



City Research Online

City, University of London Institutional Repository

Citation: Aleo, G., Bagnasco, A., Watson, R., Dyson, J., Cowdell, F., Catania, G., Zanini, M. P., Cozzani, E., Parodi, A. and Sasso, L. (2019). Comparing questionnaires across cultures: Using Mokken scaling to compare the Italian and English versions of the MOLES index. *Nursing Open*, doi: 10.1002/nop2.297

This is the published 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/22539/>

Link to published version: <http://dx.doi.org/10.1002/nop2.297>

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.


City Research Online:

<http://openaccess.city.ac.uk/>

publications@city.ac.uk

RESEARCH ARTICLE

Comparing questionnaires across cultures: Using Mokken scaling to compare the Italian and English versions of the MOLES index

Giuseppe Aleo¹ | Annamaria Bagnasco¹  | Roger Watson²  | Judith Dyson² |
 Fiona Cowdell³ | Gianluca Catania¹ | Milko Patrick Zanini¹  | Emanuele Cozzani¹ |
 Aurora Parodi¹ | Loredana Sasso¹

¹Department of Health Sciences, University of Genoa, Genoa, Italy

²Faculty of Health Sciences, University of Hull, Hull, UK

³School of Nursing and Midwifery, University of Birmingham, Birmingham, UK

Correspondence

Roger Watson, Faculty of Health and Social Care, University of Hull, Hull, HU6 7RX, UK.
 Email: R.Watson@hull.ac.uk

Abstract

Aim: The aims of this study were (a) to translate the MOLES index from English to Italian and (ii) to compare the two versions using non-parametric item response theory.

Design: An online survey was used to gather data.

Methods: Forward and back translation was used to prepare the Italian version of the MOLES which was then analysed using the non-parametric item response theory of Mokken scaling.

Results: Mokken scales were found in both the English and the Italian versions of the MOLES index. However, the two scales—while the total scale score was not significantly different—showed different properties, and Mokken scaling selected different items from each scale.

KEYWORDS

cancer, melanoma, Mokken scaling, self-examination

1 | INTRODUCTION

A range of methods exists to study the dimensional properties of questionnaires. Such dimensions are known as “latent” traits as, essentially, they are hidden within the items of questionnaires and may not be obvious without specific multivariate analysis or, when they are purported to exist, require specific multivariate analysis to demonstrate this. A simple example of a commonly used questionnaire that has demonstrable dimensions is the Hospital Anxiety and Depression Scale (HADS). The HADS is comprised of 14 items and seven of these purportedly measure depression as distinct from the seven items that purportedly measure anxiety. Indeed, the two-dimensional nature of the HADS can be demonstrated by

appropriate multivariate analysis which in the case of the HADS is factor analysis.

A range of multivariate techniques exists to study the dimensional properties of questionnaires, and these fall under two broad umbrellas: classical test theory (CTT) and item response theory (IRT) and these will, briefly, be considered. Classical test theory is, essentially, based on correlation—a measure of the common variance between two or more variables. Therefore, multivariate statistical techniques such as Cronbach's alpha, principal component analysis and factor analysis—both exploratory and confirmatory—fall under this umbrella. Factor analysis, of which there is a range of similar methods, is the method mainly used to establish dimension in questionnaires and it can be used to examine

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2019 The Authors. *Nursing Open* published by John Wiley & Sons Ltd.

whether or not there are underlying dimensions to questionnaires (exploratory factor analysis) or to test whether or not an hypothesized set of dimensions exists in a questionnaires (confirmatory factor analysis).

An alternative set of methods exists to study the dimensional nature of questionnaires, and these fall under the umbrella of IRT. These methods are so called because, rather than analysing the relationship between items, they primarily analyse the behaviour of individual items and, based on their properties, they then investigate how they relate to other items. However, individual items must meet certain minimum criteria—to be discussed—to be included in a questionnaire. IRT can be seen to offer some advantages over CTT in that they establish a more precise relationship between the score on an item and the score on the latent trait. In other words, while items will respond across the whole range of a latent trait, they will most accurately measure a region of the latent trait. For example, take two items purporting to measure depression: 1. “I do not feel it is worth getting out of bed in the morning” and 2. “I feel like ending my life”. Clearly, both are related to depression but item 1 measures a much lower range of the latent trait of depression than item 2 which represents a more serious level of danger to the individual. IRT posits that the relationship between the score on an item and the score on the latent trait is stochastic, in other words based on probability and, in the case of the items above, there is a much higher probability that someone will score high on item 1 before they score on item 2. This indicates another aspect of IRT which follows this assumption, and that is that items are ordered along the latent trait. IRT thereby becomes useful as we should, in theory, be able to tell how far along the latent trait and individual lies by only knowing the score on a single item. CTT is insensitive to the relationship between items and the latent trait.

