Improving measurement in nutrition literacy research using Rasch modelling: examining construct validity of stage-specific 'critical nutrition literacy' scales

Øystein Guttersrud[1],*, Jorån Østerholt Dalane[2] and Sverre Pettersen[2]

[1]Norwegian Centre for Science Education, University of Oslo, Naturfagsenteret, Postboks 1106, Blindern, 0317 Oslo, Norway: [2]Department of Health, Nutrition and Management, Oslo and Akershus University College of Applied Sciences, Kjeller, Norway

Abstract

Objective: Critical nutrition literacy (CNL), as an increasingly important area in public health nutrition, can be defined as the ability to critically analyse nutrition information, increase awareness and participate in action to address barriers to healthy eating behaviours. Far too little attention has been paid to establishing valid instruments for measuring CNL. The aim of the present study was to assess the appropriateness of utilizing the latent scales of a newly developed instrument assessing nursing students' 'engagement in dietary habits' (the 'engagement' scale) and their level of 'taking a critical stance towards nutrition claims and their sources' (the 'claims' scale).

Design: Data were gathered by distributing a nineteen-item paper-and-pencil self-report questionnaire to university colleges offering nursing education. The study had a cross-sectional design using Rasch analysis. Data management and analysis were performed using the software packages RUMM2030 and SPSS version 20.

Setting: School personnel handed out the questionnaires.

Subjects: Four hundred and seventy-three students at ten university colleges across Norway responded (52% response rate).

Results: Disordered thresholds were rescored, an under-discriminating item was discarded and one item showing uniform differential item functioning was split. The assumption of item locations being differentiated by stages was strengthened. The analyses demonstrated possible dimension violations of local independence in the 'claims' scale data and the 'engagement' scale could have been better targeted.

Conclusions: The study demonstrates the usefulness of Rasch analysis in assessing the psychometric properties of scales developed to measure CNL. Qualitative research designs could further improve our understanding of CNL scales.

Critical nutrition literacy

Citizens encounter nutrition issues in their daily lives. Most such encounters involve social media and require citizens to address nutrition-related issues on personal, national or even global levels. The outcome of these encounters probably depends on citizens' 'nutrition literacy'. Nutrition literacy can be defined as 'the capacity to obtain, process and understand nutrition information and the materials needed to make appropriate decisions regarding one's health'[1]. This definition has a clear link to the definition of health literacy made by Nutbeam[2]. Furthermore, Pettersen[3] has added a 'critical dimension' to the definition of nutrition literacy: 'the ability to critically assess nutritional information and dietary advice'. Pettersen[3] and Silk *et al.*[1] have described three cumulative levels of nutrition literacy referred to as 'functional', 'interactive' and 'critical' nutrition literacy.

Functional nutrition literacy (FNL) refers to proficiency in applying basic literacy skills, such as reading and understanding food labelling and grasping the essence of nutrition information guidelines. Interactive nutrition literacy (INL) comprises more advanced literacy skills, such as the cognitive and interpersonal communication skills needed to interact appropriately with nutrition counsellors, as well as interest in seeking and applying adequate nutrition information for the purpose of improving one's nutritional status and behaviour. Critical nutrition literacy (CNL) refers to being proficient in critically analysing nutrition information and advice, as well as having the will to participate in actions to address nutritional barriers in personal, social and global perspectives.

CNL is part of scientific literacy[4] – 'the capacity to use scientific knowledge, to identify questions and to draw

Corresponding author: Email oystein.guttersrud@naturfagsenteret.no

evidence-based conclusions,[5] i.e. proficiency in describing, explaining and predicting scientific phenomena, and understanding the processes of scientific inquiries as well as the premises of scientific evidence and conclusions[6].

The aim of the present study was to use Rasch modelling to examine the construct validity of a new instrument developed for measuring nursing students' CNL. The emphasis is on interpreting statistical misfit in terms of substantive inconsistency with a view to improving the CNL instrument. To date, the CNL instrument has only been assessed using classical test theory[7–9].

