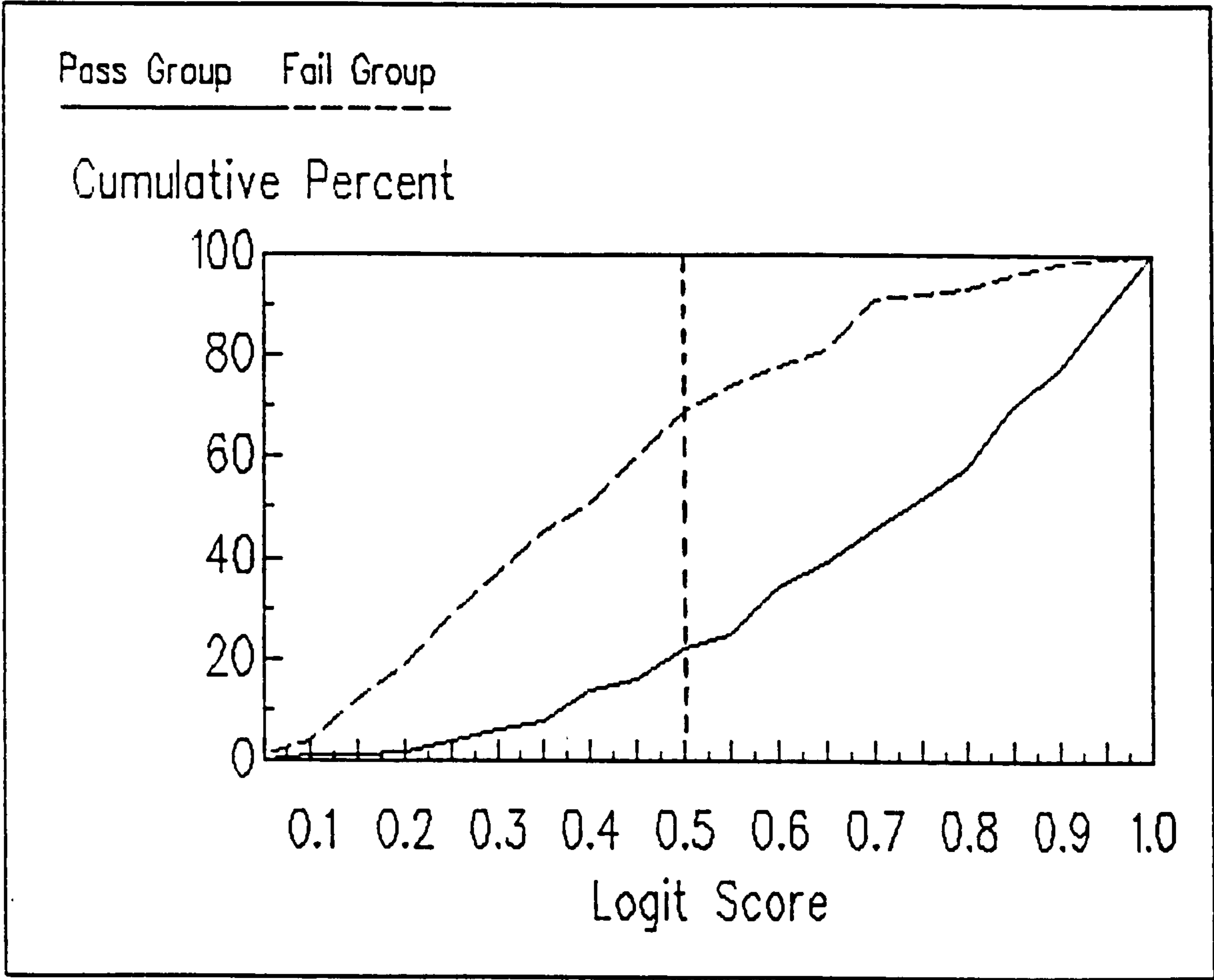


Figure 6.1 : Model 1 Criterion Groups' Logit Scores

Pass Group Fail Group

Cumulative Percent

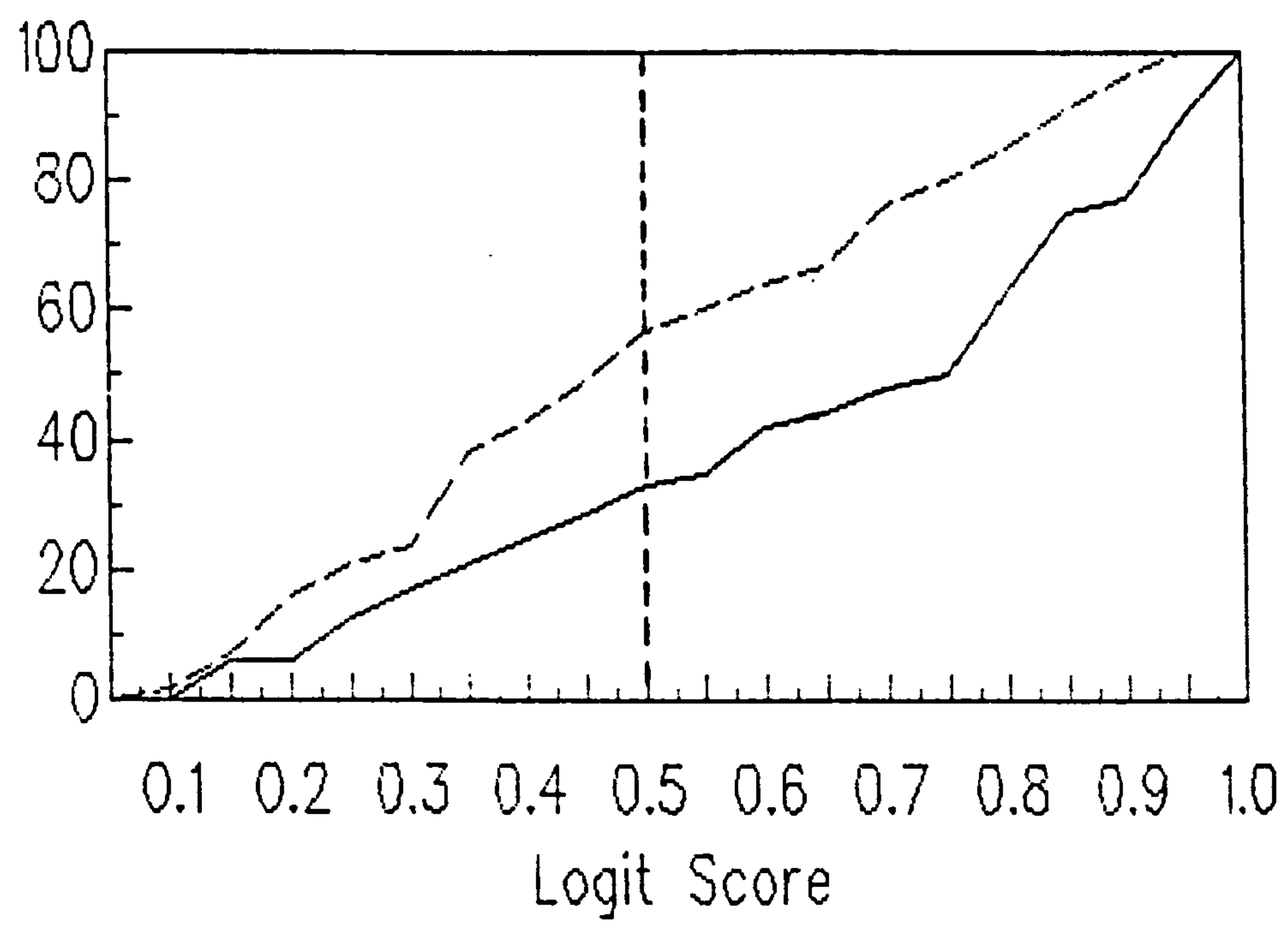


Figure 6.2 : Model 1 Validation Groups' Logit Scores

Table 6.2
Logit Model 1 Score Statistics

Group	Mean	Standard Deviation	N
Development Sample			
1985/6 All	0.57	0.26	229
1985/6 Pass	0.69	0.22	131
1985/6 Fail	0.41	0.22	98
Validation Sample			
1987 All	0.51	0.27	137
1987 Pass	0.64	0.27	48
1987 Fail	0.46	0.22	89

Table 6.3 provides the classification matrix, based on a 0.5 cut-off, and the associated hit-rates. The hit-rate is 74% overall, representing 70% of correct Fail predictions and 78% correct Pass predictions. Type I errors (predicting Fails but Pass outcome) are 22% and Type II errors (the converse) are 30%.

Table 6.3
Logit Model 1 Derivation Sample Hit-Rate*

	Actual Group		Total
	Pass	Fail	
Predicted Group			
Pass	102 (78%)	30 (30%)	132
Fail	<u>29</u> (22%)	<u>68</u> (70%)	<u>97</u>
Total	131	98	229

* Overall hit-rate 74 %

Applying Morrison (1969), which takes prior probabilities of group membership into account, the probability of correct classification is 0.74, which is significant² at better than the $\alpha = 0.01$ level³.

Table 6.4 provides the equivalent classification matrix for the 1987 validation sample showing an overall hit-rate of 61 %. The probability of correct classification, taking into account the differential Pass and Fail rates, is also 0.61, which is again significant at and beyond the $\alpha = 0.05$ level.⁴ Type I and 2 error rates are 33% and

² Calculations of critical values of t and χ^2 are derived from Fisher and Yates : *Statistical Tables for Biological, Agricultural and Medical Research*, reproduced in Siegel (1956)

³ $t = 7.01$, 227 df, $p < 0.01$

⁴ $t = 2.34$, 135 df, $p < 0.05$

43% respectively. However, due to problems associated with the one-off collapse in pass rates in 1987 discussed in section 6.7, which was largely associated with the change in GCC syllabus, true validation cannot be assessed using this sample. More reliable validation, relating to the 1988-90 entrants is described in Chapter 8.

Table 6.4
Logit Model 1 Validation Sample Hit-Rate*

	Actual Group		Total
	Pass	Fail	
<hr/>			
Predicted Group			
Pass	32 (67%)	38 (43%)	70
Fail	<u>16</u> (33%)	<u>51</u> (57%)	<u>67</u>
Total	48	89	137
<hr/>			

* Overall hit-rate 61 %

WAB Model 1

Table 6.5 provides the mean scores and standard deviations for both hold-out and validation groups for the WAB model. Figure 6.3 graphs the cumulative percentage frequencies at each score level for the hold-out groups (Pass = 39 and Fail = 28). England (1971) advises that the appropriate cut-off point to minimise Type I and Type II errors should be made on the basis of this distribution, at the level where maximum difference occurs between the two groups. This occurs at a score of 14 which is denoted by a slashed line on Figure 6.3.

Table 6.6 demonstrates that, at this level, the hit-rate is 68% (compared with 74% for logit Model 1) with 39% Type I errors and 25% Type II errors. In order to compare the classification results with those achieved by the logit procedure, the same treatment of the classification matrix (Morrison, 1969) is adopted for the WAB. The probability of correct classification is 0.67 (compared with 0.74 for the

Table 6.5
WAB Model 1 Score Statistics

Group	Mean	Standard Deviation	N
Hold-out Group (33%)			
1985/6 All	12.39	4.61	85
1985/6 Pass	14.62	4.64	45
1985/6 Fail	9.88	2.98	40
Validation Sample			
1987 All	11.84	4.15	138
1987 Pass	13.15	4.06	48
1987 Fail	11.21	3.88	90

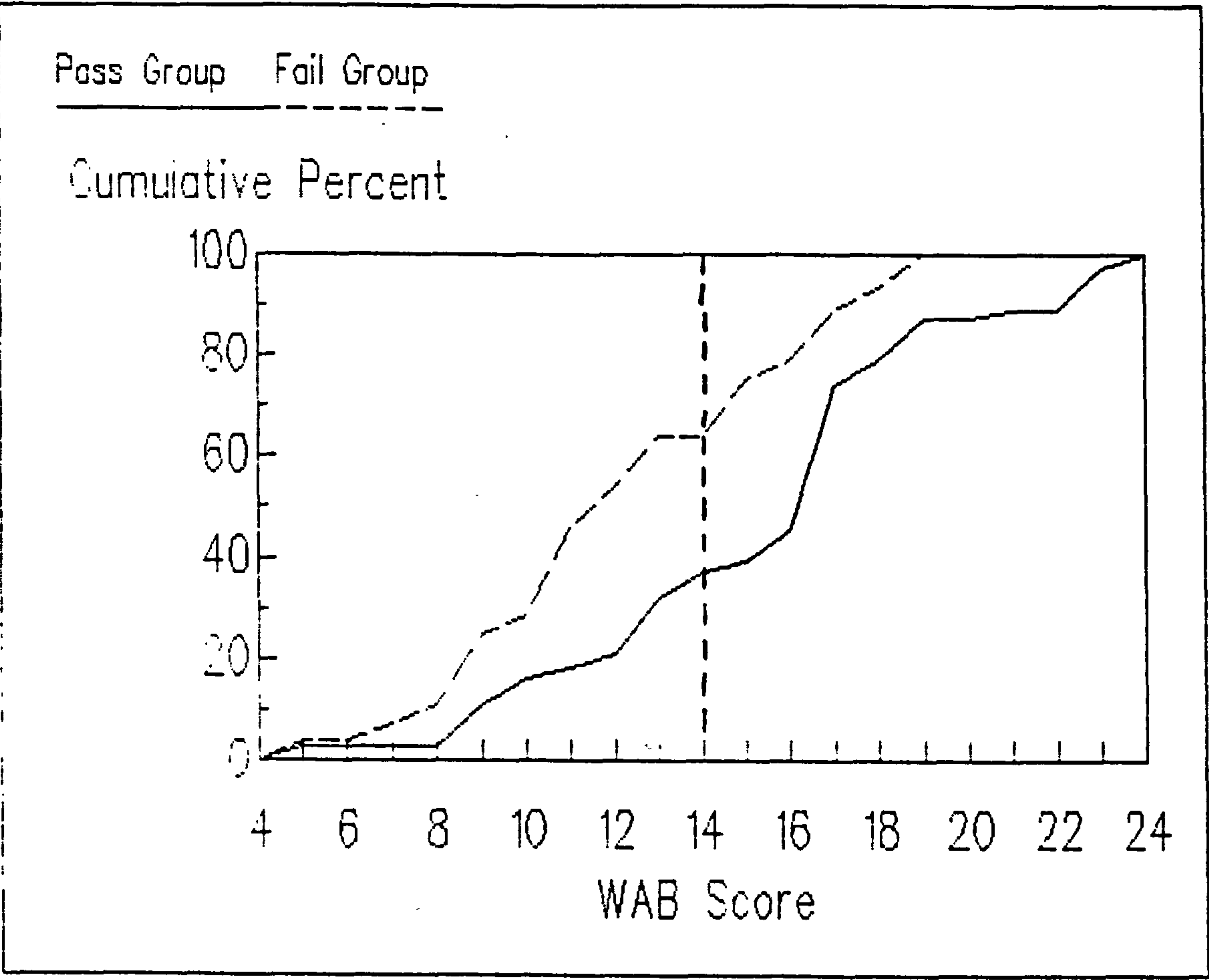


Figure 6.3 : Model 1 Hold-out Groups' WAB Scores

logit model) and this is statistically significant, albeit at the 0.05 level⁵.

Table 6.6
WAB Model 1 Hold-Out Sample Hit-Rate*

Note: This sample is not the full derivation group but a 33 % hold-out group.

	Actual Group		Total
	Pass	Fail	
Predicted Group			
Pass	23 (61%)	7 (25%)	30
Fail	16 (39%)	21 (75%)	37
Total	39	28	67

* Overall hit-rate 66 %

Table 6.1 shows the point biserial correlation coefficient for the developed model to be 0.51, which again compares favourably with other reported methods of selection and the logit model. However, the r_{pbi} for the validation sample has fallen to only 0.22, indicating a more severe shrinkage than that experienced by the logit model.

The distribution of WAB scores for the validation sample is illustrated in Figure 6.4 and Table 6.7 indicates that, applying the cut-off based on the derivation sample to the validation sample, 60% are correctly classified, with 42% Type I errors and 40% Type II errors. The probability of correct classification is 0.6⁶ and this model just fails to correctly classify the validation sample at the usual $\alpha = 0.05$ significance level.

⁵ $t = 2.43, 64 \text{ df}, p < 0.05$

⁶ $t = 1.13, 124 \text{ df}, p > 0.05$

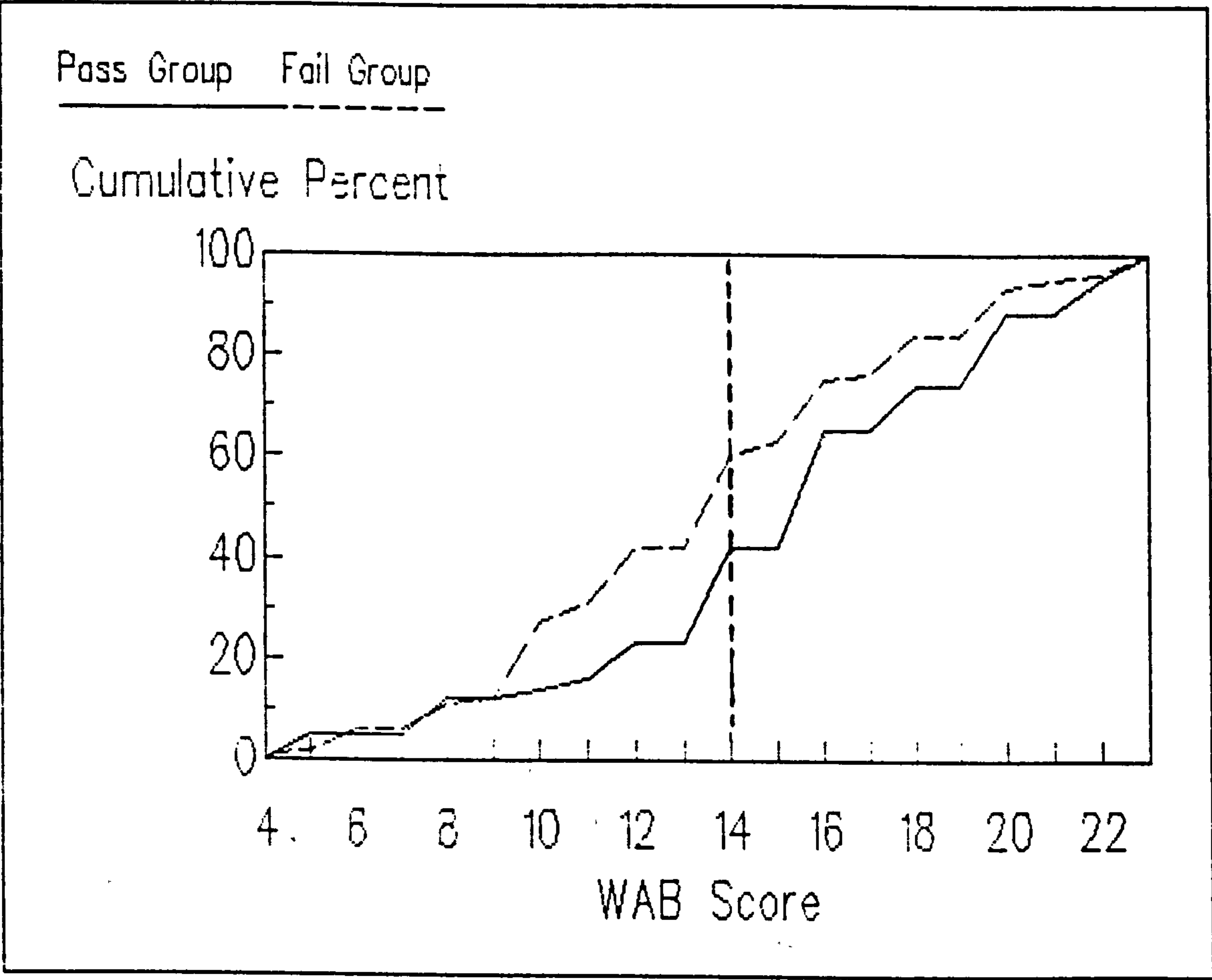


Figure 6.4 : Model 1 Validation Groups' WAB Scores

Table 6.7
WAB Model 1 Validation Sample Hit-Rate*

	Actual Group				Total
	Pass		Fail		
Predicted Group					
Pass	25	(58%)	33	(40%)	58
Fail	<u>18</u>	(42%)	<u>50</u>	(68%)	<u>68</u>
Total	43		83		126

* Overall hit-rate 60%

6.3 Comparison of Logit and WAB Results for Model 1

Examining the model derivation samples, the logit model can be seen to make 17% fewer Type I errors than the WAB (22% for logit and 39% for the WAB) and, conversely, the logit model makes 5% more Type II errors (30% to the WAB’s 25%).

However, the WAB requires more than twice the number of variables to provide a model than the logit approach and performs significantly worse on the validation sample, classification rates not achieving statistical significance. This is not surprising and reflects the crude nature of model derivation, lack of indication of the relative importance of component variables and the ‘overfitted’ and collinear model.

Not surprisingly, when comparing the r_{pbi} for the derivation and validation samples, the WAB models show a more marked fall in estimated validity.

The WAB being a linear additive model, may be tested for internal consistency using Cronbach’s α (Norusis, 1990, p190). Using the Reliability procedure in SPSS 4.0,

α is found to be 0.059, indicating that internal consistency, or reliability, of this model is low. However, the logit model is a non-linear model and variables in such models should not be highly correlated with each other, so reliability cannot be measured in this way.

6.4 Examination Model Component Variables

Table 5.3 shows the logit model to consist of a parsimonious and distinct set of only 6 variables. These variables are listed in order of size of Wald statistic, although it may be more helpful to bear in mind the $\exp(\beta)$ which indicates the amount by which the odds ratio (*probability of the successful outcome/probability of the opposite*) is increased by a one unit change in the variable. Thus each additional grade A 'O' level pass changes the odds ratio by 1.45, and having a I/II:1 degree causes a change of 3.29 in the odds ratio.

An indication of the positive/negative influence of the component variables in the WAB is much more difficult and can only be obtained from the net weights, which range from -20 to +20. The assigned weight, used to create the WAB score, ranges from 0 to 2. The assigned weight of 0 indicates a strong negative net weight of -4 or less and an assigned weight of 2 indicates a strong positive net weight of 4 or more. The intermediate assigned weights of 1 and -1 represent weaker net weights between -3 and +3. In the case of the examination performance model, the majority of the variables are found to be more effective in binary form.

Appendix C provides, asterisked, the significant differences between the percentages of the criterion groups appearing in the different predictor item categories for the examination prediction WAB. Variables which do not discriminate between groups, indicated by low net weights, are discarded. The remaining significant variables form the model. Fourteen variables were found to be successful predictors of examination success, although one was removed as it appeared to be measuring the same as another more powerful variable. The remaining 13 variables were made up

of 2 employment (background) measures, 9 achievement measures and 3 measures of commitment and these may be found in Table 5.1, with associated weights.

As expected with models developed to predict, either wholly or partially, academic achievement, a large proportion of the predictor items in both models are academic, reflecting the first of the Baehr and Williams (1967) second order factors and confirming the findings of Clarke and Sweeney (1985), Dunn and Hall (1984) and Dockweiler and Willis (1984), thus supporting the current use of such predictor variables. The majority of items included are intuitively meaningful.

The academic variables in the logit model reflect similar aspects of the data to those in the WAB. The number of grade A 'O' levels is included, as is a science/technology degree discipline. Number of 'A' level arts and languages is negatively weighted in the logit model and this variable was originally present (negatively weighted) in the WAB model but was removed since it appeared to measure the same thing as the positively weighted science and technology 'A' level variable, which was also included in the WAB model.

The retention of the Science/ Technology 'A' levels variable in preference to the Arts and Languages in the WAB was somewhat arbitrary, due to the imprecise nature of England's instructions. As key predictors, there might have been a case for including both. However, England advocates special procedures to avoid scoring the same item twice, but it is difficult to see what may be done to avoid double scoring, where the variables are negatively correlated. In the event, an examination of the net weights indicated the item which appeared to have the stronger difference between criterion group percentages. Collinearity is obviously a problem with such models.

In the logit model, good degree class entered and its inclusion does not seem inappropriate, rather it seems a surprising omission from the WAB model, where UCCA points are the only additional measure of developed academic ability to number of grade A 'O' levels. Dunn and Hall (1984), in their study investigating

the relationships between certain attributes of first time CPA examination performance, find that Grade Point Average (GPA), which is probably the best US equivalent to number of grade A 'O' levels, is the single most important attribute in predicting CPA examination scores. College entry SAT (Scholastic Aptitude Test) scores of general aptitude (g) are also significant contributory variables in their models and these probably best equate with UCCA point average.

The correlation between UCCA points and number of grade A 'O' levels is 0.39 (significant at $\alpha = 0.01$), further supporting the emphasis on academic achievement, but neither of these two measures was correlated with degree class, which was independent of all other predictors considered.

In the WAB, the number of secondary schools attended⁷, number of GCE 'O' level passes, number of mathematics and science GCE 'O' levels and 'A' levels, UCCA points and information technology skill level, represent academic components which are not present in the logit model. Since their relative contribution is unknown and as they are not present in the logit model, which has somewhat superior performance, it is likely that their effect as predictors is minimal.

The inclusion of these variables in the WAB model is not, however, surprising since research indicates that a good prior academic achievement record indicates the likelihood of success in future academic performance. The science/technology/engineering/mathematics bias is more likely to reflect the need for a numerate and logical approach in the work and professional examinations of the chartered accountant and obviously information technology skills are intrinsic to accounting practice work.

The importance of academic qualities and prior accounting experience in these

⁷ Those who had attended a sixth form college were not counted as having attended another school, but those who had attended colleges of further education or tutorial colleges had an additional school added to their total as such attendance is not beyond the control of the pupil.

models particularly uphold the findings of Eskew and Faley (1988) where SAT scores, pre-college study of accounting/bookkeeping and previous related experience were all significant factors in their model predicting student performance in Financial Accounting. Their study was undertaken in the USA where the undergraduate accounting course, being followed by the CPA examination, is broadly equivalent to the British training contract. Thus the Financial Accounting course may be compared with elements of PE1.

The importance of commitment variables, as measured by positions of responsibility, is reflected at the school level in the logit model by inclusion of the Headboy/girl (1/0) variable but not in the WAB model. Prefect, monitor variables, etc were available for selection in both the logit and WAB analyses as binary variables, but none proved significant in either model.

Obviously there are small numbers of those who have held the position of headboy or headgirl (only 6% in our sample), but the indication is that those chosen have the ability to contribute very positively to the running of the school, while maintaining an adequate study programme. These are obviously desirable characteristics in a trainee chartered accountant.

Much emphasis is placed by firms on applicants' participation in group activities, particularly active sports, and the holding of positions of responsibility for teams, organizations etc. These are regarded as signs (or indeed samples) of interpersonal skills and motivation and are used as indicators of an adequate level of social adjustment, which is very important to the practicing chartered accountant. Positions of responsibility, for example, are regarded as positive evidence of management potential. All of these variables are used by recruiters to predict likely commitment to the organization and the profession (Harvey-Cook and Taffler, 1987).

The entry of such variables in a model which predicts examination performance is perhaps surprising. Nevertheless, elements of successful practice work must enter into the subjects' approach to examination preparation, eg well-supported, secure

trainees who are enjoying a good level of success in their practice work are more likely to have a positive approach to their examinations.

However, the WAB weights do not indicate the predictor/criterion relationships that might have been expected concerning responsibility at university, counter-intuitively, small group active (ie sporting) pastimes (involving 2-6 participants) while at university are negatively weighted, while non-active small group activities (bridge, drama groups etc) are positively rated.

The work of the chartered accountant involves working co-operatively in small groups, in a non-competitive environment, so it may not be surprising that small group non-sporting experience is more important than small team active pastimes. These findings may also indicate that participation in the group activity is more important than assumption of responsibility for its organization, during training.

Approximately 35% of the sample attended fee-paying secondary schools and, although this variable is the least important variable in the logit model, its inclusion is likely to be the most contentious. In this case it may again be a proxy for a measure of social class or may reflect better examination preparation training in fee-paying schools, particularly boarding schools where levels of self-study, are closely monitored, the latter also having proved an important variable in the Dunn and Hall (1984) study predicting first time CPA examination performance.

Recruiters, have the right to choose candidates on the basis of any characteristics they see fit, as long as they do not unfairly discriminate against racial minority groups or upon the basis of sex. Since private education has significant explanatory power, it should be used on this proviso.

As biodata models are compensatory in nature, scoring badly, or not at all, on one variable may well be offset by a high score on one or more of the other variables. Thus an applicant who has attended a fee-paying school, but with an unsatisfactory academic background, in terms of the model, will not score as well as one who has

attended a state school and performed well academically.

The best possible performance on the logit model would be obtained by someone who gained a high proportion of grade A 'O' levels and took no art or language 'A' levels, attended a fee-paying school where they became headboy or girl and then went on to get a first or upper second class science or technology degree at university.

An ideal candidate as far as the WAB is concerned has more than 7 'O'/GCSE level passes, a high proportion of which are at grade A, and 4 or more in science/mathematics/technology passes. At 'A' level they are likely to have science/technology/mathematics passes and/or obtained higher than 9 (average) UCCA points. At university they have taken a science/mathematics/technology degree. They have average or above levels of information technology skill, and while at university have taken two jobs, one of which was perhaps related to chartered accountancy. While at university this candidate was not involved in any small group sports, but was a member of perhaps the chess club and the wine-tasting society. This candidate scores 26 points on the WAB.

A clear message from both models is that arts and language degree holders do not perform well in the professional examinations. For example, a trainee educated in the state system, with 5 grade A 'O' levels and an upper second class degree, scores 0.91 on the logit model if, in addition, they had studied 3 'A' levels other than arts or language and taken a science/technology degree, but only 0.34 if all their 'A' levels were arts or languages and their degree was other than in science or technology, eg Geography or Classics.

While it may be argued by the ICAEW that a graduate from any discipline may become a successful chartered accountant, it is certainly likely to be a harder road for those who have taken non-mathematical subjects at Advanced GCE and degree level.

6.5 Recruitment Success Models

These models, at the time of development, were thought to be of more interest to the practitioner than those simply seeking to predict examination success, because, however important examination performance is, the trainee also needs to exhibit an associated competence in his/her on-the-job performance. Those who do both are considered to be recruitment successes but those who leave are considered to be outright recruitment failures.

Logit Model 2

Forward stepwise logit, again using the log likelihood to control the introduction and removal of variables, was similarly used to guide the development of the second logit model. The same rule of inclusion of variables with at least 90% complete information was observed and, again, the final model results from the resubmission of the significant variables derived by the forward stepwise process, to a forced entry process.

Table 5.4 demonstrates that the model χ^2 of 44.08 (significant at <0.0001 with 6df) is sufficient to reject the null hypothesis that the model differs significantly from a perfect model. The six significant variables are shown in the table with the same statistics as for Model 1. Again, three of the variables are achievement-related and three measures commitment-related. Only one of the variables has a negative β : number of solo active pastimes undertaken while at university.

Cumulative percentage frequencies of the logit scores of the model building groups, Recruitment Success (80) and Leaver (74) are shown in Figure 6.5 and for the validation groups in Figure 6.6, with the associated mean scores and standard deviations in Table 6.8.

The classification matrix for this model is provided in Table 6.9 and is tested in the same way for significance. The overall probability of correct classification is 0.75,

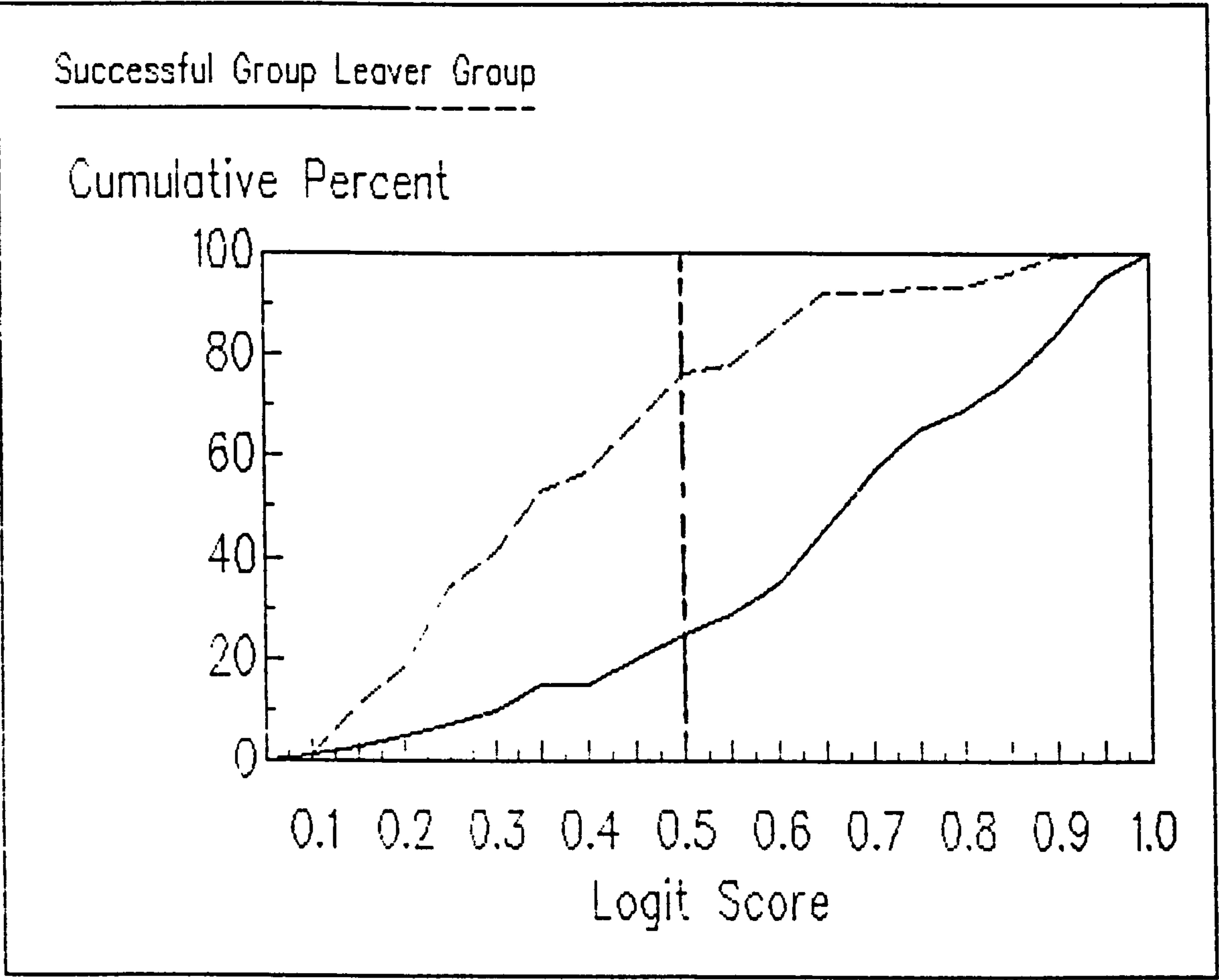


Figure 6.5 : Model 2 Criterion Groups' Logit Scores

Successful Group Leaver Group

Cumulative Percent

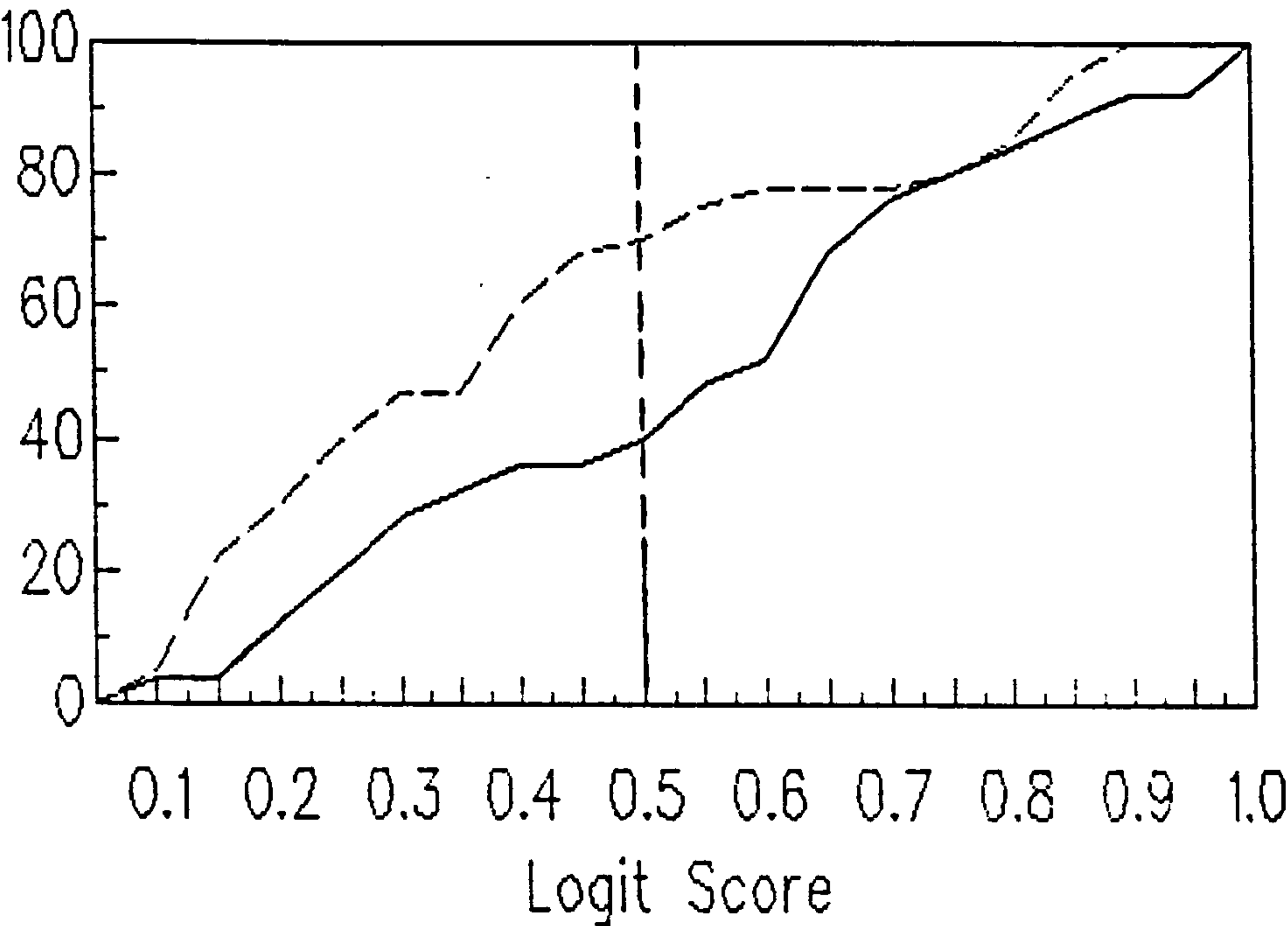


Figure 6.6 : Model 2 Validation Groups' Logit Scores

Table 6.8
Logit Model 2 Score Statistics

Group	Mean	Standard Deviation	N
Development Sample			
1985/6 All	0.51	0.25	154
1985/6 Success	0.65	0.20	74
1985/6 Leaver	0.38	0.23	80
Validation Sample			
1987 All	0.46	0.27	66
1987 Success	0.55	0.26	26
1987 Leaver	0.40	0.26	40

which is significant at better than $\alpha = 0.001^8$. The actual hit-rate is 75% with 25% Type I errors and 24% Type II.

Table 6.9
Logit Model 2 Derivation Sample Hit-Rate*

	Actual Group		Total
	Success	Leaver	
Predicted Group			
Success	60 (75%)	18 (24%)	78
Leaver	<u>20</u> (25%)	<u>56</u> (76%)	<u>76</u>
Total	80	74	154

* Overall hit-rate 75 %

The r_{pbi} calculated for this model (Table 6.1) is 0.54, which again compares favourably with other selection strategies, eg the 0.14 for interviews as predictors of supervisors' ratings⁹, the 0.10 for interviews predicting training success and the 0.03 for interviews predicting tenure, derived by the meta-analysis reported in Hunter and Hunter (1984). In comparison, the reported validities of their biodata measures for the same criteria in their analysis are 0.37, 0.30 and 0.27, respectively.

Applying Morrison (1969) to the validation sample classification matrix given in Table 6.10, the 0.66 probability of correct classification¹⁰ is again statistically

⁸ $t = 6.21, 152 \text{ df}, p < 0.001$

⁹ However, Huffcut and Arthur (1994), revisiting Hunter's and Hunter's meta-analyses, report a mean validity for predicting supervisors' ratings for entry-level jobs of 0.44, which more closely resembles our findings.

¹⁰ $t = 2.44, 64 \text{ df}, p < 0.01$

significant at and beyond the 0.01 level. Sixty-seven percent of cases are, in fact, correctly classified with 38% Type I and 30% Type II errors. However, Table 6.1 shows the r_{pbi} for this sample is reduced to 0.27.

Table 6.10
Logit Model 2 Validation Sample Hit-Rate*

	Actual Group		Total
	Success	Leaver	
Predicted Group			
Success	16 (62%)	12 (30%)	28
Leaver	10 (38%)	28 (70%)	38
Total	26	40	66

* Overall hit-rate 67%

WAB Model 2

Appendix D gives the derivation of assigned weights for WAB Model 2 groups and Table 5.2 indicates the predictor variables which were selected by the WAB technique. The mean scores and standard deviations for the holdout and validation samples are given in Table 6.11 and Cronbach’s $\alpha = 0.39$, which indicates that, although a large improvement on that for Model 1, this model also has low internal consistency.

Figure 6.7 graphs the cumulative frequency percentages at the different score levels for the hold-out groups (Recruitment Success = 31 and Leaver = 32). At a score of 7, the difference between the groups is maximized, however, below that point only 56% of the leavers would be eliminated at the expense of 33% of the pass group. For recruitment purposes the cut off of 8 is probably more realistic and both cut-offs are indicated by slashed lines on the distributions.