Item response theory describes two basic methods: parametric and non-parametric, and these are represented by Rasch analysis and Mokken scaling analysis (MSA), respectively. The difference between the methods is that parametric methods predict and, therefore, depend on a specific relationship between the score on an item and the score on the latent trait and non-parametric methods do not. The relationship between the score on an item and the score on a latent trait is represented by the item characteristic curve where the x-axis represents score on the item and the y-axis represents the probability of obtaining that score. In both methods, the ICC must be monotonously homogenous—in other words as the score on the trait increases, so does the score on the latent trait. However, the ICC in parametric IRT has a sigmoidal shape and in non-parametric IRT—provided the criterion of monotone homogeneity is met—it can assume any shape. Clearly, the two methods have different analytical features but the virtue of non-parametric IRT, represented by MSA, is that it is less conservative and tends to retain more items in an analysis. The resulting scales have high clinical utility but lack the precision of scales obtained using Rasch analysis, which is more suitable to the analysis, for example, of educational tests where greater precision is required.

1.1 | Mokken scaling

As explained above, Mokken scaling analyses the properties of individual items as described by the item characteristic curve (ICCs), which relates the score on an item to the level of the latent trait being measured. It makes no assumptions about the precise nature of that relationship requiring only that ICCs are monotonously homogeneous (they continuously increase across the range of the latent trait) and that they do not intersect (i.e., they are doubly monotonous; Mokken & Lewis, 1982). Mokken scaling assumes that the response of items to the level of the latent trait is locally stochastically independent, in other words, that the score on an item is purely a result of the level if the latent trait present and not to a score on any of the other items. Therefore, the score on one item is not dependent on the score on any other items. As stated, this is usually an assumption and is not formally tested in Mokken scaling and, currently, methods for assessment of local stochastic independence are still under development. However, inspection of items in terms of their wording can usually confirm that items are not stochastically dependent. IRT does not assume that all items have an equal level of difficulty—an assumption that is not held by classical test theory methods such as factor analysis (Mokken & Lewis, 1982). “Difficulty” means the extent to which items are endorsed by respondents with more extreme items at the upper end of the range of the latent trait being the more difficult. For example, in a scale measuring psychological morbidity, an item labelled “I want to end my life” would be more difficult than an item labelled “I don't feel like getting out of bed”. Therefore, items are arranged along the latent trait in terms of their difficulty and the properties of items can be measured using a scalability coefficient H (Loevinger's coefficient) which measures the extent to which all items are arranged as expected by their mean values along the latent trait. A Loevinger's coefficient > 0.3 is the minimum acceptable value of H indicating a weak scale; $H > 0.4$ indicates a moderate scale; and $H > 0.5$ indicates a strong scale. Items can also be analysed for violations of monotone homogeneity, and the reliability of sets of items purporting to form Mokken scales can be calculated and expressed in a reliability coefficient Rho . The coefficient Rho is preferred in Mokken scaling due to some well-known problems with Cronbach's alpha. Admittedly, Cronbach's alpha is commonly used to assess reliability in scales but it is not independent of the number of items in the scale (Agbo, 2010) and may not be accurate for relatively small numbers of respondents (Sijtsma, 2009). Rho —also known as the Molenaar Sijtsma statistic—was especially developed for use in Mokken scaling (van der Ark, Straat, & Koopan, 2018). Finally, a desirable although not essential feature of a Mokken scale is invariant item ordering (IIO) whereby the order of items along the latent trait is the same for all respondents at all levels of the latent trait. This is investigated primarily by plotting ICCs and inspecting for non-intersection—which clearly violates IIO—and then by investigating IIO mathematically to look for significant violations and then calculating the accuracy of IIO as expressed in a coefficient H_{trans} or H^T . Values of H and H^T exceeding 0.3 indicate acceptably strong scales and acceptable accuracy of IIO, respectively. For both coefficients,

values exceeding 0.4 indicate moderate levels and values exceeding 0.5 indicate high levels of strength and accuracy (Mokken & Lewis, 1982; Watson et al., 2012).