The unidimensional simple logistic Rasch model

The unidimensional simple logistic Rasch model (SLM), expressed as

$$P(X_{ni} = 1) = \frac{e^{\beta_n - \delta_i}}{1 + e^{\beta_n - \delta}},$$

models the probability that a respondent will affirm a dichotomous item[10]. The probability (P) is modelled as a function of the distance between the two independent parameters 'person location' (β_n) and 'item location' (δ_i)[11]. The graphical representation of the SLM is referred to as the item characteristic curve (ICC).

The person parameter and the item parameter represent certain locations on the underlying construct, i.e. the latent variable that the instrument is intended to measure. Person location typically refers to proficiency – the ability a person possesses – and item location to difficulty – the amount of ability associated with endorsing a certain item. Items located at zero, i.e. $\delta_i = 0$, measure moderate level of the latent variable.

The assumptions of unidimensional Rasch models are that: (i) the response probability depends on a dominant dimension (unidimensionality) – not only one factor – with the possible presence of minor dimensions[12,13]; (ii) the responses to items are independent (local independence); (iii) the raw scores contain all of the information on person location regardless of which items have been endorsed (sufficiency); and (iv) the response probability increases with higher values of person location (monotonicity). In Rasch analyses, raw scores are converted to a logit scale in the estimation process.

The unidimensional polytomous Rasch model

The unidimensional polytomous Rasch model (PRM), expressed as

$$P\{X_{ni} = x\} = \frac{1}{\gamma} e^{[\kappa_x + x(\beta_n - \delta_i)]},$$

where $\gamma = \sum_{k=0}^{m} e^{[(\kappa_k + k(\beta_n - \delta_i)]}$ is a normalization factor ensuring $\int_{-\infty}^{\infty} P \cdot d\beta = 1$, models the probability for person n with location β_n scoring x points or ticking off response category x on a polytomous item i with location δ_i[14]. κ refers to category coefficients.

Parameterizations of the polytomous Rasch model

If the observed distance between the response categories is the same across all items, e.g. the distance between 'agree strongly' and 'agree partly' for a Likert-scale item is equal to the distance between 'agree strongly' and 'agree partly' for another item, the data fit the rating scale parameterization[15] of the PRM best. If the distance is not the same across the items, the partial credit parameterization[16] is indicated.

Response categories and ordered thresholds of the polytomous Rasch model

A threshold is defined as the person location at which the probability of responding in one of two adjacent response categories reaches 0·50. A polytomous item with an $m + 1$ number of response categories has m ordered thresholds (τ_k) where $k \in \{1, 2, \ldots, m\}$ and $x \in \{0, 1, \ldots, m + 1\}$. The score x indicates the number of m ordered thresholds a respondent has passed[14].

The succeeding ordered thresholds reflect successively more of the latent ability or attitude. The ordering of thresholds is a property of the data and not the Rasch model. Disordered thresholds are clear evidence of problems in the data[17], but statistical analysis cannot determine the cause of the disordering[14].

Constructing invariant measures – over- and under-discriminating items

When an item provides data which sufficiently fit a unidimensional Rasch model, the item provides an indication of relative ability or attitude along the latent variable. In Rasch analysis, this information is used to construct measures. If the data approach a step function, the item is said to over-discriminate. If the data approach a constant function, the item is said to under-discriminate.

Strongly over-discriminating items tend to act like 'switches' which stratify the persons below and above certain ability estimates, but they are not measuring devices. Under-discriminating items tend to neither stratify nor measure.

Model fit

Fit residuals and item χ^2 values are used to test how well the data fit the model[18]. Negative and positive item fit residuals indicate whether items over- or under-discriminate. A person fit residual indicates how well a person's response pattern fits the 'Guttman structure'[19,20].

Large χ^2 indicates that persons with different locations do not 'agree on' item locations, thus compromising the required property of invariance. To adjust χ^2 probabilities for the number of significant tests performed, the probabilities are Bonferroni adjusted[21] using the software package RUMM2030[22].

Reliability

Cronbach's α is an index of internal consistence reliability[23]. When the index is calculated using estimates

from Rasch models, it is referred to as the person separation index (PSI).

