# Young adolescents' engagement in dietary behaviour – the impact of gender, socio-economic status, self-efficacy and scientific literacy. Methodological aspects of constructing measures in nutrition literacy research using the Rasch model

Øystein Guttersrud[1],* and Kjell Sverre Petterson[2]
[1]Norwegian Centre for Science Education, University of Oslo, Naturfagsenteret, Postboks 1106, Blindern, 0317 Oslo, Norway: [2]Faculty of the Health Sciences, Department of Health, Nutrition and Management, Oslo and Akershus University College of Applied Sciences, Oslo, Norway

## Abstract

*Objective:* The present study validates a revised scale measuring individuals' level of the 'engagement in dietary behaviour' aspect of 'critical nutrition literacy' and describes how background factors affect this aspect of Norwegian tenth-grade students' nutrition literacy.
*Design:* Data were gathered electronically during a field trial of a standardised sample test in science. Test items and questionnaire constructs were distributed evenly across four electronic field-test booklets. Data management and analysis were performed using the RUMM2030 item analysis package and the IBM SPSS Statistics 20 statistical software package.
*Setting:* Students responded on computers at school.
*Subjects:* Seven hundred and forty tenth-grade students at twenty-seven randomly sampled public schools were enrolled in the field-test study. The engagement in dietary behaviour scale and the self-efficacy in science scale were distributed to 178 of these students.
*Results:* The dietary behaviour scale and the self-efficacy in science scale came out as valid, reliable and well-targeted instruments usable for the construction of measurements.
*Conclusions:* Girls and students with high self-efficacy reported higher engagement in dietary behaviour than other students. Socio-economic status and scientific literacy – measured as ability in science by applying an achievement test – did not correlate significantly different from zero with students' engagement in dietary behaviour.

The ultimate goals of public health nutrition research are to describe nutrition-related public health issues by applying valid and reliable instruments and enabling citizens and policy makers to take advantage of the findings. The present Norwegian study is a quantitative study aimed at describing and understanding, using reliable and valid measures, tenth-grade students' attitudes towards nutrition-related public health issues. The first objective is to validate a revised scale assessing individuals' engagement in dietary behaviour at the personal, social and global level[1]. The second objective is to describe how different factors affect responses to the engagement in the dietary behaviour scale.

## Background

### Domains of nutrition literacy and the engagement in dietary behaviour scale

Health literacy is 'the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions'[2]. Health literacy is claimed to be a stronger predictor of health than age, income, employment, education and cultural background[3].

Nutrition literacy, being an important dimension of people's health literacy, has been defined as 'the ability to find and elaborate on nutrition information and make

conclusions regarding health issues'[4,5]. There are three major domains of nutrition literacy, which are referred to as functional nutrition literacy (FNL), interactive nutrition literacy (INL) and critical nutrition literacy (CNL)[5].

FNL refers to proficiency in applying basic literacy skills, while INL comprises the cognitive and interpersonal communication skills needed to seek nutrition information and interact appropriately with nutrition counsellors. The CNL domain covers the broad topics of 'critically evaluating nutrition information and advice' and 'engagement in dietary behaviour'.

The first of these topics comprises the skills to 'justify premises for and evaluate the sender of nutrition claims' and 'identify scientific nutrition claims'. Being critical nutrition literate therefore means being proficient in evaluating scientific enquiry and interpreting data and evidence scientifically, which actually means being scientifically literate as described by the Organisation for Economic Co-operation and Development's Programme for International Student Assessment (PISA)[6]. The second topic covered by the domain CNL includes the capability to 'be concerned about dietary behaviours' and 'engage in processes to improve dietary behaviours'[4,7]. The engagement in dietary behaviour (EDB) scale was developed to assess the EDB part of individuals' CNL.

By hypothesising individuals' scientific literacy as a predictor that facilitates the forming of persons' CNL[8], and viewing scientific literacy as a mediator that helps implement the ideas on what scientific knowledge 'is' and how scientific knowledge forms and develops, we conducted analyses of an achievement test assessing ability in science and a scale measuring self-efficacy in science.

### Self-efficacy in science, scientific literacy and socio-economic status

Self-efficacy (SE), being part of individuals' self-regulation[9], represents the personal perception of external social factors[10,11]. In social-cognitive models of health behaviour change (see e.g. Schwarzer and Fuchs[12]), SE is viewed as a predictor that facilitates the forming of intended behaviour, as a mediator that helps implement the intentions and as a moderator to help achieve the intended behaviour[13,14]. Consequently, different dimensions of SE tend to correlate. In education SE is viewed as part of individuals' self-regulated learning[9].