2 | BACKGROUND

2.1 | The MOLES index

The MOLES index is an instrument designed to test the motivation of individuals to self-examine their skin for lesions which may indicate that they have skin cancer (Cowdell & Dyson, 2014). The MOLES index is comprised of 20 items and was developed from the perspective of the Theoretical Domains Framework which is designed to make behaviour change accessible to health practitioners other than psychologists.

The MOLES index was developed, as described by Dyson and Cowdell (2014) through a combination of literature review, qualitative work and psychometric testing. As such, the MOLES index resulted from a three-stage process with a sample of members of the public and involving: (a) identifying items from the barriers to SSE identified in the literature and through a survey of members of the general population ($N = 261$); (b) categorization of barriers to theoretical framework by experts in the fields of dermatology and psychology ($N = 11$); and (c) validity and reliability testing (face validity, internal consistency, factor analysis and test-retest reliability) ($N = 314$).

Examples of items in the MOLES index include the following: "I believe examining my skin leads to better health"; "If I examine my skin I may prevent cancer"; and "I am able to make checking my skin a regular routine". Four items are negatively worded, for example: "Remembering to check my skin is difficult" and "I cannot be bothered with skin self-examination". The items are scored on a 7-point Likert type scale running from "Strongly agree" to "Strongly disagree". Therefore, higher scores indicate lower endorsement of skin self-examination and the negatively worded items are reverse scored before using the total score on the scale and before the analysis conducted in this study.

The result of the initial psychometric analysis of the MOLES (Dyson & Cowdell, 2014) was a five-factor structure: (a) Outcome expectancies; (b) Intention; (c) Self-efficacy; (d) Social influences; and (e) Memory, 20-item instrument which tested well for reliability and construct validity.

The value of this theoretically based instrument is the ease with which behaviour change techniques can be mapped (Michie et al., 2013) to the five factors (behavioural determinants) allowing theory-based pragmatic and tailored interventions to be developed to support SSE (Cowdell & Dyson, 2014). In this paper, we build on the existing MOLES index in two ways: We translated the MOLES index into another language (Italian), and we analysed the MOLES index (English and Italian versions) exploratory using Mokken scaling. We suspected that the items in the MOLES index may be suitable to MSA because they were likely to form a hierarchy. For example, some of the questions require only a belief (e.g., "I believe examining

my skin leads to better health"), whereas some require knowledge (e.g., "I could explain the correct method for skin self-examination") and some require commitment (e.g., "I am able to make checking my skin a regular routine"). Therefore, it is possible that people endorse beliefs (which are relatively easy and require no action) before they endorse knowledge and actions and, indeed, that belief and knowledge are prerequisites to action.

2.2 | Research question

How well can Mokken scaling be used to compare to version of the same scale (the MOLES) in two languages (English and Italian) and how do these versions compare when analysed using Mokken scaling?

3 | METHODS

3.1 | The translation process of the MOLES index

The developers of the MOLES were part of the present team. One member of the team is bilingual, and local Italian experts were on hand to assist. Two expert native Italian translators separately conducted the English-Italian forward translation. The two Italian versions were compared, and the differences between the two versions were resolved following a discussion by the research team. The resulting Italian version was then back-translated into English by a third expert bilingual English-Italian translator. The differences between the original English version and the English translated version of the MOLES index were discussed and resolved directly with the original authors.

3.2 | Face validity of the Italian version of the MOLES index

In September 2015, the final draft of the Italian version of the MOLES index was piloted with 30 2nd- and 3rd-year nursing students to check face validity and language clarity. All the students easily understood the questionnaire, and no further amendment was required.

3.3 | Data collection

Italian data were collected in October 2016, after presenting the study and illustrating the MOLES index to all the 1st-year nursing students, during a general assembly on their first day at a university in the north of Italy. The students were given the URL to an online version of the MOLES index and invited to complete the questionnaire by the end of October and to encourage their family members' friends to do the same thing. The questionnaire was anonymous, and its completion was voluntary. By accepting to complete the questionnaire, respondents automatically expressed their consent to take part in the study. Privacy was ensured, and data were handled exclusively for use in this study. The UK data were collected in 2014

and 2016, and ethical permission obtained as previously described (Dyson & Cowdell, 2014).