Targeting

Comparing the mean location of persons with the mean location of items provides an indication of how well the items are targeted to the persons. When items are well targeted to the person locations, the measurement error is reduced.

Differential item functioning and invariance

Rasch models are the only item response theory (IRT) models that provide invariant measurements if the data fit the model. Criterion-related construct validity, sufficiency and reliability are also provided if the data fit a Rasch model. Invariant measurement is not guaranteed if the data fit a two-parameter IRT model because these models also model item discrimination in addition to person and item location.

Differential item functioning (DIF) between a person factor's categories, e.g. male and female for gender, is evident when, for a given estimate of the latent trait, the mean scores of the people in the gender categories are 'significantly' different from each other. This means that an item has different location estimates for males and females, i.e. the observed values for males and females are described by two different ICC.

If these ICC do not intersect, the item discriminates equally strongly across the continuum for both groups and the DIF is uniform[24]. Non-uniform DIF is an important factor for non-invariant measures. Uniform DIF might be resolved[25,26] by using the 'person factor split' procedure in RUMM2030[22] while items with non-uniform DIF must be discarded.

Local trait dependence

Trait dependence violates unidimensionality and causes 'dimension violations' of local independence[27–29]. Trait dependence appears when person factors other than ability or attitude influence response, e.g. ability to guess[30] or DIF related to gender and ethnicity.

The result is usually 'less' Guttmann structure in the response patterns and under-discriminating items showing DIF that will lower construct validity[25,26]. Multidimensionality results in a decreased variance of person estimates and a decreased reliability coefficient[28].

Large variations in the percentage variance explained by each principal component (PC) is one way of generating a hypothesis about multidimensionality in the data[12,13]. The assumption of unidimensionality might be tested using the t-test procedures in RUMM2030[22] and by estimating the latent correlation between possible sub-dimensions[31].

Local response dependence

Response dependence violates statistical independence and causes 'response violations' of local independence[27–29], meaning that the entire correlation between the items is not captured by the latent trait. This might take place when a previous item gives hints or clues that affect responses to a subsequent (dependent) item, causing deviations of the thresholds of the dependent item[32].

The result is 'more' Guttmann structure in the response patterns and consequently over-discriminating items, which result in an increased variance of person estimates and an increased reliability coefficient[29,33].

A high correlation between a pair of item residuals is one way of generating a hypothesis about whether two items show response dependence[12,29]. The magnitude of the response dependence might be estimated using the 'item dependence split' procedure in RUMM2030[22]. The estimate helps test the hypothesis[27,32].

Method

Frame of reference

Using email advertisements, 473 people (response rate 52 %), of whom 8 % were males, were recruited from ten of the twenty-eight Norwegian university colleges offering nursing education, covering urban and rural areas. Almost all respondents (96 %) were third-year nursing students aged between 20 and 54 years with a mean age of 26·4 (sd 6·9) years. More than a quarter of those surveyed (28 %) lived with one or more children.

Data collection

The data collection took place during autumn and winter 2010 by means of a paper-and-pencil questionnaire handed out by school personnel. Participation was voluntary and the questionnaire was completed anonymously in the classrooms within 20 min.

The critical nutrition literacy instrument

The assessed CNL instrument consists of two scales measuring separate aspects of critical nutrition literacy (see Tables 1 and 2): (i) the 'engagement in dietary habits' scale (the 'engagement' scale) consisting of eight items and (ii) the 'taking a critical stance towards nutrition claims and their sources' scale consisting of eleven items (the 'claims' scale). A five-point Likert scale with all of the response categories anchored with a phrase was applied for all items: 'disagree strongly' (1), 'disagree partly' (2), 'neither agree nor disagree' (3), 'agree partly' (4) and 'agree strongly' (5).