Table 6.11
WAB Model 2 Score Statistics

Group	Mean	Standard Deviation	N
Hold-out Group (33%)			
1985/6 All	10.08	3.44	63
1985/6 Success	12.16	3.20	31
1985/6 Leaver	9.53	3.49	32
Validation Sample			
1987 All	12.32	3.73	71
1987 Success	13.17	3.04	29
1987 Leaver	11.74	3.46	42

Successful Group Leaver Group

Cumulative Percent

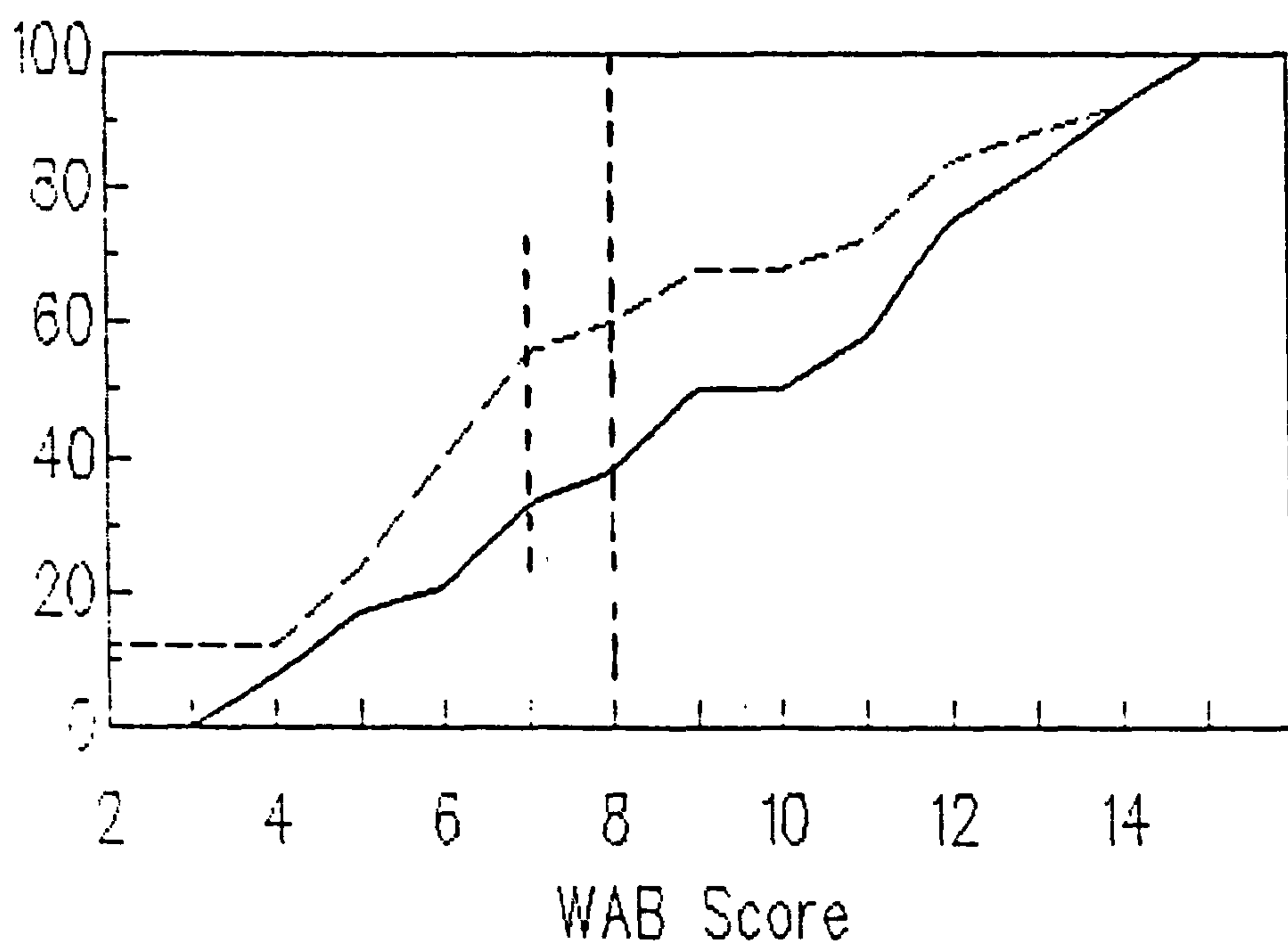


Figure 6.7 : Model 2 Hold-out Groups' WAB Scores

The higher cut-off would reject 60% of the Leavers with 38% of the Success group and this level was used to categorise the validation sample of 1987 entrants. Table 6.12 shows the classification matrix indicating a hit rate of 61% with 37% Type I and 40% Type II errors.

Table 6.12
WAB Model 2 Hold-Out Sample Hit-Rate*

	Actual Group		Total
	Success	Leaver	
Predicted Group			
Success	15 (63%)	10 (40%)	25
Leaver	9 (37%)	15 (60%)	24
Total	24	25	49

* Overall hit-rate 61%

Using the procedure previously described and a cut-off of 8, the probability of correct classification for the hold-out group is 0.61¹¹. The model’s predictive power was found not to be statistically significant at an acceptable level. Table 6.1 shows the point biserial correlation coefficient computed for this model is 0.38.

Figure 6.8 graphs the distribution of cumulative percentages of WAB scores for the 1987 validation sample (29 Success, 42 Leaver). Using the cut off of 8 points and below, 28% of the Success group are removed with 45% of the Leavers. At this level only 55% of cases are actually correctly classified with 28% Type I and 55% Type II errors. Table 6.13 provides the correlation matrix for the validation sample and the Morrison technique demonstrates that the probability of correct classification

¹¹ $t = 1.57, 47 \text{ df } p > 0.05$

Successful Group Leaver Group

Cumulative Percent

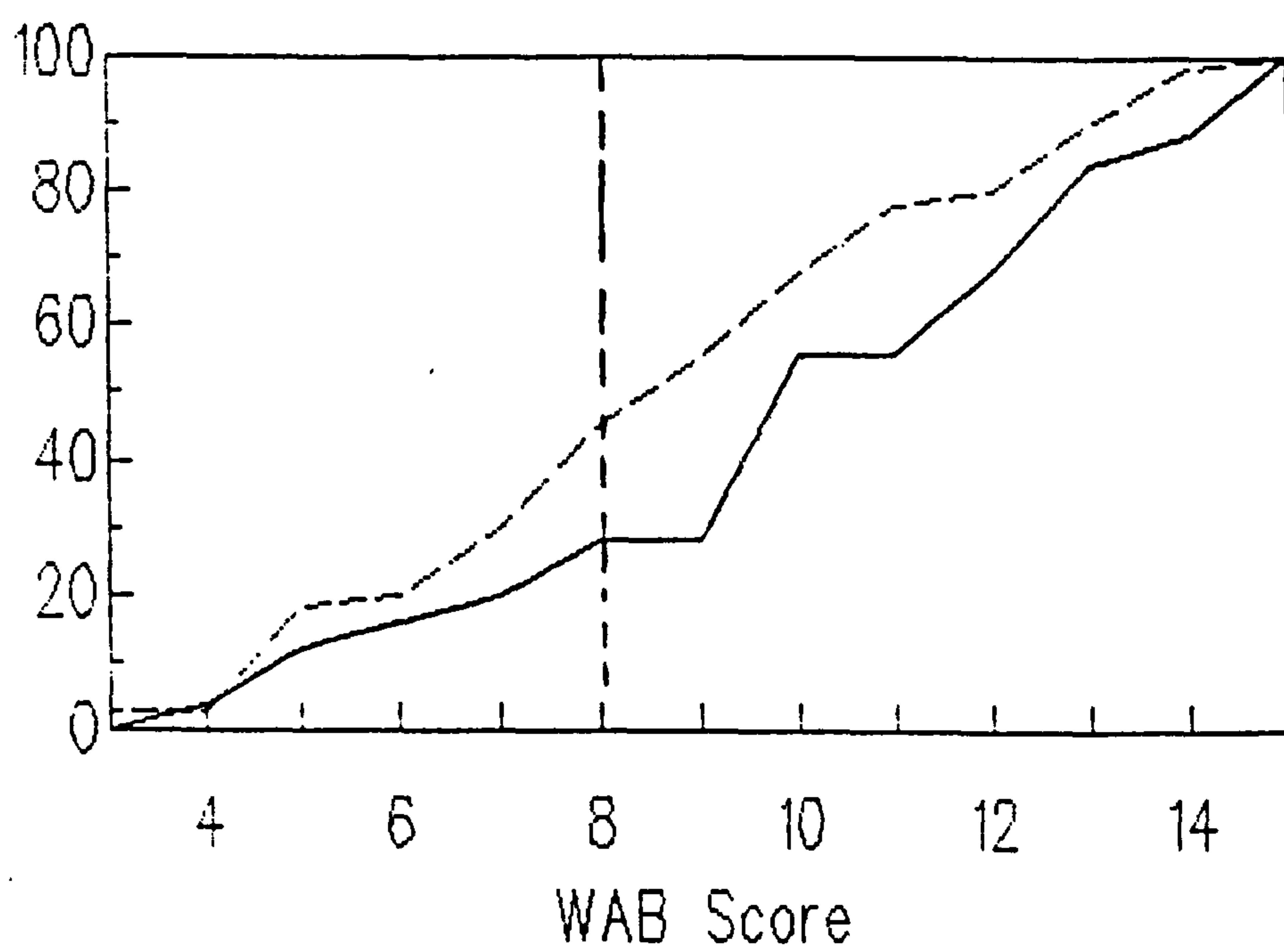


Figure 6.8 : Model 2 Validation Groups' WAB Scores

(0.55) again is not better than chance.¹²

Table 6.13
WAB Model 2 Validation Sample Hit-Rate*

Predicted Group	Actual Group		Total
	Success	Leaver	
Success	18 (72%)	22 (55%)	40
Leaver	<u>7</u> (28%)	<u>18</u> (45%)	<u>25</u>
Total	25	40	65

* Overall hit-rate 55 %

The estimated validity coefficient for this sample is 0.19, again indicating substantial shrinkage.

6.6 Comparison of the Logit and WAB Results for Model 2

In this case the logit model, correctly classifying 75% of the 1985-6 sample, maintains a significantly high level of prediction (67% accuracy) on validation.

However, the WAB model is not a good predictor, with 61% of 1985-6 sample cases being correctly classified originally but, on validation, correctly classifying only just over half of the cases (55%), including less than half of the Leavers (18/40). Type I and II error rates for the logit development sample, 25% and 24%, and validation sample 38% and 30%, compare favourably with 37% and 40% for the WAB development sample and 28% and 55% for the WAB validation sample.

¹² $t = 0.38, 63 \text{ df}, p > 0.05$

Table 6.14

Model 1 Samples 1985-7 : Comparative Statistics

Criterion Group	Development N	Sample %	Validation N	Sample %
Pass	134	30	48	22
Fail	106	24	90	40
Intermediate	202	46	85	38
	442	100	223	100

$\chi^2 = 19.65, 2 \text{ df}, p < 0.001$

Neither WAB model provides significant classification results.

6.7 Recruitment Success Model Component Variables

The logit model again is composed of six variables. The achievement variables are percentage exemption from the GCC examination, number of science and technology GCE 'A' levels and degree class and the commitment variables are headboy/girl, number of solo active pastimes at university and number of school teams/societies. There are no background variables in the logit model.

The most important variable as measured by the Wald statistic is number of 'A' level mathematics, science and technology subjects, again reflecting the need for numeracy and an analytical approach in the academic background.

The percentage of exemption from the GCC (Foundation) examination may have effects on both the examination and performance components of this model. In examination terms, the more exemption there is from the GCC examination, the more likely the first time pass at GCC and, following that, PE1 (Intermediate)¹³.

Table 5.2 lists the component variables for the WAB model. The background variables included here are area of family residence and number of jobs while at school. Achievement variables include number of GCE 'O' and 'A' level science and technology subjects, degree discipline and information technology skill level. The commitment category is represented by the number of school and university sports teams and level of responsibility achieved at school.

The component variables of the WAB model again match some of those in the logit model, viz: number of GCE 'A' level science and technology passes, headboy/girl,

¹³ An additional interesting finding in exploring the data was that only 16 of the 122 (ie 13%) failing or being referred in GCC at the first attempt subsequently passed PE1 at the first attempt.

number of teams and societies at school and degree class. However, the intuitively acceptable GCC exemption variable is not present. This is surprising for, in addition to the lower likelihood of failing the GCC examination, trainees with some familiarity with aspects of the training programme have the advantage of job information not available to those without and are thus in a better position to match themselves to the characteristics of the job (Meglino and De Nisi, 1987). In addition, having made a more informed choice, they are less likely to leave before their training contract expires and this model is concerned with predicting leavers.

Degree class is included in both models and, as in Model 1 this may be seen to uphold the findings of other studies which predict examination performance. It is, however, interesting to note that degree discipline in the WAB model has broken into the three possible weight categories in this model. The high positive rating of 2 goes, as expected, to science and technology degrees, the intermediate weighting of 1 goes to arts and combined studies degrees but the negative weight of 0 assigned in particular to social studies and humanities degrees, the latter category including Economics.

This may well reflect the high likelihood that many such graduates will have difficulty with the quantitative aspects of the GCC and later examinations, although Economics graduates do merit a one unit exemption from the GCC examination. Logit Model 1 contains the science/technology degree variable, providing further evidence that numeracy is highly desirable for examination success, which forms part of the criteria for recruitment success in Model 2. Nevertheless the weights assigned to the degree discipline variable in the WAB model appear to be counter-intuitive and inexplicable.

The commitment-related variables in the logit model reflect similar areas to the WAB, with headboy/girl (although in this case a binary variable, with monitor not significant) and number of school teams and societies contributing to its explanatory power. Once again the level of responsibility at school is shown to be important, supporting the intuitive weight given to the variable in recruitment procedures. In

the case of the WAB, 'monitors' (prefects with a special area of responsibility, eg, deputy headboy/girl, librarian, etc) have been given the same weight as headboys /girls.

The apparently surprising lack of significance of the prefect category may arise from the fact that, at most state comprehensive schools, until the recent initiatives in state schools to provide vocational education for post-16 students in addition to Advanced Level education, the sixth form represented quite a small proportion of the school population and therefore many members were made prefects. 'Prefect', perhaps, thus became a nominal category only, whereas monitors and heads of prefects do have specific functions.

Where the proportion of sixth formers to other levels is higher, less of the students are likely to be made up to prefect. In fee-paying schools and grammar schools, one would expect this to be the case, so the likelihood of being a prefect would decrease. Unfortunately this explanation also leads to the conclusion that some of the variables may be proxies for socio-economic status measures.

The second logit commitment-related variable not found in the WAB model is the negatively weighted 'number of solo active pastimes at university'. These include jogging, weight training, solo swimming, etc. This variable may again support the intuitive judgements of recruiters who feel that team or at least group participation are most important as background characteristics for trainees. It may be that those who feel the need to challenge **themselves** physically, rather than to combine their efforts with those of others, in active or non-active settings, are not suited *inter alia* to the team-orientated nature of chartered accountancy.

The commitment-related measures which do feature in the WAB model are to do with leisure and sports pursuits at university, the membership of teams and societies at school and level of responsibility at school. It is reasonable to speculate that the development of group interaction skills is the key here, even though a competitive element is present in some of the above.

However, those who have not been involved in any teams at university are rated similarly to those who have participated in more than one, even though the weight is only 1. Those who have been in only one team are rated 0. This too is inexplicable.

The remaining variables which form part of the WAB model and not the logit model are background related. They relate to area of family residence and previous employment.

Area of family residence has been used as a binary variable here: those whose family address was in London or the South East v the rest. This variable is derived from the home address given on the application form, which is usually completed during the final year of university and this study examined data which are relates to a period within 18 months of that time. It was considered unlikely that large numbers of families had changed address in the interim and so the variable was included.

The geographical location of the sample, being predominantly in the South East and London, is reflected in this split. Forty-six percent of the sample gave home addresses within the area, with the remainder split between the North (16%), the Midlands and Anglia (16%), Wales and South West (12%), Scotland and Ireland (5%) and Overseas (4%).

This variable may provide evidence that those whose families are far removed from their place of work may suffer from the lack of associated support. The decision to apply to a place of work distant from the family home is not beyond the control of the candidate but firms should perhaps be aware that those who do so may need counselling and support such as that available at university. Working and living in London can place strain on those who are not from the area, or who have not studied there, and this may cause stress which impinges on performance.

The number of part-time jobs held while at school, again which is not represented in the logit model, may indicate the subject's ability to make good job choices or a

low-priority need to work, indicating that this, again, may be a social class proxy measure.

Despite the intuitive acceptability of the many of the component variables, the WAB model is, nevertheless, a poor predictor of successful recruitment, with only a 61% chance of correct classification for the hold-out group and 55% for the validation groups, neither being statistically significant. The logit model, derived from the same data pool and the same sample but composed of less than half the number of variables can successfully predict the likelihood of leaving to an acceptable level.

6.8 Validation

At the beginning of the study it was decided that the 1985 and 1986 entrants in the sample were to be used for model building and the 1987 entrants would, *inter alia*, because of the likelihood of their examination results being unavailable when required, be used as an inter-temporal validation sample.

As already noted, the students of that year of entry were subject to a changed syllabus for the GCC examination. The ICAEW were of the opinion that the change would not materially impact on examination pass rates.

This expectation could not be tested directly since the ICAEW does not keep data on the number of attempts at the GCC examination made by their student members. Only the training firms have this information and do not report it. However, some insight can be gained by comparing the Pass/Fail/Intermediate status of the development sample (1985/6) and the validation sample (1987) intakes.

Table 6.14 provides comparative figures drawn from Table 4.1. The χ^2 test demonstrates the highly significant difference in performance of the two intakes¹⁴.

¹⁴ $\chi^2 = 19.65$, 2 df, $p < 0.001$

This may reflect the changes which took place in the examination syllabus of the GCC and not necessarily a change in the characteristics of the sample. However, the 16% increase in the proportion of the Fail group for 1987 entrants is particularly associated with poor GCC performance.

Many reasons have been suggested for the increase in the failure/referral rate, including lack of preparation on the part of the training firms and severity of marking caused by inexperience in the new subject areas. If the pass rate differential was due to problems associated with preparation for the new examination, rather than a sudden dip in the calibre of entrants, it would appear likely that some of those who should have passed, in fact, failed.

The implication for this study is that the students of the 1987 entry may not directly be compared to those of the previous two years on the basis of the GCC because the unavailability of overall data concerning the first-time pass rates does not allow us to confirm the difference, if any, between 1987 GCC results and other years. We only have the differences in the criterion group sample sizes to indicate change. Lack of stationarity of the environment must effect the performance of predictive models, particularly when, as in this case, it has a direct effect on the criterion measure. Therefore, any conclusions relating to the validity of these models must be speculative.

Using the statistical technique (Morrison, 1969), incorporating the prior probability of group membership, to examine logit Model 1 and WAB Model 1 (Tables 6.3 and 6.6) in terms of probability of correct classification, demonstrates that the logit model at the 0.5 level is superior to the WAB at the optimum cut-off (0.74 for the logit model to 67% for the WAB).

However, when the validation sample is considered (Tables 6.4 and 6.7), because of the collapse in the pass rate, neither model performs well, but the logit model still outperforms the WAB with results significant at the 0.05 level.

For Model 2, Tables 6.9 and 6.12 show that, again, the two models perform differently, with the probability of correctly classifying the development sample for the WAB being 61 % and for the logit model 75 %. However, even in the face of the collapse in the pass rates, the probability of correct classification of the validation sample for the logit model only drops to 0.66, whereas that for the WAB is 0.51. The WAB model's *ex ante* predictive ability is thus not significant at the usual acceptance level of 0.05 but the logit model's is significant at the 0.01 level.

The clear message is that, even in the face of a major change in the environment affecting the outcome of one of the key criterion variables, examination performance, the logit models are more robust over time. The superiority of these models to predict successfully, not only examination performance, but the key issue of tenure, from a limited number of highly significant hard biodata predictors, and on validation, is demonstrable. We may speculate that the weaker performance of the WAB may reflect both the less sophisticated statistical methodology and, particularly, the potential 'overfitting', resulting from using too many collinear measures leading to sample bias and poor validation sample performance.

6.9 "Empirical" versus "Rational" Methodologies

Mitchell and Klimoski (1982) are clearly in favour of using empirically derived methods for scoring biodata, rather than other methods, and their reasons for being so may be considered here, although, there seems to be confusion over their use of the term 'rational'.

The Pocket Oxford Dictionary defines rational as '...of or based on reasoning, rejecting what is unreasonable or cannot be tested by reason'. England's advice to use as many variables as possible as '...a large proportion of them will not differentiate between the criterion groups' (p15) was issued to practitioners and not to behavioural scientists *per se*. This 'suck it and see', blanket approach to item compilation is what invokes the criticism of 'blind empiricism' (Baehr and Williams,

1967). Indeed, in a footnote, Mitchell and Klimoski emphasize that 'Here, it is emphasized that an empirical approach should not be equated with non-rationality.' (p416)

In effect, their 'rational' approach simply means that they attempt to explain the criterion in terms of factors derived from the predictor data. This is undertaken in order to compare the models resulting from the use of the factors as proxy variables with an old established 'empirical' method.

They report better cross-validity associated with the WAB than their 'rational' methodology and they recommend the use of the WAB 'where the practitioner simply wants to maximise the prediction of the job criterion, which is very often the case.' Examining their table for rational v. empirical predicted criterion status (Table 4, p416), it may be seen that their chosen WAB cut-off point eliminated 59% of the failures while retaining 85% of the successes.

The probability of correct classification for their WAB (p41) is 76% and, for their so called rational method, 72%. In short, there is very little to choose between their models, which they admit ('...the practical differences between the two approaches appear to be modest.') (p416)

The appropriateness of the non-parametric empirical approach is questioned in this study. Here logit Model 1 outperforms the WAB model both on development and validation and the WAB technique failed to produce a statistically significant model for predicting recruitment success, based on the same data used by the logit approach to provide a good model which maintained statistical significance in its classification of the validation sample.

Mitchell and Klimoski, while reporting better cross-validation results for the WAB, also report much higher 'shrinkage' between development and validation for the WAB method than the 'rational' approach, even over a period of less than 1 year, which indicates that the WAB procedure may not be reliable. The findings of this

study are similar but with much more marked differences in validity found between validity coefficients for the derivation and validation samples.

Tested on a different sample, even though the logit models developed here suffered slight shrinkage, this method provided models that were more stable over time than the WAB models, even in the face of a major change in the measurement of the criterion variable.

Their concern with constructing meaningful predictor variables with clear relationships to the model criteria may well have resulted in Mitchell and Klimoski developing unreliable models, since, as previously discussed, their retained factors only accounted for 31% of the variance in their sample data. Having reduced the variable set by this strategy, their regression model cannot validly be compared with their WAB, derived using the whole data set.

The observed differences in their study results may therefore be directly related to the complexity of their experimental design. Siegel (1956), in support of the use of non-parametric tests, stresses their ease of computation and their usefulness where only small samples are available, eg for pilot studies. However, he also emphasises that

'...researchers and students in the behavioural sciences need to spend more time and reflection in the careful formulation of their research problems and in collecting precise and relevant data.' (p vii),

thus obviating the need for complex experimental design. Neither here nor in the Mitchell and Klimoski study is there restriction to small samples.

In this study we restricted the data to verifiable items and to those variables which have been reported in the literature as being valid predictors of our criteria and, therefore, our empiricism cannot be deemed 'blind'.

The usefulness of the non-parametric WAB technique, however, is what is questioned here. The parametric logit approach, using the same data set, provides

viable, reliable empirical models. Had Mitchell and Klimoski used a more reliable empirical approach, the experience of this research is that adequate predictive models, which are capable of indicating the strength and nature of the relationships between variables and predictors, may be developed without recourse to more complex approaches.

Mitchell and Klimoski agree that ' [the rational approach's]...principle contribution lies in its enhanced explanatory power and parsimony.' These contributions are also to be found in our empirical parametric approach. The necessity of adopting such complex 'rational' measures is therefore questionable.

This study cannot accord with their recommendation that where the employer wishes to maximise the prediction of a job criterion, the WAB should be used. The WAB technique is simple but not effective, suffering severe shrinkage problems and providing cumbersome, uninterpretable models. Logit is not a particularly complex procedure but requires intensive computing. However, even though Siegel expresses general concern over the mathematical sophistication of behavioural scientists (p vii), this process, like many others, is thankfully available on personal computer packages.

6.10 Conclusion

A comparison between two different model building approaches, the WAB and logit has been made. The two sets of models developed to predict

- (i) success in the graduate conversion and first professional examinations: and
- (ii) success in recruitment terms viz: adequate or better performance in the first six months in addition to Model 1 success criteria v. leaving within two years,

are reviewed here and their validity discussed.

The WAB technique provides complex working models with many variables and is not able to indicate the relative contribution of the component variables. The logit models are simpler and yet provide high levels of accurate prediction with fewer variables. In addition, they provide information on the strength of the relationship between predictor and criterion.

Only the logit technique is successful in producing viable models for both (i) and (ii) above. The logit models are also superior in terms of validity and reliability, although the validation results from 1987 are treated speculatively, due to the effects of environmental change. This study concurs with Schwab and Oliver (1974) who find that the usefulness of the WAB technique is overstated.

Because of the nature of the sample, which is drawn from a representative group of accounting firms, our results suggest that, although these models do not reflect inter-firm differences, they are capable, as stand alone devices, of making significant improvements in recruitment in medium-sized firms across the whole profession. The results achieved here compare very favourably with studies investigating other means of raising the validity of the recruitment process.

Based on the results of this chapter, our logit models are those used in the piloting exercise which was undertaken to ascertain the most appropriate practical methods of using such biodata models for professional recruitment purposes.

The next chapter, Chapter 7, specifically reports on issues relating to implementing the logit models in an industrial setting and discusses the most appropriate methods of use. For example, although the calculations given hitherto for the classification of trainees using the logit models have been derived from a 0.5 cut-off, this is not an efficient use of the probability measure represented by the logit. A more useful strategy for practical application, working directly with the applicant's probability of success, is described and tested.

7.1 Introduction

The previous chapter demonstrates, in principle, that the developed logit models can be successful in identifying appropriate trainees at the pre-selection stage. The purpose of this chapter is, *inter alia*, to explore their optimum use in practice in recruiting firms and to provide evidence of their reliable performance over time.

The literature indicates that the majority of biodata models are implemented by the use of a biodata inventory, which is completed by the applicant and then scored by the employer. The need for such a process derives from the use of soft data items which require responses from the subject.

Since the models developed in this study rely solely on the verifiable data, which is readily available on application forms, there was no need to develop an inventory. Implementation in the recruitment round immediately following development was via a new approach to model application, viz: a computer-aided scoring system.

This enables the recruiter to score any standard application form or cv, without the need for special forms or, indeed, for the firms to appear to be treating their applicants in a different manner from other firms. Where information is not yet available (eg, degree class) or missing, the recruiter may score the applicant for any likely outcome, before making an offer of an interview. Where the applicant has a very high probability of success, even when the least advantageous scenario is scored, the interview may be used to complete missing information.

The logit models developed here depend on very few variables and this results in rapid processing time and a reduced likelihood of processor input errors. The greater complexity of the weighting formula, compared with the WAB, is easily

accommodated by the program. Thus, vetting the application, however presented, theoretically should take no longer than the usual pre-sift and yet be of greater predictive power than current pre-sift plus interview strategies.

The logit score represents the percentage likelihood of a successful outcome, measured in model terms. Whereas a simple dichotomous classification of Pass/Fail around the 0.5 level severely limits the information provided by the logit score, working directly with the applicant's probability of success substantially enhances utility, particularly when a large proportion of those in the development samples do not fall neatly within the criterion groups.

To reduce the number of applicants, in a systematic manner, to those highly likely to succeed, model predictions are used, rather than the usual brief pre-sift. In this way the number of interviews offered is reduced and the interview is no longer concerned with selection but an assessment of the candidates individual fitness for the particular firm. Thus model use shifts the focus of selection to Stage II rather than Stage III of the process graphically described in Figure 3.1.

The first model is used to reduce the likelihood of accepting an applicant who will subsequently prove a poor examinee and the second model to reduce the likelihood of recruiting an applicant who will leave before the end of the training contract, thus failing to offset recruitment costs and causing disruptions in staff scheduling.

Three firms agreed to implement the models in the recruitment round immediately following the pilot exercise described in this chapter. Practitioners in these firms were asked if, given their own recruitment needs and the volume of applications, their need was for a bald cut-off score indicating outright rejection, versus acceptance, or whether a more interactive approach was called for and they opted for the latter.

The second section of this chapter therefore discusses the setting of decision strategies based on suitable cut-offs, which reflect firms' needs and their advice on

the costs of misclassification. The following sections then report and discuss the fairness of the models, when applied to groups at risk of discrimination, and the outcome of applying those strategies to the recruited sample groups, as if they represented the full applicant sample, to ascertain whether the predictive models do provide a significant improvement in selection validity.

This chapter also reviews the results of using such models as replacements for selection interviews. Section 7.5 discusses the theoretical results of replacing the selection interview with a single, and a tandem, model-based strategy, again making inferences about the applicant pool from the recruited sample data. The nature of the scoring software is described and the benefits associated with using its use for scoring applications.

The need to assess the likelihood of the applicant pool being able to provide sufficient numbers of candidates indicated to be of high calibre by the model-based strategy, is addressed in section 7.7, where a brief pilot exercise, involving the population of applications arriving at three medium-size firms during the latter part of 1990, is reported and conclusions are drawn concerning the implications of practical implementation of a model-based strategy to reduce recruitment errors.

7.2 Developing Suitable Cut-off Points

Discussion with staff partners and directors of recruitment in the sampled firms indicated the interactive approach to model use was considered to be most useful. The ability to be able to put a figure on a candidate's likelihood of being successful appealed to such staff and, in addition, they felt that this approach would be less likely to alienate lower levels of recruitment staff who, unless directly able to interact with the models, might feel that they were being 'deskilled', and more senior staff, who rightly feel that candidate interviews form an integral part of recruitment practice.

To this end, the score distributions of the model samples are examined and a lower cut-off, below which the candidate's chance of a successful outcome are very poor, and an upper cut-off, above which the candidate's chances of success are very good, are set.

The cumulative distributions of the logit scores for Models 1 and 2, including the scores for the intermediate groups are illustrated in Figures 7.1 and 7.4 below.

7.2.1 Model 1 Cut-off Points

Examining the distribution for Model 1 in Figure 7.1, it can be seen that only 22% of the Pass group scores fall below the 0.5 level but 69% of the Fail group do so.

Table 7.1 indicates the outcome probabilities for the 1985/6 entrants on Model 1 at score intervals of 0.2. For those who score below 0.2 in Model 1 (9% of the population) the probability of being a member of the Pass group is 5% and of the Fail group 49%. Thus those who score below 0.2 are nearly 10 times more likely to be examination failures than examination successes. The distribution of these probabilities may be examined in Figure 7.2 where the clear differentials in probabilities of the successful and unsuccessful outcome are demonstrated.

In order to provide cut-off points below which the likelihood of success is unacceptably low and above which the likelihood is acceptably high, the logit model scores of the 1985/6 groups were examined. The groups' distributions indicate that, for those who score below 0.3 (17% of the population), the probability of being a member of the Pass group is 11%, of the Fail group 49% and of the Intermediate group 40%. At the other end of the score distribution, the chance of being a member of the Pass group in the 37% of those scoring above 0.7 is 46%, of being a Fail group member only 6% and an Intermediate group member, 48%. Those scoring

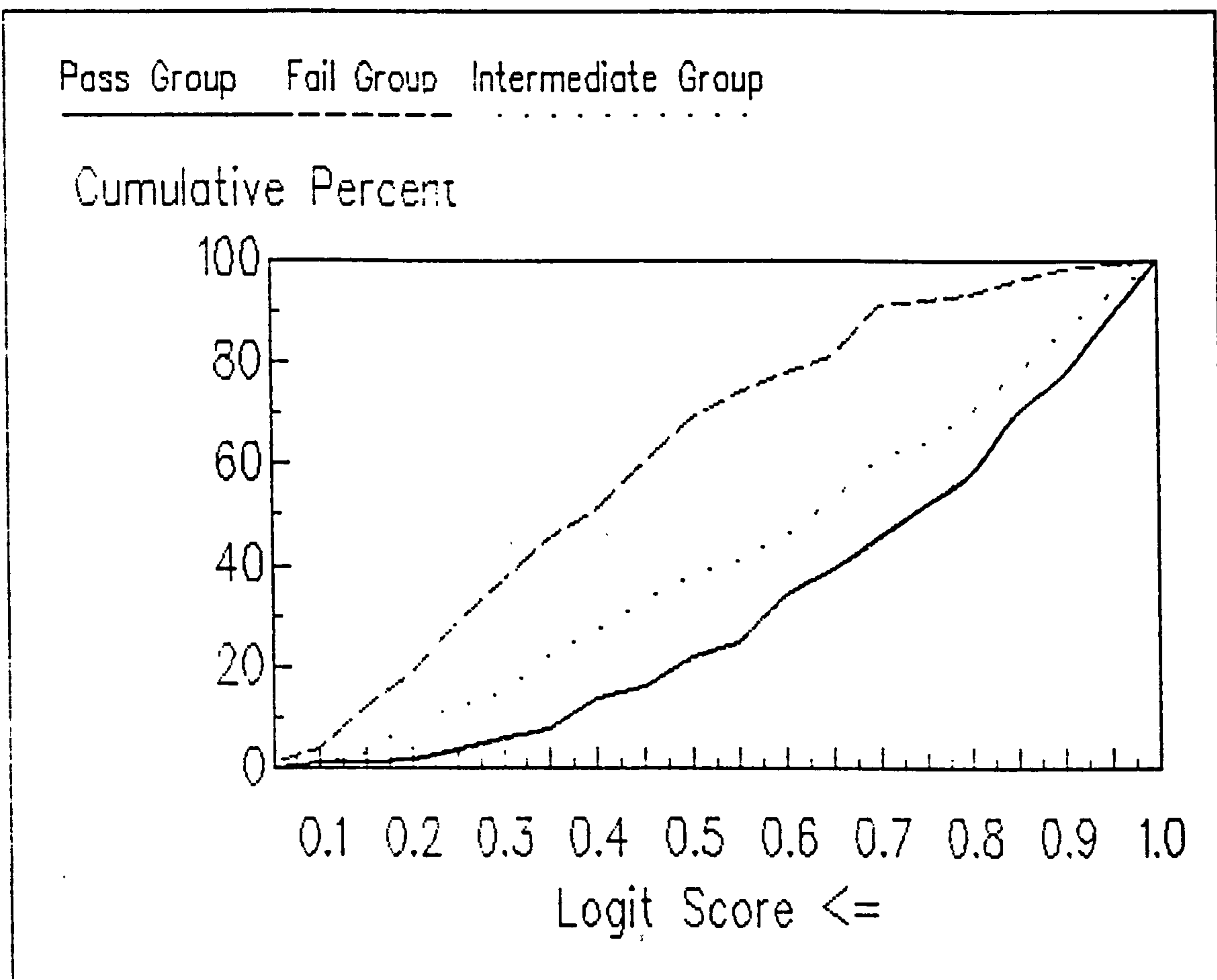


Figure 7.1 : Model 1 1985/6 Full Sample Logit Scores

Table 7.1
Model 1: 1985/6 Entrants' Logit Outcome Probabilities

P	Population Percentage	Candidate Outcome Probability		
		Pass	Fail	Intermediate
0.00 - 0.20	9	0.05	0.49	0.46
0.21 - 0.40	20	0.20	0.38	0.42
0.41 - 0.60	21	0.30	0.30	0.40
0.61 - 0.80	22	0.35	0.16	0.49
0.81 - 1.00	28	0.46	0.06	0.48
	100%	31%	23%	45%

Table 7.2
Model 2: 1985/6 Entrants Logit Outcome Probabilities

P	Population Percentage	Candidate Outcome Probability		
		Success	Leaver	Intermediate
0.00 - 0.20	11	0.08	0.28	0.62
0.21 - 0.40	24	0.08	0.29	0.63
0.41 - 0.60	28	0.14	0.19	0.67
0.61 - 0.80	22	0.30	0.07	0.63
0.81 - 1.00	15	0.38	0.08	0.54
	100%	19%	18%	63%

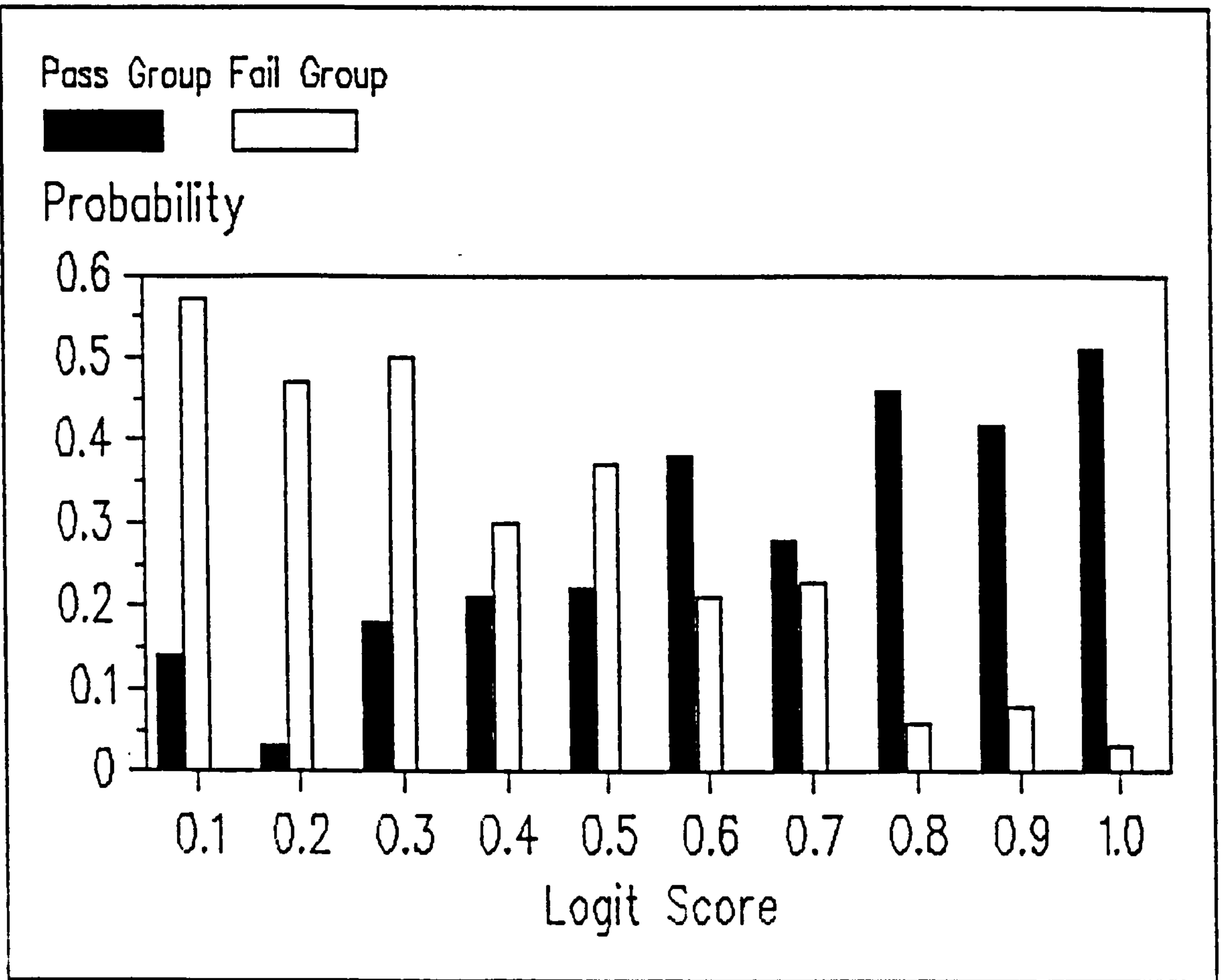
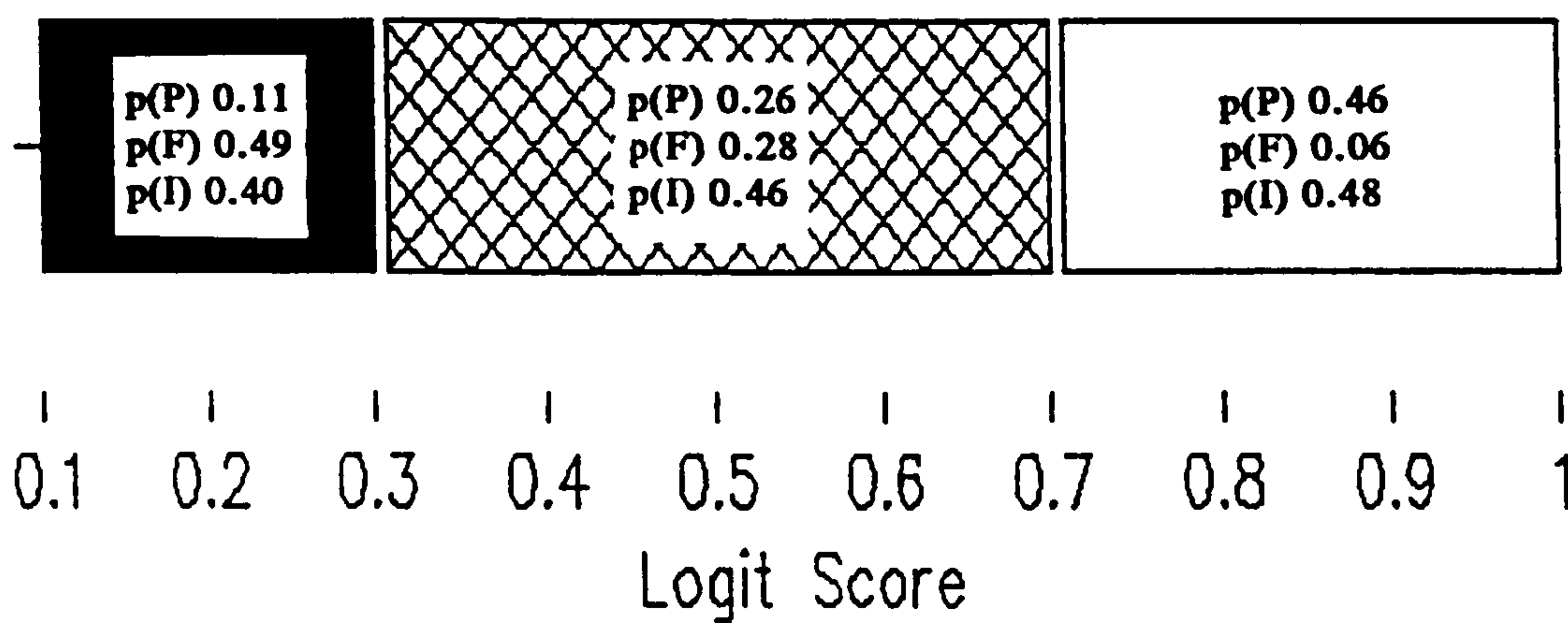


Figure 7.2 Model 1 Group Outcome Probabilities

Reject 17% Inspect 46% Accept 37%



Probability of Pass $p(P)$
Probability of Fail $p(F)$
Probability of Intermediate $p(I)$

Figure 7.3 Model 1 Selection Decision Strategy

above 0.7, therefore, are > 7 times more likely to be Pass group members than Fail group members. A decision strategy based on these probabilities may be examined in Figure 7.3, where the bar indicates the probabilities of group membership in three score areas.