3.4 | Analysis

Package “mokken” (<https://cran.r-project.org/web/packages/mokken/mokken.pdf> last accessed 20 May 2017) from the online public domain statistical software *R* (<https://www.r-project.org/> last accessed

12 April 2016) was used to analyse the data. Data were entered into *R* by converting from SPSS files into .Rdata files using package “foreign” in *R* and then analysed in the following sequence: the automated item selection procedure “aisp” was used, with default settings, to investigate how many putative scales were present in the data; the resulting scales were then analysed to see whether the items were likely to form a Mokken scale using “coefH” to establish the scalability of items, item pairs and the total scales; items were

TABLE 1 Mokken scaling of Italian and UK MOLES data

Item	Descriptor	Mean item scores [H_i (SE)]			
		Italy (N = 619)	UK (N = 460)		
1.	I believe examining my skin leads to better health	2.06^a	[0.46 (0.023)]	4.40 ^a	[0.37 (0.031)]
2.	I could describe the moles and marks on my skin	3.02		3.85 ^a	[0.40 (0.028)]
3.	My doctor/nurse encourages me to self-examine my skin regularly	3.61		5.37 ^a	[0.31 (0.033)]
4.	If I examine my skin I may prevent cancer	2.36^a	[0.47 (0.026)]	2.38	
5.	Remembering to check my skin is difficult	4.23		5.25 ^a	[0.42 (0.030)]
6.	I can make the effort to examine my skin each month	2.48 ^a	4.41		
7.	My friends encourage me to examine my skin regularly	4.34		3.87	
8.	It does not occur to me to examine my skin	4.24		5.06 ^a	[0.47 (0.027)]
9.	The risk of skin cancer is exaggerated by the medical profession	4.29		2.58	
10.	I could make a habit of skin self-examination	2.58^a	[0.52 (0.024)]	2.90 ^a	
11.	If I had a skin lesion, self-examination and early reporting may prevent it getting worse	3.23^a	[0.53 (0.022)]	2.68	
12.	I am able to make checking my skin a regular routine	2.74^a	[0.52 (0.023)]	2.91	
13.	I could explain the correct method for skin self-examination	3.68 ^a		4.52	
14.	I know someone who had skin cancer	3.86	2.54		
15.	I cannot be bothered with skin self-examination	4.51		4.15 ^a	[0.34 (0.030)]
16.	Examining my skin will make me feel more control over my health	2.53^a	[0.50 (0.031)]	2.79	
17.	I would be able to explain the benefits of skin self-examination to somebody else	3.11^a	[0.33 (0.033)]	3.13	
18.	I am confident about my ability to examine my skin	3.66		2.45	
19.	I feel confident that (with the help of someone else if needed) I could examine my skin thoroughly	2.45		2.45 ^a	[0.31 (0.033)]
20.	My family encourages me to examine my skin regularly	3.57		4.67	
	Mean total	3.32	3.60		
	H_s^a	0.47 (0.021)	0.38 (0.023)		
	H^{Ta}	0.27	0.35		
	Rho^a	0.85	0.79		

^aFor items included in scale 1. Those shown in bold are the items included in the Mokken scales.

then checked to exclude any items violating monotonicity using “check.monotonicity”; item pairs were then plotted using “plot(check.iio(FileR))” and the item pairs examined for intersection, floor and ceiling items and any items lying far from the main cluster to decide whether they were suitable for analysis of IIO using “iio.results <- check.iio(FileR)” followed by “summary(check.iio(FileR, item.selection = FALSE))”; and reliability of resulting scales was checked using “check.reliability”. SPSS version 22.0 was used to perform an independent samples *t* test.

3.5 | Ethical approval

The original ethical application in the UK referred to above was subject to a minor modification in 2015, and then, this study was approved by the Academic Board of the Italian university.

4 | RESULTS

4.1 | Demographics

The total number of participants in the present study was 1086:620 from Italy (340 females; 278 males [2 non-responses]; age range 18–70) and 466 from the UK (381 females; 85 males; age range 18–85). Any items with non-responses were removed before running the analysis.