Results

Comparing the parameterizations of the polytomous Rasch model using summary statistics

The first step in the analysis was to determine the appropriate model that fitted the data (see Table 3). The rating scale parameterization fitted the data from the engagement scale better when comparing overall χ^2

Table 1 Wording of items

Item	Item phrasing (engagement scale)	Rev
18.	I am concerned that the price of food that is considered to be healthy may get too high	
15.	I am concerned that there is not a wide selection of healthy food at the grocery stores where I usually shop	
17.	I try to influence others (for instance family members and friends) to eat healthy food	
13.	I require my college/university, workplace, etc. to offer healthy food	
16.	I welcome any initiative aimed at promoting a healthy diet for children and adolescents	
14.	I engage actively in initiatives aimed at promoting a healthier diet (for instance at my college/university)	
12.	I get engaged in issues contributing to the provision of a healthier diet for most of the people in this country	
19.	I would like to get involved in political issues directed at improving the population's diet	
	Item phrasing (claims scale)	
24.	I have confidence in the various diets that I read about in newspapers, magazines, etc.	x
21.	I am critical of the dietary information that I receive from various sources in society	
20.	I am concerned that the dietary information that I read may not be based on science	
25.	I believe my body tells me what it needs in terms of nutrients, regardless of researchers' opinions about this	x
29.	I am confident that the media's presentation of new scientific findings concerning a healthy diet is correct	x
23.	I am familiar with the criteria for scientifically based content in health claims	
22.	I often refer to newspapers and magazines if I discuss diet with others	x
26.	I am influenced by the dietary advice that I read about in newspapers, magazines, etc.	x
27.	I am confident that some of the methods within alternative medicine (such as health foods) provide me with credible dietary advice	
28.	I find it hard to distinguish scientific nutritional information from non-scientific nutritional information	x
30.	I base my diet on information that I get from scientifically recognized literature (for instance, the journals published by the Norwegian Medical Association and the Norwegian Directorate of Health)	

Reverse-scored items (rev) are indicated by 'x'.

Table 2 Items in order of location (loc)

Item	Scale	Cluster	Loc	Res	df	χ^2	df	$P(\chi^2)$	Thresholds
18.	Engagement	Concern	−1·66	2·5	398·3	13·1	7	0·070	
15.	Engagement	Concern	−1·23	−0·7	400·9	6·7	7	0·464	
17.	Engagement	Concern	−0·27	0·9	400·9	7·5	7	0·379	
13.	Engagement	Concern	−0·01	1·7	402·6	5·3	7	0·627	
16.	Engagement	Democracy	0·20	−1·1	400·0	12·7	7	0·079	
14.	Engagement	Democracy	0·83	−3·6	400·0	2·8	7	0·902	
12.	Engagement	Democracy	0·87	0·7	399·2	4·6	7	0·706	
19.	Engagement	Democracy	1·28	1·1	393·1	2·9	7	0·898	
24.	Claims	Evaluating	−0·61	0·3	412·8	23·6	7	0·001	Disordered
21.	Claims	Evaluating	−0·50	−0·6	411·9	7·7	7	0·359	Disordered
20.	Claims	Evaluating	−0·46	−0·1	411·0	4·0	7	0·774	Disordered
25.	Claims	Evaluating	−0·06	4·5*	408·3	30·9	7	0·000*	Disordered
29.	Claims	Evaluating	0·03	0·8	406·5	10·2	7	0·176	
23.	Claims	Identifying	0·12	1·0	391·2	14·2	7	0·047	Disordered
22.	Claims	Identifying	0·12	0·0	406·5	6·5	7	0·477	Disordered
26.	Claims	Evaluating†	0·25	−1·3	412·8	15·1	7	0·034	Disordered
27.	Claims	Identifying	0·31	0·6	404·7	4·0	7	0·781	
28.	Claims	Identifying	0·39	1·0	393·9	2·6	7	0·919	
30.	Claims	Identifying	0·40	1·6	404·7	8·1	7	0·326	Disordered

Table 2 refers to scale ('engagement in dietary habits' (engagement) and 'taking a critical stance towards nutrition claims and their sources' (claims)), cluster ('concern about dietary habits' (concern), 'willingness to engage in democratic processes to improve dietary habits' (democracy), 'justifying premises for and evaluating the sender of nutrition claims' (evaluating) and 'identifying scientific nutrition claims' (identifying)), loc (item location), res (fit residual), df (degrees of freedom), chi-square (χ^2), chi-square probability ($P(\chi^2)$) and ordering of thresholds.
*Item 25 under-discriminates.
†Item 26 overlaps with the identifying items.

while the partial credit parameterization provided the best fit for the data from the claims scale.