While SE is a measure of students' self-reported future expectations about achievement at the present time, an achievement test measures part of students' scientific literacy. The assessment items in the applied achievement test operationalised the Norwegian natural science curriculum, which focuses on five main areas: 'the budding researcher', 'diversity in nature', 'body and health', 'phenomena and substances' and 'technology and design'[15].

The achievement test items were also distributed across the cognitive domains 'knowing', 'applying' and 'reasoning'. While knowing covers facts, concepts and procedures, applying involves direct application of knowledge and conceptual understanding. Items categorised as reasoning assess proficiency in evaluating scientific enquiry, interpreting data and evidence scientifically in unfamiliar situations and complex contexts.

Socio-economic status (SES) reflects social position in relation to others and the traditional indicators at the individual level have been income, education and occupation[16]. These are often used interchangeably even though they are only moderately correlated[17,18].

The PISA survey[19], assessing 15-year-old students, included several measures of SES in the student questionnaire. Different measures of economic, cultural and social capital at home were applied. Among all these indicators, the number of books at home had the most powerful individual correlation with science ability[19]. The number of books at home is also highly correlated with parental education and income[20].

### The unidimensional logistic Rasch model for polytomous data (PRM)

In the mathematical representation of the Rasch model for polytomous data (hereafter, 'the polytomous Rasch model'; PRM), $P\{X_{ni} = x\} = 1/\gamma\{\exp[\kappa_x + x(\beta_n - \delta_i)]\}$, where $\gamma = \sum_{k=0}^{m} \exp[\kappa_k + k(\beta_n - \delta_i)]$ is a normalisation factor ensuring $\int_{-\infty}^{\infty} P \cdot d\beta = 1$, a person's attitude is described by a single, unidimensional latent variable $\beta_n$ defined so that $-\infty < \beta_n < \infty$[21,22]. The graphical representation of the PRM, referred to as the item characteristic curve, relates the probability $P$ of person $n$ with attitude $\beta_n$ ticking off response category $x$ on a polytomous item $i$ with affective level $\delta_i$[23]. The different $\kappa$ refer to category coefficients. In the case of the achievement test, $\beta_n$ refers to a person's ability and $\delta_i$ to item difficulty.

Invariant measurement is not guaranteed if the data fit a two- or a three-parameter item response theory model. Only Rasch models provide invariant measurements and support construct validity if the data fit the model. Reliability and sufficiency are also provided when data fit a Rasch model.

The requirements of unidimensional Rasch models are that: (i) the raw scores contain all of the information on a person's attitude (sufficiency); (ii) the response probability increases with higher attitude (monotonicity); (iii) the responses to items are independent (local independence); and (iv) the response probability depends on a dominant dimension (unidimensionality)[24,25]. If factors other than the dominant dimension influence item responses, the data are biased.

### Measurement bias – differential item functioning

Differential item functioning (DIF) means that an item has different affective levels for different groups of individuals such as males and females. Then the observed values for males and females are best described by two different curves similar to the theoretical item characteristic curve.

If these curves are parallel the item discriminates similarly across the continuum for both groups and the DIF is said to be mainly uniform[26]. Non-uniform DIF is an important factor for non-invariant measures. Items that show non-uniform DIF should be discarded while items mainly showing uniform DIF might be resolved[27,28] by using the 'person factor split' procedure in the item analysis package RUMM2030[29].

### The requirement of local independence

The local independence requirement implies that there are no dependencies among items other than those that are attributable to the latent trait. This means that after taking into account the person's attitude (latent trait), responses to the questionnaire items should be independent. Likewise, taking into account the person's ability (latent trait), responses to the achievement test items should be independent. Violations of local independence have been formalised as 'response dependence' and 'trait dependence', where the latter is also referred to as 'multidimensionality'[30].

### Response violations of local independence

Response dependence between items appears when two items share something more in common than can be accounted for by the latent trait. One example of response dependence is when two questionnaire items ask for more or less the same information, causing redundancy in the data. Another example is when a previous achievement test item offers clues that affect responses to a subsequent, dependent item[31,32]. Response dependence violates statistical independence and causes 'response violations' of local independence[30,33,34], meaning that the entire correlation between the items is not captured by the latent trait. The result of response dependency is deviations of the thresholds of the dependent item[32].