4.2 | Mokken scaling analysis

The outcome of the *aisp* indicated for the Italian and UK samples showed that in both Italy and the UK, eight items clustered on a single scale; the remainder either did not scale or formed other clusters with too few items to form a meaningful scale. The focus of the subsequent analysis was, therefore, on the items clustering on scale

1 in both the Italian and the English samples. Inspection of the relative item ordering by mean values suggested that the Italian and UK samples were insufficiently similar to merit combining the samples; the two scales only have one item in common. From both samples, one further item was removed from the scale due to violating monotonicity, leaving seven items in each scale.

All 20 questions from Section B of the MOLES index are shown in the order in which they appear in the questionnaire along with their mean values for the Italian and the UK samples (Table 1). The difference in total mean scores—tested using a *t* test—was not significantly different between the Italian and UK samples. For clarity, the values of *H_i* and the respective standard deviations are only shown for the items which scale. The values of *H_s* along with their respective standard deviations, the values of *H^T* and the values of *Rho* are given at the foot of each column. Inspection of item pair plots for the combined sample showed that items were quite closely clustered with minimal intersection and no items showing either a “floor” or a “ceiling” effect or lying far from the cluster. None of the seven items remaining in either the Italian or UK scales violated IIO. Using the standard errors, the 95% confidence intervals around *H_s* and *H_i* were inspected and they did not include the lower bound value of 0.30. The seven items from the Italian data formed a moderate Mokken scale which was reliable, but *H^T* was not strong enough to show IIO. The seven items from the UK data formed a weak Mokken scale which was reliable, and *H^T* was strong enough to show weak IIO.

Items are ordered according to their mean value in Table 2. Higher mean scores indicate lower endorsement of the item and, therefore, greater difficulty. In this light, the least difficult item in the Italian data was: “I believe examining my skin leads to better health” and the most difficult item was “I would be able to explain the benefits of skin self-examination to somebody else” and, in the UK data, the least difficult item was: “I feel confident that (with the help of someone else if needed) I could examine my skin thoroughly” and

TABLE 2 Items in scale 1 ordered by increasing mean value

Item	Italy	Item	UK
1	I believe examining my skin leads to better health ^a	19	I feel confident that (with the help of someone else if needed) I could examine my skin thoroughly ^c
11	If I had a skin lesion, self-examination and early reporting may prevent it getting worse ^a	2	I could describe the moles and marks on my skin ^c
4	If I examine my skin I may prevent cancer ^a	15	I cannot be bothered with skin self-examination ^{*e}
16	Examining my skin will make me feel more control over my health ^a	1	I believe examining my skin leads to better health ^a
10	I could make a habit of skin self-examination ^b	8	It does not occur to me to examine my skin ^e
12	I am able to make checking my skin a regular routine ^b	5	Remembering to check my skin is difficult ^{*e}
17	I would be able to explain the benefits of skin self-examination to somebody else ^c	3	My doctor/nurses encourages me to examine my skin regularly ^d

Abbreviation(s): NB, higher means lower endorsement; *, reverse scored.

^aOutcome expectations factor.

^bIntentions factor.

^cSelf-efficacy factor.

^dSocial influences factor.

^eMemory factor.

the most difficult item was: “My doctor/nurses encourages me to examine my skin regularly”. Only one item: “I believe examining my skin leads to better health” was common to both scales.

5 | DISCUSSION

The results show that Mokken scales exist in both the Italian and the English versions of the MOLES index. The same number of items formed a Mokken scale in the Italian and the English versions. There was only one item in common between the English and the Italian versions meaning that the two scales were insufficiently similar to combine the samples and analyse for a single Mokken scale.

The two scales indicate that different constructs within the MOLES index are important in Italy and the UK. Items ordered by Mokken scaling in Italy relate mainly to belief about the value of SEE in terms of the “Outcome expectations” and “Intentions” factors previously identified (Dyson & Cowdell, 2014). Items in the UK scale mainly relate to the “Social influences” and “Memory” factors previously identified (Dyson & Cowdell, 2014). Both scales share items from the “Self-efficacy” scale. There is no overall significant difference in the total scale scores and looking for significant differences between individual items is prone to type I error; therefore, an explanation must be sought for the very different Mokken scales formed in the two samples and what the implications are for the use of the MOLES index.