The two scales' item fit residual mean and standard deviation deviated from their expected values, i.e. 0 and 1, as their values were 0·17 (sd 1·92) and 0·61 (sd 0·89), respectively (see Table 3).

The functioning of response categories and ordering of thresholds

The pattern of responses for items 20, 21, 23 and 30 indicated that the response category 'disagree partly' (2) did not have the highest probability of being selected by any attitude level. The response patterns to items 24 and 26 indicated that the 'neither agree nor disagree' (3) response category was not the most likely response for any attitude level. These observations implied disordered thresholds in the data.

Item discrimination, item fit and person fit

According to the item fit residuals and χ^2 statistics, the under-discriminating item 25 did not fit the model. Individual person fit residuals showed that eight people had a z-fit residual outside the range $z = \pm 2.5$.

Table 3 Summary statistics for the engagement scale and claims scale applying partial credit parameterization and rating scale parameterization

Scale	Model	χ^2	df	$P(\chi^2)$	PSI	z	SD	S	K	Loc	SD
Engagement	Partial credit	65	56	0·19	0·77	0·03	1·15	0·25*	1·90*	0·72	0·99
Engagement	Rating scale	56	56	0·49	0·77	0·17	1·92	0·69*	0·79*	0·90	0·96
Claims	Partial credit	127	77	0·00	0·71	0·72	1·51	1·18	1·09	0·30	0·61
Claims	Rating scale	127	77	0·00	0·71	0·47	2·66	0·44	1·27*	0·34	0·61
Claims	Partial credit†	87	70	0·08	0·71	0·61	0·89	0·35*	1·47*	0·30	0·66
Claims	Rating scale†	106	70	0·00	0·71	0·49	1·97	0·19	1·59*	0·35	0·66
Claims	Partial credit‡	113	70	0·00	0·69	0·64	0·76	0·67*	0·76*	0·27	1·23

Table 3 refers to total item chi-square (χ^2), degrees of freedom (df), chi-square probability ($P(\chi^2)$), person separation index (PSI), mean fit residual (z) with its standard deviation, skewness (S), kurtosis (K) and mean person location (loc) with its standard deviation.
*Negative values.
†Analyses where item 25 was deleted.
‡Analyses where items were rescored.
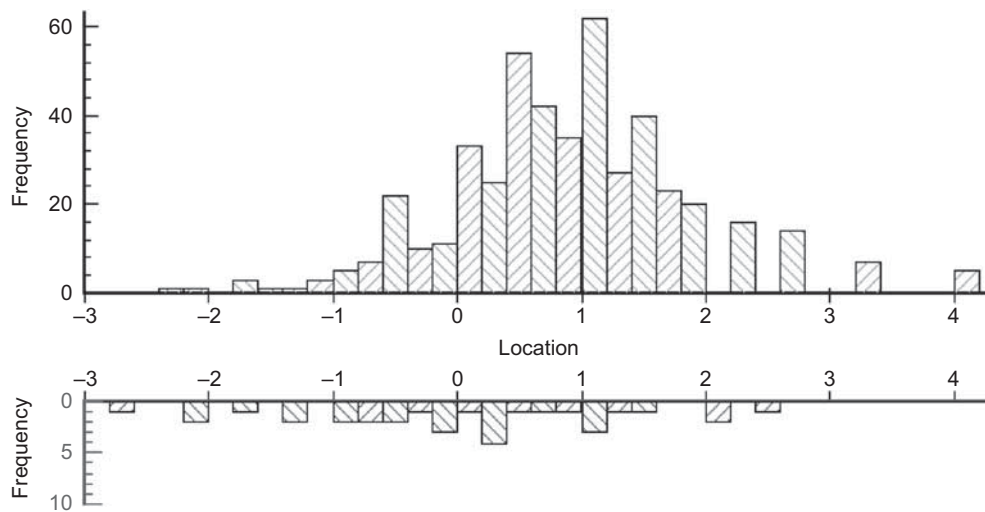
Fig. 1 The distribution of respondent attitudes (upper bars) and item threshold affective levels (lower bars) for the engagement scale.