7.2.2 Model 2 Cut-off Points

The cumulative frequencies of the logit scores for the three groups of entrants, Success, Leaver and Intermediate, in Model 2 are presented in Figure 7.4. Twenty-five percent of the Success group score below 0.5 compared with 76% of the Leaver group.

The outcome probabilities for the 1985/6 groups are given in Table 7.2. In this model, those scoring above 0.6 (38% of the population) are nearly 5 times as likely to be Successful as Leavers, those scoring below 0.20 (11% of the population) are more than 3 times as likely to be Leavers as Successful.

In addition, the probabilities associated with group membership at each 0.1 score interval are again graphically displayed in Figure 7.5. Examining Table 7.2, it is clear that this model is less refined than Model 1 as the 'Intermediate' group make up 63% of the total sample, rather than the 45% in Model 1. This is due to the proportion of examination successes who had poor performance ratings (11% of the group), indicating commitment, rather than ability-related problems, those delayed in passing their examinations (12%), the numbers of outright examination failures who did not leave the firm (14%) and those partially successful, having passed GCC but failed PE1 (55%) or failing GCC but passing PE1 (6%) (Table 4.2).

Thus Model 2 may be regarded as a rather 'blunt' instrument. However, the crucial emphasis on the use of the model is focused upon predicting and thereby the avoidance of leavers. So, for example, accepting only those

Successful Group Leaver Group Intermediate Group

Cumulative Percent

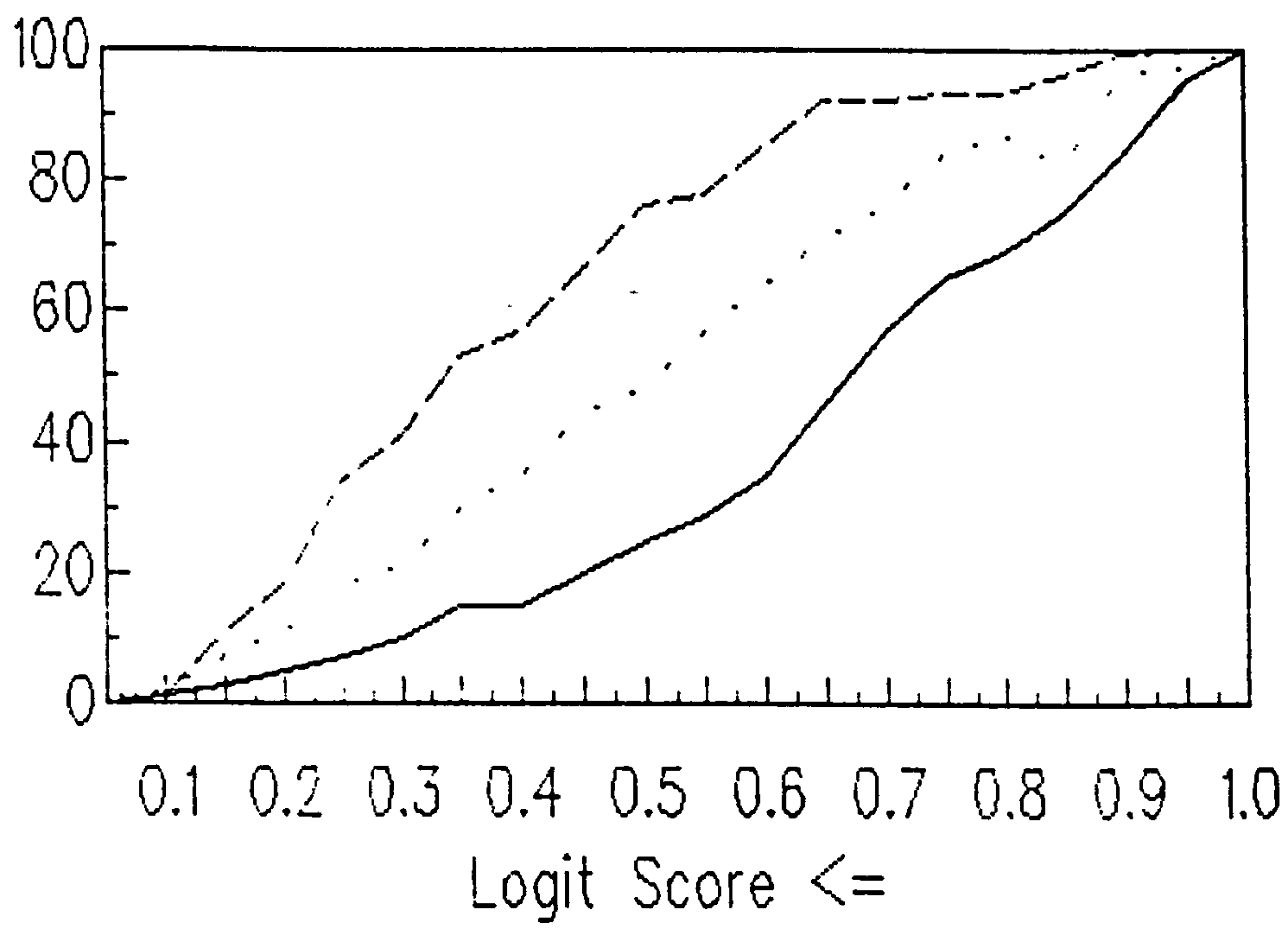


Figure 7.4 : Model 2 1985/6 Full Sample Logit Scores

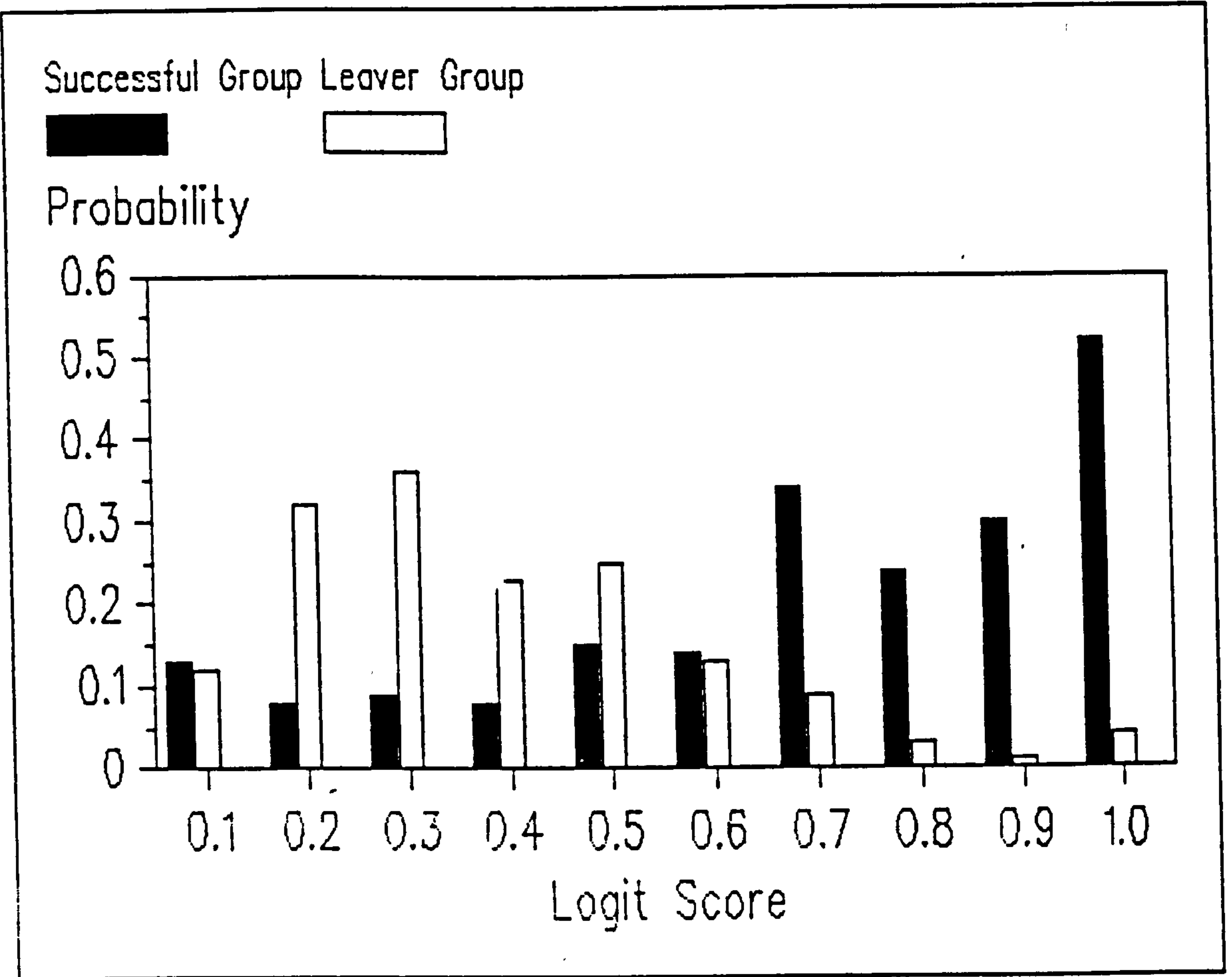


Figure 7.5 Model 2 Group Outcome Probabilities

who score above 0.6 reduces the likelihood of recruiting a leaver to 7%. A similar decision strategy to that given in Figure 7.3 for Model 1 may be found in Figure 7.6 for Model 2. At score levels below 0.4 (35% of scores), there is nearly 4 times the likelihood of recruiting a leaver rather than a successful entrant, indeed there is only an 8% chance of being successful with a score below this level. However, at score levels above 0.6 (38% of scores) the position is reversed, with almost 5 times the likelihood of recruits being successful to that of their leaving.

7.3 Assessing the Impact of the Models on Sensitive Groups

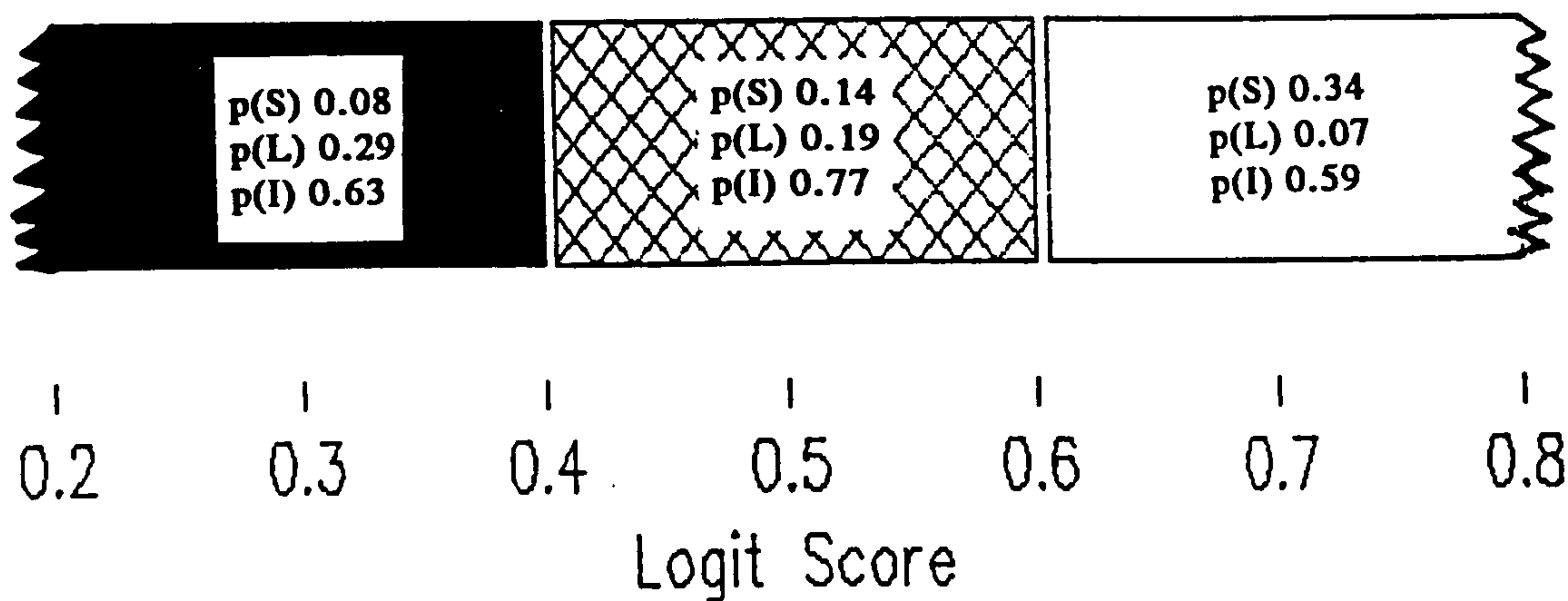
Owens (1976) and Hunter and Hunter (1984), *inter alia*, provide evidence that biodata scoring is less likely to have unfavourable impact on minorities and women than other methods of assessment. In this section the variables in the models are examined to ascertain that none represent an imbalance in treatment for such groups.

In each case the mean model scores and component variable scores of females and minorities were compared with males and Europeans respectively.

7.3.1 Sex

Recruitment to the chartered accountancy profession is heavily weighted in favour of males and ICAEW statistics indicate that women entrants only made up approximately 35% of the intakes between 1984 and 1994 (ICAEW, 1995). The profession is rightly still perceived to be male-dominated as the statistics also indicate that the female percentage of the total membership of the ICEAW has only increased from 6.5, to 14.3 in the same period. It is important to stress that the figures discussed below relate to a recruited sample, which reflects existing recruitment bias, when discussing whether there may be unfair discrimination in using the models.

Reject 35% Inspect 27% Accept 38%



Probability of Success $p(S)$
Probability of Leaver $p(L)$
Probability of Intermediate $p(I)$

Figure 7.6 Model 2 Selection Decision Strategy

Table 7.3 shows that in Model 1, three variables have significantly different mean scores for males and females. Women trainees have significantly higher numbers of grade A GCSE/'O' level passes, with a mean score of 3.2 and 2.3 for males ($t = 3.1$, $p < 0.01$). They are more likely to have taken arts subject 'A' level GCEs, with a mean score of 0.55 compared with men, 0.29 ($t = 3.12$, $p < 0.01$). They are less likely to have attended an independent school, with a mean score of 0.27 cf for males 0.41 ($t = -2.48$ $p < 0.05$).

As already discussed, state examinations measure developed ability and should provide an unbiased method of comparing between groups of people. However, the choice of subject taken at Advanced level is usually, while restricted for all students in terms of what the educational establishment can offer, within the control of the student. The fact that they do have a choice of subject means that a variable reflecting that choice cannot be considered unfair as a predictor. The choice of arts subjects at Advanced level is not a positive indicator of professional examination success for any entrant and recruited females appear more likely to have made it. In the event, scoring on this negative variable does not necessarily mean a low overall score.

Surprisingly, there is no corresponding significant difference between the sexes on the number of science/technology 'A' level passes variable, which would have been expected due to the low number of 'A' levels taken. There is also no significant difference between mean scores on the science/technology degree variable.

The lower mean score on the independent school variable means that women trainees are less likely to have come from such backgrounds. This may reflect a lower female population in independent schools, recruiters' perception of the importance of this variable for women, or the lack of such women applying to the profession. Since this variable will be scored by a pc-based program, before, presumably other accounting staff see the

Table 7.3

Model 1 Item Scores by Sex of Entrant

Model Item Description	Female Mean	Female Std Dev	Male Mean	Male Std Dev	t	Sig
Number of Grade A GCE 'O' level passes	3.18	2.84	2.3	2.16	3.10	0.002
GCE 'A' level Arts Subjects	0.55	0.73	0.29	0.67	3.12	0.002
Independent School	0.27	0.44	0.41	0.50	2.48	0.014

Note : Other model variables indicate no significant difference between scores.

application form, there should be no adverse effect on females applying.

When the sexes are compared in terms of their model score, women have higher mean scores than men ($t = 1.61$, $p = 0.06$). Women trainees are, on average, slightly less likely to fail their first two examinations than their male colleagues (51 % of females are in the Pass group, compared with 46 % of males). It is possible, because recruiters have been more stringent when recruiting women, their higher average score on the A grade GCE variable (the most significant variable in the model in terms of its impact on the logit score) contributes to this slight superiority. However, the present difference is not significant.

In Model 2 the only significant difference between male and female variable scores was found to be on the solo active pastimes variable, where the group means were 0.8 for females and 0.5 for males ($t = 2.59$, $p < 0.01$). Women undergraduates recruited to the profession appear to have taken their exercise alone, by choice, more often than men. Again, there is no reason to suppose that an item which is dependent upon choice is an unfair discriminator. There is no significant difference between mean Model 2 scores for the sexes.

7.3.2 Race

There are no significant differences between Afro-Caribbean trainees' and Asian trainees' mean scores and those of the Europeans, on any variable in Model 1. Afro-Caribbeans, representing only 0.8 % of our sample, do not differ from Europeans in terms of logit score for the model, but the mean score of 0.89 for Asians, representing 1.5 % of the sample, is significantly higher than that for Europeans, 0.58 ($t = 2.71$, $p < 0.01$). This is primarily accounted for by their A grade 'O' level GCE mean score of 4.2, compared with that of Europeans, 2.7. Taken together with the percentage of entry cohort figures (0.05 % Afro-Caribbeans and 0.08 % Asians), the evidence is

not consistent with strong negative bias.

There are no differences between these racial minorities and Europeans in terms of Model 2 variables or logit score.

7.4 Testing the Decision Strategy on the Validation Groups

The results of adopting the same cut-off points suggested in Figures 7.3 and 7.6 may be examined in Tables 7.4 and 7.5 below.

7.4.1 Model 1 Validation Groups

Taking the lower cut off of 0.3 for Model 1 (17% of all scores as for 1985/6), the validation group details given in Table 7.5 and based on Figure 3, show that 17% of the Pass group, 24% of the Fail group and 11% of the intermediate group fall within this score area. Those scoring above 0.7 (34% of cases), include 52% of the pass group, 24% of the Fail group and 34% of the intermediate cases.

Table 7.5 demonstrates that, on validation, the probability of being a member of the Fail group below a score of 0.3, is 0.55 and that of being a Pass group member is 0.21. These compare with 0.49 and 0.11 respectively for 1985/6 (Table 7.4). However, at the other end of the scale, for those scoring above 0.7, the probability of being a Pass group member in the validation sample is 0.34, of being a Fail group member is 0.28, compared with 0.46 and 0.06 for the model development sample of 1985/6. Either the effect of the change in the examination structure in 1987 is responsible for a poorer performance for 1987 or the model is not reliable.

7.4.2 Model 2 Validation Groups

Table 7.4

Model 1: Logit Score Outcome Probabilities
1985/6 Groups at Suggested Cut-off Points

Scores	Population Percentage	Candidate Outcome Probability					
		Pass <i>p</i> (P)	% Pass Group	Fail <i>p</i> (F)	% Fail Group	Int. <i>p</i> (I)	% Int. Group
0.00 - 0.30	17	0.11	(06)	0.49	(37)	0.40	(15)
0.31 - 0.70	46	0.26	(39)	0.28	(54)	0.46	(46)
0.71 - 1.00	36	0.46	(55)	0.06	(09)	0.48	(39)
100%		31%		23%		45%	

Table 7.5

Model 1: Logit Score Outcome Probabilities
1987 Groups at Suggested Cut-off Points

Scores	Population Percentage	Candidate Outcome Probability					
		Pass <i>p</i> (P)	% Pass Group	Fail <i>p</i> (F)	% Fail Group	Int. <i>p</i> (I)	% Int. Group
0.00 - 0.30	17	0.21	(17)	0.55	(24)	0.24	(11)
0.31 - 0.70	49	0.14	(31)	0.44	(53)	0.42	(51)
0.71 - 1.00	34	0.34	(52)	0.28	(24)	0.38	(34)
100%		22%		41%		37%	

The results of applying the cut-off points illustrated in Figure 7.6 to the 1985/6 sample are again described in Table 7.6 below. Thirty-five percent of the cases score below 0.4. This segment of the sample contains 15% of all successful group cases, 57% of all leavers and 35% of intermediate cases. The segment scoring above the upper cut-off of 0.6, which includes 38% of all cases, contains 65% of all successful cases, 15% of leavers and 36% of intermediate cases.

Adopting a similar approach to Model 2 subjects as that shown above for Model 1, an examination of those in Table 7.7 scoring below 0.4 in the validation sample (41% of the population) indicates that this segment consists of 60% of the Leaver group, 35% of the Successful group, and 38% of the Intermediate group.

Those scoring above 0.6 (34% of cases) contain 42% of the Success cases, 22% of the Leavers' group and 35% of the Intermediate cases. The probability of being Success, below a score of 0.4, is 10%, and of being a Leaver, 27%, compared with the 1985/6 sample figures of 8% and 29%. Above the 0.6 logit score, the probability of being a Success group member is 15% and of being a Leaver 12%. These compare with 34% and 7% for 1985/6.

Again, either the problems associated with the instability of the examination criterion variable are manifest or the model does not prove to be reliable. However, in the case of Model 2, the very high likelihood of being a member of the Intermediate group reduces the effectiveness of the model.

7.5 Discussion of Validation Results

Although the lower cut-offs hold well for the 1987 groups, as far as identifying likely Fail and Leaver group members, there is a loss of predictive power in both

Table 7.6

**Model 2: Logit Score Outcome Probabilities
1985/6 Groups at Suggested Cut-off Points**

Candidate Outcome Probability							
Scores	Population Percentage	Success $p(S)$	%Success Group	Leaver $p(L)$	%Leaver Group	Int. $p(I)$	%Int. Group
0.00 - 0.40	35	0.08	(15)	0.29	(57)	0.63	(35)
0.41 - 0.60	27	0.14	(20)	0.19	(28)	0.67	(29)
0.61 - 1.00	38	0.34	(65)	0.07	(15)	0.59	(36)
100%		19%		18%		63%	

Table 7.7

**Model 2: Logit Score Outcome Probabilities
1987 Groups at Suggested Cut-off Points**

Candidate Outcome Probability							
Scores	Population Percentage	Success $p(S)$	%Success Group	Leaver $p(L)$	%Leaver Group	Int. $p(I)$	%Int. Group
0.00 - 0.40	41	0.10	(35)	0.27	(60)	0.63	(38)
0.41 - 0.60	25	0.11	(23)	0.13	(18)	0.77	(27)
0.61 - 1.00	34	0.15	(42)	0.12	(22)	0.73	(35)
100%		12%		19%		69%	

models at the higher cut off. This may at least partially reflect the inevitable result of problems associated with examination changes, manifested by the resulting drop in Pass group trainee numbers from 31% to 22%, and the increase of the Fail group proportion from 23% to 41% of the sample for Model 1. The results, in turn, reduce the numbers of successful trainees defined by Model 2 from 19% to 12%, although the proportion of leavers remained steady.

In Model 1, the segment scoring below 0.3 in the 1985/6 sample contained 6% of Pass group members and 37% of the Fail group (Table 7.4). In the validation sample, these percentages changed to 17% and 24% respectively. However, above the upper cut off of 0.7, 44% of 1985/6 Pass group are correctly classified and 9% of Fail group incorrectly classified compared with 52% and 24% respectively in the 1987 sample.

In Model 2, even though the percentage of Success students correctly classified at the 0.6 cut-off drops from 69% in the 1985/6 sample to 42% for 1987, they still outweigh the incorrectly classified Leavers by roughly 2:1 (Table 7.5). The proportion of Leavers scoring above 0.6 is 21% compared with 15% for 1985/6 and the Intermediate group remained the same. The lower cut-off still rejects 60% of the Leavers.

It would therefore seem prudent to adopt a policy of rejection adhering to the lower cutting points suggested and basing acceptance for interview upon a system where the higher the score above the suggested cut-off points, the more readily a personal approach is made. A further consideration, reflecting a problem common to all such models, is that the *samples used here will not reflect the nature of the total population as they are representative of only those selected by the sample firms.*

The decision 'bars' (Figures 7.3 and 7.6) indicate that those who score in the 'grey area' between upper and lower cut-offs should be carefully inspected. An obvious recommendation is to only approach those who score above the upper cut-off, but where numbers applying drop (an extremely unlikely contingency), the higher

scoring intermediate group may be considered. Those who fall in the lower band should never need to be considered.

After discussion with staff partners in the three firms agreeing to implement the models (Finnies, Kidsons Impey and Stoy Hayward) it was decided that the 0.3 lower cut-off for Model 1 should be raised to 0.4. This decision was based on the low likelihood of success of a candidate so rated and the high volume of applications received, which have to be reduced to manageable numbers.

7.6 Replacing the Selection Interview

This thesis explores the potential effects of using biodata models to replace selection interviews, but with the necessary caveat that the recruited sample represent trainees selected by the interview based approach from the full applicant pool, whereas the logit models are developed and validated on already recruited samples only.

Considering Model 1 Table 7.8 below demonstrates that applying a cut-off logit score of 0.7 to 1985/6 recruits would result in a group composed of 46% examination successes, 6% examination fails and 48% who refer or fail one examination. These compare with the figures for the full 1985/6 sample recruited by pre-selection and two selection interviews : 30% Pass, 24% Fail and 46% Intermediate cases.

The group scoring above 0.7 accounts for 55% of all recruited Pass group members, 9% of Fails and 39% of all Intermediate cases¹⁵, and provides evidence of the superiority of the logit scoring approach to selection.

Adopting a similar cut-off for the 1987 validation cases results in a group made up

¹⁵ χ^2 for 1985/6 recruited Pass/Fail/Intermediate v 1985/6 / high scoring group = 27.84, $p < 0.001$

Table 7.8

Results of Replacing Interviews with Logit Model 1

	Sample Characteristics			
	Recruited 1985/6	Sample 1987	High Scoring Groups 1985/6	High Scoring Groups 1987
Pass	30%	22%	46%	34%
Fail	24%	40%	6%	28%
Intermediate	46%	38%	48%	38%
	100%	100%	100%	100%

Table 7.9

Results of Replacing Interviews with Logit Model 2

	Sample Characteristics			
	Recruited 1985/6	Sample 1987	High Scoring Groups 1985/6	High Scoring Groups 1987
Success	24%	12%	34%	15%
Leaver	23%	19%	7%	12%
Intermediate	53%	69%	59%	73%
	100%	100%	100%	100%

of 34% Pass, 28% Fail and 38% other cases, compared with the actual recruited sample figures of 22%, 40% and 46%, respectively. This group accounts for 52% of all recruited Pass group members, 24% of all Fails and 34% of Intermediate cases¹⁶. In this case, the difference between the high scoring group and the original recruited sample falls just short of significance.

Although only around one third of those originally accepted score 0.7 or above (36% in 1985/6 and 34% in 1987), shortfalls in numbers recruited need to be made up by the uptake of more of the high scoring suitable applicants not interviewed by firms, for whatever reason, and offset by the reduced need to 'over-recruit' to allow for drop-out and failure rates of approximately one third. However, the quality of the full graduate sample available to firms was not known. Evidence supporting the feasibility of reducing the pool to only the high scoring applicants is supplied by pilot implementation in the three participating firms reported in section 7.8.

Similar results pertain for the second model, used to reduce the likelihood of leaving. Table 7.9 below demonstrates that applying the 0.6 cut-off to the 1985/6 recruited sample, results in a group composed of 34% Success, 7% Leavers and 59% Intermediate cases, the majority of these refer or fail examinations but do not leave (77%), 12% delay their examinations and the remaining 11%, despite passing examinations, do not perform well in practice work in their first six months. The group scoring above 0.6 includes 65% of all those rated Success in terms of the model and 15% of all Leavers.

These figures compare favourably with those for the recruited sample for those years, which are 24% Success, 23% Leavers and 53% Intermediate cases¹⁷. This again provides evidence of the superior performance of the model-based approach compared with the judgementally based selection system.

¹⁶ χ^2 for 1987 recruited Pass/ Fail/Intermediate v high scoring group = 5.54, $p > 0.05$

¹⁷ χ^2 for recruited Success/Leave/Intermediate v high scoring group = 19.81, $p < 0.001$

When the cut-off is applied to the 1987 recruited sample, the group scoring above 0.6 are 15% Success, 12% Leavers and 73% Intermediate cases, compared with the recruited sample figures of 12%, 19% and 69%¹⁸. This high scoring group includes 42% of all Success group members, 21% of all Leavers and 35% of Intermediate cases but, once again, the differences fall short of being significant.

In both cases the results obtained by simply using model scores to differentiate between trainees provide evidence of their superiority over the interview process which recruited them.

Model 2 does not seem to be particularly useful since such a large proportion of the sample fall in the Intermediate category. Table 4.2 indicates that those who are successful in examination terms but not in performance terms represent only approximately 10% of the intermediate cases, the rest may be viewed as examination failures. Thus the major difference between the examination model and the overall performance model is the ability to predict leavers.

7.7 Tandem Use of Models

A recruitment strategy based on dual model use is investigated here, whereby the acceptance of successful examinees and the avoidance of likely leavers are the combined recruitment criteria.

Seventy-nine cases in the 1985/6 development sample meet the criteria for both membership of the Pass group in Model 1 and Success group in Model 2 and have adequate data for scoring on both models (out of a possible 85 cases). Of these, 37 (47%) score above the upper cut-off in both models and 19 (24%) score below the cut-off on both models. The remaining cases scored high on Model 1 and low on

¹⁸ χ^2 for 1987 recruited sample Success/Leave/Intermediate v high scoring group = 5.61, $p > 0.05$

Model 2 (11%) and low on Model 1 and high on Model 2 (18%).

Table 7.10 demonstrates the score grouping for the Pass/Success and Pass/Not Success groups. The policy of recruiting only those scoring above both upper model cut-offs would result in the acceptance of 38% of the full sample. Eighty percent of this group are members of both the Model 1 Pass group and the Model 2 Success group, the remaining members being unsuccessful in recruitment terms.

Using the models in tandem can only be of use where recruitment staff have clear goals. At the beginning of the recruitment period, when volumes of applicants are high, there may be many suitable applicants who score above the recommended cut-offs on both models and attracting them to the firm will not be difficult. As the round progresses, however, less of the better candidates may remain unplaced and more of those applying may be unsuitable. The judgement of the recruitment partner, weighted by the numbers already recruited, numbers outstanding and risks associated with acceptance of less advantageous qualities, will be supported by the information provided by the models.

For example, where the models indicate a high likelihood of examination success coupled with a score lower than the recommended 0.6 for the recruitment model, the partner must judge the effects of recruiting such a candidate, who may well, if Model 2 is reliable, have motivational problems and leave. Similarly, where the score on Model 1 falls just below the cut-off, but the score for Model 2 is high, the indications are that the applicant may well not pass examinations at the first attempt, but will be well motivated and perform well.

In reality, it is clear that recruiting risks become unacceptable the lower the scores fall in the scoring scale. Thus, successful recruitment will be reflected in maintaining as high an average score as possible.

When faced with large numbers of application forms and certain targets to achieve, it is tempting to recruit a large proportion of the desired intake number as soon as

Table 7.10

Results of Tandem use of Models on 1985/6 Cases

Group Membership	*HiHi	Model Scores			Total
		HiLo	LoHi	LoLo	
Pass/Success	37 (47%)	9 (11%)	14 (18%)	19 (24%)	79
Pass/Not Success	9 (21%)	11 (26%)	6 (14%)	16 (38%)	42
	46 (38%)	20 (17%)	20 (17%)	35 (29%)	121

* HiHi >0.7 Model 1 and >0.6 Model 2
 HiLo >0.7 Model 1 and <=0.6 Model 1
 LoHi <=0.7 Model 1 and >0.6 Model 2
 LoLo <=0.7 Model 1 and <=0.6 Model 2

the peak numbers are received. This not only reflects the desire to have the following year's recruitment needs taken care of, but to 'snap up' what appear to be the best candidates before other firms do so.

Using an effective scoring method to identify the most likely successful trainees requires the courage of one's convictions. Numbers applying to the profession show no sign of dropping and if firms are keen to reduce quantity and improve quality, thus avoiding expensive wastage, it would seem prudent to use the higher cut-off firmly and recruit on a more progressive basis, rather than the euphemistic 'wide trawl' approach described by one recruitment partner, with the implicit 'cream rises' philosophy, which has previously prevailed.

In practice, although Model 2 is not such a refined model, both models may be used, so that, in cases where the recruiter is uncertain concerning a candidate, even though the score on Model 1 indicates a satisfactory likelihood of good examination performance, Model 2 may be used to ratify judgements of professional suitability but, perhaps more importantly, the likelihood of leaving.

It has been noted that the need to 'over-recruit' as a compensation for habitual drop-out figures will be diminished and that fewer candidates will be considered suitable than previously, using the a model-based decision strategy. Again, the need to assess the availability of different calibre applicants within the complete pool is emphasized. Use of the models, either alone or in tandem, relies on adequate cases being available for selection.

7.8 Scoring Software

Within the three piloting firms applications were pre-selected by a clerk to remove those who did not meet ICAEW minimum entry criteria (passes at GCE 'O' level Mathematics and English), and/or the firms' own criteria of a minimum of 9 UCCA points and those requiring work permits. The remaining applications were then

handed to a partner or personnel manager who decided which applicants to call in for interview.

Use of the logit models requires the weighting formula to be applied to each application and then a decision made as to whether the applicant meets the set model criteria for acceptance. Since these decision criteria are pre-set, staff judgemental input is not necessary.

Computer-aided scoring and selection was the obvious choice to increase the speed and efficiency of the coding process and to make it available to all interested parties within the piloting firms. LOTUS-123 based software was therefore developed to score the application forms entering the firms and to provide the necessary information for those arranging interviews.

The software program is tailored to each firm in that the order of the questions follows the order of their application form information. In addition, areas which are of specific concern to the firm, eg facility with foreign languages, are incorporated in the program. Examples of the output of this program may be found in Appendix F, where the scores for both models are given.

When the application arrives in the firm, the clerk checks for the minimum requirements and then processes the form, using the LOTUS program. Each applicant having been scored, the processor can reject outright those scoring below the appropriate cut-off and go no further on the program, simply marking the form for a rejection letter.

The majority of applicants are undergraduates and unable to give their degree class. This is a key variable in both models. Firms have in the past had to depend on predictions of this variable based on tutors' reports, yearly examination results and 'guesstimates'. Expected degree class, the most widely used proxy, is correlated at $r = 0.41$, in our data, (significant at $\alpha = 0.01$) with actual degree class but though significant can hardly be considered an accurate measure.

Using the software, the coder may score the applicant without degree class and the program will generate scores for both good and poor degree eventualities. This allows the firm to decide on whether the possession of the lower class degree will still place the candidate in the 'accept' bracket or involve unacceptable risk. The firm can decide whether it will offer a conditional place, subject to gaining the better class degree, which may result in the applicant accepting the place provisionally and subsequently taking another firm offer of unconditional acceptance, or outright rejecting those who have an associated risk factor.

In order to cover the eventuality of missing data, another fact of life governing the use of multiple application formats, the processor can, using the same process as for unknown degree class, score the applicant with nothing in the missing section and then again with one or more different scores in that cell. There is a provision for the coder to leave a message for the next assessor, explaining what the differing scores mean. If the applicant is subsequently selected for interview, the interviewer can ascertain the correct cell contents and if necessary, rescore the applicant.

For those scoring above the reject score, the program generates a facing sheet which details all the information on the candidate upon which the models depend. The application form, with facing sheet attached, is then ready to pass on to the appropriate member of staff for further consideration. (Examples of facing sheets may be found in Appendix F, with a copy of the summary sheet described below.)

The program also generates a summary sheet which contains all the details of the applicants processed in the session, so that a record of applicants is to hand when the inevitable queries arrive concerning lost applications or explanations of the reasons for rejection.

This new approach to biodata model use has much to recommend it. The summary enables the second level of the recruitment team to check that those selected have been correctly coded and random checks of rejected applications can assure the firm that they should indeed have been rejected.

7.9 Pilot Exercise

Differences between a full sample of applicants and a sample of recruits are to be expected, but the extent to which these differences will affect the future use of models has to be assessed. A brief pilot exercise to assess the quality of the applicant pool, prior to implementing the models as part of the pre-selection process is reported here.

In September 1990, firms providing data for the 1985-87 sample were approached with regard to taking part in a pilot of the developed models prior to implementation. Three agreed to do so, conveniently a large medium (Stoy Hayward), a medium medium (Kidsons Impey) and a small medium firm (Finnies), in terms of size.

We analyzed all the applications from non-accounting graduate/ undergraduate applicants entering the Firms between 1/1/90 and 31/12/90 and scored these on both models. This took place in January 1991 when the bulk of applications for the following September had been received by Firms.

When the scoring exercise was complete, Firms were visited and were apprised of the differences between their selection decisions and those resulting from a model-based strategy and these differences were discussed.

Tables 7.11 and 7.12 break the application forms received down into those for whom degree class is known and those for whom it is unknown. Model 1 scores are examined in Table 7.11 and Model 2 scores in Table 7.12.

Seventy-three percent of the applications studied relate to graduates, indicating that most had either decided to take a year off before applying to the Firms or had been employed elsewhere in the interval between graduating and professional entry.

Table 7.11
Model 1 Decision Strategy Applied to Piloting Firms

7.11.1 Graduates

Score Level	Rejected	Interviewed	Total
> 0.7	29 (18%)	42 (45%)	71 (27%)
0.4 - 0.7	64 (40%)	35 (35%)	99 (38%)
< 0.4	69 (42%)	23 (23%)	92 (35%)
	162 (62%)	100 (38%)	262 (100%)

7.11.2 Undergraduates (If I/II.1 obtained)

Score Level	Rejected	Interviewed	Total
> 0.7	22 (58%)	33 (55%)	55 (56%)
0.4 - 0.7	13 (34%)	20 (33%)	33 (34%)
< 0.4	03 (08%)	07 (12%)	10 (10%)
	38 (39%)	60 (61%)	98 (100%)

7.11.3 Undergraduates (If I/II.1 not obtained)

Score Level	Rejected	Interviewed	Total
> 0.7	6 (16%)	15 (25%)	21 (21%)
0.4 - 0.7	18 (47%)	20 (33%)	38 (39%)
< 0.4	14 (37%)	25 (42%)	39 (40%)
	38 (39%)	60 (61%)	98 (100%)

Of these 262 applicants, 71 (27%) score above 0.7 on Model 1, indicating their desirable nature and of those, 42 were accepted for interview (60%) and 29 rejected (40%). Ninety-two people (35% of sample) score below 0.4 in Model 1, ie are totally unacceptable in terms of the examination model, and of these 69 (75%) were rightly rejected but 23 (25%) were accepted for interview. Thus 23 people who should not have been considered were interviewed (23% of all interviews) and 29 people who should have been offered an interview, were not.

These details are tabulated in Table 7.11.1, where there are significant differences between the high risk and low risk applicants accepted for interview¹⁹, with more interviews being offered to the latter. Firms are however rejecting 40% of the high scoring applicants along with 75% of the very poor applicants, but still interviewing 23 of the latter. In fact 58% of all interviews were with those scoring below 0.7.

Ninety-eight applications relate to undergraduates. These are again examined in the three areas of the suggested decision strategy, in Table 7.11.2, as if they hold a I/IIi degree and in Table 7.11.3, as if they have a lower class of degree. In this case, there are no significant differences between the choice of acceptance or rejection for interview between prospective high calibre entrants and the rest²⁰. In other words, where the recruiting firm did not have the degree class information upon which to base their decision to interview, they were unable to distinguish between good and poor applicants.

Examining Table 7.11.2 there are 55 applicants (56%) who, with a good degree would have scored above the 0.7 cut-off and 33 were rightly interviewed, leaving 22 who also represent missed opportunities for the Firms. Of the 43 applicants (44%) who would have scored under 0.7, even if they gained a I/II.i, 27 were interviewed. In fact, 7 of the 10 people who would have scored below 0.4 on the examination model and be deemed totally unacceptable, even with a good degree,

¹⁹ $\chi^2 = 20.38, 2 \text{ df}, p < 0.001$

²⁰ $\chi^2 = 0.36, p > 0.05$

were interviewed²¹. Again the differences between those interviewed and those not are not significant, reinforcing the inability of the recruiting staff being able to judge good from poor applicants, particularly when degree class was not available to them.

7.9.1 Individual Firm Results for Model 1

The results of the selection processes of each of the three Firms is briefly discussed here. Only the graduate applicants are considered in the light of their scores on Model 1.

Stoy Hayward

In Stoy Hayward, the largest firm, there was a policy of interviewing the majority of applicants, except those with very poor qualifications. Considering the examination model scores, of their 144 graduate candidates (75% of those applying to Stoy Hayward), high scoring applicants represent 29% of the sample. First interviews were given to 73% of them but also to 58% of the intermediate scorers and 38% of the low scorers.