First, the differences in the items included in the scales could indicate differences in the perception of risk of melanoma between the Italian and the UK samples. Items that are ordered in Mokken scales are likely to be those that respondents largely respond to consistently relative to one another. Therefore, it appears that respondents in the Italian sample more consistently responded to a set of items related to belief about SSE and the UK sample responded more consistently to a set of items about actions related to SSE. Due to the there being no statistically significant difference between the two samples and no consistent difference in the pattern of responses to the MOLES items, the apparent difference in the two scales probably does not indicate the importance ascribed to any particular aspects of SSE. Thus, the differences may not have utility in designing interventions or targeting specific aspects of SSE. However, the potential utility of the scales is that these items may also respond consistently to health education and health promotion about melanoma and SSE. Thus, they may have utility—separately—in measuring the outcome of SSE interventions in Italy and the UK, respectively.

It is possible that larger sample sizes may lead to inclusion of more items and greater congruence between the two scales. Thus, a future line of research is suggested by repeating the study with larger samples, possibly in the region of $N = 1,000$ per country (Straat, Ark, & Sijtsma, 2014). It would also be valuable to replicate the confirmatory factor analysis in an Italian sample. A useful indication of the utility of the MOLES—which is about motivation—would be to relate actual practices related to SSE with the MOLES index

in individuals. In that light, the present study suggests a clear line of research related to SEE in different populations.

5.1 | Limitations

Fewer than 50% of the items in the MOLES were included in either of the Mokken scales. This raises the question of the purpose of the remaining items and the possibility of construct underrepresentation. The implication could be that some items in the MOLES are redundant, but it should also be noted that the sample sizes in the present study are relatively small according to our most recent understanding of sample size requirements for Mokken scaling (Straat et al., 2014).

6 | CONCLUSION

The significance of this study lies in its originality in applying Mokken scaling to the MOLES index according to rigorous analytical criteria. The study provides additional psychometric insight into the MOLES index and augments the original work which used factor analysis. An immediate line of inquiry is suggested that could further test the construct validity of the MOLES index by comparing the latent structure that is apparent in the Mokken scales with a measurement of actual practices—frequency and efficacy—of skin self-examination.

CONFLICT OF INTEREST

No authors have any conflict of interest to declare.

ORCID

Annamaria Bagnasco  <https://orcid.org/0000-0002-9079-8460>

Roger Watson  <https://orcid.org/0000-0001-8040-7625>

Milko Patrick Zanini  <https://orcid.org/0000-0002-1081-6279>

REFERENCES

- Agbo, A. A. (2010). Cronbach's alpha: Review of limitations and associated recommendations. *Journal of Psychology in Africa*, 20, 233–239. <https://doi.org/10.1080/14330237.2010.10820371>
- Cowdell, F., & Dyson, J. (2014). A novel intervention for skin cancer prevention. *Dermatology of Nursing*, 13(3), 45–49.
- Dyson, J., & Cowdell, F. (2014). Development and psychometric testing of the ‘Motivation and Self-Efficacy in Early Detection of Skin Lesions’ index. *Journal of Advanced Nursing*, 70, 2952–2963. <https://doi.org/10.1111/jan.12436>
- Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., ... Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of Behavioral Medicine*, 46, 81–95. <https://doi.org/10.1007/s12160-013-9486-6>
- Mokken, R. J., & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological*

- Measurement*, 6, 417–430. <https://doi.org/10.1177/014662168200600404>
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74, 107–120.
- Straat, J. H., van der Ark, L. A., & Sijtsma, K. (2014). Minimum sample size requirements for Mokken scale analysis. *Educational and Psychological Measurement*, 74, 809–822. <https://doi.org/10.1177/0013164414529793>
- Van der Ark, A., Straat, J. H., & Kooipan, L. (2018). Package 'mokken'. Retrieved from <https://cran.r-project.org/web/packages/mokken/mokken.pdf>
- Watson, R., van der Ark, L. A., Lin, L.-C., Fieo, R., Deary, I. J., & Meijer, R. (2012). Item response theory: How Mokken scaling can be used in

clinical practice. *Journal of Clinical Nursing*, 21, 2736–2746. <https://doi.org/10.1111/j.1365-2702.2011.03893.x>

How to cite this article: Aleo G, Bagnasco A, Watson R, et al. Comparing questionnaires across cultures: Using Mokken scaling to compare the Italian and English versions of the MOLES index. *Nursing Open*. 2019;00:1–7. <https://doi.org/10.1002/nop2.297>