Reliability estimates

Cronbach's α was estimated using the SPSS statistical software package version 20. Cronbach's α for the engagement scale was 0·80. Cronbach's α for the claims scale increased from 0·69 to 0·70 when item 25 was deleted. The PSI were 0·77 for the engagement scale and 0·71 for the claims scale (see Table 3).

Targeting – mean person attitude

The average person location value was 0·90 for the engagement scale and 0·30 for the claims scale (see Table 3). Figure 1 shows the distribution of person locations and item threshold locations for the engagement scale. Item locations are reported in Table 2.

Resolving the item with uniform differential item functioning

The responses to item 16 were influenced by whether or not the person lives with children. Item 16 showed uniform DIF as it discriminated equally strongly for both categories. By splitting item 16, the two virtual items'

affective values differed by more than 0·6 logits between the person factor categories 'live with children' (location −0·24, Fig. 2 left curve) and 'do not live with children' (location 0·38, Fig. 2 right curve).

Possible dimension violations of local independence

The correlation coefficient between the residual of each item on the engagement scale and the first PC was positive for the 'concern' items and negative for the 'democracy' items. These sets of items are identical to the clusters initially formed based on a qualitative judgement of item content (see Table 2).

The correlation coefficient between the residual of each item on the claims scale and the first PC was positive for items 20, 21, 23 and 30 and negative for items 22 and 24–29. The PC summary in RUMM2030 indicated more variations in the amount of percentage variance explained by each component for the claims scale than for the engagement scale. These analyses indicated that items 20, 21, 23 and 30 might tap into a subscale of the claims scale while items 22 and 24–29 might form a second subscale of the claims scale.
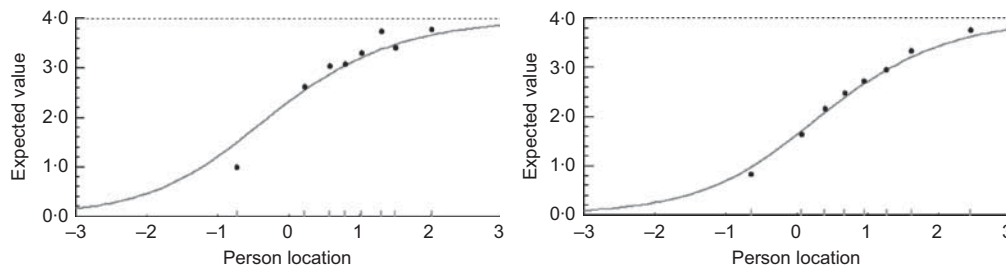
Fig. 2 The item characteristic curves for the two virtual items appearing after splitting item 16 for the person factor categories 'live with children' (left curve) and 'do not live with children' (right curve) for the person factor 'parenthood'. The dots represent the observed values in eight class intervals. The mean person location of each interval is marked at the x-axis. The dotted line is an asymptote

A subtest analysis based on clusters of items with positive and negative correlation coefficients with the first PC was performed. The latent correlation between the two possible subscales of the engagement scale was $r = 0.82$ (the concern items and the democracy items), while the latent correlation between the two possible subscales of the claims scale was $r = 0.31$ (items 20, 21, 23, 30 and 22, 24–29).

Applying the equating tests procedure in RUMM2030, the percentage of persons with 'significantly' different scores to a 5% level on the two possible subscales of the engagement scale and the two possible subscales of the claims scale was 5% and 13%, respectively. The *t*-test procedures in RUMM2030 indicated that the engagement scale had acceptable unidimensionality while the claims scale might have problematic dimensionality.

Possible response violations of local independence

The residual correlation between items 20 and 21 slightly exceeded 0.3. The magnitude of the response dependence was not estimated as the items had disordered thresholds indicating problematic data.