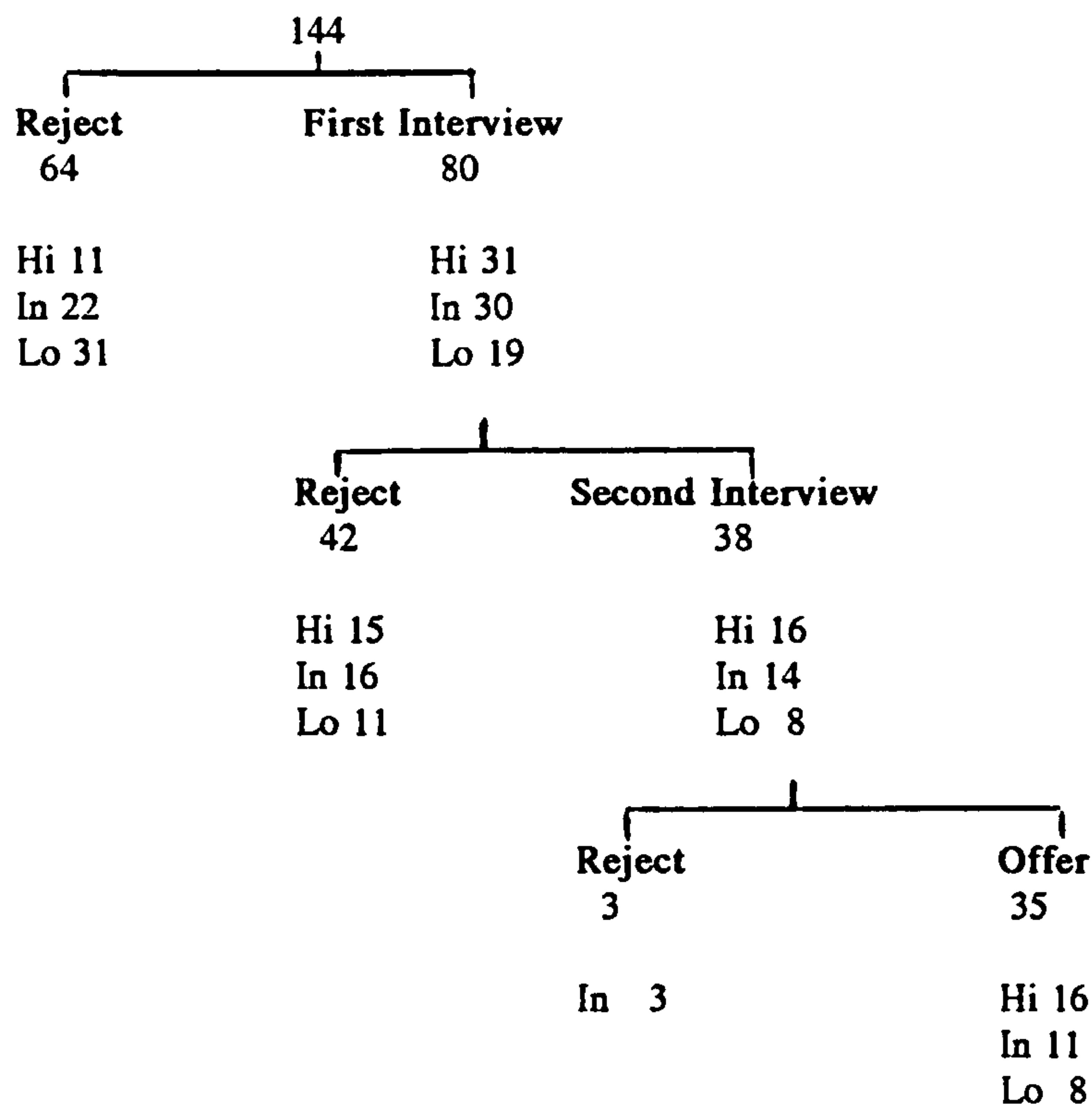
The reduction process by which suitable applicants are selected is illustrated in Figure 7.7. Of the 35 offers made, 46% were to highly rated applicants, 31% to risky applicants and 23% to those with a very high likelihood of examination failure. Twenty-six good applicants are rejected during the process (49% of all high calibre graduates) and 8 poor candidates are offered places (16% of all poor candidates).

Kidsons Impey

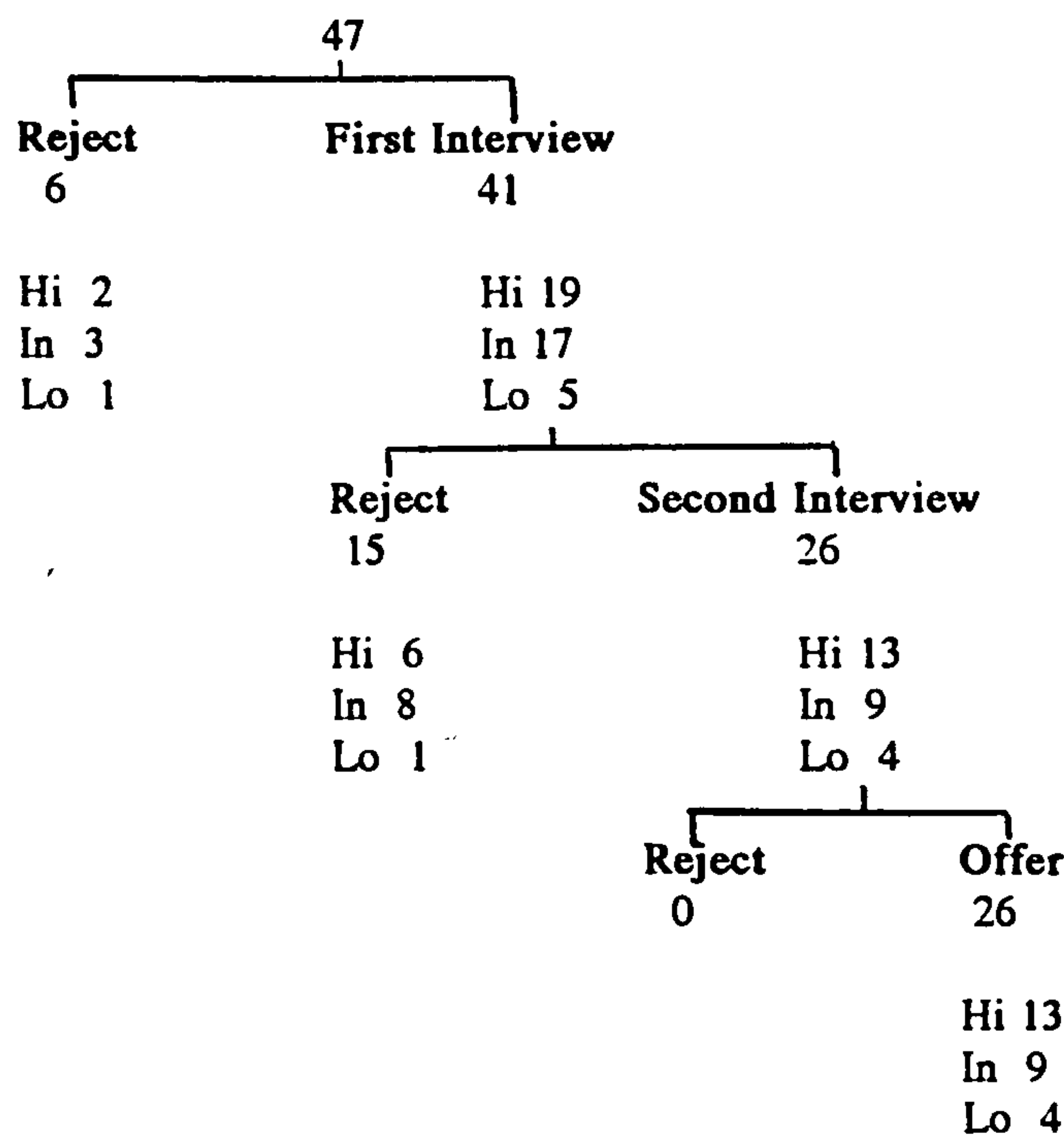
Kidsons Impey, the second largest firm, interviewed far less often. In their case, 88 applicants were graduates (68%) and only 14 of these (16%) were pre-selected for interview. Fifteen (65%) of the 23 desirable high scoring applicants were rejected, along with 30 of the 33 intermediate scoring

²¹ $\chi^2 = 1.93, p > 0.05$

7.7.1 Graduate Applicants



7.7.2 Undergraduate Applicants



* Hi > 0.7, In 0.4 - 0.7, Lo < 0.4

Figure 7.7 Stoy Hayward Decision Tree (Based on Model 1 Scores)

applicants and 29 of the 32 low scorers.

The reduction process is described graphically in Figure 7.8 which indicates that the Firm's pre-selection stage, which appears satisfactory in removing poor candidates, is, in fact, also removing more than two thirds of the good applicants. The interview process removes 88% of the remaining good candidates, either by rejection, or causing withdrawal, and results in only two offers, only one of which was to a high scoring applicants. This emphasises the extremely wasteful nature of selection interviews, which, in the case of Kidsons Impey resulting in the loss of 96% of those highly rated in terms of Model 1.

Finnies

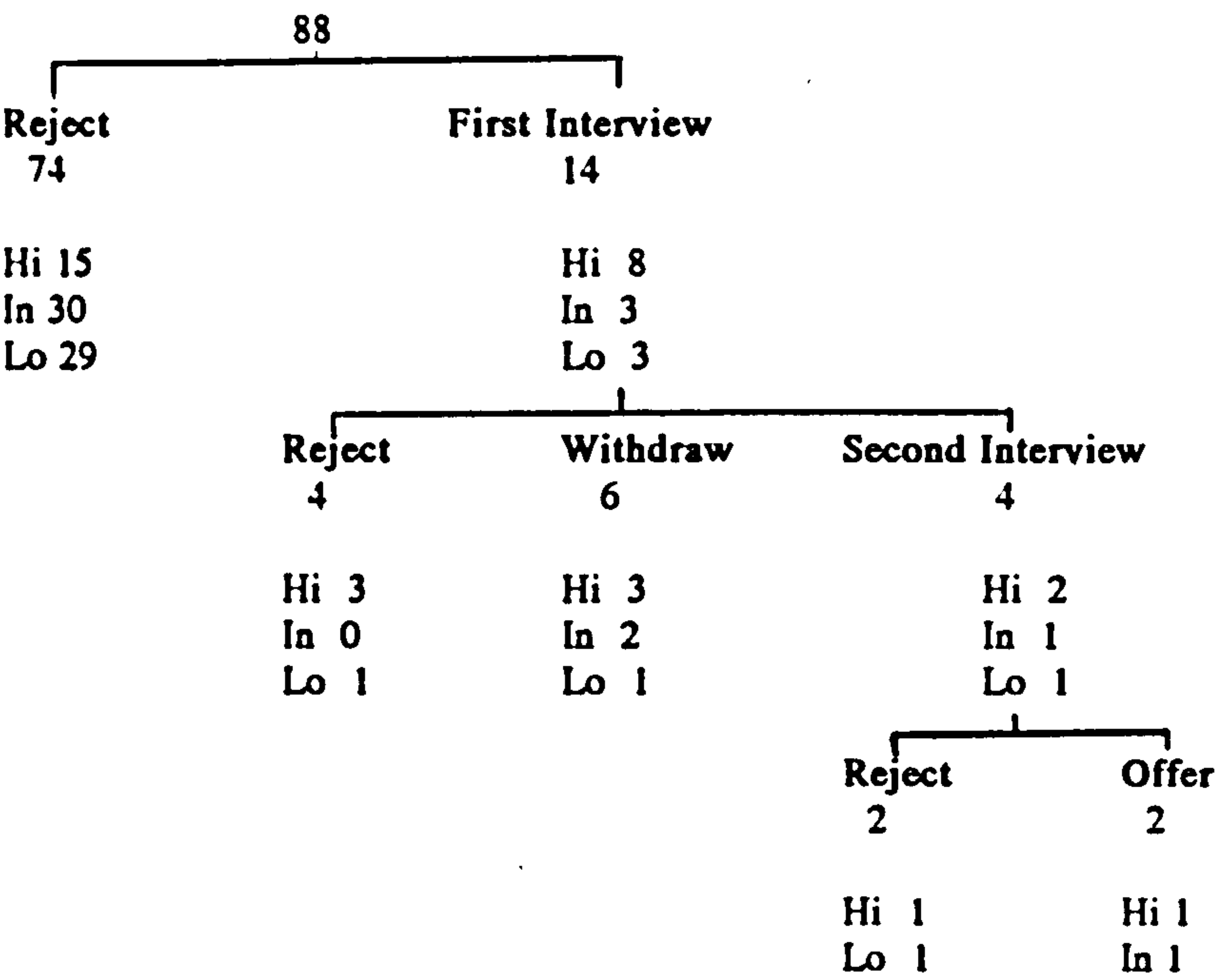
The last and smallest firm, Finnies, only conducted 6 interviews from their pool of 30 non-accounting applicants (Figure 7.9).

The 6 applicants consisted of 3 with high scores, 2 with intermediate scores and 1 with a low score. Two of the high scorers were accepted at interview but one subsequently withdrew after the first interview; the remainder were rejected. Only one of the 10 applicants yet to graduate was interviewed but subsequently withdrew. There are insufficient cases here to draw any strong conclusions, however, only one sixth of the good quality candidates available actually were made offers.

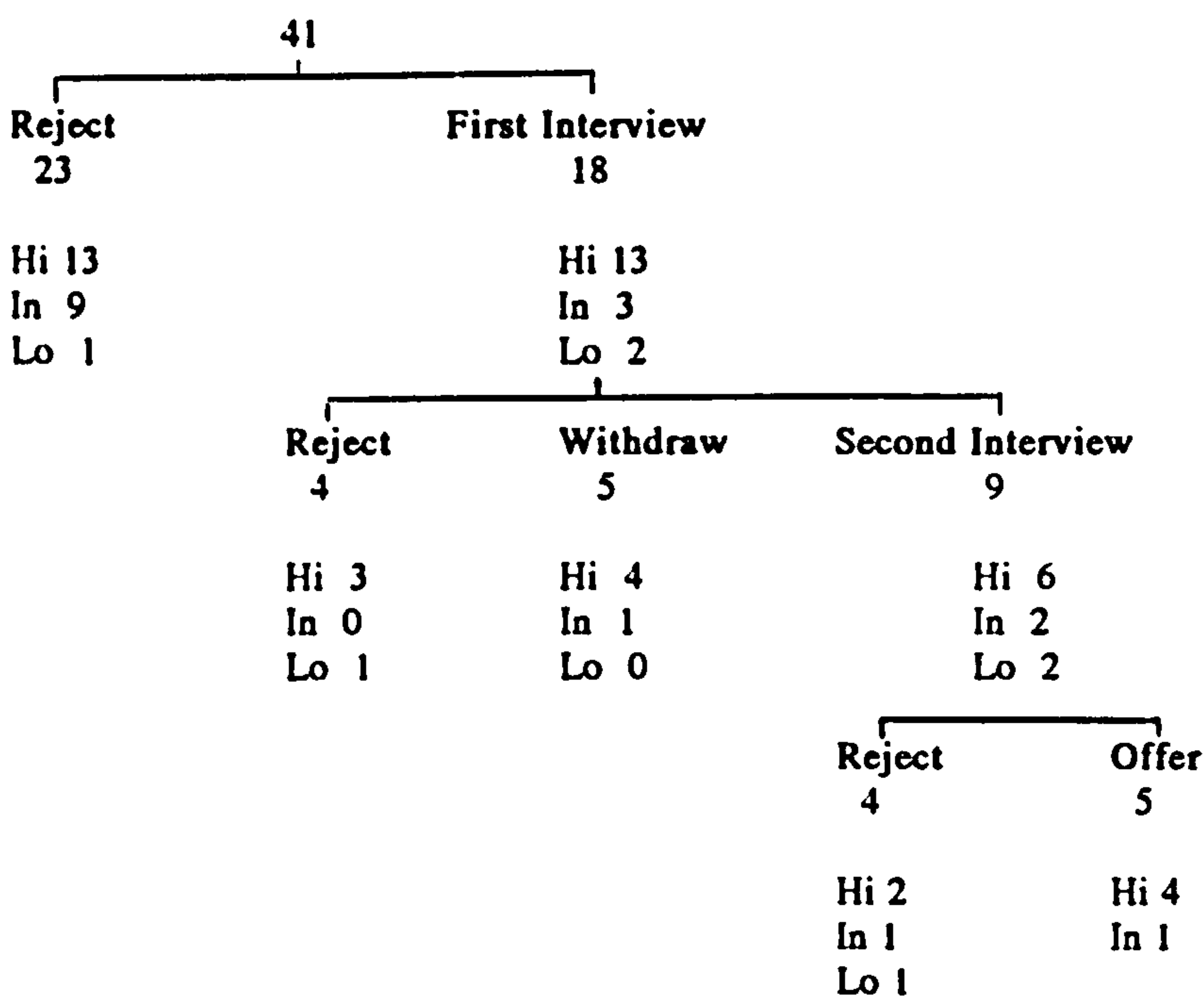
7.9.2 Individual Firm Results for Model 2

Table 7.12 gives equivalent data for Model 2. In this case, where applicants' degree class is known, 34% of cases have model scores above the suggested 0.6 cut-off. Of these 89 applicants, 32 only (36%) were interviewed. Sixty-one people had scores between 0.4 and 0.6, and of these 27 (44%) were interviewed. In the lowest scoring group of 112, ie those who should not be considered, no less than 41 (37%) were offered interviews. In short,

7.8.1 Graduate Applicants



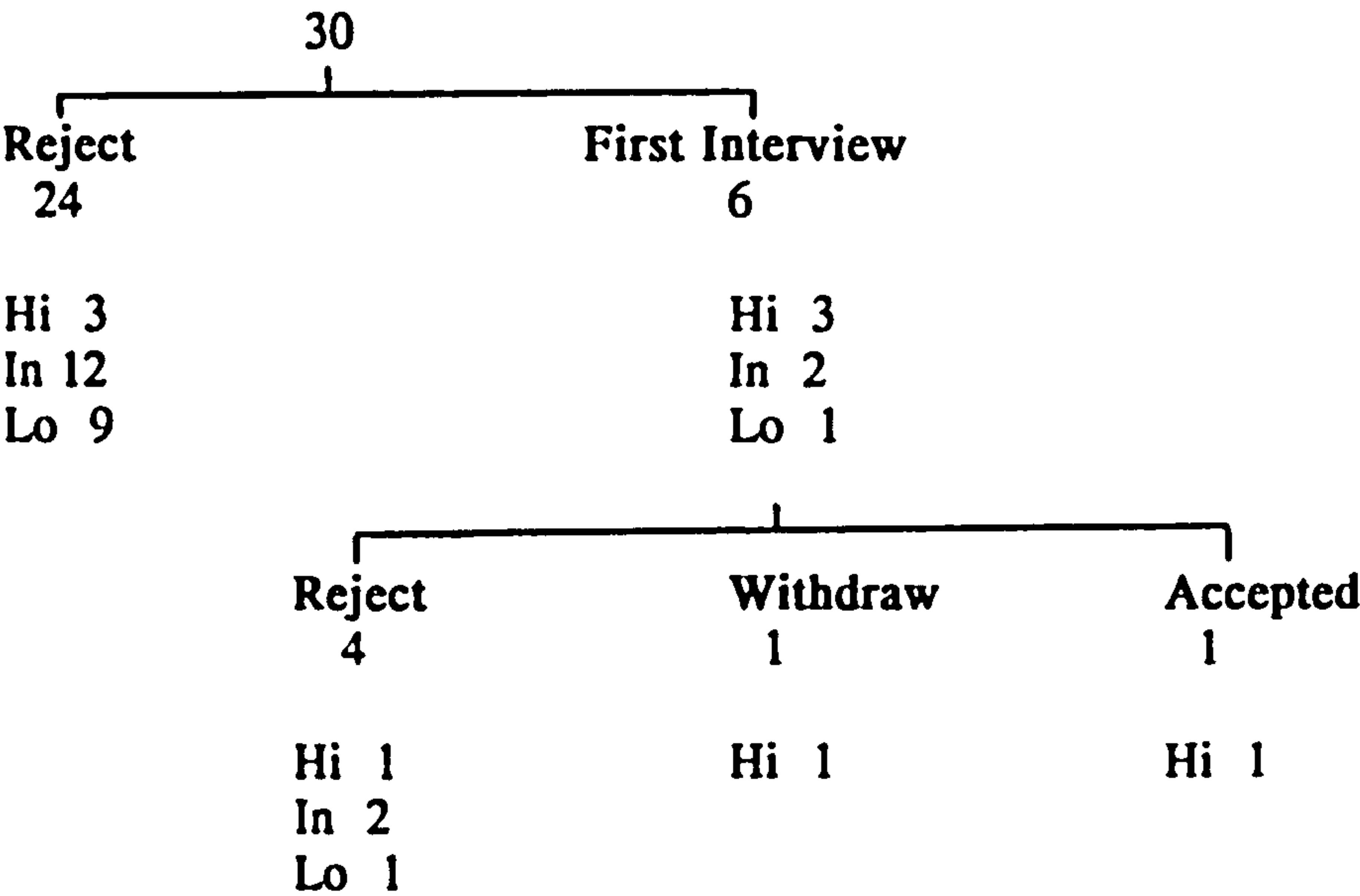
7.8.2 Undergraduate Applicants



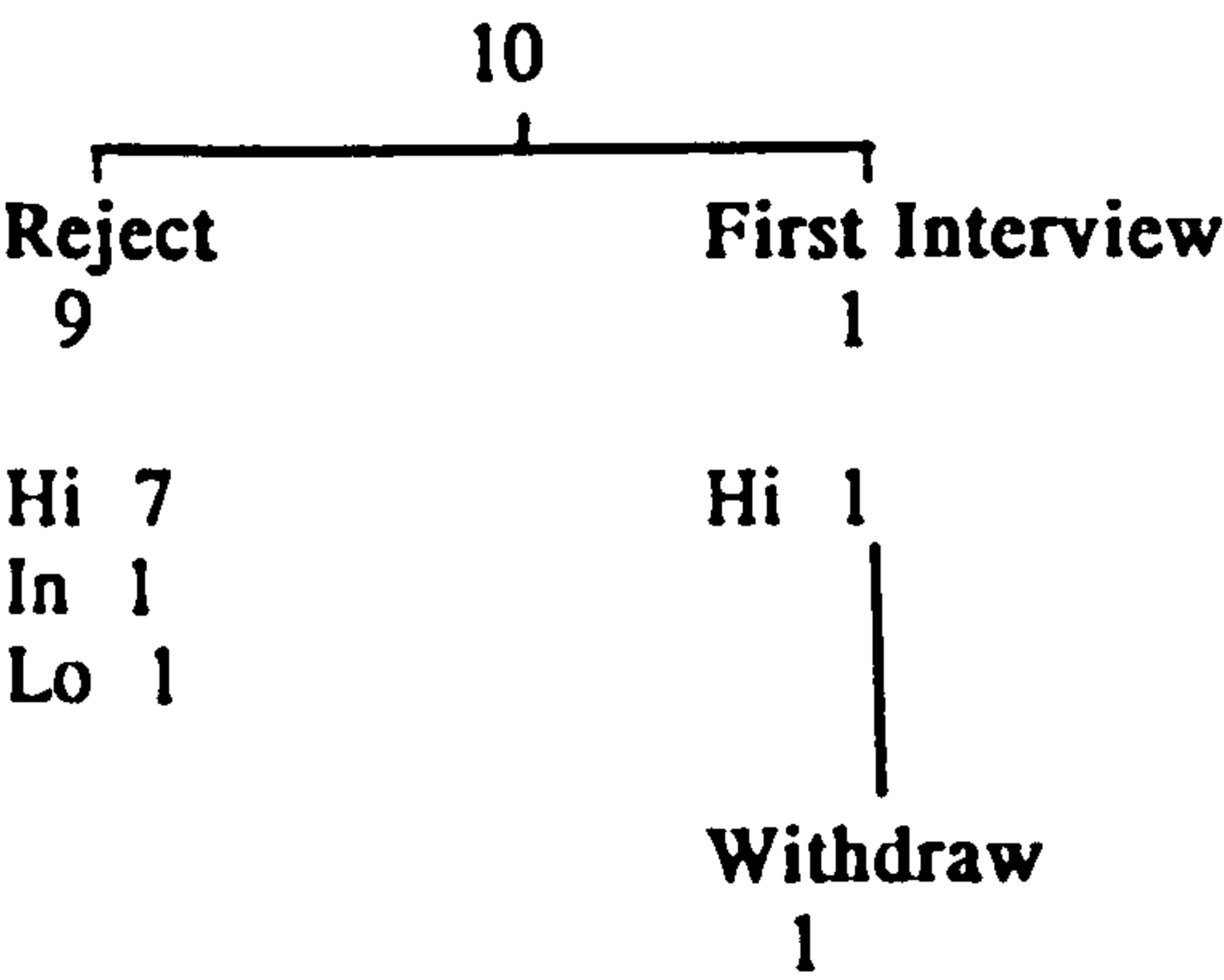
* Hi > 0.7, In 0.4 - 0.7, Lo < 0.4

Figure 7.8 Kidsons Impey Decision Tree (Based on Model 1 Scores)

7.9.1 Graduate Applicants



7.9.2 Undergraduate Applicants



* Hi > 0.7, In 0.4 - 0.7, Lo < 0.4

Figure 7.9 Finnies Decision Tree (Based on Model 1 Scores)

as many interviews are being offered to those graduate applicants who are unsuitable, in terms of the model, as to those indicated as being suitable.

Tables 7.12.2 and 7.12.3 again examine the scenarios attendant upon achieving a I/IIi degree or not. Thirty seven applicants would have scored below 0.6, even with a good degree and of these, 24 were offered interviews. However, of those who would have achieved a high score, even with a lower class degree, 34 cases, 18 (53%) were not offered an interview.

7.9.3 Parallel Use of the Models

Although the Model 2 may well not be as helpful in selecting successful candidates as Model 1 (section 7.4), it is a good indicator of the likelihood of leaving. Possible tandem use of the models has been discussed in section 7.7 above, using the Model 2 score to indicate the latter likelihood for those with borderline scores on Model 1.

The software described in section 7.8 indicates that the results of both models are available on the software program. In order to provide an example of the usefulness of the computer generated pre-selection information, particularly for tandem use of the models, the pilot sample applications are grouped into those who scored well on both models, those who scored well in one and not in the other (an intermediate category) and those who scored badly on both (poor candidates).

Table 7.13 shows the numbers of application forms from the sample of 262 received during the pilot period. To simplify presentation, these are grouped into three categories: High (scoring above 0.7 on Model 1 and above 0.6 on Model 2), Intermediate (scoring above 0.7 on Model 1 but scoring below 0.6 on Model 2 or scoring below 0.7 on Model 1 but scoring above 0.6 on Model 2) and Low (scoring below 0.7 on Model 1 and below 0.6 on

Table 7.12

Model 2 Decision Strategy Applied to Piloting Firms
7.12.1 Graduates

Score Level	Rejected	Interviewed	Total
> 0.6	57 (35%)	32 (32%)	89 (34%)
0.4 - 0.6	34 (21%)	27 (27%)	61 (23%)
< 0.4	71 (44%)	41 (41%)	112 (43%)
	162 (62%)	100 (38%)	262 (100%)

7.12.2 Undergraduates (If I/II.1 obtained)

Score Level	Rejected	Interviewed	Total
> 0.6	25 (66%)	36 (60%)	61 (62%)
0.4 - 0.6	8 (21%)	16 (27%)	24 (25%)
< 0.4	5 (13%)	8 (13%)	13 (13%)
	38 (39%)	60 (61%)	98 (100%)

7.12.3 Undergraduates (If I/II.1 not obtained)

Score Level	Rejected	Interviewed	Total
> 0.6	18 (48%)	16 (27%)	34 (35%)
0.4 - 0.6	10 (26%)	20 (33%)	30 (30%)
< 0.4	10 (26%)	24 (40%)	34 (35%)
	38 (39%)	60 (61%)	98 (100%)

Table 7.13

Combined Model Scores for Firms' Application Forms

7.13.1 Graduates

Score Category	Firms' Decision		Totals
	Interview	Reject	
High	19 (19%)	19 (12%)	38 (15%)
Intermediate	40 (40%)	47 (29%)	87 (33%)
Low	41 (41%)	96 (59%)	137 (52%)
	100 (38%)	162 (63%)	262 (100%)

7.13.2 Undergraduates

Score Category	Firms' Decision				Total
	Interview (If I/II.1 achieved)	Reject	Interview (If <II.1 achieved)	Reject	
High	19 (70%)	8 (21%)	14 (23%)	13 (48%)	27 (26%)
Inter	21 (34%)	16 (42%)	17 (28%)	14 (37%)	31 (32%)
Low	20 (33%)	14 (37%)	29 (49%)	11 (29%)	40 (42%)
	60 (61%)	38 (39%)	60 (61%)	38 (39%)	98 (100%)

Model 2).

The first striking observation is that equal numbers of graduates (19) scoring highly on both models appear in the 'accept for interview' and 'reject' groups. That is, half the high calibre candidates rejected. In addition, in the intermediate group the case is almost the same with 40 being accepted for interview (46% of score group) and 47 (56% of score group) being rejected.

However, the Firms were rather better at identifying the more disastrous students who score in the Low area (52% of known degree class scores), since 70% of these applicants were rejected at the outset. This still means that the Firms conducted 41 (41% of graduates interviewed) costly interviews which should not have been undertaken and ignored the opportunity of assessing 19 very good candidates.

Firms normally work on the assumption that undergraduate applicants will get a good degree, tempered by *ad hoc* interpretations of the end of year examination results to date. Examining the numbers of applicants for whom degree class was not known (98 cases), 19 score above the upper cut off on both models if they achieve a good degree. They are fairly well identified with a >2:1 ratio of interview to rejection.

For the 27 undergraduates who would have scored High on both models, even with a poor degree, there is almost equal acceptance and rejection for interview. In other words, 13 applicants, almost half of the undergraduates who are highly likely to be all round successes, have not been interviewed.

Again, those in the Intermediate groups are more or less equally treated with marginally more being offered an interview than not. However, in the case of those undergraduate candidates who will score badly even if they gain a good degree, the ratio of acceptance to rejection is 3:1. The latter is an

indication of the importance of having some measure of assessing the likely impact of low degree class as there is no point in approaching those who will have little chance of success in accountancy, even if they gain a I/II:i. An additional 21 interviews were wasted in this way, on the basis of model predictions.

7.9.4 Offers Made To Applicants

It was possible to ascertain from the Firms the number of offers they had made on the basis of the interviewing process. It is probably better to examine these on a firm by firm basis, as the various combinations of possible scores for the undergraduates are complex.

Stoy Hayward, on the basis of their interviewing process, which involved two interviews, both with senior members of staff, had offered places to 61 applicants. Of these, 37 had already graduated and 26 had not yet graduated.

Table 7.14 shows the score areas for both models and only 5/35 (14%) offers made to graduates are made to those with High scores on both models. In contrast 13 offers (37%) are made to those scoring below the upper cut-off on both models. For the undergraduates, only 2 applicants scoring above the upper cut-off on both models, independent of whether they gain a I/III degree, are offered places and this contrasts with the 8 applicants who score below the cut-offs, even if achieving a I/III.

Kidsons Impey offer details are given in Table 7.15. Only 4 offers are made to graduates and 5 to undergraduates. Of the two offers made to graduates, one is made to an applicant who scores above the upper cut-off on both models (25%). The other offer is made to a candidate with an intermediate score. Only one undergraduate accepted by the Firm scores above the upper cut-offs on both models, regardless of degree class and 3

Table 7.14

Stoy Hayward: Offers of Training Contract to 1990 Applicants

Graduates		Undergraduates			
		If I/II.1		If < II.1	
High*	5	High	8	High	2
Intermediate	17	Intermediate	10	Intermediate	3
Low	13	Low	8	Low	21
Total Offers	35		26		26

* High >0.7 Model 1 and >0.6 Model 2

Intermediate = Above upper cut-off on one model and below upper cut off on the other

Low <0.7 Model 1 and <0.6 Model 2

Table 7.15

Kidsons Impey Offers of Training Contract to 1990 Applicants

Graduates		Undergraduates			
		If I/II.1		If < II.1	
High*	1	High	3	High	1
Intermediate	1	Intermediate	1	Intermediate	1
Low	0	Low	1	Low	3
Total Offers	2		5		5

* High >0.7 Model 1 and >0.6 Model 2

Intermediate = Above upper cut-off on one model and below upper cut off on the other

Low <0.7 Model 1 and <0.6 Model 2

applicants score below the cut-offs on both models even if obtaining a good degree.

Finnies, the smallest firm of the three, only interviewed 6 people and offered one place to a High/High scorer who subsequently withdrew. This Firm was being very cautious in its recruitment due to very poor professional examination records for students in the preceding years and general need to cut back on staff. Only 3 of the 30 graduates applying score above 0.7 on the Model 1 and 7 score above 0.6 on Model 2. Two score above the cut-off on both models, one of whom was interviewed and accepted.

It is useful, at this point to consider the question of whether the samples differ between Firms. In Table 7.16 it is noticeable that *Finnies*, the smallest firm, receive the lowest proportion of high quality candidates, in terms of examination success, and have the largest proportion of mediocre applicants. All three Firms have roughly one third of applicants predicted by the model as being examination failures. The differences between the Firms are not significant²².

Equivalent information regarding the full sample Model 2 characteristics is found in Table 7.17 where again the difference between firms is not significant²³.

7.9.5 Discussion With Firms

The software instructions are explicit in that the usual standards of academic achievement should be met before the applicant is scored and Firms' pre-sifts should remove those not meeting the firm's basic academic

²² $\chi^2 = 1.62$, 4 df, $p > 0.05$

²³ $\chi^2 = 1.99$, 4 df, $p > 0.05$

Table 7.16

Total Applications Received Categorised by Model 1 Score

	Applications Received			
	Stoy Hayward	Kidsons Impey	Finnies	Total
H*	42 (29%)	23 (26%)	6 (20%)	71 (27%)
I*	52 (36%)	33 (38%)	14 (47%)	99 (38%)
L*	50 (35%)	32 (36%)	10 (33%)	92 (35%)
	144 (100%)	88 (100%)	30 (100%)	262 (100%)

* H >0.7, I 0.4-0.7, L <0.4

Table 7.17

Total Applications Received Categorised by Model 2 Score

	Applications Received			
	Stoy Hayward	Kidsons Impey	Finnies	Total
H*	50 (35%)	32 (36%)	7 (23%)	89 (34%)
I*	33 (23%)	19 (22%)	9 (30%)	61 (23%)
L*	61 (42%)	37 (42%)	14 (47%)	112 (43%)
	144 (100%)	88 (100%)	30 (100%)	262

* H >0.6, I 0.4-0.6, L <0.4

requirements, in the form of UCCA points and the ICAEW's requirement of basic C grades for 'O' level/GCSE Mathematics and English. This is an important point, since the recruits upon whom the models are based have already been through this original filter. (There were, however, a number of exceptions to the basic 9 UCCA point rule in the trainee sample.) For this reason, the figures relating to the sample of applications used above do not include subjects with less than 9 UCCA points.

When the pilot scoring was completed, senior recruitment staff were asked to pre-sift again and explain the apparent anomalies. In the case of those seemingly desirable applicants rejected, the Firms were unable, in the majority of cases, to produce a reasonable explanation for not calling the applicant for interview, other than, for a few cases, the very pertinent one of the applicant apparently not meeting the Firm's individual pre-sift criteria, eg, B grades for GCE 'O' level Mathematics and English.

If they are not directly involved in choosing applicants for interview, partners tend towards the belief that obviously missed opportunities are the result of errors occurring during the pre-sift or post pre-sift, eg the assessor had simply been faced with too many forms and had adopted an out-of-hand rejection strategy to clear the backlog.

Where directly involved, they offered reasonable explanations for very few cases. For example, in the case of one person applying to Finnies, who had attended university one year early but had extremely high scores on both models, the recruitment partner deemed her too young. Other excuses included C grade English Language 'O' level for those whose grasp of English might not be powerful enough (non-English mother tongue) and poorly written application form.

Nevertheless, the irregularity of the performance of the pre-sift and the almost random nature of selection for interview was forcibly demonstrated.

In Stoy Hayward it appears that a huge effort was made at the outset of the recruitment period, to interview and accept as many candidates as possible and that after the peak time for receipt of applications there would be no places available for even very good candidates applying late. The firm were certainly telling applicants in the January immediately after this pilot that there were no more places available for 1991/2 entry.

7.10 Implications of the Pilot

An examination of the recruitment system in current use, unfortunately, lends much weight to the theory of chance being prevalent in recruitment in these three representative firms. Such a hypothesis makes a good deal of sense when viewed in the light of the drop-out rates which cause the ICAEW and member firms such concern. The use of any statistically developed scoring system, able to predict subsequent criterion variable performance must be viewed as highly advantageous both for applicants and employers (section 6.8), providing it is suitably demonstrated to be valid and reliable.

The pilot exposes the realities of the pre-selection function in chartered accountancy firms. Whatever rules were being operated within our three representative Firms, pre-sifting was failing to identify high risk applicants, putting them forward for interview at the expense of high quality, low risk applicants.

The problem was exacerbated by unknown degree class as many of those who had a very high risk of failing examinations and/or leaving, even after obtaining a good degree, were offered interviews and, conversely, those who were low risk applicants even if they obtained a lower class degree, were not identified and rejected.

One of the key points here is that professional recruiters, working directly with the applicants' probability of success, using either one or both of the models, have the ability to gauge the risk attached to candidates and judge them accordingly, in terms

of the odds for and against their success and survival, before they meet them face-to-face.

It has been suggested earlier in section 3.4, that, for those candidates who have been highly rated by the models, the interview should concentrate upon assessing their ability to fit in to the firm's working environment, concentrating on 'relationship' variables. For, whereas these models are derived from a professional sample and intended for professional use, each organization will have individual differences which will effect entrant' success. The fit being confirmed, then a selling exercise should take place to convince the applicant that they should accept an offer from the firm.

7.11 Summary and Conclusions

The most appropriate strategy for using the developed models has been discussed here. The score distributions for the 1985/6 model development groups of both models were examined with a view to deciding which cut-off points, in conjunction with interaction from recruitment staff, would produce the best results in terms of the predicted success of the selected sample.

Both models are considered in terms of their relative predictive validity, but the results of applying the suggested strategies to the validation sample are tempered by the knowledge that these represent a recruited sample and not the full applicant sample, which is reduced normally by pre-sifting and selection interviews.

For the Model 1, an upper cut-off of 0.7 was established, those scoring above this level being >7 times more likely to be Pass than Fail group members. The upper cut-off for the Model 2 was set at 0.6, above which score the likelihood of being a member of the Leaver group is reduced to 7%.

Model 2 proved to be a less 'sharp' instrument, due to the large sample numbers

falling within the Intermediate category and of limited use. Nevertheless, Model 2's value may lie in its ability to predict leavers.

Applying the suggested cut-offs to the validation sample of 1987 trainees reveals that the results of selection by the models, while still offering very much better prospects than were obtained by current recruitment strategies, do not compare favourably with those of the 1985/6 validation sample. However, this sample may be, for reasons already discussed in section 6.6, not directly comparable with recruits of previous years. The conclusion is that the results are either directly affected by the non-stationarity of the environment or are not reliable.

The use of both models in tandem has also been discussed, adopting a strategy of recruiting only those scoring above the upper limits on both models. This would result in recruiting 38% of the trainee sample, 80% of whom were actually successful in terms of both models. The need for clear recruitment goals is emphasised in order for recruitment personnel to make judgements of the wisdom of accepting those who have not achieved high scores in both models, even where apparent recruitment number shortfalls may result from rejecting them.

It is suggested here that firms using a model-based strategy should only ever consider recruiting those who score above the cut-offs on the models, since there are likely to be adequate applicants of high calibre, in model terms, available without the need to consider those associated with higher failure/ leaving rates.

Discussion of the strategy of using the model-based approach to replace the selection interview, thus switching the emphasis of selection to Stage II of Figure 3.1, reveals that just over one third of entrants actually score above the 0.7 Model 1 cut-off and not all, of course, are totally successful in model terms. However, it is suggested that firms should be able to recruit sufficient numbers of such low risk candidates if they reduce the numbers recruited to reflect the reduction in likelihood of poor performance and loss through leaving. Nevertheless, an inspection of the applicant pool needs to be undertaken to confirm the likely availability of enough high calibre

applicants.

This chapter also reports on such an exercise which involved analyzing application forms from graduates and undergraduates arriving during the peak time in three medium sized firms. The analysis involved pre-sifting and model-scoring all applications received within a three month period. A software-driven scoring system was developed for the purpose of scoring applications and this provides the applicant's score for both models.

This pilot demonstrated that the three Firms were indeed missing the opportunity of recruiting high calibre graduate applicants by rejecting half at the pre-selection stage and wasting interviews on 38% of those scoring below the set cut-off of either one or the other model. One third of undergraduate interviews were wasted on those who would have a very poor chance of success even if they were to gain a good degree class, while half of those who would score well on both models even with a II:ii degree were not interviewed.

On the basis of the results of the pilot, all three Firms expressed interest in continued use of the models and decided to implement them on a proper basis, commencing with all applications from non-relevant graduates reaching the firms from September 1991, ie the 1992 entrants.

Firms were approached in the summer of 1992, after one year of implementation, to discover how they had found the system and what improvements might be possible. Also, bearing in mind the warnings concerning the shrinkage problems associated with biodata models, and the results of the 1987 validation, a full validation exercise, based on the three years subsequent to the model development and validation years of 1985-7 was suggested, to assess the reliability of the models well out of the sample time period.

The following chapter will describe the validation exercise which resulted from this consultation. This represents Part II of this study and also involves not only an

assessment of the utility of the derived models in practice, but a test of the effects of applying the models to accounting graduates, not part of the original sample.

Part II

Validation and Further Development of Logit Models

8 Validation and Further Development of Logit Models

8.1 Introduction

The specific focus of this research is the improvement of recruitment to the chartered accountancy profession using predictive hard biodata. This improvement is seen to benefit both the employer and the employee. The former benefits from increased productivity and reduced recruitment costs and the latter from the reduced likelihood of becoming a Type I or II error statistic with the associated loss of confidence and precious time. Chapter 3 examined the extant recruitment procedures adopted by the profession which were found to be unfavourably viewed by the literature. Poor recruitment practice, in particular, dependence on the unstructured selection interview, has led to the unacceptable recruitment error rates which triggered the research reported in Part I.

Chapters 5, 6 and 7 described the development and validation of predictive recruitment biodata models, Chapter 7, in particular, providing evidence of the usefulness of the developed models in practice. However, the issue of long term validation needs to be addressed as the original intention to validate the models developed from 1985/6 data on the 1987 trainee sample was compromised by changes in a criterion variable common to both models.

In Part II of this study, the issue of reliability is again addressed. Validation is conducted on a sample of entrants from the 1988, 1989 and 1990 entries. The three firms used to pilot the models on their current trainees, also provide details on the three years subsequent to those used to develop the original models. In the interim period, Finnies had merged with Stoy Hayward and both firms agreed to take part in the validation exercise.

In addition to the data on entrants with non-accounting degrees, data was also

collected for trainees with full exemption from GCC, those with accounting degrees. The sample, therefore, constitutes the total intakes to these firms and is considered to be representative of entrants in these three years across the broad spectrum of medium size firms.

Reported reduction in validity of biodata models over time (Shuh, 1967; Owens and Schoenfeldt, 1979, Eberhardt and Muchinsky, 1982b; Mitchell and Klimoski, 1982; Drakeley, 1989b) may have been variously due to the inclusion of soft data items in models, changes in either the predictor or criterion variables, sampling error, model over-fitting or methodological weaknesses.

The logit biodata models developed in this study are based on hard data items drawn from reasonably large, representative samples. The logit methodology provides robust models and should not result in over-fitting. The procedure offers an indication of the strength of the relationships between predictor and criterion variables and provides a score which represents the probability of the successful outcome in model terms. Nevertheless, further validation is crucial to determine the underlying reliability of the models.

It has been suggested that biodata model keys, unlike cognitive test items, are not transportable (Hunter and Hunter, 1984), ie, they are organization-specific. Furthermore, Theyer (1977) argues that biodata model validity is moderated by 'age, organizational practices and procedures, criterion and temporal changes in the nature of the job among other factors.' Rothstein *et al* (1990), however, find that:

'... biodata questionnaires are capable of capturing general characteristics of people that conduce to success or failure on the job in a wide variety of settings, organizational climates, technologies and so on.'

Part II of this research specifically tests the thesis that appropriately developed biodata tools are, indeed, not only stable over time but can generalize across a profession, treating it as one large organization.

In addition, if the evidence demonstrates that the models originally developed for one segment of the intake, the non-accounting graduates, transfer to another subsample, accounting graduates, then this would support the arguments of Rothstein *et al* (1990) concerning the generalizability of biodata models across similar populations. This is a feature of the various cognitive tests which are used by many different types of firm on a routine basis, for scoring applicants on a wide range of criteria. The biodata models developed here, it is argued, may well have similar properties to those of a cognitive test and should be applicable to any graduate applicant to the accounting profession and, perhaps, other professions attracting similar types.

If Holland's theory of vocational preference (Holland, 1976) holds true, the likelihood is that models developed to predict professional success should apply to all members of the profession and be transferable to different groups within that vocational type.

Although accounting graduates are exempt from the GCC (Foundation) examination, they still sit the two professional examinations and they also enter into the same three year training contracts. They are considered to be at a higher level of accounting competence when entering their contracts than the non-accounting graduates, but when non-accounting graduates have successfully passed their Foundation examination, accounting firms consider the two groups to be of equal ability in approaching PE1.

Nevertheless, accounting graduates have pre-entry accounting experience and are less likely to suffer from motivation-related withdrawal. They have made the decision to become accountants much earlier than non-accounting graduates, as such, their commitment levels are likely to be higher and they are thus less likely to leave, at least within the first two years.

The criteria the profession hold to be of greatest importance for the non-accounting graduate applicants, in terms of prediction, described in section 4.2, do not all hold

for these entrants. Recruiters in the profession recognise that previous accounting experience is likely to result in lower ORS and therefore a lower likelihood of leaving while under contract (Wilson, 1989).

Recruiters in the Firms participating in the pilot study viewed Model 2 as an indicator of motivation for the non-accounting graduates, because of the initial performance criterion adopted (first six months), and therefore not relevant to the accounting graduates. However, the importance of professional examination success for both types of entrant remains a key issue for these firms. Thus, only the results of applying Model 1 to the accounting graduate trainees are examined in this chapter.

The emphasis of the empirical research described in Part I is on theoretical issues but practical issues within firms directly affect the use of such devices. As these models were specifically developed for practical use, the opportunity was also taken to discuss experience of using the models in a practical situation and several interesting points emerged. For example, no consideration was made by firms of the implications in terms of changes in the need for recruitment staff which would result from the use of a model-based approach.

In the traditional unstructured pre-sift system, the sifter simply counts the UCCA points and ensures that the applicant has reasonable grades for 'O' level (GCSE) Mathematics and English (grade C in both the latter being requirements of the ICAEW). Application forms meeting these requirements are then vetted by a more senior colleague, who decides which applicants should be interviewed, using judgemental or arbitrary, and often inconsistent, criteria.

The arguments put forward for a model-based approach include, *inter alia*, using such models to reduce the number of applicants approached to those only with an good chance of success, thus reducing the time input of expensive senior staff. However, the demands on clerical staff at the pre-sift stage may increase and those concerned may well be fully committed in other aspects of work. The result may be

an overload, particularly at busy times, resulting in unacceptable turnaround times for applicants.

The above scenario was, in fact, mirrored in Stoy Hayward (now including Finnies), when approached in 1992, by the expected considerable reductions in interviews and expensive partner time, through adoption of the models. However, as no further provision was made for staff at the pre-sift stage, the processor had not been able to keep up with flow of incoming applications.

On investigation in 1992 it was apparent that Stoy Hayward were employing arbitrary measures to reduce the number of forms to be processed. Despite the acknowledgement that the pre-sifting system and interview selection procedures of the Firm had been found wanting, the Firm reverted to using their previous entry criteria, with which their processor was familiar and content. The forms passing this pre-sift were then examined by a more senior staff member who selected applicants using judgement based degree class and subject. Those successful at interview were then scored before being offered a second interview.

Kidsons Impey conducted very few interviews anyway and their policy was to score all applicants with the required UCCA points at their National Recruitment office in London, recommending those who score above the two upper model-based cut-offs to the office of their choice. It was left to the recruiting offices to make the decision whether to approach the applicant or not.

Even though computerised selection takes only a few minutes, it still takes more time than clerical staff took for the original vetting. In addition, there are two items which require more than a simple count : type of school and number of solo active pastimes at university. The latter also requires judgement and concentration and associated time, which neither of the processors at Kidsons Impey doing the pre-sift could spare. This again caused a backlog to build up but the firm continued to use the models as recommended for the 1991 applicants.

However, whereas the models are explicitly aimed at non-relevant graduate applicants with average (9) UCCA points, in 1992 Kidsons Impey, as a result of a senior partner reviewing ICAEW statistics after a particularly poor set of PE1 examination results in the Firm, raised their cut-off to 11 UCCA points. This was despite there being no evidence that this is any more valid a method of reducing numbers of poor applicants than 9 UCCA points or 12 UCCA points. The models indicated that subject studied at 'A' level are important and UCCA point counts do not take these into consideration.

To deal with these problems, the possibility of developing an efficient pre-sift based on one or two key model variables, which would significantly reduce the numbers to be processed without unacceptable loss of potentially good candidates, is examined here.

Firms also expressed an interest in the real implications of selection model use for them in financial terms, and these are briefly examined, although, of necessity, the analysis depends on estimated costs and revenues provided by the firms.

To summarise, the aims of Part II of the study are

- (i) to confirm the reliability of both models by testing their inter-temporal stability on a sample of trainees from the three years following those used for model development and validation,
- (iii) to test the generalizability of such devices a) across different organizations and b) across different samples
- (ii) to develop an efficient method of reducing numbers of applicants to be scored without unacceptable loss of desirable candidates, in order to offset the effects of model use on recruitment staff time and
- (iv) to assess the costs and benefits associated with recruiting successful and unsuccessful trainees of both types, in order to provide evidence of the

financial utility of using such models for recruitment purposes.

8.2 Model Validation

8.2.1 Sample Details

Original model development used 1985/86 entrant data from the 23 firm offices identified by the ICAEW and initial validation of these models used the 1987 intake data from the same offices. Further validation reported here is carried out on 1988, 1989 and 1990 trainee performance in the piloting firms: Stoy Hayward, Fannies and Kidsons Impey. Stoy Hayward and Fannies trainees are drawn from their London offices. Kidsons Impey students come from their London, Birmingham and Manchester offices.

This sample is again treated as if it were a group of applicants, whereas, of course, they are recruits. Since recruitment to the profession, based on the interview, may be viewed as not all that much better than a random process, given average UCCA points, viewing the sample as representative of the applicants to the population of medium size accounting firms is, arguably defensible.

This is supported by the details of the applicant pool given in Table 7.11.1, where the distribution of scores from the application forms indicates a similar spread to those obtained from the recruited sample. Combining the graduate applicant numbers with the undergraduates (treated as if they gain a good degree) (360 cases), provides a total of 126 scoring above 0.7, 102 scoring ≤ 0.4 and 132 with scores between 0.4 and 0.7. These proportions do not differ significantly from the equivalent figures for the full 1985/6 sample of 154, 120 and 145, respectively ²⁴.

²⁴ $\chi^2 = 0.4$, 4 df, $p > 0.05$

8.2.2 Data

Data for all available accounting (exempt) and non-accounting (non-exempt) trainees from the above offices is used in Part II. Table 8.1 provides details of the 323 cases broken down by firm and exempt/non-exempt status. Exempt graduates make up 24% of the sample. Stoy Hayward provide 211 cases (65% of sample), Kidsons Impey, 83 cases (26%) and Finnies the remaining 29 cases (9%).

Additional items of biographical data were collated from the personnel files and application forms of the students in the sample, in addition to the information required by the models, to provide appropriate information to rework the models if required.

8.3 Model 1 Validation Results

Details of the 1988-90 validation sample (244 non-relevant trainees) may be found in Table 8.2. There is no significant difference between the proportions recruited in the different categories compared with the original 1985/6 sample (Table 4.1)²⁵. Figure 8.1 provides Model 1 score distributions for the Pass and Fail groups of non-exempt 1988-90 entrants. This sample is used to test the out-of-sample predictive ability of the two developed models, well outside the original training period, using the cut-offs of 0.7 (upper) and 0.4 (lower) discussed in section 7.13.

Table 8.3 provides the outcome probabilities by Model 1 score for the 1988-90 entrants and Figure 8.2 graphs the equivalent probabilities of Pass and Fail group membership alone. The probability of criterion group membership for the 1988-90 sample, within the three areas illustrated in Figure 7.3, is tabulated in Table 8.4. This may be compared with those in Table 7.1 and Figure 7.2. There is obviously

²⁵ $\chi^2 = 4.97, p > 0.05$

Table 8.1
Validation Sample Details

Firm	Exempt Graduates	Non-Exempt Graduates	Total
Stoy Hayward	57 (72%)	154 (63%)	211 (65%)
Kidsons Impey	19 (24%)	64 (26%)	83 (26%)
Finnies	3 (4%)	26 (11%)	29 (9%)
	79 (24%)	244 (76%)	323(100%)

Table 8.2

**1988-1990 Non-exempt Entrants
Model 1 Validation Sample**

Criterion Group	Characteristics	Number	
Pass	Pass GCC and pass PE1 at first attempt	92	
Fail	Fail/refer GCC and fail/refer PE1	18	
	Fail/refer GCC and leave	26	
	No GCC sat	16	60
Intermediate	Pass GCC and fail/refer PE11	68	
	Pass GCC and no PE1 sat	12	
	Fail/refer GCC and pass PE1	12	92
			<u>244</u>

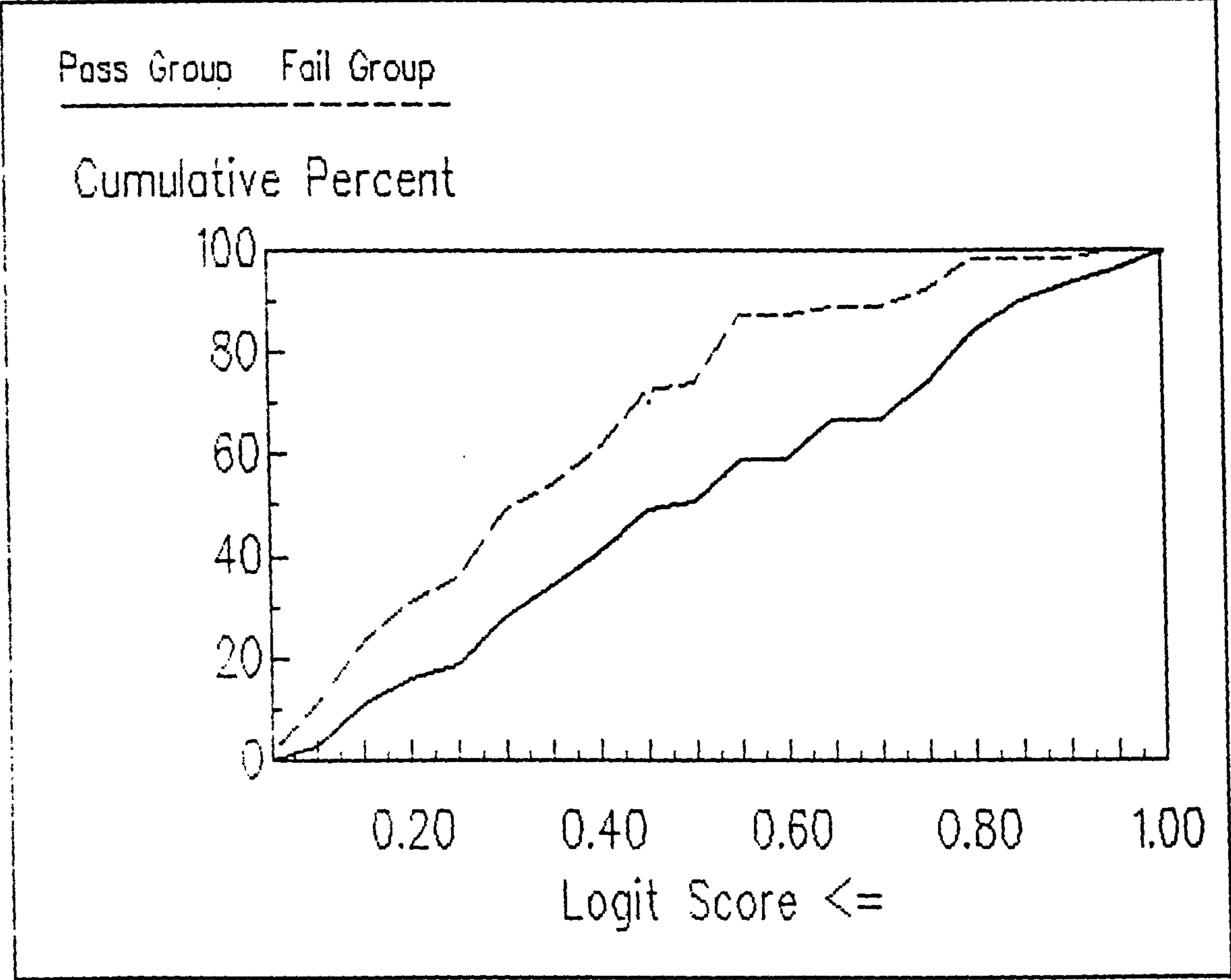


Figure 8.1 : Model 1 1988/90 Groups' Logit Scores

Table 8.3**1988-90 Non-Exempt Validation Sample
Model 1 Outcome Probabilities**

Score	Population % (244)	Candidate Outcome Probability		
		Pass (92)	Fail (60)	Intermediate (92)
0.00 - 0.10	6%	0.20	0.47	0.33
0.11 - 0.20	17%	0.25	0.28	0.47
0.21 - 0.30	18%	0.25	0.26	0.49
0.31 - 0.40	14%	0.35	0.21	0.44
0.41 - 0.50	10%	0.42	0.33	0.25
0.51 - 0.60	9%	0.30	0.35	0.35
0.61 - 0.70	5%	0.54	0.08	0.38
0.71 - 0.80	12%	0.52	0.21	0.27
0.81 - 0.90	4%	0.89	0.00	0.11
0.91 - 1.00	5%	0.55	0.09	0.36
100%	0.37	0.25	0.38	

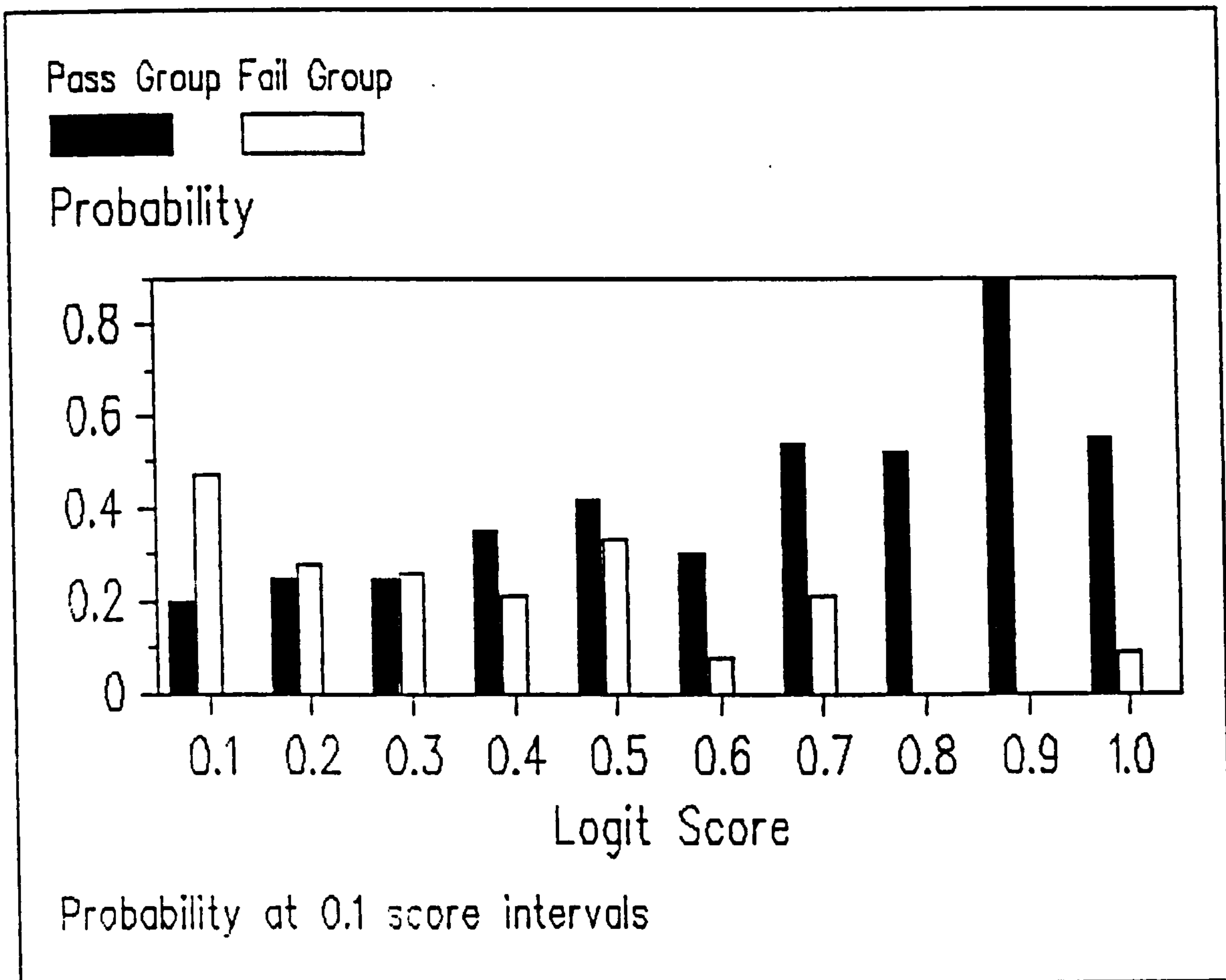


Figure 8.2 Model 1 1988-1990 Group Outcome Probabilities

Table 8.4

1988-90 Non-Exempt Validation Sample

Model 1 Decision Group Probabilities

Score Level	All	Probability of Group Membership		
		Pass	Fail	Intermediate
<= 0.4	55%	0.28	0.28	0.44
0.41 - 0.70	25%	0.40	0.28	0.32
>0.70	20%	0.59	0.14	0.27
(1988-90)	100%	37%	25%	38%
(1985/6)		31%	23%	46%

still clear differentiation between groups' scores.

The percentages of the full 1988-90 sample (n= 241 for whom scoring data was available) represented by each of the Pass, Fail and Intermediate groups is compared with the corresponding groups of the 1985/6 development sample at the foot of Table 8.4. Very little difference between the development sample (31%, 23% and 46% respectively) and the 1988-90 validation sample (37%, 25% and 38%) may be observed²⁶. However, the differences between the 1987 validation sample (22%, 41% and 37% respectively) and the 1985/6²⁷ and the 1988-90 samples²⁸ are much more marked and significant.

For 1988-90 entrants, the likelihood of being an examination failure with a score higher than 0.7 is only 14%, but of being successful is 59%, over four times as high. However, for the 1987 validation sample, the probability of being a Pass group member for the 34% of the sample who score >0.7 is only 0.34.

Eighty percent of the 1988-90 validation sample scores lie below the 0.7 cut-off (compared with 63% for the 1985/6 sample and 66% for 1987) and within this segment of the sample, the likelihood of being a member of the pass group is 0.31 (compared with 0.20 for 1985/6 and 0.37 for 1987).

These results support the argument that the 1987 validation results were directly affected by the change in the GCC examination and that many of those who would have passed this examination prior to that year, and perhaps in subsequent years, did not pass in 1987. Subsequently, perhaps when the training firms had come to terms with the new syllabus, the GCC first time pass rate appears to have returned to former levels.

²⁶ $\chi^2 = 3.45$, 4 df, $p > 0.05$

²⁷ $\chi^2 = 21.07$, 2df, $p < 0.001$

²⁸ $\chi^2 = 17.46$, 2df, $p < 0.001$

This model continues to exhibit satisfactory ability to select suitable applicants and clear inter-temporal predictive ability.

8.4 Model 2 Validation Results

The breakdown of the 1988-90 validation sample (244 non-relevant trainees) may be found in Table 8.5, where the non-exempt Success, Leaver and Intermediate groups of this sample are described. Unfortunately, of the 244 non-relevant cases available for analysis, 44 were lost due to missing data problems. Nevertheless, the number of cases falling within the criterion groups for the 1988-90 sample may be compared with those in Table 4.2 for the 1985/6 sample.

The 1985/6 scored sample is made up of 24% Successful, 22% Leaver and 54% Intermediate, the 1988-90 sample figures being 25% Successful, 24% Leavers and 51% Intermediate. There is no significant difference between these two samples on this basis²⁹.

The frequencies of 1988-90 sample scores for Model 2 may be examined in Figure 8.3, where again clear differentiation is seen between groups. Table 8.6 provides the group outcome probabilities by Model 2 score and Figure 8.4 graphs the equivalent probabilities of being Successful or Leaver.

An examination of Figure 8.4 and Table 8.6 reveals that the likelihood of being a Success group member at the lower end of the score range is very low (0.05), compared with 0.37 of being a Leaver. Those scoring below 0.4, the suggested lower cut-off, have a probability of 0.29 of leaving and only 0.16 of being rated successful in model terms. These figures compare with the development sample (0.29 and 0.08 respectively) and 1987 validation sample (0.27 and 0.10 respectively).

²⁹ $\chi^2 = 0.2$, $p > 0.05$

Table 8.5

**1988-1990 Non-Exempt Entrants
Model 2 Validation Sample**

Criterion Group	Characteristics	Number	
Success	Pass GCC and PE1 at first attempt and have an average or above assessment rating at six months		49
Leaver	Leave within two years		48
Intermediate	Fail GCC and fail PE1 but not leaving	15	
	Fail either GCC or PE1 but pass the other	65	
	Pass both GCC and PE1 but low six month rating	23	103
			200

Table 8.6

**1988-90 Non-Exempt Validation Sample
Model 2 Outcome Probabilities**

Score	Population % (200)	Candidate Outcome Probability		
		Success (48)	Leaver (49)	Intermediate (103)
0.00 - 0.10	10%	0.05	0.37	0.58
0.11 - 0.20	14%	0.14	0.29	0.57
0.21 - 0.30	14%	0.18	0.32	0.50
0.31 - 0.40	16%	0.24	0.21	0.55
0.41 - 0.50	16%	0.30	0.24	0.46
0.51 - 0.60	8%	0.27	0.07	0.66
0.61 - 0.70	8%	0.35	0.24	0.41
0.71 - 0.80	10%	0.37	0.21	0.42
0.81 - 0.90	2%	0.33	0.00	0.67
0.91 - 1.00	2%	0.60	0.00	0.40
	100%	0.25	0.24	0.51

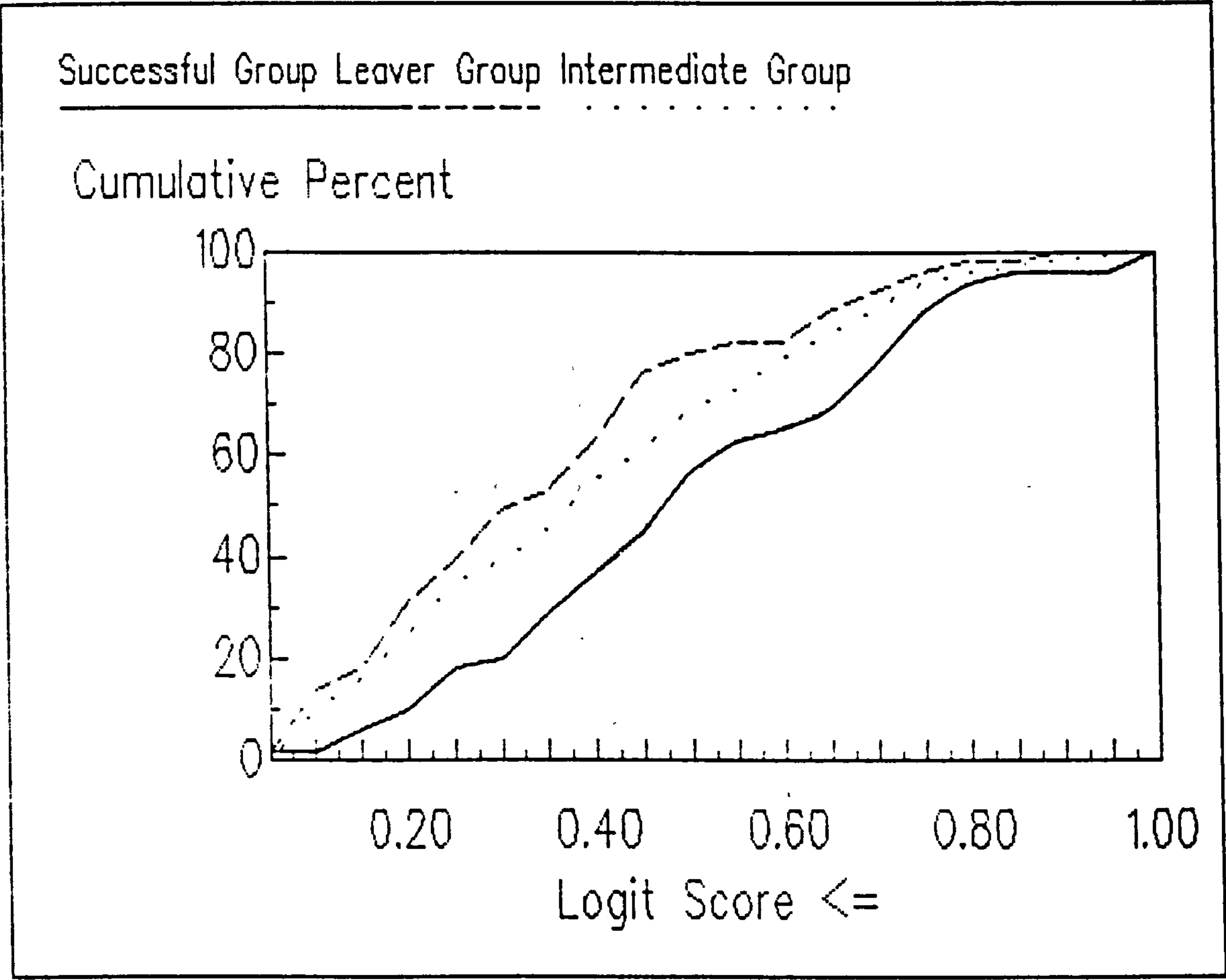


Figure 8.3 : Model 2 1988/90 Groups' Logit Scores

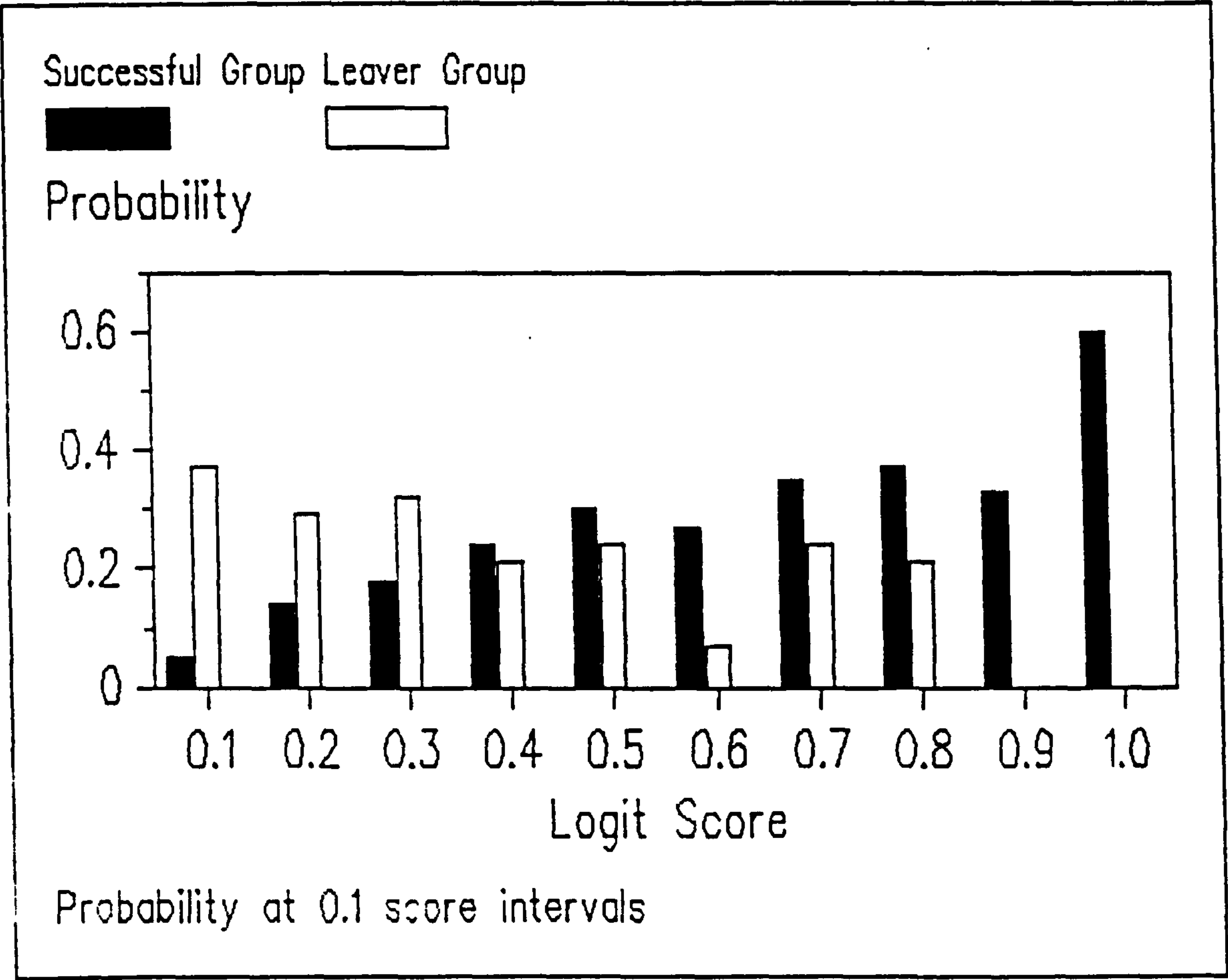


Figure 8.4 Model 2 1988-1990 Group Outcome Probabilities

Table 8.7
1988-90 Non-Exempt Validation Sample
Model 2 Decision Group Outcome Probabilities

Score Level	All	Success	P r o b a b i l i t y	
			Leaver	Intermediate
< = 0.4	54%	0.16	0.29	0.55
0.41 - 0.60	24%	0.29	0.18	0.53
> 0.60	22%	0.39	0.19	0.42
(1989-90)	100%	(25%)	(24%)	(51%)
(1985/6)		(24%)	(22%)	(54%)

For those in the 1988-90 sample scoring above 0.8, the likelihood of being a Leaver has completely vanished (compared with 0.08 for 1985/6 and 0.12 for 1987), although only 4% of the sample were so rated, compared with 15% of 1985/6 trainees and 9% of 1987.

The probability of Success with a score above the recommended 0.6 cut-off is 0.39% and of Leaving is only 0.19. These compare with 0.34 and 0.07 for 1985/6 and 0.15 and 0.12 for the 1987 validation sample, respectively. For the 1988-90 sample, the respective probabilities for those who score below 0.4 are 0.16 (Success) and 0.29 (Leaver), compared with 0.08 and 0.29 for 1985/6 and 0.1 and 0.27 for 1987.

It should be noted how many of this validation sample fall in the Intermediate category, 51% are so categorized, compared with 38% for Model 1. Thus, because of the high likelihood of trainees being members of the Intermediate group, this model is arguably of limited use, even though the cut-off of 0.6 removes 83% of the leavers.

8.5 Summary of Validation on the Non-Exempt 1988-90 Sample

The inter-temporal stability of both the original models has been examined by using a validation sample well separated in time from the original model derivation sample. Having established that the 1988-90 validation sample does not differ significantly from the 1985/6 model derivation sample, the examination prediction model is found to maintain validity for selecting non-exempt trainees to the chartered accountancy profession. Maintenance of the same acceptance cut-off score of 0.7 results in a 59% chance of recruiting an applicant successful in examination terms and a 14% chance of recruiting a likely examination failure.

However, even though Model 2 is useful for predicting likely 'leaver' trainees, the criteria originally selected by firms for group membership have subsequently proved

not to be particularly useful. In addition, the processing time associated with coding percentage of exemption (which requires consulting the ICAEW's university course lists) and the two commitment variables, in order to provide the prediction of success v. leaving, resulted in Model 2 being less popular with participating firms.

The hard data component of the criterion measure of Model 2, ie passing the Foundation and first Professional Examinations at the first attempt, is successfully predicted by Model 1. Nevertheless, predicting the likelihood of a trainee leaving is very important to firms because such losses are very costly in both in terms of finance and disruption. An appropriate parallel use of the model might be as a predictor of the likelihood of leaving for those rated highly on Model 1.

8.6 Extension to Accounting Graduates

To test the theory that such models may be 'transportable', ie, generalizable across samples, the results of applying the developed models to the accounting graduates are examined here.

Accounting graduates, classed as 100% exempt from GCC (Foundation) examination, sit both Professional Examinations. Examination failure by such trainees, although not so costly for the Firm, is still to be avoided, and, therefore, the ability to predict their likelihood of examination failure is desirable.

However, by nature of their degree, they cannot score on the science/technology degree variable in Model 1 (Table 5.3), and thereby all scores will be biased downwards. As such, *ceteris paribus*, application of this model needs to be modified for these applicants.

Since accounting graduates must score 100% on the GCC exemption variable of Model 2 (Table 5.4), and such graduate trainees do not tend to leave within the first two years, the criteria for this model are inappropriate for such applicants and

modification of the model is not attempted.

The equivalent criteria for Model 1 pass group membership for exempt graduates is first-time passing of PE1 within the first two years. This is because exempt graduates usually take their PE1 examination approximately 18 months into their contract but deferred sittings are taken 6 months later. The time horizon has to allow for a deferred sitting. Fail group membership is determined by leaving before PE1 or failing or being referred at the first sitting of PE1. Table 8.8 gives the numbers in each group (there is obviously no Intermediate category for these cases) and mean scores and standard deviations are given in Table 8.9.

As expected, the overall mean score of 0.31 compares unfavourably with that of the full 1985/6 development sample (0.57) being 0.26 lower. Thus a likely cut-off for acceptance might be approximately 0.5 as opposed to the 0.7 cut-off for the non-exempt sample.

The frequency distribution of the exempt students' scores on the Model 1 is represented by Figure 8.5, where the negative shift of the scores is apparent when compared with the distribution in Figure 8.1 (non-exempt trainees). However, this model still successfully differentiates between passing and failing trainees.

In Table 8.10, the outcome probabilities of these trainees at the 10% score intervals confirms that the cut-off needs to be lower to make the model work efficiently for this group of entrants. The cut-off of 0.7, used for non-exempt applicants, is not appropriate, as only 9% (7 cases out of the 78) of the trainees score above this level and the lower cut-off for those non-exempt applicants considered most inappropriate, 0.4, removes no less than 65% of the exempt sample, including 55% of the pass group with, however, 78% of the fail group.

The ratio of the probability of success to failure for a score above 0.5 is 7:1 (88% : 12%) but for a score ≤ 0.5 there is a greater probability of failure than success at PE1 (55% : 45%). This cut-off would, however, remove 94% of the Fails at the

Table 8.8

**1988-1990 Exempt Entrants
Model 1 Validation Sample**

Criterion Group	Characteristics	Number
Pass	Pass PE1 at the first attempt within 2 years	42
Fail	Fail / refer PE1 at first attempt or leave before taking PE1	<u>36</u> <u>78*</u>
* One successful trainee was removed from the analysis as 'A' level information was missing.		

**Table 8.9
Exempt Trainees Model 1 Score Statistics**

Group	Mean Score	Standard Deviation	N
All	0.31	0.21	78
Pass	0.37	0.23	42
Fail	0.24	0.15	36

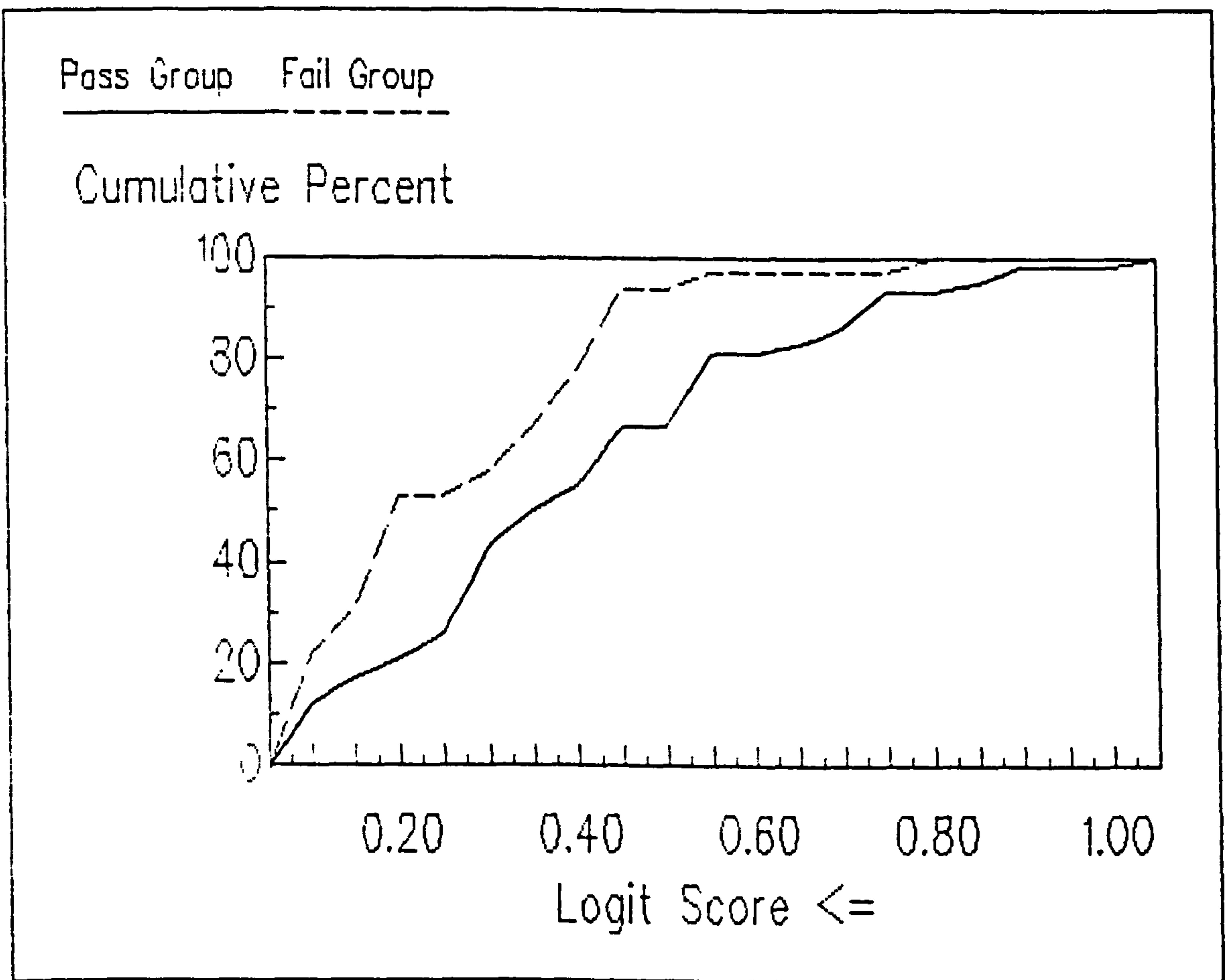


Figure 8.5 : Model 1 Exempt Graduate Groups' Logit Scores

Table 8.10

**1988-90 Exempt Validation Sample
Model 1 Outcome Probabilities**

Score	Population Percentage (78)	Outcome Probability	
		Pass (42)	Fail (36)
0.00 - 0.10	17%	0.38	0.62
0.11 - 0.20	19%	0.27	0.73
0.21 - 0.30	14%	0.81	0.19
0.31 - 0.40	15%	0.41	0.59
0.41 - 0.50	14%	0.45	0.55
0.51 - 0.60	9%	0.86	0.14
0.61 - 0.70	3%	1.00	0.00
0.71 - 0.80	5%	0.75	0.25
0.81 - 0.91	3%	1.00	0.00
0.91 - 1.00	1%	1.00	0.00
	100%	0.54	0.46

expense of 67% of the Pass group and represents 79% of the accounting degree trainees, which compares with those 80 % of non-accounting trainees removed by the 0.7 cut-off, where the ratio of probability of Pass to Fail is 4:1.