Discussion

Our qualitative categorization of the engagement scale items into sets finds support in the quantitative empirical data. What is really interesting in the data is that the location of the engagement scale items is clearly differentiated by stages, meaning that the 'concern' items require lower 'engagement in dietary habits' to overcome than the 'democracy' items.

Items measuring 'global perspectives' should be developed to incorporate the global aspect with the engagement scale, and phrases making clearer reference to personal or social perspective should be added to items 16 and 18 accordingly. Further, the references to children and adolescents in item 16 needs to be revised to avoid DIF for the person factor 'parenthood'.

Except for item 26, the locations of the claims scale items are also differentiated by stages. The 'evaluating' items require lower levels of 'taking a critical stance towards nutrition claims and their sources' to overcome

than the 'identifying' items. A rather low latent correlation between subsets of the claims scale items indicates possible multidimensionality in the data.

The collapsing of the adjacent response categories of the claims scale items was based on the disordering of the threshold estimates in the Wright map. The patterns of response for the items with disordered thresholds on the claims scale suggest that these items might function like four-point response formats.

Items 20 and 21 should be rephrased to avoid the items collecting redundant information. Items 22 and 27 should be rephrased to further distinguish them from the 'evaluating' items. Item 25 must be discarded as it is phrased like a true–false item and under-discriminates. Item 26 could be rephrased to alter its affective level as we seek to differentiate items by stages.

New and revised items should be field-trialled to ensure that the CNL instrument can measure students' CNL invariantly across student years without disordering the stage-specific sets of items.

Cronbach's α and PSI are valid measures of reliability only when items are independent. Response violations (item 21) and dimension violations (items 16 and 25) of local independence cause non-invariant measures and affect the reported reliability estimates of the scales.

The engagement scale could have been better targeted with the mean item location at a higher affective level corresponding to persons' mean attitude levels. For example, response category 1 'disagree strongly' is out of range for most items.

Our finding that the rating scale parameterization fits the data of the engagement items better indicates that the distance between the response categories is the same across all items. This is not the case for the claims scale items.

Conclusions

Taken together, these results suggest that further psychometric analyses on similar and different samples should be carried out and complemented by qualitative focus group interviews in order to ensure that the sets of items make sense both conceptually and empirically

against the Rasch model. Our method of examining construct validity using item response modelling has important implications for the future development and validation of quantitative public health research.