In Table 8.11, for the exempt trainees scoring below 0.25, the ratio of probability of Fail to Pass is 1.7:1 (63%:37%), however, this cut-off, removing 38% of the exempt trainees, would eliminate 53% of the Fails at the expense of 26% of the Pass group. This would represent a more appropriate lower cut-off, if one were required.

The middle range of those scoring between 0.25 and 0.5 includes 40% of the Pass group and 42% of the Fail group. Within this range, there is a probability of 0.53 of being a Pass group member and 0.47 of being a Fail group member.

8.7 Summary of Application to Accounting Graduates

Even though the Foundation-exempt trainee sample is relatively small (N=78), preliminary indications are that Model 1 may be used for these accounting graduates who were not part of the original sample. To accommodate such applicants, the cut-off score of 0.5 is used in contrast to 0.7 for the non-exempt applicants, where greater stringency is required. This cut-off retains 21% of the sample, which is equivalent to that retained by the use of the 0.7 cut-off for non-accounting graduate sample members and the probability of being successful in examinations is very high, 0.88.

However, because of the nature of the component variables, Model 2 cannot be applied to accounting graduates. In addition, Model 2 criteria are specific to non-accounting graduates.

Thus, where the criterion is equally applicable to the target samples, in this case performance in the professional examinations, the predictive model is transportable and there is evidence supportive of it being neither organization nor sample specific.

Table 8.11
1988-90 Exempt Trainee Validation Sample
Model 1 Decision Group Outcome Probabilities

Score Level	All	Pass	Probability of Fail
< = 0.25	38%	0.37	0.63
0.26 - 0.50	41%	0.53	0.47
> 0.5	21%	0.88	0.12
	100%	(54%)	(46%)

8.8 Development of a Pre-Sifting Strategy

Firms' experience in the use of the models indicates an application form can take up to 5 minutes to process, additional to the initial sift for those requiring work permits or with inadequate or non-standard qualifications. Notwithstanding the considerable savings resulting from reduced interview commitment, increased validity, etc, none of our three firms were prepared to commit additional clerical resources to the application processing stage. Therefore, to maintain acceptable turnaround response times at peak periods, while still maximizing the benefits associated with use of the logit models, the number of candidates to be scored needs to be effectively reduced.

Such a process needs to be conducted in a valid and scientific manner so that the maximum number of candidates who would subsequently be poorly rated by the model-based system are removed, at the cost of a minimal number of high calibre applicants, ie minimising misclassification.

Model 1 is used to investigate the feasibility of using model components for pre-sifting. Examination of the partial correlation statistic R in Table 5.3 indicates that the number of grade A 'O' levels is the most important single contributor to the power of this model and it is significantly correlated both with 'A' level performance as measured by UCCA points ($r=0.44$, $p<0.001$) and, more importantly, with professional examination passes at the GCC and PE1 levels ($r=0.25$, $p<0.001$), which is to be expected from its inclusion in the model.

Table 8.12 tests whether this variable is suitable for pre-sifting and provides the numbers of trainees removed and retained by its use as a selection criterion. This table provides the numbers of those in the Pass, Fail and Intermediate categories of Model 1 retained and rejected³⁰, applying cut-offs of 3 or more and 4 or more A grade 'O' level/GCSE passes to the non-exempt trainee majority, with the usual

³⁰ > 5 grade A 'O' levels was investigated but proved impractical as too many cases are removed.

Table 8.12

Pre-Sift on ≥ 3 and ≥ 4 Grade A 'O' Level Passes

8.12a Results of adopting ≥ 3 Grade A 'O' levels as a cut off for processing

Model Group	Accept	% of Accept Area	Grey Area	% of Grey Area	Reject	% of Reject Area	Total	% of Total
Pass	23	(66)	16	(43)	6	(27)	45	(48)
Fail	3	(8)	4	(11)	3	(14)	10	(11)
Intermediate	9	(26)	17	(46)	13	(59)	39	(41)
	35		37		22		94	

8.12b Results of adopting ≥ 4 Grade A 'O' levels as a cut off for processing

Model Group	Accept	% of Accept Area	Grey Area	% of Grey Area	Reject	% of Reject Area	Total	% of Total
Pass	23	(66)	9	(34)	4	(44)	36	(53)
Fail	3	(9)	3	(12)	1	(11)	7	(7)
Intermediate	9	(25)	14	(54)	4	(44)	27	(40)
	35		26		9		70	

minimum 9 UCCA (18 UCAS) points.

Table 8.12a indicates that adopting the strategy of scoring only those with >3 grade A 'O' level passes, 94 cases are scored on Model 1 out of 237, ie a reduction of 62%. Such a cut-off results in the rejection of 24% of all recruited trainees who were unsuccessful and the retention of 50% of those who were successful. The 'accepted' group of 35 are 23 (66%) Pass group, 3 (8%) Fail group and 9 (26%) Intermediate group. This compares with the actual proportions of 37% Pass, 25% Fail and 38% Intermediate trainees in the full non-exempt trainee sample of 244 (Table 8.2) and of those resulting from scoring the full sample on Model 1: 59%, 14% and 27% (Table 8.4).

However, a more efficient pre-sifting strategy is to process only candidates with at 9 or more UCCA points and at least 4 grade A 'O' level/GCSE passes. Table 8.12.b shows that processing numbers after this pre-sift would be reduced to 70 of those with non-accounting degrees, ie 30% of cases. The 35 who are provisionally accepted on this basis are the same as those accepted by the 3 grade A 'O' level criterion, with 24 fewer cases scored.

Figure 8.6 demonstrates the result of adopting the latter pre-sift strategy to the non-accounting graduate entrants and indicates the proportions of those 'rejected' or 'accepted', who actually passed or were unsuccessful in their examinations. One hundred and eighty of the recruits had 9 or more UCCA points, out of the original 237 with 'A' level data. Seventy of these had, in addition, 4 or more grade A 'O' level passes and are scored. Thirty-five (50%) score below 0.7 and are rejected. The remaining 35 cases, who, if they were applicants, would be provisionally accepted, are proportionately, 66% Pass group, 8% Fail group and 26% Intermediate group.

Table 8.13 compares this strategy (iii) with (i) model scoring only and (ii) of using a 9 UCCA point sift plus model scoring. The use of grade A 'O' levels, in addition to the 9 UCCA point sift, reduces numbers to be processed by 71%, and 28% of Pass group members are retained, with 4% of the Fails and 7% of the Intermediate

Figure 8.6a

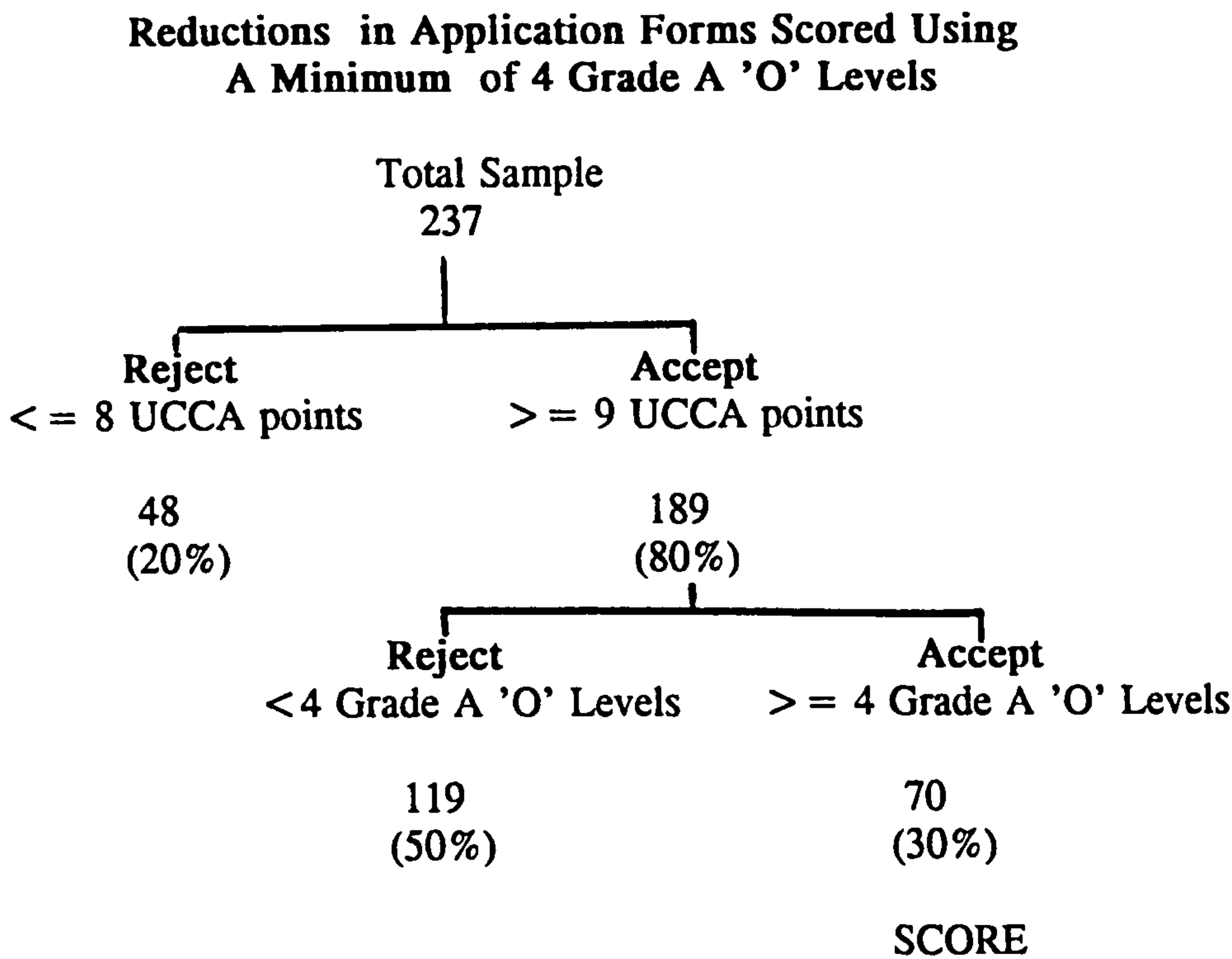


Figure 8.6b

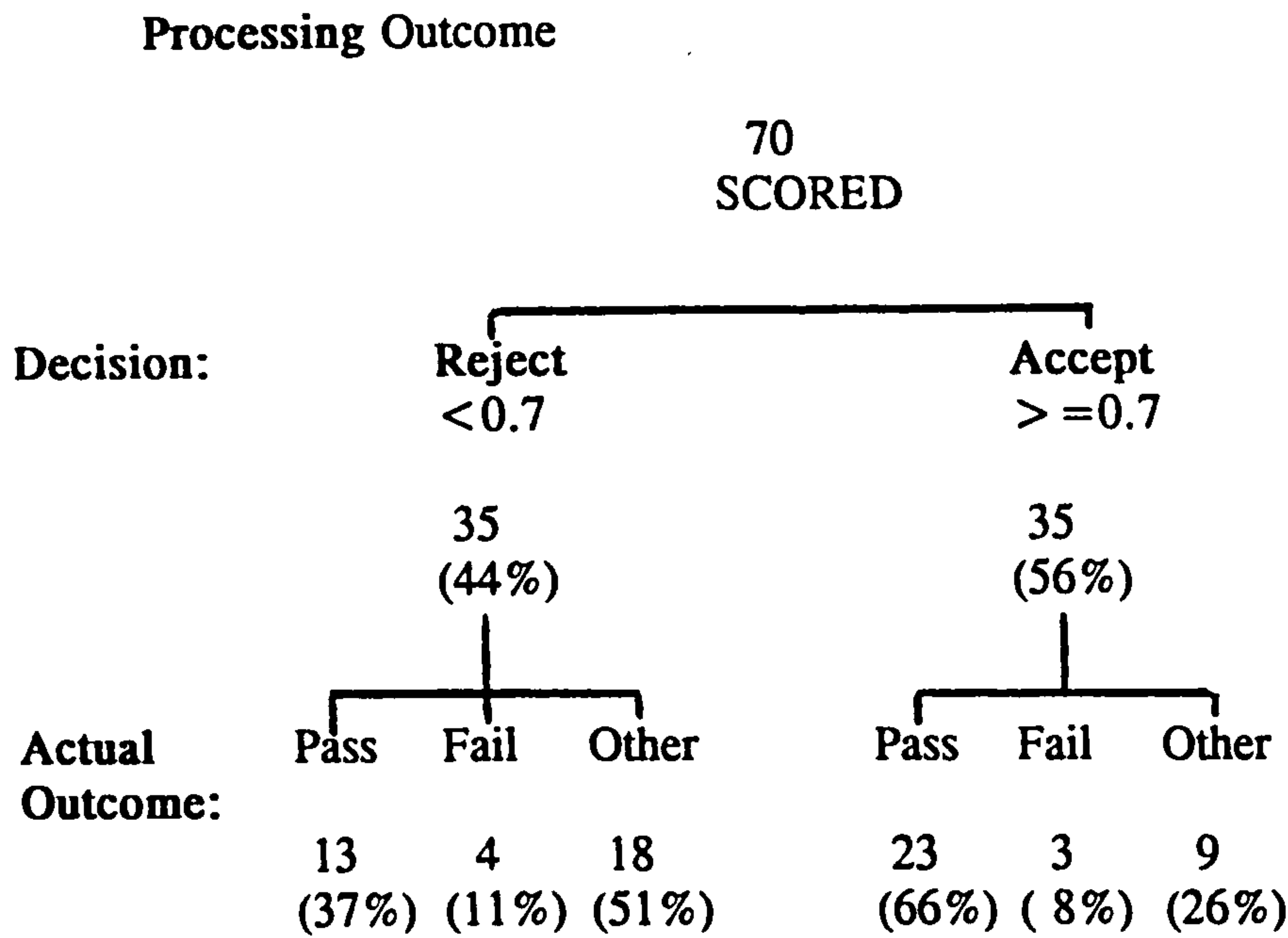


Figure 8.6: Pre-sift Strategy Results

Outcome of Alternative Pre-Sifting Strategies

Strategy	Total	Number "Rejected"			Total	Number "Accepted"		
		Actual Outcome	Actual Outcome	Actual Outcome		Actual Outcome	Actual Outcome	Actual Outcome
		Pass	Fail	Other		Pass	Fail	Other
(i) Model only (237 scored)	181 (76%)	56 (31%)	26 (14%)	99 (55%)	56 (24%)	27 (48%)	45 (7%)	25 (45%)
(ii) 9 UCCA points + model (189 scored)	142 (75%)	45 (32%)	22 (12%)	75 (53%)	47 (25%)	26 (55%)	4 (9%)	17 (36%)
(iii) 9 UCCA + 4 A 'O'levels + model (70 scored)	35 (50%)	13 (37%)	4 (11%)	18 (51%)	35 (50%)	23 (66%)	3 (8%)	9 (26%)

group.

The recommended strategy of Figure 8.6 leads to only 30% of the 237 cases being scored, 119 fewer than with the 9 UCCA point screen and 167 fewer than with no pre-screening. This reduction in processing costs is bought at the price of only 4 fewer subsequently successful trainees 'accepted' compared with no strategy other than the logit model or 3 fewer than using the model with a 9 UCCA point pre-sift. There is an added advantage, however, of accepting 1 less Fail and 16 fewer Intermediate cases associated with no pre-sift and 1 less Fail and 8 fewer Intermediate cases than with a 9 UCCA point pre-sift.

The models are intended to be applied in practice to the population of application forms and our results relate to an already recruited sample. Unfortunately, it is not possible to assess the effect of using the pre-sift and model score on the 1990 applicant sample. At the time of the pilot, the possibility of developing a pre-sift was not being considered and the forms were merely scored and later returned to the supplying firms after discussion. The indications are, however, that the recruited sample does not differ significantly from a sample of applicants (sections 7.1 and 8.2), and therefore there should be adequate numbers of applications to successfully adopt this prescreen.

However, the effect of adopting the 4 grade A 'O' level pre-sift strategy for the accounting graduate trainees in the validation sample, would reduce the number processed by 82%. Accounting graduates (not included in the model building process), by nature of their degrees, have both practical and theoretical experience unavailable to their non-accounting graduate colleagues which may counteract their evident poorer performance at 'O' level. For this reason, the number of A grade 'O' levels might not be a valid pre-sift for this group.

Since there are fewer Foundation-exempt applicants than non-exempt (only 25% - see ICAEW statistics) and using a suggested Model 1 cut-off of 0.5 alone provides a very high probability of a successful examinee (88% Pass group, 12% Fail group),

a set pre-sift criteria is probably not necessary for such applicants, who may be directly processed by the software.

The minimum UCCA point criterion of 9 reduced numbers for exempt trainees to be scored by a quarter, but leads to an acceptance of 2% fewer Fails and 5% fewer Intermediate cases, with a 10% enhanced pass rate (69%). However, in effect there is no difference in the number of Pass group trainees 'accepted' by this process but the unsuccessful trainees 'accepted' are reduced from 41% to 31%.

8.9 UCCA (UCAS) Point Pre-sift Strategies

There is a prevalent belief within the profession, that 'A' level performance, measured by UCAS points, is an efficient pre-screen for applications. This may stem from the ICAEW's yearly digests of Education and Training Statistics, in which professional Examination performance is tabulated by UCAS points. These tables certainly indicate that those with higher points do better in their professional examinations. Firms appear to have interpreted this to mean that UCAS points are an efficient performance predictor, whereas they may only reflect recruitment preferences over time.

Firms generally restrict entry to those who have 9 or more UCCA points (18 UCAS), ie average or higher grades for their Advanced Level GCE's, and all three piloting firms preferred their applicants to have a minimum of 9 UCCA points, or the equivalent of 3 C grades. ICAEW Statistics (1986-1990) indicate that, of those graduate entrants with Advanced Level qualifications in the years 1985-91, on average 78% have more than 9 UCCA points.

In this study, UCCA points are not found to be significantly related to professional examination performance, whereas the subjects studied at 'A' level are found to be a component of both developed models. In fact, in this sample, although UCCA points are positively correlated with number of grade A 'O' level GCE's ($r=0.44$,

$p < 0.001$), they are negatively correlated with degree class ($r = -0.15$, $p < 0.01$). Thus good grades at 'O' level are likely to indicate good grades at 'A' level, but 'A' level performance is not necessarily a good indicator of degree class.

Table 8.14 compares the results of applying two purely UCCA point based pre-sift strategies: pre-screening on 9 or more UCCA points and screen on a minimum of 11 UCCA points (to which Kidsons Impey changed), on all the graduate entrants to the three piloting firms.

The table demonstrates that, although the number of cases 'accepted' for interview falls to 115 from 189, as minimum acceptable UCCA points moves from 9 to 11, the percentage pass rate is largely unchanged (38% to 40%) and not statistically different from using no pre-sift at all (36%). The result is purely a reduction in numbers with approximately the same Type I and II errors.

Ignoring the UCCA points, but using the > 3 grade A Ordinary Level pass criterion alone, only 78 members of the non-exempt 1989-90 entry sample would have been scored, 39 (50%) Pass group, 7 (9%) Fail group and 32 (41%) Intermediate group.

8.10 Summary of Pre-Sift Development

Section 8.7 above indicates that a minimum 9 UCCA points requirement is not an efficient pre-sift but in our three middle-size firms, this and B grade passes at English and Mathematics 'O' level/GCSE were the major criteria for consideration for interview. In this study we found a significant negative correlation ($r = 0.15$, $p < 0.001$) between degree class and UCCA points.

Whereas both logit models contain the I/III degree class variable, Model 1 also contains the positively weighted science/technology degree variable and Model 2 contains the negative arts 'A' level variable. Those who have studied such 'A' levels will not go on to take science/technology degrees. Thus depending on degree class

Table 8.14
UCCA Point Scoring Strategy

Strategy	Total	Number "Rejected"			Total	Number "Accepted"		
		Pass	Fail	Other		Pass	Fail	Other
(i) None					237	83 (35%)	30 (13%)	124 (52%)
(237)								
(ii) Minimum 9 UCCA points	48 (20%)	14 (29%)	2 (4%)	32 (67%)	189 (80%)	71 (37%)	26 (14%)	92 (49%)
(iii) Minimum 11 UCCA points	122 (51%)	39 (32%)	12 (10%)	71 (58%)	115 (49%)	46 (40%)	16 (14%)	53 (46%)

in such cases will probably not prove particularly efficient as a pre-sift, as no account is taken of 'A' level disciplines.

The logit models emphasize the crucial importance of the subject studied at the GCE Advanced level, rather than the standard of performance, which is of crucial importance at Ordinary level. The number of grade A 'O' level/GCSEs provides a valid pre-sift. A strategy of accepting only applicants who have 9 or more UCCA points and 4 or more grade A 'O' level passes would have resulted, in this sample, in a 70% reduction in the number of applications processed.

Of the 70/237 cases meeting these conditions and scoring above the Model 1 cut-off, 66% are Pass, 8% Fail and 26% Intermediate.³¹ The full recruited sample of 244 non-accounting graduates, selected by the interview-based procedure, provided 37% Pass, 25% Fail and 38% Intermediate trainees.

Thus the pre-sift satisfactorily reduces processing numbers and subsequent application of Model 1 results in a high likelihood of recruitment success in examination performance terms. However, the use of the 9 UCCA point pre-sift is unnecessary, as the 'O' level sift is more efficient.

8.11 Analysis of Costs and Benefits Deriving from Model Use

The empirical research reported in Part I was motivated by the desire of the ICAEW Research Board to improve recruitment practice and, in particular, to reduce the costs to the firm and student associated with examination failure and drop-out of trainees. However, the Benveniste *et al* (1986) ICAEW study reports that '...students who fail examinations do not make as much difference to the costs borne by training offices as might have been expected.' They point out that the real

³¹ Seven of the 244 recruits did not have sufficient information on 'A' level performance to be used in this analysis

impact of such recruitment errors on the firm lies with the revenue side of the equation, together with the disruptive effects on office/client routine and staff scheduling; the latter effects being unquantifiable, but usually severe.

Utility Analysis has developed from the practitioner's need to quantify selection decision outcomes, in order to gain maximum utility from decision processes.

Boudreau (1992) (referring to Cronbach and Gleser, 1965) points out that, while, scientifically, measurement is regarded in terms of its ability to reduce uncertainty about the true value of some quantity, in practical applications, the value of a measure is assessed in terms of its ability to aid the user in correctly allocating those tested, ie to enhance the qualitative decision. '...(T)he usefulness of a selection system depends on its ability to provide information that will improve decisions, where decision improvements are measured in terms of value decision outcomes.' (Boudreau, p631)

Whereas most recruitment and selection processes have their predictive validity reported in terms of validity coefficients or coefficients of determination, Boudreau points out that only very large improvements in validity can have a substantiable impact on organizational outcome. Utility Analysis allows the recruiter to measure the potential benefits associated with a new procedure but its ultimate utility depends upon the situation in which it is to be used.

The apparent performance of the developed models of this thesis compares favourably with selection interviews (structured and unstructured) and other evaluative techniques, in terms of summary statistics. However, the potential implications, in terms of net financial gain to firms using these models, has not been assessed.

Unfortunately a major component of Payoff Utility Analysis (Boudreau, 1992), the SD_y (difference in £ value associated with one standard deviation of the criterion measure) cannot be calculated here. The cost per applicant of each process has to be quantifiable for Payoff UA. Using model scoring to replace the selection interview

certainly reduces associated costs, but does require an increase in time commitment at lower staff levels. Neither the reduction in the former nor the increase in the latter are quantifiable from the estimates given by the firms. For example, the two pilot firms estimated their recruitment costs to be as low as £480 per entrant, although these figures ignore the major part of their costs incurred.

However, the differences in net gain to the firm resulting from the use of the developed strategies can be calculated for discrete levels of employee progress, using figures provided by the pilot firms, and some estimate of savings to all similar size firms in the profession can be made. The following analysis is therefore based on the representative figures supplied by both Stoy Hayward and Kidsons Impey.

Costs

The costs associated with recruiting students relate solely to the recruitment function, ie advertising, promotional material, partner and accompanying staff time for interviewing and presentations, and application processing by clerical staff; general personnel function costs and office overheads are ignored. Once recruited, the trainee incurs fees from the private training firms and examination fees from the ICAEW. In addition to these, the trainee receives a salary commensurate with the size of the firm, the size of the employing office and his/her level of experience.

Revenues

Trainees, in turn, generate income from client firms, who are charged at an hourly rate applicable to the trainee's level of experience. Nevertheless, during the first year of the contract trainees spend relatively less time in client organizations, since much of the time is taken up with block release for training, in-house courses, non-chargeable work, and holidays and examinations.

Our two firms agreed that a fair estimate of the chargeable hours actually recoverable from the client is 80% on average, per year. They provided us with salary tables relating to the levels of training and examination performance from which an average salary for each of the stages was derived. A further 15% has been

added to these, to account for the employers' National Insurance contribution, etc.

Assumptions Relating to Costs and Revenues

The analysis, therefore, is based on the following assumptions:

- (i) Forty-six and a half weeks are available for work after holidays, in-firm courses and training courses are are deducted.
- (ii) Firms agreed to pay trainees' resit costs only for the Foundation examination, which results in a loss of a further 4 weeks chargeable hours (£3,039) and a resit fee (£500). Later resit fees are passed on to the student. Nevertheless, the firm cannot charge so highly for a trainee who has not passed examinations and this results in lower income, eg,

Resit Foundation, Resit Intermediate =	£3,049 loss in year 1
	<u>£6,216</u> loss in year 2
	<u>£9,265</u>

- (iii) Although salary in both Firms is related to performance, an average representative salary for stage of training and examination status was used to calculate costs.

Year 1 salary :	£13,750
Year 2 salary :	£15,079
Year 3 salary :	£17,875

- (iv) The ICAEW examination fees and estimates given by firms of the costs of training courses are also used to calculate the following costs:

Foundation	: £1,368 (+ £500 resit)
Intermediate	: £1,585 (no resit fee paid)
Final	: £1,315 (no resit fee paid)

- (v) Recruitment costs, based on Firms' *ad hoc* analysis of recruitment material, administration costs and non-chargeable partner time for interviews per year, divided by the resulting number of recruits, are used to calculate a one-off recruitment cost figure of £480. Again it should be noted that this figure does not include overheads and should be considerably higher.
- (vi) Approximately 80% of possible chargeable hours, after deduction of above, are recoverable from clients and the hourly charge-out rate reflects the trainees' level of experience.

Year 1 income :	£26,314 (no retake)
Year 2 income :	£33,359 (no retakes)
Year 3 income :	£46,232 (no retakes)

Figure 8.7 summarises estimated revenues and direct costs of students at different stages during their training contract (3 years) and Table 8.15 lists the overall contributions associated with the different 'branches' of the 'tree'. The diagram is divided into two parts, 8.7a for accounting graduates and 8.7b for non-accounting graduates.

Net contributions for accounting graduates range from £63,500 for those who pass both PE1 and PE2 at the first attempt within the expected time frame, to £55,500, for those who have to resit both examinations. However, the fully successful non-accounting graduate who passes the Foundation and both Professional Examinations at the first attempt, makes a contribution of only £54,500, a difference of £9,100 over the three years.

It has been noted that, for non-accounting graduate trainees, the likelihood of passing PE1 at the first attempt after failing GCC is remote and thus, in Table 8.15, Resit/Resit/Pass and Resit/Resit/Fail are the most likely outcomes of such failure. These paths make an unsuccessful non-accounting graduate trainee between £9,000 and £15,000 less profitable than a successful trainee.

Whereas the numbers actually applying to the firms and resulting in the recruited samples are unknown, the pilot exercise allows an examination of the full spectrum of applicants. Firms' current procedures (sifting application forms and two interviews) applied to those for whom full information was available in the pilot (the graduates), resulted in 38 offers being made. These offers, if all taken up, in terms of their probability of success, based on the characteristics of the 1988-90 sample, will result in 18 Pass, 12 Fail and 8 Intermediate entrants. The success ratio, in terms of Pass/Fail would be 1.5. Using Figure 8.7 to extrapolate, the associated net gain to two the firms at the end of the two year period covered by the model criteria, provided no entrant left, would be £858,800.

However, without interviewing, using the Model 1 cut-off of 0.7 to score applicants, the recruited sample of 71 would be made up of 33 Pass, 4 Fail and 34

Foundation	Stage of Training Intermediate	Final	Total Revenue (£000)	Total Costs (£000)
------------	-----------------------------------	-------	----------------------------	--------------------------

Figure 8.7a NON-EXEMPT

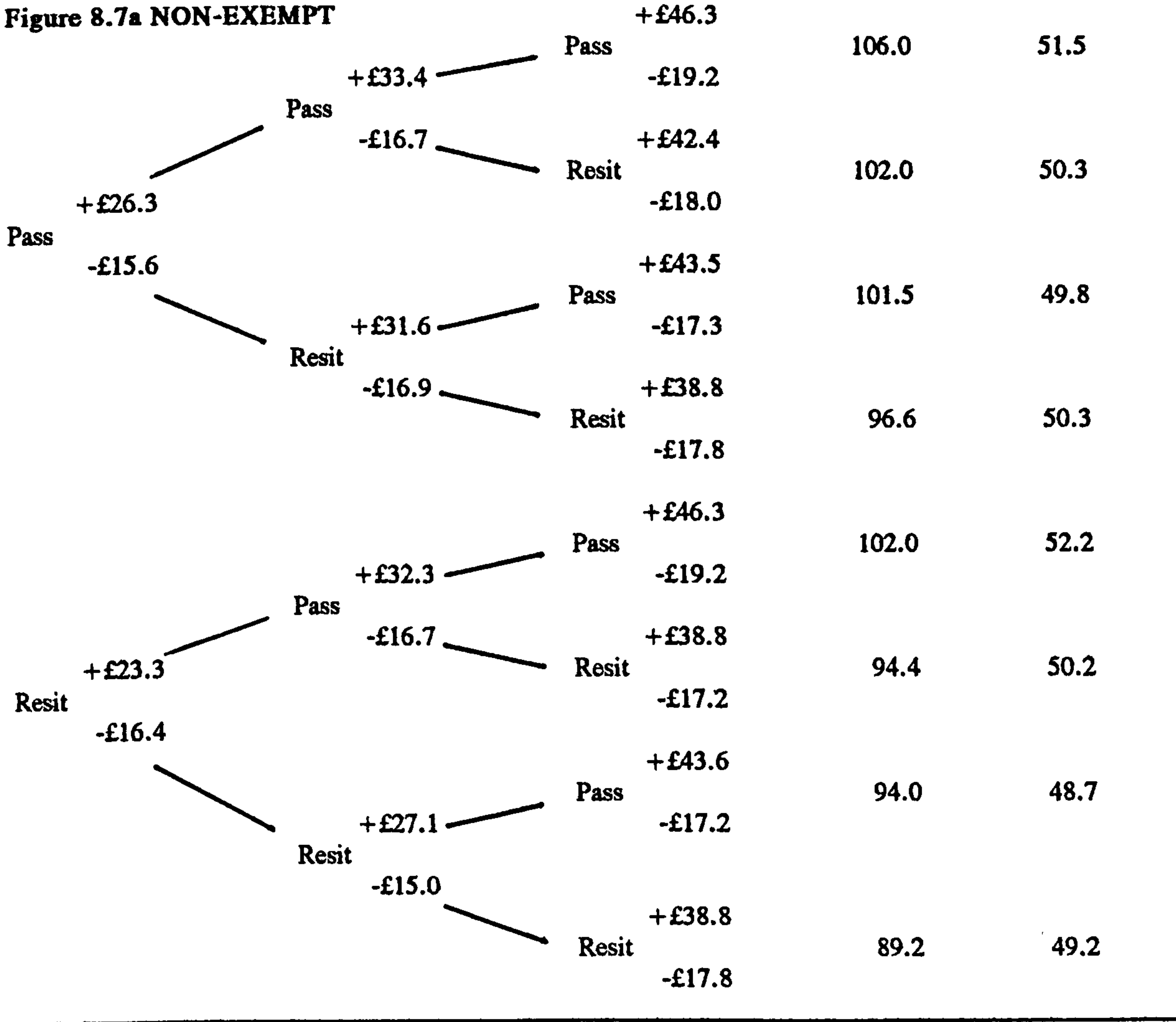
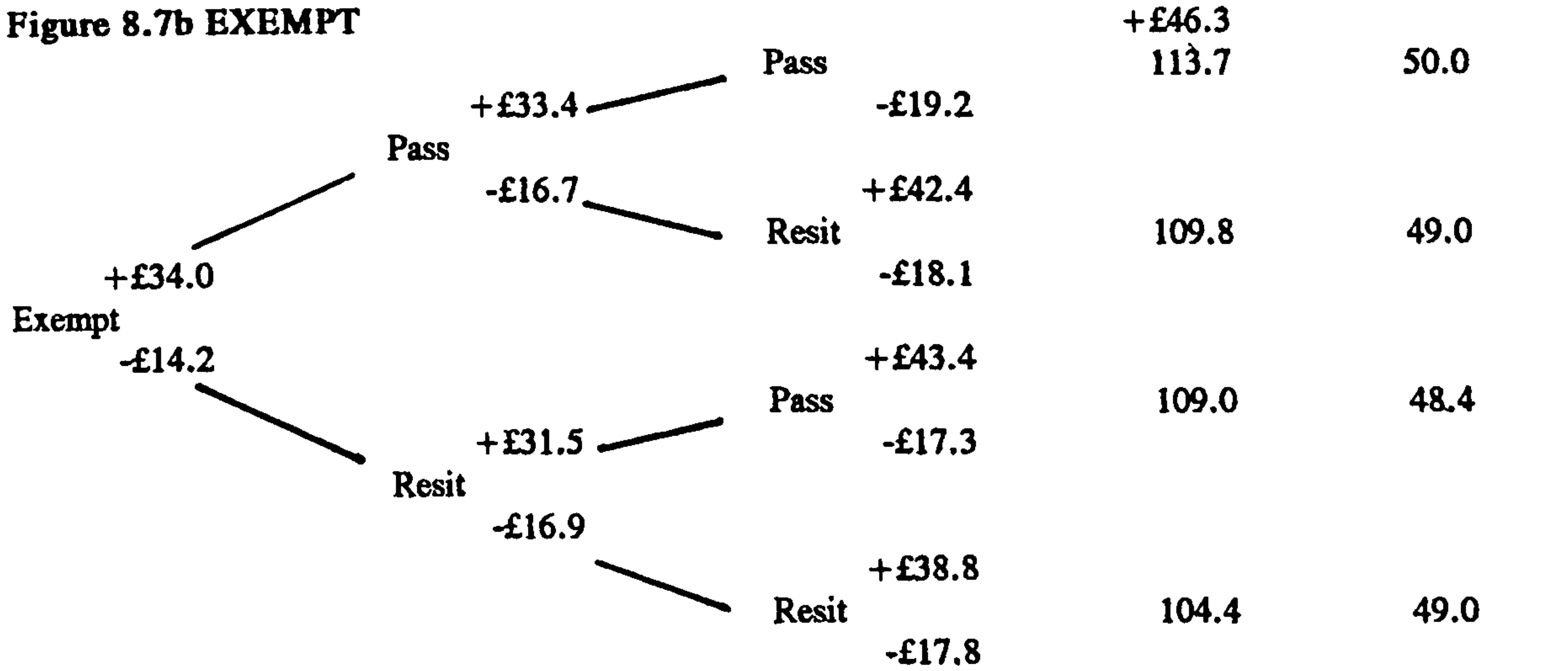


Figure 8.7b EXEMPT



Note : + denotes revenue
 - denotes costs

Figure 8.7 : Estimated Contribution from Graduate Entrants by Stage of Training

Table 8.15

**Total Contribution Towards Overheads Associated
with Student Training**

Path Taken (Foundation*, Intermediate, Final)	Contribution £'000
Non-Relevant Graduates	
Pass Pass Pass	54.5
Pass Pass Resit	51.7
Pass Resit Pass	51.7
Pass Resit Resit	45.4
Resit Pass Pass	49.7
Resit Pass Resit	44.1
Resit Resit Pass	45.3
Resit Resit Resit	40.0
Relevant Graduates	
Pass Pass	63.6
Pass Resit	60.8
Resit Pass	60.6
Resit Resit	55.5

* If non-relevant

Intermediate, a comparable success ratio of 8.25. This creates a net gain of £1,680,000 (a difference of £821,200) should all take up their offers and assuming none leave. Leavers are, however, associated with examination failure, and the model selection contains only 4 Fails, whereas the firms' offer sample contains 12 Fails.

The wisdom of meeting prospective recruits has been noted, although firms are recommended to only interview those scoring above the model cut-offs, thus only 71 interviews would have been conducted, not 142, considerably reducing recruitment costs.

Leavers

Losses associated with those who cancel contracts are, for the firm, concerned with lost revenue for the remainder of the contracted term. For example, a non-accounting graduate trainee who leaves after unsuccessful GCC performance represents a maximum loss of £15,600 at the Intermediate (PE1) stage and £27,000 at the Final (PE2) stage, presuming that the trainee proves to be successful in PE1 and PE2, and minimum losses of £12,000 and £21,000 respectively, presuming they are unsuccessful in both.

Population Utility Estimates

In 1990, the pilot entry year, medium size firms offices with between 21 and 100 partners, recruited 1,778 trainees and this represents 31 % of the graduate intake for that year (ICAEW, 1993). While it may not be the case that poor examination performance is evenly distributed over the profession, for the purposes of the analysis here, 31 % of the graduate leavers from medium size firms at each stage, have been used to provide the approximate loss figures for medium size firms reported below.

Approximately 175 of the non-accounting graduates, cancelled their contracts during the first year, or prior to PE1. This figure, multiplied by £12,000 (the smallest loss scenario for the Intermediate Stage [pre PE2]), indicates a loss of expected trainee

revenue of £2.1m, and then, multiplied by £21,000, a minimum loss in the third year, of £3.7m.

In the second year, a further 34 entrants, who had not succeeded at the Foundation level, cancelled their contracts, representing a loss of £0.7m. Also in year 2, 140 entrants left after taking PE1, representing lost revenue of £2.9m in year 3. In total, the estimated minimum loss incurred by medium sized firms, due to 1990 entry non-accounting graduate leavers, would be, on this basis, no less than £9.4m.

Accounting graduates did not leave during 1990-91, but 84 left after PE1. These 84, multiplied again by the minimum loss of income associated with the poorest performance in year 3 (£21,000), provide an estimated loss of £1.8m.

These very approximate estimates of loss attributable to leavers total £11.2m and, of course, provide no indication of the less tangible, but more significant, in-firm effects of such wastage.

8.12 Summary of Costs and Revenues

One important result of this brief analysis is that accounting graduates are more cost-effective than non-accounting graduates with net contribution on average 17% higher for equivalent stages of training. This is largely because they do not have to sit the GCC (Foundation) examination, with associated loss of revenues to the firm, and do not leave within two years.

Thus, an exempt graduate trainee, rated highly by Model 1, is a much more profitable proposition than a good non-exempt graduate trainee. In addition, such entrants are less likely to leave during the early stages of training with reduced unquantifiable costs.

A second important observation is that the difference in measurable benefits

associated with different examination histories is not as significant as might have been expected. The point made by Benveniste *et al*, concerning the importance of the readily quantifiable costs associated with examination failures and drop-outs is a very germane.

However, while examination failures are costly in terms of disruption and lower revenues, avoiding leavers is still seen to be of paramount importance.

8.13 Discussion

This chapter reports the validation undertaken to provide evidence of the reliability of the predictive selection models developed for medium size accounting firms in Part I as the long term validity of biodata models in selection has been frequently called into question. Hunter and Hunter (1984) suggest that poor long term results may be a function of inadequate sample size and/or methodological problems and Barrett and Doverspike (1992) provide evidence that predictive validity may be stable over time. The key to successful long term validity may rely upon the stability of the criterion variables.

On the basis of previous research studies, this study did not expect to find any particularly noticeable shrinkage in the performance of the models because they are based on hard data item predictors and are not subject to problems associated with faking. In addition, the samples are large enough to provide reasonable assumptions about the population and two different methodologies have been compared, the most favourable being the parametric logit technique.

However, where the stability of the criterion measure was undermined, in this case by the Foundation examination syllabus change, the validation results for the 1987 sample were almost certainly adversely affected, as pass rates for that year were much lower than for previous years. The more recent validation sample, drawn from 1988-90 entrants, provides the means to test whether poor validation results were,

indeed, the result of the change in the criterion variable.

There is no significant difference between the proportions of the 1985/6 sample within the examination prediction model criterion groups and those in the 1988-90 validation sample (section 8.4) but these samples differ significantly from the 1987 validation sample. This supports the assumption that 1987 entrants are not directly comparable with the derivation sample from 1985/6 and confirms the need to validate on a separate sample.

Model 1, when applied to the 1988-90 sample, provides very acceptable results, although the mean score for this sample is 0.41 compared with that of 1985/6, ie, 0.57. Those of the 1988-90 sample who score below the model cut-off, and who would be rejected by the software scoring system, have a 0.31 probability of being successful in examination terms, compared with that of the 1985/6 sample of 0.2.

Eighty percent of the 1988-90 trainees' scores lie below the 0.7 cut-off, compared with 63% of the 1985/6 model derivation sample, but the ratio of the probability of Pass (0.59) to Fail (0.14) group membership for the remaining 20% scoring above that cut-off is still greater than 4:1. Whereas, for those in the 1987 sample scoring above 0.7, it is very poor, with the probability of Pass group membership being 0.34 and Fail group membership, 0.28.

The mean Model 2 score for the 1989-90 sample is 0.39, compared to 0.51 for the 1985/6 sample. For those scoring above the 0.6 cut-off, the model provides a 2:1 probability of success to leaving, compared with the 5:1 ratio of the 1985/6 sample, although only 22% of scores lie above the cut-off compared with 38% of the derivation sample.

Bearing in mind the inevitable loss of predictive power resulting from fitting models to separate samples, these models are still performing well on trainees entering up to 5 years after members of the derivation sample. Since no severe reduction in predictive power is noted, the criticism of long term unreliability may not fairly be

directed at these models.

The nature of the prediction criteria chosen by firms results in large numbers of the sample falling into the Intermediate group of Model 2 (63% in 1985/6, 69% in 1987 and 51% in 1988-90) and it is noted that this model is not as 'sharp' as the examination performance model. However, it may provide firms crucially with a probability of leaving for the non-accounting graduate applicant, examination performance being successfully predicted by Model 1. Thus an applicant highly rated on Model 1 may be scored on Model 2 to assess the likelihood of leaving within two years.

This chapter also provides evidence that biodata models, and therefore their keys (or weights) may be used successfully on subjects drawn from different organizations with similar characteristics. The models are successfully validated here on a sample of three representative medium sized professional firms.

In addition, the Model 1, with a suitably adjusted cut-off, is likely to be as useful for accounting graduates as for the intended sample of non-accounting graduates. This directly supports Rothstein *et al*'s (1990) comments regarding the appropriate scaling of the instrument to allow different subjects to be scored and the selection of criteria which are applicable across organizations and samples.

A lower cut-off of 0.5 retains a similar proportion of cases (21%) to those of the non-accounting graduates retained by the 0.7 cut-off (20%). The likelihood of selecting a successful accounting graduate examinee above this level is 0.88, compared with 0.59 for non-accounting graduates. As with a cognitive test, our evidence suggests that the Model 1 may be applied to any suitable applicant sample within the profession. The model, based on hard data, is therefore 'transferable' to a sample with different characteristics.

However, the criteria for development of Model 2 are not applicable to accounting graduates, as it was specifically developed to predict the likelihood of a trainee

leaving during the first two years of the training contract. Such graduates approach the profession with previous experience not available to the non-accounting graduates and, since they are far better informed concerning the nature of the work, they are far less likely to find themselves mis-matched. Furthermore, as they sit no examinations during their first 18 months, the second and most likely reason for leaving, ie examination failure, is also unlikely to affect them. For these reasons no attempt is made to apply this model to the accounting graduates.

Nevertheless, both models discussed here are developed and validated on a sample from 23 different sized accounting firms' offices, and thus, although they may be profession specific, neither are organization specific.

An exploration of issues deriving from use-in-practice revealed that model use in the piloting firms resulted in an unforeseen pressure on recruitment staff who were faced with extra work when they were already fully committed. Although the firms reaped benefits deriving from reduced interview commitment, there were no plans to increase processor time. The necessity of reducing the pool to the most suitable applications, so that only those very likely to succeed in examination model terms are scored, became apparent and an efficient and speedy method of pre-screening was investigated.

The number of grade A 'O' level/GCSE passes proves extremely useful for the purpose. Applying a cut-off of ≥ 4 grade A 'O' levels/GCSE's to the recruited sample of non-accounting graduates entering the pilot firms between 1989 and 1990, reduces the number of cases to be processed by 70% from 237 to 70 with the loss of only 4 of the Pass group trainees who would have been accepted using the model cut-off of 0.7 on all 237 cases. This loss is, however, associated with a rejection of a further 42 Fail group members and 16 Intermediate cases.

The group 'accepted' by this strategy is composed of 66% Pass group, 8% Fail group and 26% Intermediate cases, whereas the actual figures for the recruited sample are 40%, 30% and 30%. Clearly this strategy prevents arbitrary pre-sifting

procedures from compromising the efficiency of the developed models. It is also more valid and more reliable than the current 9 UCCA point pre-sift, retains a sufficiently high proportion of good calibre applicants and provides far more reliable results than the previous interview-based methods.

The estimated quantifiable costs of individual unsuccessful performance in firms, in terms of reduced revenues, may be mapped by following the progress of poorer trainees in Figure 8.7. Although it is unlikely that larger firms would allow students who have failed an examination more than once to remain in the firm to take further examinations, this may not pertain in the smaller medium size firms. Larger firms may well be better able to accommodate the loss of a few trainees, although workload re-scheduling may cause disruption, indeed, they have actively over-recruited to allow for such adjustments.

The analysis of costs and benefits, based on the 'ball-park' figures from piloting Firms, indicates that accounting graduates are more profitable than non-accounting graduates. Even those who are successful in neither PE1 nor PE2 are more profitable than non-accounting graduates who pass GCC (Foundation), PE1 and PE2 at the first attempt.

Approximations of losses incurred by the size of firm studied here, and based on ICAEW leaver statistics for the 1990 entrants, indicate that net losses of expected revenue of around £9m result from the non-accounting graduate leavers, and £2m from accounting graduates leaving after PE1, for the segment of medium sized firm offices with 21-100 partners for their 1990 intake.

Whereas the quantifiable costs of recruitment failure may be approximated, the unquantifiable costs referred to by Benveniste *et al* (1986) are not available to analysis. They must, however, have great impact on the smooth running of the firm. Trainees who are not performing well and are poorly motivated do not make a good contribution to the firm, making poor impression upon the client and, in some cases, working counter-productively. Trainees who leave cause manpower problems and

incur severe loss of projected revenue. The benefits of reliable prediction of both examination success and the likelihood of leaving are manifest.

9.1 Introduction

This study is divided into two parts, the first concerned with the empirical research and development of predictive models and the practical application in an industrial setting, and the second with the long term validity of such models and their 'generalizability' between different organizations and samples. The data and samples are provided by medium size chartered accountancy firms and the empirical research was part-funded by the ICAEW, the professional body.

Attention is drawn in Chapter 1 and Chapter 3 to the recruitment procedures used generally within the accountancy profession, which are not all viewed favourably in the research literature and face criticism for their poor validity and reliability. The profession's almost total reliance upon judgemental selection methods, ie, the unstructured selection interview, causes particular concern. This thesis does not question the need for the employer and prospective employee to meet but supports the suggestion of Herriot (1984) that this meeting does not necessarily have to be used for selecting recruits, but can be a two-way assessment of employment 'fitness'.

Within the profession, recruiters, faced with unacceptable wastage levels (ICAEW, 1995) are understandably anxious to improve current procedures. Chapter 2 discussed the use of biographical data for predicting occupational success. Owens (1968) provides evidence that individuals can be classified on the basis of similar life histories and Holland (1976) demonstrates that those with similar backgrounds are likely to make similar vocational choices. Thus, identifying the background characteristics which are the best predictors of future success, within a given occupation, should provide the basis for accurate predictive models for selection to that occupation.

Methods used by employers to select employees are discussed in Chapter 1 and, while the literature suggests cognitive tests generally provide the highest validity coefficients, scored biodata has been found to provide comparative validities while not exhibiting the same adverse effects on sensitive groups. Unstructured interviews are rightly criticised for their low validity but structured interviews, particularly those which are behaviourally based (situational interviews), have been found to be very useful as selection tools, also providing high validity.

Using a representative sample of medium sized firms, the feasibility of using scored biodata for selecting graduate entrants to the profession, rather than relying mainly on the interview as the prime selection tool, is investigated. Such a strategy would result in the employer approaching only those whose biodata score indicates a high likelihood of success in the job. The resulting meeting may be of any format favoured by the recruiter, where both parties assess person/organization fit.

This study provides evidence that there is sufficient hard biodata contained in any standard application format to provide far higher levels of accurate prediction, in terms of success in the training period measured by professional examination progress and tenure, than the present system, based on clinical judgement, operating in medium size chartered accountancy firms.

An original contribution to occupational psychology theory is made by the results of the empirical research, reported here in Chapters 5 and 6. In these chapters, appropriate methods of biodata model-building are explored and the widely used non-parametric Weighted Application Blank technique (England, 1971) is specifically compared with the parametric logit procedure.

The logit procedure proves superior and two final models are derived using this technique. Each uses only six key variables available on the overwhelming majority of applications and may be used to give a clear indication of the applicant's probability of success in the Graduate Conversion and first Professional Examination (PE1), or of leaving within two years.

The work reported here also implements the predictive models via a software scoring system specifically developed for the purpose. The program does not require any particular skill to use and may be applied to any format of biographical information presented by applicants. The logit score, generated by the software, represents the probability of the applicant achieving the successful criterion outcome and allows the recruiting firm to gauge the risks involved in recruiting the candidate. Advantages of use include fairer and more reliable selection procedures, a reduction in the number of interviews conducted and associated costs and an increase in the number of high calibre recruits thereby enhancing the profession as a whole.

A further contribution to knowledge is made by examining the feasibility of replacing the error-prone clinical predictions of applicant performance, deriving from the current unstructured interviews, with those derived by a simple statistical model. The model-based approach is shown to reduce recruitment errors by restricting entry to those with a very high likelihood of success and compares very favourably with judgemental methods and other recruitment procedures.

However, such model-based approaches are criticised, notably by Mitchell and Klimoski (1982), for their susceptibility to shrinkage. Models were fitted on 1985 and 1986 'non-relevant' entrant data and validation, using 1987 trainee data, is discussed in Chapter 6 but, due to problems associated with a change in the examination syllabus, directly affecting the criterion variable of both sets of fitted models, the issue of reliability was not satisfactorily resolved.

Using a later sample of 1988-90 entrants to 3 of the firms from the original sample, the issue of the long-term reliability of biodata models is addressed in Part II. Very positive evidence of stability of the biodata models is provided. The models were found to perform well on subjects recruited up to 4 years after members of the derivation group.

The latter sample is also used to demonstrate the robust nature of the logit models and provides original evidence of their ability to 'transfer' to trainees with

accounting degrees who are very different from the non-accounting graduate population upon which the models were developed. This 'transferability', which is usually considered to be a feature only of cognitive tests (Hunter and Hunter, 1984) suggests that there may well be further applications for hard biodata models within professions.

In addition, the validation results suggest the models developed here from sample data pertaining to 22 different firms of different sizes within the 'medium-size' category, are suitable for use in any medium sized firm, refuting suggestions of the organizational specificity of biodata models. The implication is that biodata models, which can be applied to application form data, may provide predictions of future job-related behaviour across professions made up of many different organizations.

The practicalities of implementing such models in an industrial setting, via a pilot study and subsequent 'live' use of the models, are investigated in Chapters 7 and 8, with particular reference to the impact of such procedures on recruitment staff and recruitment policy. In particular, while accurate Utility Analysis cannot be carried out because of the availability of only estimated costings, the latter and estimated revenues supplied by the piloting firms are used to calculate estimates of the differences in net contribution associated with different combinations of examination result and leaving. This analysis allows the recruiter to judge the HRM implications of the use of such approaches. The use of the model-based system significantly increases the 'hit' rate, decreasing both Type I and II recruitment errors.

It is noted that firms actually 'over-recruit' to accommodate an expected wastage rate of approximately one third and not all of those who remain in the firms can be deemed 'successful'. Although this study examines the more 'tangible' losses associated with the poor validity of current recruitment methods, the 'intangible' costs, eg, lower productivity associated with poorly motivated or failing trainees, are not amenable to analysis.

In this concluding chapter, the specific objectives set in Chapter 1 will be discussed

and suggestions for areas of future research are made. Among the principal contributions of this thesis are :

- (i) a valid comparison made between the generally accepted WAB technique and an appropriate parametric alternative (logit),
- (ii) the provision of a software driven scoring system for applications to provide rapid, fair processing,
- (iii) the provision of evidence that biodata models developed on a representative sample from a profession may be generally applied to member organizations, ie that they are not organization-specific, and
- (iv) the provision of evidence that, like cognitive tests, appropriate biodata models are not sample-specific and may be applied to samples differing from the development sample, but to whom the criterion is applicable.

9.2 The Development of Selection Models

It is the contention of Mitchell and Klimoski (1982) that non-parametric methods of model derivation, like the WAB, are superior for the development of biodata models designed for practical use. They argue that such methods require fewer assumptions (eg, that of linearity of the data) than parametric multivariate methodologies and therefore capitalise less on chance.

However, their major concern was to address the criticism that empirical methods do not provide adequate explanation of criterion/predictor relationships, by developing a 'rational' approach, deriving their model component variables by factor analysis. They conclude that the WAB approach is superior to their more rational approach using parametric procedures³² and recommend it for practical use.

The WAB methodology develops 'over-fitted' models, which require large numbers of variables to be scored and, therefore, increase the likelihood of data loss due to

³² However, as indicated in Chapter 6, their adopted methodology is empirically flawed and, as a result, their conclusions are questionable.

missing information, and processor error. It is also impossible to determine which component variables are of greatest importance to the model's predictive power. Furthermore, there is no way of dealing with items which are contingent upon each other or collinear and the scores are likened to 'an unravelled network of information' (Drakeley (1989a) p152, quoting Guion (1965)).

It seems surprising that the WAB methodology is still considered to be useful, when there are so many multivariate alternatives available for dichotomous and polychotomous criteria. The popularity of the WAB may well arise from its ability to be developed from categorical data, which has no apparent ranking structure, without imposing the developer's assumptions of rank or worth. In this study, categorical variables were turned into 'dummy' binary variables for building the logit models, in order not to impose arbitrary ranks.

Another small but important issue may be raised by examining the weighting tables in the WAB approach. Despite thorough research, no original record of the derivation of Strong's (1926) Tables could be found. Strong refers to Kelly and Cowdery but does not provide references for them. Strong reports having found a near linear relationship between a net weight calculation derived by Kelley (which is provided, but this is mystifying since none of the terms are defined!) and the differences between the groups' percentages for the predictor variables. It is this relationship which generated the tables. Detailed library search prior to 1927 did not reveal any reference to Cowdery and Kelly relating to this subject.

This thesis compares the WAB with logit and argues that the multivariate empirical methodology, applied to verifiable data drawn from sources known to be valid predictors of the criterion measure, produces superior models, within which the strength of those relationships, and their positive/negative nature, is indicated by the available statistics.

Predictive Validity

The classification matrices of the developed models are used to assess predictive

validity in terms of the probability of correct classification. Results for logit Model 1 application to both the derivation and original validation samples are significant at and beyond the 0.01 level. However, the WAB results are significant at the 0.01 level when classifying the held-out sample, but only just achieve significance at the 0.05 level on the validation sample data.

For the Recruitment Success models, the logit model classification of the derivation sample is significant beyond the 0.001 level and of the validation sample at beyond the 0.01 level. The WAB methodology, however, provides a model with results not significant at the 0.05 level when applied to either the hold-out sample or the validation sample.

It might be argued that a fair comparison is not made here between the two methodologies. The use of the probability of correct classification to compare between the logit and WAB procedures may seem questionable, as the 0.5 cut-off is used in logit as in LDA, but the practitioner sets the cut-off for the WAB.

Mitchell and Klimoski (1982) actually used linear discriminant analysis (LDA) to set their optimum WAB score and found it was identical to that provided by England's 'index of greatest differentiation', recommended to the practitioner for setting the cut-off. The Morrison technique, used here to compare predictive ability, was developed for LDA. It is reasonable to suppose that comparing the results of a cut-off derived in the way England suggests, with logit, should be valid.

For our purposes, ie, the development of a scoring device not intended for applicant use, the logit procedure provides models which are superior to the WAB models. They are parsimonious of data, each of the two being only composed of 6 variables, but, more importantly, they maintain a significant level of accuracy, even in the face of a change in the nature of the criterion, viz, examination syllabus changes in 1987, which the WAB model failed to accommodate in Model 2.

Shrinkage

It should be noted that the both logit models and WAB models suffered shrinkage in 1987 and, on the basis of a comparison between the two methods, the logit models were chosen for implementation. However, there was no attempt made to validate the WAB models on the 1988-90 sample, and it is possible that the WAB was more adversely affected by the criterion change than the logit and might have fared better on later validation. Nevertheless, the complexity of the approach, problems in variable interpretation, the over-fitted nature of the models and the need to input so much data (increasing the likelihood of processor error) also contributed to the decision to use the logit models.

The ICAEW is restructuring the professional examination system and it may be that the models will lose predictive ability over time as the changes take effect. However, in theory, the subject matter examined will remain the same but the order of presentation and the manner of examination will change. Since the likelihood of failing the Final examination after passing the Intermediate examination at the first attempt is very low, it would appear that changing the order of presentation of the subject matter will not make material difference to model predictions. Changing the examination structure from the present formal system to a modular system with a case study and partly multiple-choice, may, however, change the pass rate and compromise the model.

It will be necessary to revisit the profession when the new system has been in place for a suitable length of time to assess the impact of the changes, if any, on model performance.

Faking

Since these models are designed to be used to score application forms, faking responses should not be an issue. Unless the information given on the application is false, the applicant has no opportunity to falsify the model outcome. This study took the approach of using only those items, not only known to be good predictors of success in terms of the criteria investigated, but verifiable and available on the

majority of applications as predictor data. However, there is still a minimal risk that an applicant, becoming aware of what firms are looking for (eg, a high number of grade A GCSE passes) will attempt to falsify his/her information.

Labelling data 'verifiable', in principle begs the question as to whether it can indeed be verified. Firms using such models need to exercise their right to ensure that it is. Many firms ask interviewees to bring their certificates or copies to the interview and this practice will minimise the risks associated with the academic data. However, there are items, like the number of solo pastimes at university, which, if their use as predictors became known, might easily be compromised.

In a sense the latter variable is not, in principle, readily verifiable. How can an employer be sure of what activities anyone undertakes unless these are corroborated by another? The subjects used to develop the measures were not aware that the study was being undertaken. However, if applicants become aware that forms are being vetted by a scoring system, they may change their data to 'improve' it in line with what they feel recruiters are looking for. This might have the effect of falsely enhancing or reducing their score.

The only sure way to prevent the data becoming contaminated (or at least remain as uncontaminated as that provided by the sample subjects) is to use the scoring system on a confidential basis. However, if the data is to be kept on a data base, this must be declared at the outset and permission to retain the information sought from the applicant, in line with legal requirements.

In order to ensure that the applicant is aware that the data may be checked, it has been suggested that instead of referees, the firm should ask for a witness to the correctness of the information (Anderson and Shackleton, 1986). Rather than dismayed at the prospect, experience suggests that most academics would rather do this, knowing the information is readily available from student files, than prepare a reference for a student they may have met very infrequently and do not know very well.

The impact upon applicants of using such a scoring system is not assessed here but some applicants, dismayed by firms refusing them interviews may ask for feedback on the cause of rejection. This is not a new problem for firms but it is a difficult area. However, providing a diplomatic response should not compromise the firm and should reduce adverse impact on the applicant. Applicants offer their information to the firm in the knowledge that they are going to be judged upon it. There should be no reason why they would object to its being scored appropriately, rather than being subject to a cursory pre-sift which produces random results.

Nevertheless, while applicant awareness of a scoring system should not compromise the hard data items, it does increase the need for firms to sample applications and make factual checks of their accuracy.

9.3 Model Performance versus Judgemental Methods

Examination of the results of conventional recruitment methods in the profession, based on our representative sample of firms, suggests they are no better than random selection in their results. This is not surprising, since the recruitment methods depend upon judgemental selection. This thesis provides evidence that model-based approaches provide more valid results than judgemental methods and are arguably fairer to applicants and more cost-effective for the profession.

Evidence from the pilot study and the recruited samples indicates that the gloomiest predictions of the efficiency of the unstructured selection interview are demonstrated in the chartered accountancy profession (Dipboye *et al*, 1984; Herriot, 1984), at least within the medium sized firms we examined. For example, in the later validation years (1988-90), interviews have only been 37% successful in terms of recruiting examination successes, whereas the entrants from the same years scoring above 0.7, the chosen cut-off, on Model 1 have a 0.59 (59%) probability of success, 0.27 (27%) of failing or referring one examination (most likely GCC) and only a 0.14 (14%) likelihood of failing outright.

The probability of correct classification for 1988-90 Pass and Fail groups is 0.77³³, whereas the clinical judgemental methods (ie, actual recruitment decisions) were only 60% successful on the same terms, ie, where the Intermediate cases are ignored. Using the suggested presift provides even greater differentiation, being 88% successful when only Pass and Fail outcomes are considered.

Whereas some firms in the professional may well be using structured, and perhaps situational interviews, none of the representative sample firms were doing so. However, they do not represent the very smallest firms nor the very largest. It is most unlikely that small firms, who struggle to find recruitment resources, will have been able to develop and use such tools. Large firms may well be an exception, since they apply large resources to the HRM function and employ many personnel and occupational psychology professionals. It is not the purpose of this thesis to explore this issue further.

Is it justified here to treat the recruited trainees as if they are a group of applicants in order to draw conclusions about the likely performance of the models versus interviews? The evidence, provided by the pilot study model scores, indicates that the distribution of applicant sample logit scores did not differ significantly from that of the recruited trainees and, on this basis, we are justified.

The logit provides an unbiased estimate of risks associated with the applicant in terms of general suitability, but the likelihood that the applicant will be successful in terms of the individual firm's environment must be assessed by other means. Both the applicant and the firm's representatives must meet to confirm their choices. However, using the logit approach, the number of interviews undertaken by firms falls dramatically and, simultaneously, the validity of the selection process rises.

³³ $t = 5.75, p < 0.01.$

9.4 Long Term Validity

Biodata models have been criticised for their susceptibility to shrinkage or loss of validity over time. This failing is not, however, confined to biodata measures. Such shrinkage may be associated *inter alia* with soft data items, which may fluctuate in ability to impact on the criteria over time and attention to this problem is drawn by Owens and Schoenfeldt (1979) and Eberhardt and Muchinsky (1982b). Loss of validity may also be associated with changes in the criterion variable as in our case (Guion, 1992).

The predictor items included in the models developed here are all hard data items. To a great extent they are concerned with specific academic achievement measures which are probably the most valid available indicators of academic success in professional examinations and, in addition, are standardized. Little change in the efficacy of Model 1 is expected over time as the criterion measure is purely academic (section 9.2).

However, Model 2 incorporates a 'soft' criterion measure, the performance rating, which may render this model more unreliable, as organizations have differing perspectives and focuses which lead to differentials in the perceived importance of behaviours. They also have different methods of measuring performance. Guion (1992) indicates that performance measurement '...may be the most intractable measurement problem of all' (p338). Nevertheless, firms within a profession are likely to agree on the important elements of acceptable performance and be able to recognise poor or unacceptable performance when it is observed.

Both models contain variables other than those directly related to academic achievement, and to use Drakeley's (1989) definitions, both motivational and background predictors are present. These items may well be those which contribute to the prediction of the motivational aspects of successful examination preparation and performance and to the prediction of commitment-related withdrawal, although this has not been specifically tested here.

Although the models do not perform as well on the original validation sample subjects, it is not clear how much of this loss is attributable to the criterion variable and how much is due to the 'shrinkage' which might be expected, since predictive models do not generally perform as well on inter-temporal validation samples as they do on the derivation sample.

Anecdotal evidence from the accountancy profession suggests that the professional training firms did not adjust well to the change of syllabus in the Foundation examination in 1987, and this led to the unusually low first time pass rates. It was supposed that training firms had remedied the situation during the following year, as pass rates returned to former levels. Our data certainly demonstrates a collapse in pass rates for 1987 but, of course, we cannot firmly conclude that training firms were to blame.

Examination of the proportions of criterion group members in the 1988-90 sample, reveals them not to be significantly different from those in the 1985/6 model development sample, but both are markedly different from 1987, the increased proportion of those falling in the Fail/Leave categories being significant. The full validation exercise undertaken in Part II of the study, using the 1988-90 entrants, concludes that the models demonstrate high levels of validity in the longer term.

Thus, for whatever reason the 1987 sample differed from the two previous years and the three subsequent years, model performance was adversely affected in that year. As the models work well on the latter years' entrants, it is reasonable to suppose that the examination syllabus change was, to a great extent, responsible for this.

The long term results must relate to the stability of the criterion measure. The findings of this study support the proposals of Rothstein *et al* (1990) that the validity of predictive biodata models may remain stable for many years, providing the criterion measure remains stable.

9.5 Validity Generalization

Entrants with accounting degrees are not included in the derivation sample and participating firms expressed a wish for a similar scoring device for such applicants. The models developed for the non-accounting graduate majority are applied here to the accounting graduates to test the hypothesis that they will be valid predictors and may be 'transferred' across different populations.

The results suggest that the examination prediction biodata model predicting, as it does, virtually the same criteria for accounting and non-accounting graduates, may, with a scale adjustment be applied successfully to the former. These results also lend support to Rothstein *et al* (1990), who argue that such models, appropriately developed, may be as generally applicable as cognitive tests. Indeed it seems highly likely that the predominantly academic indicators of professional examination success are valid across samples.

The need for an adjustment in scaling arose from the different academic profiles of the two target groups. The non-accounting graduates had a higher average number of grade A GCE 'O' level passes than the accounting graduates and the latter could not score on the science/technology degree variable. In addition, accounting graduates are likely to be more committed to the profession, having taken their career decision on leaving school and having had three years of related experience unavailable to the non-accounting graduates.

The criteria for the second model, predicting Recruitment Success v. Failure (leaving within two years) are not applicable to the Foundation (GCC) exempt graduates and thus they are not scored on Model 2.

Since the models are developed using a very representative sample of 23 offices, drawn from 22 medium sized firms, ie, a multi-organizational approach as recommended by Rothstein *et al* (1990), the results are certainly applicable to the whole population of medium size accounting firms and potentially, although not

tested here, to the smaller and larger firms. They are neither designed nor intended to be organization specific, but are intended to be applied in all medium size accounting firms to improve recruitment across the board. The success of these models further supports the arguments of Rothstein *et al* (1990) who maintain that biodata models are not necessarily organization specific.

Many of the elements of successful performance will be present in all work situations and some will be more obviously situation specific. It may well be that these models would be equally useful in another profession, where professional examination performance and 'acceptable' work-related performance are applicable predictive criteria. Certainly the ability-related items are highly likely to be 'global' predictors of professional success.

9.6 Implementation Issues

Whereas most selection techniques are applied to applicants who have been pre-selected on the basis of some set criteria, unbiased scoring, in this case, occurs for all applicants meeting accepted minimum professional requirements.

The disadvantages of using statistical computer-based systems obviously include the animosity of those confident, however falsely, that personal judgement in the selection process is vital and that such models 'dehumanize' the whole selection process. Unfortunately, as section 1.3.3 indicates, it is the input of the human judgement aspects in the selection interview, which render the process unreliable.

The present system in the accounting profession, involving as it does, one or more unstructured selection interviews, provides a recruitment function with results no better than chance, but which still appeals to practitioners for all the reasons cited in section 1.3.3 (Tversky and Kahnemann, 1974). Whether these results are echoed by the experience of other graduate recruiters or not, casual acceptance of such human resource wastage does no credit to any organization and cannot be justified

simple because the methods appear to be 'acceptable' to the participants.

Applicants who were previously rejected because of their poor qualifications will still be rejected by the use of the suggested pre-sift and Model 1, although, in practice, there is no need to use the pre-sift, nor an UCCA point constraint, as the model will remove the unsuitable applicants. However, the time constraints of processors in the pilot firms forced the issue of using a model variable to define a suitable pre-sift to remove the bulk of the applications before scoring.

The numbers of applicants filtered out by the developed process are great and firms might feel concern that not enough good candidates will get through. However, the evidence of the brief pilot exercise, reported in Chapter 7, highlights the number of highly desirable applicants not even interviewed and very poor applicants given interviews. The indications are that there are likely to be adequate high calibre applicants within the pool, given that firms using predictive selection models are relieved of the need to over-recruit to insure against high wastage rates and should therefore be recruiting fewer, higher quality trainees.

For example, using the pre-sift and the 0.7 cut-off for Model 1, firms may assume that those non-accounting graduates selected will contain 8% at risk of failing or being referred in both Foundation and Intermediate examinations, 26% are those who will fail or be referred in either one or the other of these examinations, but the remaining 66% of trainees will prove successful in terms of examinations. A marked difference from the 1988-90 trainee record of 25%, 38% and 37% respectively³⁴. Examination of Model 2 scores will also give an indication of the applicant's likelihood of leaving, which, simply in cost terms, is to be avoided.

The overload on processors which led to the adoption of a pre-sift for the piloting firms, begs the question: Why are such large numbers of inappropriate candidates submitting applications? This may be due to a combination of factors. The

³⁴ $\chi^2 = 19.17$, 2 df, $p < 0.001$.

recruitment literature and presentations may not provide accurate or adequate information about the job, thus forcing errors of choice in prospective applicants. It is certainly clear that the observations of Wilson (1989) and Dean *et al* (1988), concerning the information available on the accounting profession, are upheld. Obviously, those applying to the firms are not sufficiently informed to self-select accurately (Williams, 1984; Herriot, 1986; Meglino and De Nisi, 1987; Wilson, 1989).

The nature of the introductory literature and content of presentations may reflect the firms' major recruitment concerns. For example, while the emphasis of firms' chosen recruitment criteria in this study is placed firmly on performance and tenure, there was no suggestion that these might not be adequate, having no specific focus on the more employee-centred criteria. It is certainly possible for a trainee to perform satisfactorily in the job but to be unhappy in the firm. If this is the case, the trainee is unlikely to wish to remain with the firm at the end of the contracted period and may leave before that time. This may result from organizational, rather than occupational, commitment problems.

There are a number of reasons why trainees' characteristics may not match with those required by the firm. They may not have adequate general academic ability to perform well in professional examinations and to develop sound job-knowledge. They may not have the necessary social skills to cope with the demands of professional work, and develop poor relationships with colleagues and/or with clients. They may not be of the 'conventional' type or find the size of work group or type of work setting suits him/her.

Those whose personality fits Holland's "conventional" vocational type have attributes which are congruent with those required by chartered accountancy. They may, however, be more suited to another conventional profession, for example, law or banking. Firms in the accountancy profession generally do not find this analysis helpful. They believe that a conventional image is not likely to attract the 'right' applicants, as their profession is neither "stuffy" nor "staid", although it may be

"conservative". This is missing the point, whether they wish it or not, those who are successfully employed as accountants fit the type (Aranya *et al*, 1978). If firms wish to recruit successful trainees, a circumspect job-description will not help.

The demonstrable necessity of standardizing the selection system, to reduce errors, encouraged participating firms to implement the models in their offices in the recruitment round following the pilot, ie the 1991/92 round recruiting 1992 entrants. Reluctance to use statistical methods in place of selection interviews was, in practice, overcome by the pilot exercise drawing firms' attention to the lack of efficiency of their pre-selection methods, compared with those deriving from a simple model, and the 'ones that got away'.

Concerning the validity and fairness of such approaches, this study clearly shows that, treating the entrants as a group of applicants, stand-alone use of the developed models is far more valid than the full selection interview-based strategy currently adopted by accounting firms. There is no evidence, on the basis of our data, that applicants scored by the models are not all being treated fairly, regardless of race or sex, which may not have been the case in the past.

In effect, using these models to replace selection interviews would result in a substantial increase in overall recruitment validity, decreasing the loss of likely successes and increasing the rejection of likely failures. The estimated financial implications of this increase, even assessed in the lowest terms, argue strongly for the use of the models to be incorporated by firms into their recruitment process.

9.7 Limitations of this Study and Further Research

The selection process, however refined, will have diminished effect where what happens before and after its application is not adequately standardised nor based on clearly defined objectives, so the need for proper analysis and development of the primary and tertiary stages of recruitment is emphatic.

This study is certainly limited by its chosen restriction to the use of hard verifiable predictor data. For example, unquantifiable data may provide the key to predicting the more motivational aspects of performance (Drakeley, 1989a), which may not be predicted by hard biodata. Certainly a biodata inventory might be considered for the profession, utilizing both types of data to gain maximum predictive ability.

The WAB approach yielded some interesting insights into the variables available on the application form which indicate significant differences between criterion groups. While the WAB approach does not offer the same explanatory power as the multivariate approach in this research, the results may well form the basis of a more qualitative biodata instrument, such as a self-report inventory.

The ecology model of development (Mumford *et al*, 1990) emphasises the importance of the choices an individual makes which result in his/her differentiation from others in his/her subgroup. The individual's developmental trace can be likened to the progress of a ball in a pin-ball machine: at each developmental marker a choice reflecting a personal vision of the future is made and the individual moves on to the next check. Thus progress through life and the development of the individual can, in theory, be predicted.

Biodata inventories with multiple choice response categories allow the user to capture information which may give insight into the 'causal influences' and 'developmental markers' which contribute to identifying the individual's developmental trajectory (Stokes and Reddy, 1992). While the WAB approach was unable to provide a reliable, sophisticated model, the items which differentiate between the successful and unsuccessful recruit may reflect important developmental markers for both types of model.

However, soft data is not easily amenable to statistical analysis, since arbitrary rankings of response categories are imposed by researchers. Mitchell and Klimoski (1982) address this issue by performing factor analysis to provide predictor variables in the form of factors. They admit to arbitrary ranking on hypothetical continua,

ie, assuming interval-scale relationships where they may not exist. The use of such data may be, in considerable part, responsible for biodata's reputation for excessive 'shrinkage' (Shuh, 1967; Owens and Schoenfeldt, 1979; Eberhardt and Muchinsky, 1982b; Mitchell and Klimoski, 1982).

Considerable time is taken up by developing instruments with multiple choice formats and checking for test/retest reliability. It would not be prudent to devote high levels of resources to such practices where the criteria had not been optimally set. The models developed here may be seen as an interim measure for increasing recruitment validity while the profession assesses its long term recruitment goals and sets criteria which reflect them.

A key area for concern in this study must be the lack of job analysis used to develop the recruitment criteria. Herriot (1986) is clear that the entire process of recruitment should be based on a job analysis to focus the recruiter on what (s)he is really seeking and throughout the literature this is a recurring theme. In this study the recruitment professionals identified what they believed to be their recruitment criteria and, as these reflected valid concerns and echoed those of the professional body (ICAEW), they were adopted.

However, it has already been noted that no account has been taken of more essential trainee-centred criteria like job satisfaction. As firms aim to make a profit from any trainee entering the firm, their limited focus is understandable, their priorities are based on maximising profit. However, as members of a profession, it is in their interest to promote recruitment goals which lead to a membership enriched in other ways.

For example, recruiters in this study stressed the importance to them of a high likelihood of first-time pass rates in professional examinations and yet the brief cost/benefit analysis undertaken in Part II indicates that the differences between those who have examination failures and those who are successful are not so great in direct cost terms as might have been expected. Firms obviously believe that first time

passing is more important than it is in these narrowly defined profit terms.

The 2 year time criterion time-frame reflects concerns about wastage, the bulk of trainees who leave doing so within the first two years. The statistics indicate that those who manage to survive the first two years are most unlikely to leave during the third (ICAEW, 1992) and, whereas the majority of those who leave are assumed to be examination failures, this was not tested here. However, no account was taken of the effects of the second professional examination (PE2), nor of performance later in the trainee's career, particularly after the training period is over.

Narrowing the criterion variables down to those relating to the first two years assumes again that successful recruitment is only concerned with the training period. The fact that over half of the members of the ICAEW work outside accountancy should cause as much, if not greater, concern.

Firms' recruitment priorities should, perhaps, lie in identifying those who will make successful partners and generate new business and who may not be those who will pass examinations at the first attempt. They may be 'innovators' whose characters have made them appear unsuitable in the past, rather than the 'adaptors' which are often perceived as more suitable (Kirton, 1976).

A full analysis of the skills and characteristics required to perform well in chartered accountancy will clarify recruitment objectives and provide invaluable information on which areas provide the key constituents of successful and unsuccessful performance within the training period and beyond. This information, appropriately disclosed to applicants, will improve the likelihood of efficient self-selection (Williams, 1984; Herriot, 1986; Meglino and De Nisi, 1987; Wilson, 1989).

This study also reveals a pressing need in the profession for an overall approach to identifying comparative levels of behaviour in the generality of work situations encountered by trainees. Even though salary structures are related to performance in many of the firms, their appraisal systems are inadequate and poorly prepared.

In this study we were unable to develop comparative scales for performance data and had to rely on a rating which is completed at 6 months on an ICAEW TR (training record). This is completed by the training partner who may or may not have actually worked with the trainee and, under scrutiny for reliability, this measure cannot be deemed satisfactory. Model 2 may only maintain validity because the criterion depends on the simple dichotomy of acceptable v. unacceptable performance.

The measure used here, rating at 6 months into the training contract, may well be insufficient to reflect future practice performance, as non-accounting graduate trainees spend very little of this period out in client firms, or even in their own. It may, however, be that these six months represent a crucial stage during which the entrant makes the adjustments to expectations to reduce ORS, those failing to make the necessary adjustments then being more likely to leave. Unfortunately, this cannot be tested here, since there is not adequate performance data for comparison with later periods.

Until such time as the range of acceptable behaviours is identified and methods developed to score them fairly, no direct comparison can be made between levels of practice work performance of trainees within the same firm, let alone between trainees from different firms. The development of Behaviourally Anchored Rating Scales [BARS] might prove very useful in raising the validity and reliability of performance appraisal for the profession as a whole, providing invaluable guidance for raters making assessments and for trainees qualified accountants making self-assessments.

This study provides models to increase the validity of the initial stage of recruitment but firms' representatives will still meet those selected before a firm offer of employment is made. Bearing in mind the likelihood of interviewer bias affecting the outcome of this meeting, some control is still required. Firms should perhaps consider testing interviewers' ability to pass on and elicit relevant information without prejudice and develop a list of items which must be covered and rated by

each decision-making party. The nature of the items which contribute to successful job-person match should be the subject of future research (Rynes and Gerhardt, 1990).

Having provided evidence of the 'generalizability' of the biodata models developed in this research, another obvious area for further work is the applications of these models to larger and smaller firms than were represented in the derivation sample.

In addition, the 'transferability' of the models to a different sample (the accounting graduates) and the ability of the models to generalize across a profession, provides optimistic evidence that it may well be possible to develop one predictive biodata model for scoring applicant information for a number of different professions with similar recruitment criteria.

Recent work within a Big Six accounting firm has already successfully developed models to predict overall professional examination performance (including the second Professional examination) and superlative ratings of performance after completion of the training contract. The evidence of this thesis indicates that these models are likely to be applicable to other very large accounting firms (although for commercial reasons they would not be released) and, again, probably other professions with large graduate entries.

Bibliography

ACCOUNTANT, THE (1988): Top UK Firms Ranked by Fee Income. June 1988: 14.

ACKERMAN, P.L. (1989): Within task intercorrelations of skilled performance: Implications for predicting individual differences. (A comment on Henry and Hulin, 1987.) Journal of Applied Psychology 74: 360-364.

ANDERSON, N. and SHACKLETON, V. (1986): Recruitment and Selection : A review of developments in the 80's. Personnel Research 15 (4): 19-26.

ARANYA, N., EKHARAN, I.M. and AVRAHAM, B. (1978): An empirical examination of the stereotype accountant based on Holland's theory. Journal of Occupational Psychology 51: 139-145.

ARVEY, R.D. and CAMPION, J.E. (1982): The Employment Interview: A summary and review of recent research. Personnel Psychology 35: 281-322.

ASHER, J.J. (1972): The Biographical Item: Can it be improved? Personnel Psychology 27: 519-521.

BAEHR, M. and WILLIAMS, G.B. (1967): Underlying dimensions of personal background data and their relationships to occupational classification. Journal of Applied Psychology 51: 481-490.

BAEHR, M. and WILLIAMS, G.B. (1968): Prediction of sales success from factorially determined dimensions of personal background. Journal of Applied Psychology 65: 662-671.

BARBER, A.E., HOLLENBECK, J.R., TOWER, S.L. and PHILLIPS, J.M : The effects of interview focus on recruitment effectiveness : A field experiment. Journal of Applied Psychology 6: 886-896.

BARRETT, R.A.A. and DOVERSPIKE, D. (1992): The implications of apparent declines in predictive validities over time: A critique of Hulin, Henry and Noon. Personnel Psychology 45: 601-617.

BECKER, T.E. and COLQUITT, A.L. (1992): Potential versus actual faking of a biodata form: An analysis along several dimensions of an item type. Personnel Psychology 45: 389-406.

BENVENISTE, I., SHERER, M. and WARING, R. (1986): An evaluation of front end loaded education for ICAEW students. (University of Sussex).

- BETHEL-FOX, C.E., CURETON, R.N. and TAYLOR, J.A. (1988): The effectiveness of biodata in pre-selection for the Appointments-in-Administration Competition. (Civil Service Selection Board Report No.37).
- BEVAN, S. (1989): Staff selection: art or science? Accountants' Magazine June: 33-34.
- BLANKSBY, M., and ILES, P. (1990): Recent developments in assessment centre theory, practice and operation. Personnel Review 19 (6): 33-44.
- BLINKHORN, S. and JOHNSON, C. (1990): The insignificance of personality testing. Nature 348 (20th December): 671-672.
- BRENNON, J. and MCGEEVOR, P. (1990) : Ethnic Minorities in the Graduate Labour Market. (London:CRE).
- BRITISH PSYCHOLOGICAL SOCIETY (1993) : Graphology in Personnel Assessment (Leicester, BPS).
- BOUDREAU, J.W. (1992): Utility Analysis for Decisions in HRM. In Handbook of Industrial and Organizational Psychology, Vol 2 eds Dunnette, M. and Hough, L.M. (Palo Alto : Consulting Psychologists Press Inc).
- CAMPION, J.E. and ARVEY, R.D. (1989): Unfair Discrimination in the Employment Interview. In The Employment Interview, eds Eder, R.W. and Ferris, G.R. (New York : Sage).
- CAMPION, M.A., PURSELL, E.D. and BROWN, B.K. (1988) : Structured interviewing : Raising the psychometric properties of the employment interview. Personnel Psychology 41: 25-42.
- CARLSON, R.E., SCHWAB, D.P. and HENNEMAN, H.G. (1970) : Agreement among selection interview styles. Journal of Industrial Psychology 5 (1): 8-17.
- CARLSON, R.E, THAYER, P.W., MAYFIELD, E.G and PETERSON, D.A. (1971): Improvements in the selection interview. Personnel Journal 50 (4): 268.
- CASCIO, W.F. (1975) : Accuracy of verifiable biographical information responses. Journal of Applied Psychology 60 (6): 767-769.
- CHILDS, A. and KLIMOSKI, R.J. (1986) : Successfully predicting career success : An application of the biographical inventory. Journal of Applied Psychology 71: 3-8.
- CLARKE, R.L. and SWEENEY, R.B. (1985) : Admission to accounting programs : Using a discriminant model as a classification procedure. The Accounting Review, LX (3): 508-518.