Acknowledgements

References

1. Silk KJ, Sherry J, Winn B *et al.* (2008) Increasing nutrition literacy: testing the effectiveness of print, web site, and game modalities. *J Nutr Educ Behav* **40**, 3–10.
2. Nutbeam D (2000) Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. *Health Promot Int* **3**, 259–267.
3. Pettersen S (2009) Kostholdsinformasjon og annen helseinformasjon. In *Mat og helse i skolen. En fagdidaktisk innføring* (*Food and Health in Schools. An Introduction to Education*), pp. 87–100. Bergen: Fagbokforlaget.
4. Pettersen S (2007) *Health Claims and Scientific Knowledge. A Study of How Students of Health Sciences, their Teachers, and Newspaper Journalists Relate to Health Claims in Society.* Oslo: University of Oslo.
5. Organisation for Economic Co-operation and Development, Programme for International Student Assessment (2003) *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills.* Paris: OECD Publications.
6. Organisation for Economic Co-operation and Development, Programme for International Student Assessment (2004) *Learning for Tomorrow's World. First Results From PISA 2003.* Paris: OECD Publications.
7. Dalane JØ 2001. *Nutrition literacy hos sykepleierstudenter* (*Nursing Students' Nutrition Literacy*). Kjeller: Oslo and Akershus University College.
8. Aarnes SB (2009) *Utvikling og utprøving av et spørreskjema for å kartlegge nutrition literacy: assosiasjon til kjønn, utdannelse og fysisk aktivitetsnivå* (*Development and Testing of a Questionnaire to Assess Nutrition Literacy: Association with Gender, Education and Physical Activity Level*). Akershus: Akershus University College Norway.
9. Kjøllesdal JG (2009) *Nutrition literacy. Utvikling og utprøving av et spørreskjema som måler grader av nutrition literacy* (*Development and Testing of a Questionnaire Measuring Levels of Nutrition Literacy*). Kjeller: Oslo and Akershus University College.
10. Andrich D (1988) *Rasch Models for Measurement.* Beverly Hills, CA: SAGE Publications.
11. Rasch G (1960) *Probabilistic Models for Some Intelligence and Achievement Tests.* Copenhagen: Danish Institute for Educational Research. Expanded edition 1983. Chicago, IL: MESA Press.
12. Smith EV (2002) Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Measure* **3**, 205–231.
13. Linacre JM (1998) Detecting multidimensionality: which residual data-type works best? *J Outcome Meas* **2**, 266–283.
14. Andrich D, de Jong JHAL & Sheridan BE (1997) Diagnostic opportunities with the Rasch model for ordered response categories. In *Applications of Latent Trait and Latent Class Models in the Social Sciences*, pp. 59–70 [J Rost and R Langeheine, editors]. New York: Waxmann.
15. Andrich D (1978) A rating scale formulation for ordered response categories. *Psychometrika* **43**, 561–573.
16. Wright BD & Masters GN (1982) *Rating Scale Analysis: Rasch Measurement.* Chicago, IL: MESA Press.
17. Fisher RA (1958) *Statistical Methods for Research Workers*, 13th ed. New York: Hafner.
18. Smith RM & Plackner C (2009) The family approach to assessing fit in Rasch measurement. *J Appl Measure* **10**, 424–437.
19. Andrich D (1985) An elaboration of Guttman scaling with Rasch models for measurement. In *Sociological Methodology*, pp. 33–80 [N Brandon-Tuma, editor]. San Fransisco, CA: Jossey-Bass.
20. Andrich D (1982) An index of person separation in latent trait theory, the traditional KR-20 index, and the Guttman scale response pattern. *Educ Res Perspect* **9**, 95–104.
21. Bland JM & Altman DG (1995) Multiple significance tests: the Bonferroni method. *BMJ* **310**, 170.
22. RUMM (2009) *Extending the RUMM2030 Analysis*, 7th ed. Duncraig, Australia: RUMM Laboratory Pty Ltd.
23. Traub RE & Rowley GL (1981) Understanding reliability. *Educ Measure Issues Pract* **10**, 37–45.
24. Andrich D & Hagquist C (2001) *Taking Account of Differential Item Functioning through Principals on Equating. Research Report* no. 12. Perth, Australia: Social Measurement Laboratory, Murdoch University.
25. Brodersen J, Meads D, Kreiner S *et al.* (2007) Methodological aspects of differential item functioning in the Rasch model. *J Med Econ* **10**, 309–324.
26. Looveer J & Mulligan J (2009) The efficacy of link items in the construction of a numeracy achievement scale from Kindergarten to Year 6. *J Appl Measure* **10**, 247–265.
27. Andrich D & Kreiner S (2010) Quantifying response dependence between two dichotomous items using the Rasch model. *Appl Psychol Measure* **34**, 181–192.
28. Marais I & Andrich D (2008) Effects of varying magnitude and patterns of response dependence in the unidimensional Rasch model. *J Appl Measure* **9**, 105–124.
29. Marais I & Andrich D (2008) Formalising dimension and response violations of local independence in the unidimensional Rasch model. *J Appl Measure* **9**, 200–215.
30. Ryan JP (1983) Introduction to latent trait analysis and item response theory. In *Testing in the Schools New Directions for Testing and Measurement*, pp. 48–64 [WE Hathaway, editor]. San Francisco, CA: Jossey-Bass.
31. RUMM (2009) *Interpreting RUMM2030 Part IV Multidimensionality and Subtests in RUMM*, 1st ed. Duncraig, Australia: RUMM Laboratory Pty Ltd.
32. Andrich D, Humphry SM & Marais I (2012) Quantifying local, response dependence between two polytomous items using the Rasch model. *Appl Psychol Meas* **36**, 309–324.
33. Smith EV (2005) Effect of item redundancy on Rasch item and person estimates. *J Appl Measure* **6**, 147–163.