- COHEN, J. and LEFKOWITZ, J. (1974) : Development of a biographical inventory blank to predict faking on personality tests. Journal of Applied Psychology 59 (3): 404-405.
- COLLIER, P. and MAYER, C. (1985): An investigation of university selection procedures. Economic Journal 163-170.
- COLLINS, S.H. (1987) : Recruitment. Journal of Accountancy 163 (2): 52-60.
- CRAWLEY, B., PINDER, R. and HERRIOT, P. (1990): Assessment centre dimensions, personality and aptitudes. Journal of Occupational Psychology 63 (3): 211-216.
- DAVIS, K.R.D. (1984) : A longitudinal analysis of biographical subgroups using Owens's developmental-integrative model. Personnel Psychology 37: 1-4.
- DAWES, R.M. (1977): Case-by-case versus rule generated procedures for the allocation of scarce resources. In: Human Judgement and Decision Processes in an Applied Setting, ed. by Kaplan, M.F. and Schwartz, S. (Academic Press).
- DAWES, R.M. and CORRIGAN, B. (1984): Linear models in decision making. Psychological Bulletin 81: 95-106.
- DEAN, R.A., FERRIS, K.R. and KONSTANS, C. (1988) : Occupational reality shock and organizational commitment : evidence from the accountancy profession. Accounting, Organizations and Society 13 (3): 235-250.
- DIPBOYE, R.L., ARVEY, R.D. and TERPESTRA, D.L. (1977) : Sex and physical attractiveness of raters and applicants as determinants of resume evaluations. Journal of Applied Psychology 62: 288-294.
- DIPBOYE, R.L., FONTENELLE, V. and GARNON, K. (1984) : Effects of previewing the application on interview process and outcomes. Journal of Applied Psychology 69 (1): 118-128.
- DIPBOYE, R.L., FROMKIN, H.L. and WIBACK, K. (1975) : Relative importance of applicant sex, attractiveness and scholastic standing in evaluation of job applicant resumes. Journal of Applied Psychology 60: 39-43.
- DIPBOYE, R.L., STRAMLER, G.A. and FONTENELLE, G.A. (1984): The effects of application recall of information from the interview. Academy of Management Journal 27 (3): 561-575.
- DOCKWEILER, R.C and WILLIS, C.G. (1984) : On the use of entry requirements for undergraduate accounting programs. Accounting Review 59 (3): 496-504.
- DOWNS, S. (1970): Predicting Training Material. Personnel Management 2: 26-28.

- DRAKELEY, R.J. (1989a) : Achievement, Background and Commitment: Classification of Biodata in Personnel Selection. PhD Thesis: University of London (Berkbeck College), Faculty of Science.
- DRAKELEY, R.J. (1989b) : Biographical Data. In: Handbook of Assessment in Organizations, ed. Herriot, P. (London: Wiley & Sons.
- DRAKELEY, R.J., HERRIOT, P. and JONES, A.P. (1988) : Biographical data, training success and turnover. Journal of Occupational Psychology 61 (2): 145-152.
- DREHER, G.F., ASH, R.A. and HANCOCK, P. (1988) : The role of the traditional research design in underestimating the validity of the employment interview. Personnel Psychology 41 (2): 315-328.
- DUNN, W.M. and HALL, M.W. (1984) : An empirical analysis of the relationships between CPA examination candidate attributes and candidate performance. Accounting Review LIX (4): 674-689.
- EBERHARDT, B.J. and MUCHNINSKY P.M. (1982a) Biodata determinants of vocational typology : An integration of two paradigms. Journal of Applied Psychology 67 (6): 714-732.
- EBERHARDT, B.J. and MUCHINSKY, P.M. (1982b) : An empirical investigation of the factor stability of Owens's Biographical Questionnaire. Journal Of Applied Psychology 67 (2): 138-145.
- EBERHARDT, B.J. and MUCHINSKY, P.M. (1984) : Structural validation of Holland's Hexagonal Model : Vocational classification through the use of biodata. Journal of Applied Psychology 69 (1): 174-181.
- EFRON, B. (1975) : The efficiency of logistic regression compared to normal discriminant analysis. Journal of the American Statistical Association 70: 898-899.
- ENGLAND, G.W. (1971): The Development and Use of Weighted Application Blanks (revised edition). (University of Minnesota Industrial Relations Centre: Minneapolis).
- ESKEW, R.K. and FALEY, R.H. (1988) : Some determinants of performance in the first college level Financial Accounting course. Accounting Review 63 (1) 137-147.
- FERGUSON, E. and HATHERLEY, D. (1991): The work environment in the accountancy firms : A comparison of student expectations and trainee perceptions. British Accounting Review 23: 123-132.
- FESTINGER, A.L. (1957): A Theory of Cognitive Dissonance. (Row Peterson: Evanston, Illinois)

FLYNN, R.R. (1989): Rushton, Evolution and Race: An essay on intelligence and virtue. The Psychologist 2 (9): 363-366.

GABLE, M., HOLLON, C. and DANGELLO, F. (1989): The relationship of application form information and performance to managerial trainee turnover. International Journal of Management 6 (3): 289-295.

GHISELLI, E.E (1966): The Validity of Occupational Aptitude Tests. (New York: Wiley and Sons)

GHISELLI, E.E. (1973): The validity of aptitude tests in personnel selection. Personnel Psychology 55: 491-2.

GILBERT, E.S. (1968) : On discriminant analysis using qualitative variables. Journal of the American Statistical Society 63: 1399-1412.

GILMORE, D.C. and FERRIS, G.R. (1989) : The effects of applicant impression management tactics and interviewer judgements. Journal of Management 15 (4): 557-564.

GLENNON, J.R., ALBRIGHT, L.E, and OWENS, W.A. (1966): A Catalog of Life History Items. (Greensboro, N.C.: The Richardson Foundation)

GOLDSTEIN, I.L. (1971) : The application blank : How honest are the responses? Journal of Applied Psychology 55 (5): 491-492.

GUION, R.M. (1992): Personnel Assessment, Selection and Placement. In Handbook of Industrial and Organizational Psychology, Vol 2 eds Dunnette, M. and Hough, L.M. (Palo Alto : Consulting Psychologists Press Inc).

HAEFNER, J.E. (1977) : Race, age, sex and competence as factors in employer selection of the disadvantaged. Journal of Applied Psychology 62: 199-202.

HAKEL, M.D. (1989): The State of the Employment Interview: Theory and Research. In The Employment Interview eds Eder, R.W. and Ferris, G.R. (New York : Sage)

HARRIS, M.M (1990): Reconsidering the employment interview: A review of recent literature and suggestions for further research. Personnel Psychology 43 (2): 691-726.

HARVEY, R.J. (1994): The Common Metric Questionnaire (London: The Psychological Corporation).

HARVEY-COOK, J.E. (1984): Recruitment procedures in the accountancy profession. M.B.A thesis: City University Business School

HARVEY-COOK, J.E. and TAFFLER, R.J. (1987) : Graduate recruitment

procedures in the UK accounting profession : A preliminary study. Accounting and Business Research Spring: 99-108.

HEILMAN, M.E. and SARUWATARI, L.R. (1979) : When beauty is beastly : The effects of appearance and sex on evaluations of job applicants for managerial and non-managerial jobs. Organizational Behaviour and Performance 23: 360-372.

HELMREICH, R., BAKEMAN, R. and RADLOFF, R. (1973) : The life history questionnaire as a predictor of performance in navy diver training. Journal of Applied Psychology 57: 148-153.

HENRY, R.A. and HULIN, C.C. (1987): Stability of skilled performance across time: Some generalizations and limitations on utilities. Journal of Applied Psychology 72: 457-462.

HERRIOT, P.(1984) : Down From the Ivory Tower : Graduates and Their Jobs. (Chichester: Wiley & Sons).

HERRIOT, P.(1985): Give and take in selection. Personnel Management 17 (May): 33-35.

HERRIOT, P. (1987) : Graduate recruitment : Getting it right. Employment Gazette, February.

HERRIOT, P. and ROTHWELL, C. (1981): Organizational choice and decision theory: Effects of employers' literature and selection interview. Journal of Occupational Psychology 54: 17-31.

HERRIOT, P. and ROTHWELL, C. (1983): Expectations and impressions in the graduate selection interview. Journal of Occupational Psychology 56: 303-314.

HOLLAND, J.L. (1976) : Vocational Preferences. In Handbook of Industrial and Organizational Psychology, ed by Dunnette, M. (Chicago: Rand McNally).

HOLLMAN, T.D. (1972): Employment interviewers' errors in processing positive and negative information. Journal of Applied Psychology 56: 130-134.

HOSKIN, K. and STEELE, A. (1991): Assessing the Business Professional (ICAEW: London).

HOUGH, L.M, EATON, N.K., DUNNETTE, M.D., KAMP, J.D. and McCLOY, R.A (1990): Criterion related validities of personal constructs and the effects of response distortion on their validities. Journal of Applied Psychology 75: 581-595.

HUFFCUT, A.I. and ARTHUR, W. (1994) : Hunter and Hunter revisited : Interview Validity for Entry Level Jobs. Journal of Applied Psychology 79 (2): 84-190.

HULIN, C.L., HENRY, R.A. and NOON, S.L (1990): Adding a Dimension: Time as a factor in the generalizability of predictive relationships. Psychological Bulletin 107: 328-340.

HUNTER, J.E and HUNTER, R. (1984) : Validity and utility of alternative predictors of job performance. Psychological Bulletin 96: 72-98.

ICAEW (1988): Digest of Education and Training Statistics 1986/7 (London: ICAEW).

ICAEW (1989): Recruitment and Retention in Small Firms. (London: ICAEW)

ICAEW (1990): Recruitment in the 1990's: A sellers' market. Accountancy (2): 143.

ICAEW (1992): Education and Training and Student Salary Statistics - 1990/91 (London: ICAEW)

ICAEW (1995): Education and Training and Student Salary Statistics - 1993/4 (London : ICAEW)

JACKSON, D. and ROTHSTEIN, M. (1993): Evaluating personality testing in personnel selection. The Psychologist 1: 8-11.

JANZ, T (1982): Initial comparisons of patterned behavior description interviews versus unstructured interviews. Journal of Applied Psychology 67: 577-580.

KAHNEMAN, D. and TVERSKY, A. (1973): On the psychology of prediction. Psychological Review 80: 251-273.

KEENAN, A. (1976) : Effects of non-verbal behaviour of interviewers on candidates' performance. Journal of Occupational Psychology 49: 171-176.

KEENAN, A. (1977) : Some relationships between interviewers personal feelings about candidates and their general evaluation of them. Journal of Occupational Psychology 50: 275-283.

KEENAN ,A. (1978) : The selection interview : Candidates reactions and interviewers' judgements. British Journal of Social and Clinical Psychology 17, 201-209.

KEENAN. A. (1985) : Graduate recruitment : How graduates 'select' companies - A note. Personnel Research 14 (1): 12-14.

KEENAN, A. and WEDDERBURN, A.A. (1980) : Putting the boot on the other foot : Candidates' descriptions of interviewers. Journal of Occupational Psychology 53: 81-89.

KIRTON, M. (1976): Adaptors and Innovators: A description and measure. Journal of Applied Psychology 61 (5): 622-629.

KLIMOSKI, R. and BRICKNER, M. (1987) : Why do Assessment Centres work? The puzzle of assessment centre validity. Personnel Psychology 40 : 243-259.

KOUTSOYANNIS, A. (1983): Theory of Econometrics (2nd Edition) (London : Macmillan Press)

KUHNERT, K.W. and RUSSELL, R.C.J. (1990): Using constructive development theory and biodata to bridge the gap between personnel selection and leadership. Journal of Management 16 (3): 595-608.

LACHENBRUCH, P.A. (1967): An almost unbiased method of obtaining confidence intervals for the probability of misclassification in discriminant analysis. Biometrics 23: December, 639-645.

LATHAM, G.P. (1989): The reliability, validity and practicality of the situational interview. In The Employment Interview eds Eder, R.W. and Ferris, G.R. (New York : Sage).

LATHAM, G.P., SAARI, L.M. and PURSELL, E.D. (1980) : The situational interview. Journal of Applied Psychology 65: 422-427.

LATHAM, G.P and SAARI, L.M. (1984) : Do People Do What They Say? Further studies on the Situational Interview. Journal of Applied Psychology 69 (4): 569-573.

LEE, R. and BOOTH, J.M. (1974): A utility analysis of a WAB designed to predict turnover for clerical employees. Journal of Applied Psychology 59 (4): 516-558.

LIDEN, R.C. and PARSONS, C.K. (1986) : A field study of job applicant interview perceptions, alternative opportunities and demographic characteristics. Personnel Psychology 39: 109-122.

MACAN, T.H. and DIPBOYE, R.L. (1990): The relationship of interviewers pre-interview impressions to selection and recruitment outcome. Personnel Psychology 43: 745-768.

MAEL, F.A. (1991): A conceptual rationale for the domain of attributes of biodata items. Personnel Psychology 44: 763-792.

MATTESON, M.T. (1978): An alternative approach to using biographical data for predicting job success. Journal of Applied Psychology 51: 155-163.

MAURER, S.D and FAYE, C. (1988): Effect of situational interviews and training on interview rating agreement: An experimental analysis. Personnel Psychology 41 (2): 329-344.

MEGLINO, B.M. and DE NISI, A.S. (1987) : Realistic Job Previews : Some thoughts on their more effective use in managing the flow of human resources. Human Resource Planning 10 (3): 157-167.

MITCHELL, T.W. and KLIMOSKI, R.S. (1982) : Is it rational to be empirical? A test of methods for scoring biodata. Journal of Applied Psychology 67: 411-418.

MONAHAN, C.J. and MUCHINSKY, P.M. (1983) : Three decades of personnel selection : A state of-the-art analysis and evaluation. Journal of Occupational Psychology 56: 215-225.

MORRISON, D.G. (1969) : On the interpretation of discriminant analysis. Journal of Marketing Research 6: 156-163.

MORROW, P.C. (1990): Physical attractiveness and selection decision making. Journal of Management 16 (1): 45-60.

MOTOWIDLO, S.J and TIPPINS, N. (1993): Further studies of the low-fidelity simulation in the form of a situational inventory. Journal of Occupational and Organizational Psychology 66: 337-344.

MUCHINSKY, P.M. (1979) : The use of reference reports in personnel selection : A review and evaluation. Journal of Occupational Psychology 52: 287-297.

MUMFORD, M.D., STOKES, G.S. and OWENS, W. (1992): Developmental Determinants of Individual Actions: Theory and practice in the application of background measures. In Handbook of Industrial and Organizational Psychology, Vol 3 eds Dunnette, M. and Hough, L.M. (Palo Alto : Consulting Psychologists Press Inc).

MUMFORD, M.D and OWENS, W. (1987): Methodology Review: Principles, procedures and findings in the application of background measures. Applied Psychological Measurement 11:1-31.

NEINER, A.G. and OWENS, W.A. (1985) : Using biodata to predict job choice among college graduates. Journal of Applied Psychology 71: 127-136.

NORUSIS, M.J. (1990): SPSS/PC+ Advanced Statistics 4.0. Chicago: SPSS Inc.)

OATES, N. (1985): Evaluation of a biographical sifting system for the Diplomatic Service. (Civil Service Selection Board) Unpublished manuscript.

O'KANE, B. (1989): Accountants: The facts and figures. Accountancy Centenary Supplement (7): 9-15.

OKANES, M.M. and TSCHIRGI, H. (1978) : Impact of face-to-face interviews on prior judgements of a candidate. Perceptual and Motor Skills 46: 322.

ORPEN, C. (1985): Patterned behaviour description interviews versus unstructured interviews: A comparative validity study. Journal of Applied Psychology 70: 774-776.

OWENS, W.A. (1976) : Background Data. In Handbook of Industrial and Organizational Psychology, ed Dunnette, M.V. (Chicago: Rand McNally).

OWENS, W.A. and SCHOENFELDT, L.F. (1979) : Towards a classification of persons. Journal of Applied Psychology 65 (5): 569-607.

PACE L.A. and SCHOENFELDT, L.F. (1977) : Legal concerns in the use of weighted application blanks. Personnel Psychology 30: 159-166.

PA Consulting Group (1993) : Summary of Graduate Salaries and Recruitment Trends (London : PA Associates).

PORTER, L.W. and STEERS, R.M. (1973): Organizational, work and personal factors in employee turnover and absenteeism. Psychological Bulletin 70:161-176.

PREMACK, S. and WANOUS, J. (1985) : A meta analysis of Realistic Job Preview Experiments. Journal of Applied Psychology 70 (4) : 706-719.

PRESS, S.J. and WILSON, S. (1978) : Choosing between logistic regression and discriminant analysis. Journal of the American Statistical Association 73 (364): 699-705.

REILLY, R.R. and CHAO, G.T.(1982) : Validity and fairness of some alternative employee selection procedures. Personnel Psychology 35: 1-61.

RITCHIE R.J. and BOEHM, V.R.(1977) : Biographical data as a predictor of men's and women's management potential. Journal of Vocational Behaviour 11: 363-368.

RHODE, J.G., SORENSEN, J.E and LAWLER, E. (1977) : Sources of professional turnover in public accounting firms revealed by the exit interview. Accounting, Organizations and Society 2 (2): 165-175.

ROBERTSON, I., GRATTON, L. and SHARPLEY, D. (1987): The psychometric properties and design of managerial assessment centres: Dimensions into exercises wont go. Journal of Occupational Psychology 60: 187-96.

ROBERTSON, I.T. and KANDOLA, R.S. (1982) : Work sample testing: validity, adverse impact and applicant reaction. Journal of Occupational Psychology 55: 171-183.

ROBERTSON, I.T. and MAKIN, P.A. (1986): Management selection in Britain : A survey and critique. Personnel Psychology 59 (1): 45-57.

- ROTHSTEIN, H.R., SCHMITT, F.L., ERWIN, F.W., OWENS, W.A and SPARK, C.P. (1990): Biographical data in employment selection: Can validities be made generalizable? Journal of Applied Psychology 75 (2): 175-184.
- RUSHTON, J.P. (1990): Race differences *r/k* theory and a reply to Flynn. The Psychologist 3 (5): 195-198.
- RYNES, S.L. and BOUDREAU, J.W. (1988) : College recruiting in large organizations : practice, evaluation and research implications. Personnel Psychology 39 : 729-752.
- RYNES, S.L. and GERHARDT B. (1990) : Interviewer assessments of applicant "fit" : An exploratory investigation. Personnel Psychology 43 (2): 13-35.
- RYNES, S.L. and MILLER, H.E. (1983) : Recruiter and job influences of candidates for employment. Journal of Applied Psychology 68 (1): 147-154.
- SACKETT, P.R., TENOPYR, M., SCHMIDT, N. and KEHOE, J. (1985): Commentary on forty questions about validity generalization and meta-analysis. Personnel Psychology 29: 79-101.
- SAKS, A.M. and CRONSHAW, S.F. (1990) : A process investigation of realistic job previews: mediating variables and channels of communication. Journal of Organizational Behaviour 11 : 221-236.
- SCHMITT, N. (1976) : Social and situational determinants of interview decisions : Implications for the employment interview. Personnel Psychology 29: 77-101.
- SCHMITT, N., GOODING, R.Z., NOE, R.A. and KIRSCH, M. (1984) : Meta analysis of validity studies published between 1964 and 1982 and the investigation of study characteristics. Personnel Psychology 37: 407-422.
- SCHMITT, N. and OSTROFF, C. (1986): Operationalizing the behavioural consistency approach: selection test development based on content-orientated strategy. Personnel Psychology 39: 91-108.
- SCHRADER, A.D. and OSBURN, H.G. (1977): Biodata faking: Effects of induced subtlety and position specificity. Personnel Psychology 30: 395-404.
- SCHWAB, D.P. and OLIVER, R.L. (1974) : Predicting tenure with Biographical Data: exhuming buried evidence. Personnel Psychology 27 : 125-128.
- SHAFFER, G.S., SAUNDERS, V. and OWENS, W.A. (1986): Additional evidence for the accuracy of biographical data: Long-term retest and observer ratings. Personnel Psychology 39 : 791-809.
- SHUH, A.J. (1967): The predicability of employee turnover: A review of the literature. Personnel Psychology 20: 133-152.

SHUH, A.J. (1978): Contrast effect in the interview. Bulletin of the Psychometric Society 11: 195-196.

SHULMAN, L. (1986): Research program to develop procedures for the assessment of teachers. Orientation Document: School for Education, Stanford University, California.

SIEGEL, S. (1956) : Non-parametric Statistics for the Behavioral Sciences (Kogakusha : McGraw Hill).

SMITH, M. (1994) : A theory of the validity of predictors in personnel selection. Journal of Occupational Psychology 67 : 13-31

SMITH, M. and ABRAHAMSEN, M. (1992): Patterns of selection in six countries. The Psychologist 5 (5):205-207.

SMITH, P. C. and KENDELL, L.M. (1963) : Retranslation of expectations : An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology 47: 149-155.

SMITHER, J.W., REILLY, R.R, MILLSAP, R.E., PEARLMAN, K. and STOFFEY, R.W. (1993): Applicant reactions to selection interviews. Personnel Psychology 46: 49-76.

STEAD, W.G. and SHARTLE, C.L. (1940): Occupational Counselling Techniques (New York : American Book Co).

STOKES, G.S., MUMFORD, M.D. and OWENS, W. (1989): Life history prototypes in the study of human individuality. Journal of Personality 57: 509-549.

STOKES, G.S., REDDY, S. (1992): Use of background data in organizational decisions. International Journal of Industrial and Organizational Psychology, 7: 285-321.

STREBLER, M. (1990) : Survey of Graduate Specialists and Clerical Staff (unpublished).

STREBLER, M. (1991): Biographical Data: Issues In Practice (London: IMS)

STRONG, E.K. (1926): An interest test for personnel managers. Journal of Personnel Research, 5 : 194-203.

SUPER, D.E. (1980) : A life-span life-space approach to career development. Journal of Vocational Behaviour 16: 282-298.

SWINBOURNE, P. (1986): Personality questionnaires as a source of learning in management education and training : A comparison of the OPQ and the 16PF. Personnel Review 15 (1): 32-35.

THEYER, P.W. (1977): Somethings old, somethings new. Personnel Psychology 30: 513-524.

TURNAGE, J.J. and MUCHINSKY, P.M. (1984) : A comparison of the predictive validity of assessment centre evaluations versus traditional measures of forecasting supervisors job performance. Journal of Applied Psychology 69 (4) : 595-602.

TVERSKY, A. and KAHNEMANN, D. (1974) : Judgement under uncertainty: heuristics and biases. Science, 185: 1124-1131.

UNIVERSITY STATISTICS RECORD (1992) (Oxford: University Statistics Record, on behalf of the University Grants Committee.

WANOUS, J.P. (1977) : Occupational entry : newcomers moving in from the outside. Psychological Bulletin, 84: 601-618.

WEISS, D.J. (1976) : Multivariate Procedures. In: Handbook of Industrial and Organizational Psychology, ed Dunnette. M.V. (Chicago: Rand McNally)

WERNIMONT, P.F. and CAMPBELL, J.P. (1968): Signs, samples and criteria. Journal of Applied Psychology 52: 372-376.

WIESNER, W. and CRONSHAW, S. (1988): A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. Journal of Occupational Psychology 61: 275-90.

WILLIAMS, A.P.O. (1984): The neglected process of self-selection. In: Annual Conference of the Military Testing Association, 26th, Munich, 1984. Proceedings.

WILSON, R. (1989) : How great expectations founder at the end of the recruitment trail. Accountancy Age (10), 18-19.

WINGROVE, J., GLENDINNING, R. and HERRIOT, P. (1984) : Graduate pre-selection : A research note. Journal of Occupational Psychology 57: 169-171.

WINTER, B.C., BLYTH, D.M. and HERRIOT, P. (1987): You're in the army now? British Psychological Society, Occupational Psychology Conference, January 1987. Proceedings.

WRIGHT, P.M., LICHTENFELS, P.A. and PURSELL, E.D. (1989): The structured interview: Additional studies and meta analysis. Journal of Applied Psychology 62 (3): 191-199.

Appendix

Appendix A : Original Predictor Data Items Collected

Academic Items

Number of written languages

Number of spoken languages

Typing Skills

Information Technology Skills :

6 possible levels of competence : 0,1,2 below average/ 3,4,5 average and above

Number and type of secondary education establishments

Type of 'A' level education (school VIth form, tertiary college, Further Education college)

Ordinary GCE 'O' Levels / GCSE Examinations

Examining Board

Number of passes

Number of Grade A passes

Disciplines of passes :

(Mathematics/science/technology, art/languages, social sciences and humanities)

Number of subjects retaken

Number of subjects failed

Further Education

Number of 'A' level passes (or Scottish Highers/ Irish Leaving Certificate)

Number of retaken subjects

Number of failed subjects

UCCA Points (and average calculated)

Disciplines of subjects (as for 'O' level/GCSE)

Academic prizes gained

Athletic prizes gained

Higher Education

Type of higher education (university [Oxford/Cambridge], polytechnic, other)

Related Degree (offering exemption from Foundation Examination)

Single or combined subject degree

Yearly undergraduate examination results

Type of degree

Title of degree

Expected degree class

Actual degree class

Percent exempt from GCC examination

Number of theses / projects related to chartered accountancy

Post-graduate qualifications

Post-graduate discipline

Any other qualifications obtained :

eg music grades, life-saving and personal survival awards, Duke of Edinburgh's Award, First-Aid certificates.

Employment Items

Number and type of jobs while at school

Number of above related to chartered accountancy

Number of jobs while at school in which responsibility taken for supervision of others

Number and type of jobs while at university (including number of placements for sandwich student)

Number of above related to chartered accountancy

Number of jobs while at university in which responsibility taken for supervision of others

Home Life Items

Area of family residence:

North East, North West, Midlands, South West, South East, Wales, East Anglia, London, Scotland, Eire/Ulster, Overseas.

Length of time at that residence :

0-5 yrs, 6-10 yrs, 11-15 yrs, 16-20 yrs, 20+ yrs

Distance from the nearest large town :

0-10 miles, 11-20 miles, over 20 miles

Number and sex of siblings

Position in family

Father's occupation

Mother's occupation

Personal Items

Sex

Date of Birth

Nationality

Preferred location of training office

Marital status

Children (1/0)

Driving licence (1/0)

Car owner (1/0)

General health

Serious illness

From which source received introduction to firm :

Personal Contact, Literature, Careers Service, Other

Social Involvement

Number of voluntary activities while at school/ at university

Responsibilities at school

Social involvement at school:

Number of teams, clubs, societies (including Scouts/ Guides) and positions of responsibility

Social Involvement at university (campus based) :

Number of teams, clubs, societies and positions of responsibility

Number of large, small and solo activities : active and non-active

Social Involvement at university (off-campus activities) :

Number of teams, clubs, societies positions of responsibility

Number of large, small and solo activities : active and non-active

Other interests or hobbies

Visits abroad in the last three years

By whom accompanied on above

Appendix B : Variables Used in Logistic Regression

Predictor Variables

Academic Items

Number of written languages

Information technology skills : 5 point competence scale

Number and Type of Secondary School establishments attended :

State

Independent

Type of educational establishment where 'A' levels taken :

Secondary School FE/ Tertiary College

Number of 'O' level GCE' passes

Number of grade A 'O' level GCE passes

Number of 'A' levels passes or equivalent

Number of 'A' level arts subjects (including languages)

Number of 'A' level science/ technology subjects

Number of 'A' level humanities/ social science subjects

Number of UCCA points

Average UCCA points

Degree Discipline :

Arts and Languages

Mathematics, Science and Technology

Humanities and Social Sciences

Combined subjects

Mathematics/ Science/ Technology degree (Binary)

Degree class I/IIi (Binary)

Percentage exemption form GCC

Employment Items

Number of jobs while at school

Number of jobs while at university

Number of jobs related to chartered accountancy

Social Items

Number of school teams

Number of responsibilities for above

Number of societies/clubs at school (in or out of school)

Number of responsibilities for above

Number of teams at university

Whether any responsibility held for above

Number of large, small and solo active pastimes at university

Number of large, small and solo non-active pastimes at university

Criterion Variables

Professional Examination progress

Foundation Examination : 3 attempts - Pass / Refer / Fail

Professional Examination I : 3 attempts - Pass / Refer / Fail

Professional Examination II : 3 attempts - Pass / Refer / Fail

Withdrawal Record

Withdrawal at any stage prior to or between the above examinations (9 possible stages) for any of the following reasons :

Firm's request, mutual agreement (counselled out), own volition, transfer.

Summary of practice work record

Aggregate supervisors' ratings at six months and 12 months for:

Technical Competence

Relationships with clients and colleagues

Communication Skills

Commercial sense and financial awareness

Judgement

Initiative

Administration

Independence of mind

Personal qualities

Professional Approach

Appendix C : Model 1 : Examination Performance Prediction

WAB Variables, Group Frequencies and Net and Assigned Weights

**Denotes significant variable*

Item	Category	% Pass	% Fail	Diff in %	Net Weight	Ass'd Weight
Sex	Female	37	37	0	0	1
	Male	63	63	0	0	1
Area of Resid- ence	NE/NW	15	15	0	0	1
	Midl/Ang	19	23	-4	-1	1
	SW/Wales	9	14	-6	-1	1
	London/SE	51	38	13	3	1
	RI/Ui/Scot	3	6	-3	-1	1
	Overseas	1	3	-2	0	1
Driving Licence	Holder	19	30	-11	-2	1
	None	81	70	11	2	1
Source of Firm Inform- ation	Contact	11	17	-6	-1	1
	Brochure	18	31	-13	-3	1
	Careers	56	48	-8	-2	1
	Other	15	4	9	2	1
Voluntary work (S)	No	90	83	-7	-1	1
	Yes	10	17	7	1	1
# Jobs while at school	0	33	27	6	2	1
	1	52	48	4	1	1
	> 1	15	24	-9	-2	1
# Jobs at University	<2	31	10	21	5	2*
	> =2	69	91	-22	-5	0*
Univ. accy experience	No	62	85	-23	-5	0*
	Yes	38	15	23	5	2*
Voluntary Work (U)	No	86	77	9	2	1
	Yes	14	23	-9	-2	1

Item	Category	% Pass	% Fail	Diff in %	Net Weight	Ass'd Weight
School	Prefect	48	48	0	0	1
Respons- ibility	Monitor	33	39	-6	1	1
	Headb/girl	10	9	1	0	1
	Other	8	3	5	2	1
Societies	0	52	48	4	1	1
/clubs at	1	33	39	-6	-1	1
school	> 1	15	13	2	0	1
School	0	49	52	-3	-1	1
Sports	1	33	39	-6	-1	1
Teams	> 1	15	13	2	0	1
School	No respons.	32	37	-5	-1	1
club/team	Responsibility	69	63	6	1	1
Sports	0	48	42	6	1	1
teams at	1	38	50	-12	-3	1
University	> 1	15	8	7	1	1
Responsib- ility	No	78	78	0	0	1
	Yes	22	22	0	0	1
University	0	31	45	-14	-3	1
club/	1	45	36	9	2	1
society	> 1	24	19	6	1	1
Responsib- ility	No	67	68	29	7	2
	Yes	33	62	-2	-7	0*
# Large gp	0	70	75	-5	-1	1
active	1	21	12	9	2	1
pastimes	> 1	9	13	-4	-1	1
# Small gp	0	43	27	6	4	2*
active	> 0	57	73	-16	-4	0*
pastimes						

Item	Category	% Pass	% Fail	Diff in %	Net Weight	Ass'd Weight
# Solo active pastimes	0 >0	61 39	52 48	9 -9	2 2	1 1
# Large gp non- actv pastimes	0 >0	94 6	91 9	3 -3	1 -1	1 1
# Small Gp non-actv pastimes	<2 >1	66 33	87 13	-21 20	-5 4	0* 2*
# Solo non-active pastimes	0 1-2 >2	32 51 17	42 39 19	-10 12 -2	-2 3 0	1 1 1
Spoken Languages	0 1 >1	43 35 22	42 27 30	1 8 -8	0 2 -2	1 1 1
Inform Technology Skill	<Ave >=Ave	51 49	72 28	-21 21	-5 5	0* 2*
Independent School	State Private	38 62	48 52	-10 10	-2 2	1 1
Number of Schools	1 >1	72 28	49 51	23 -23	5 -5	2* 0*
# 'O' levels	<=7 >7	9 91	28 72	-19 19	-4 4	0* 2*

Item	Categ	% Pass	% Fail	Diff %	Net Weight	Ass'd Weight
# Grade A	<4	56	85	-29	-7	0*
'O'levels	>3	44	15	29	7	2*
# 'O'level	<4	39	59	-20	-4	0*
Math/Sci	>3	61	41	20	4	2*
# 'O'social	0	71	69	2	0	1
Science	>0	29	31	-2	0	1
# 'O'	<3	21	27	-6	-1	1
Art/Lang	>2	79	73	6	1	1
# 'O'	0	25	29	-4	-2	1
Human-	1-2	67	66	1	0	1
ities	>2	9	4	5	2	1
# 'A'	<2	38	67	-29	-7	0*
Sci/Tech	>1	62	33	29	7	2*
# 'A'	0	66	61	5	1	1
SocSci	>0	34	39	-5	1	1
# 'A'	0	65	48	17	4	2*
Art/Lang	>0	35	52	-17	-4	0*
# 'A'	0	75	70	5	1	1
Humans	>0	25	30	-5	-1	1
UCCA	<9	27	44	-17	-4	0*
Points	>8	73	56	17	4	2*

Item	Category	% Pass	% Fail	Diff in %	Net Weight	Ass'd Weight
Degree	Science/Tech	46	21	25	6	2*
Subject	Other	54	79	-25	-6	0*
Degree	I/II.1	52	43	9	2	1
Class	<II.1	48	57	-9	-2	1
%	0	69	58	11	2	1
exempt	25	21	34	-13	3	1
from GCC	50	8	8	0	0	1
	75	2	0	2	2	1

Appendix D : Model 2 Recruitment Success Prediction

WAB Variables, Group Frequencies and Net and Assigned Weights

**Denotes significant variable*

Item	Category	% Success	% Leave	Diff in %	Net Weight	Ass'd Weight
Sex	Female	43	33	10	2	1
	Male	57	67	-10	-2	1
Area of Resid	Lond/SE	56	31	25	6	2*
	Other	44	69	-25	-6	0*
Driving Licence	Holder	18	24	-6	-1	1
	None	82	76	6	1	1
Source of Firm Info	Contact	10	14	-4	-1	1
	Brochure	12	20	-8	-2	1
	Careers	55	54	1	0	1
	Other	23	11	12	3	1
Voluntary work (S)	No	79	80	-1	0	0
	Yes	21	20	1	0	0
# Jobs at School	<2	85	68	17	4	2*
	>1	15	32	-17	-4	0*
# Jobs at Univ	<2	29	28	1	0	1
	>=2	70	72	-2	0	1
Accy exper (U)	No	85	85	0	0	1
	Yes	15	15	0	0	1
Voluntary wok (U)	No	77	86	9	-2	1
	Yes	23	14	9	2	1

Item	Category	% Success	% Leave	Diff in %	Net Weight	Ass'd Weight
School	Prefect	45	50	-5	-1	1*
Respons- ibility	Monitor	36	48	-12	-3	1*
	Headb/g	9	2	7	4	2*
	Other	9	2	7	4	2*
School	0	44	51		-7	-11
clubs/ society	1	33	37	-4	-1	1
	> 1	23	12	11	2	1
School	0	48	59	-11	-2	1*
sports	1	35	35	0	0	1*
teams	> 1	17	6	11	4	2*
Univers	0	50	35	15	3	1
sports	1	33	53	-20	-4	0
teams	> 1	17	12	5	1	1
Team (U)	No	78	73	5	1	1
Respon	Yes	22	27	-5	-1	1
Club/	0	33	40	-13	-3	1
society at university	1	36	42	-7	-1	1
	> 1	31	19	12	3	1
Club /soc	No	65	68	-3	-1	1
resp (U)	Yes	35	32	3	1	1
# Large	0	67	77	-10	-2	1
gp activ- ities (u)	1	20	17	3	1	1
	> 1	13	6	7	1	1

Item	Category	% Success	% Leave	Diff in %	Net Weight	Ass'd Weight
# Small gp	0	35	40	-5	-1	1
Activities	>0	65	60	5	-1	1
@ Univ						
# Solo	0	59	46	13	3	1
Activities	>0	41	54	-13	-3	1
@ Univ						
# Large	0	89	89	0	0	1
gp non-	>0	11	11	0	0	1
active (U)						
# Small gp	<2	47	50	-3	-1	1
non-	>1	53	50	3	1	1
active (U)						
# Solo non	0	35	40	-5	-1	1
-active	1-2	33	23	10	2	1
pastimes	>2	32	37	-5	-1	1
(U)						
Spoken	0	35	41	-6	-1	1*
Langs	1	44	28	16	4	2*
	>1	20	31	-11	-2	1*
Info	<Ave	45	66	-21	-5	0*
Technol	>=Ave	55	34	21	5	2*
Skill						
Independ-	State	44	42	-2	0	1
ent School	Private	56	58	2	0	1

Item	Category	% Success	% Leave	Diff in %	Net Weight	Ass'd Weight
Number	1	76	61	15	3	1
Schools	>1	24	39	-1	-3	1
# Grade A	<4	59	62	-3	-1	1
'O' level	>3	41	38	3	1	1
# 'O'	<4	33	53	-20	-4	0*
Math/Sci	>3	67	47	20	4	2*
# 'O'	0	66	63	3	1	1
SocSci	>0	34	37	-3	-1	1
# 'O'	<3	15	20	-5	-1	1
Art/Lang	>2	85	80	5	1	1
# 'O'	0	15	33	-18	-4	0*
Human- ities	>0	85	67	18	4	2*
# 'A'	<2	32	74	-42	-11	0*
Sci/Tech	>1	68	36	42	11	2*
# 'A'	0	60	58	2	0	1
SocSci	>0	40	42	-2	0	1
# 'A'	0	73	56	17	4	2*
Art/Lang	>0	27	44	-17	-4	0*
# 'A'	0	76	66	10	2	1
Humans	>0	24	34	-10	2	1
UCCA	<=9	30	29	1	0	1
Points	>9	70	71	-1	0	1

Item	Category	% Success	% Leave	Diff in %	Net Weight	Ass'd Weight
Degree	Sci/Tech	51	22	29	7	2*
Subject	Other	49	78	-29	-7	0*
Degree	I/II.1	69	53	16	4	2*
Class	<II.1	31	47	-16	-4	0*
% Exempt	0	69	58	11	2	1
from	25	21	34	-13	3	1
GCC	50	8	8	0	0	1
	75	2	0	2	2	1

Appendix E : Firms Participating in Part I

Blick, Rothenberg and Noble

Brebner, Allen and Trap

Clarke Whitehill

Cohen Arnold

Dixon Wilson

Finnie and Co.

Fraser Russell

Hacker Young

Hays Allen

Hodgson Impey (London)

Hodgson Impey (Hull)

Kidsons

Larking Gowan

Littlejohn Frazer

Longcrofts

Moore Stephens

Moores and Rowlands

Morrison Stoneham

Neville Russell

Reeves and Neylan

Saffery Champness

Stoy Hayward

Appendix F : Software Output Examples

City University Business School
Selection Models for Graduate Applicants

Undergraduate Applicant

Date :29/09/94

Name : Jane Smith

Ref No: 91232

Message: Late Entrant

Mature Entrant:	Y
Relevant Graduate:	N
Grade A 'O' levels:	5
'A' level Arts :	1
'A' level Science :	2
Headboy or girl :	N
Independent school :	Y
Science/Mathematics degree :	Y
Degree Class I/II_i :	N/A
Number of school teams/clubs/societies :	2
Number of SOLO active pastimes at university :	1
Percentage exemption from Foundation :	25

	If I/II_i	If < II_i
Examination Model Score :	92%	77%
Recruitment Model Score :	73%	53%

City University Business School
Selection Models for Graduate Applicants

Undergraduate Applicant **Date :29/09/94**

Name : Andrew Jones **Ref No:** 91233

Message:

Mature Entrant:	Y
Relevant Graduate:	Y
Grade A 'O' levels:	2
'A' level Arts :	2
'A' level Science :	0
Headboy or girl :	Y
Independent school :	N
Science/Mathematics degree :	N
Degree Class I/II_i:	N/A
Number of school teams/clubs/societies :	3
Number of SOLO active pastimes at university :	2
Percentage exemption from Foundation :	100

	If I/II_i	If < II_i
Examination Model Score :	69%	40%

City University Business School
Selection Models for Graduate Applicants

Graduate Applicant

Date :29/09/94

Name : Justin Case

Ref No: 91234

Message:

Mature Entrant:	N
Relevant Graduate:	N
Grade A 'O' levels:	5
'A' level Arts :	2
'A' level Science :	1
Headboy or girl :	N
Independent school :	Y
Science/Mathematics degree :	N
Degree Class I/II₁ :	Y
Number of school teams/clubs/societies :	4
Number of SOLO active pastimes at university :	0
Percentage exemption from Foundation :	33

Examination Model Score :	70%
----------------------------------	------------

Recruitment Model Score :	81%
----------------------------------	------------

City University Business School
Selection Models for Graduate Applicants

Graduate Applicant

Date :29/09/94

Name : Arthur Meaux

Ref No: 91235

Message: Father is a partner

Mature Entrant:	N
Relevant Graduate:	Y
Grade A 'O' levels:	0
'A' level Arts :	3
'A' level Science :	0
Headboy or girl :	N
Independent school :	N
Science/Mathematics degree :	N
Degree Class I/II₁ :	N
Number of school teams/clubs/societies :	0
Number of SOLO active pastimes at university :	3
Percentage exemption from Foundation :	100

Examination Model Score :	2%
----------------------------------	-----------