A high correlation between a pair of item residuals (a residual is the difference between the observed and the expected value) is one way of generating a 'post hoc' hypothesis of response dependence[24,30]. When two questionnaire items ask for the same information causing redundancy in the data, one would normally form a subtest, i.e. merge the two items into one composite item. Using the 'item dependence split' procedure in RUMM2030, the magnitude of the dependence of a pair of achievement test items, where one offers a clue for the response to the other, might be estimated[29] and used to test the hypothesis of response dependence[27,31].

### Dimension violations of local independence

Multidimensionality or trait dependence means that 'multiple' latent variables or traits play a role and that some items measure one latent variable and other items measure another latent variable. One might form subtests and study whether the latent variables measure one overarching dimension or measure unique aspects. If the latent variables measure unique aspects, the theoretical composite construct might not find support in the empirical evidence as the data are not sufficiently unidimensional.

If, for example, the overarching dimension 'ability in natural science' is measured using different clusters or subsets of items assessing knowledge in biology, chemistry, geology and physics, each subset of items represents a latent variable. If, for example, the items assessing knowledge in biology and the items measuring knowledge in physics rank the students quite differently, the different subsets of items might form subscales that contribute with unique variance to the distribution of students' score sums in natural science. Then the composite construct 'ability in natural science' is not sufficiently unidimensional and we should report one score in biology and one score in physics as opposed to a score in natural science – the overarching dimension. Therefore, if a theoretical composite construct is not sufficiently unidimensional, one might want to split the assessment instrument into as many parts as there are latent variables or subscales and do separate analyses. Principal component analysis of residuals might help investigate the dimensionality of the data.

### Principal component analysis of residuals

A principal component analysis converts a set of observations (the data) of correlated variables (the items) into a set of linearly uncorrelated variables called principal components. The first principal component has the largest possible variance, i.e. accounts for as much of the variability in the data as possible, and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to or uncorrelated with the other components. A principal component analysis therefore reveals the internal structure of the data in a way that best explains the variance in the data. Principal component analysis is closely related to factor analysis.

In a natural science achievement test the cluster of items in biology and the cluster of items in physics have 'ability in science' in common. If we remove the common latent trait from the data we are left with the residuals or the deviations from the Rasch model. If the residuals of the biology items correlate positively with the first principal component while the other items correlate negatively, the cluster of items in biology might share something else in common than the general underlying variable 'natural science' can 'explain'. If so, the items in biology represent an additional latent trait that might violate the hypothesis of unidimensional data and hence violate local independence[30,33–35].

Large variations in the percentage variance explained by each principal component is one way of generating a 'post hoc' hypothesis about multidimensionality in the data[24,25]. In principle, such hypotheses should come from theoretical and conceptual considerations. The hypothesis might be tested by applying the equating tests

and the $t$-test procedures in RUMM2030[29], and by estimating fractal indices based on a subtest analysis.

### Fractal indices and reliability indices specific to a subtest analysis

A set of $n$ items can be analysed either as $n$ items or as two composite items (subscales) where each subscale takes on the role of an item. The subtest analysis takes account of multidimensionality in the data, and fractal indices ($A$, $c$ and $r$) are estimated specific to the subtest. The value $A$ describes the variance common to all subscales, the value $c$ characterises the variance that is unique to the subscales and the variable $r$ is the correlation between the two subscales[29]. A subtest analysis performed on a data set with acceptable unidimensionality will return a high value for both $A$ and $r$ and a low value for $c$.

Reliability indices do not indicate whether a scale measures a unidimensional variable or not, but instead provide the value of the reliability on the assumption of unidimensionality[29]. In the presence of a multidimensional subscale structure, the variance of person estimates and hence the reliability indices inflate[34].

Further, comparing the overall test-of-fit index, i.e. the total item $\chi^2$, obtained when the analyses uses (i) the discrete items and (ii) the subscales as two items might indicate changes in fit to the model taking the multidimensionality into account.

### The parameterisations of the polytomous Rasch model, the thresholds and the likelihood ratio test

When the observed distance between the response categories on a rating scale is identical across the items, the data fit 'the rating scale parameterisation'[23] of the PRM best. If the distance is not the same across the items, 'the partial credit parameterisation'[36] is indicated. When applying the partial credit parameterisation, the 'thresholds' should be ordered.

A threshold is defined as the person location at which the probability of responding in one of two adjacent response categories is equal, and in the special case of dichotomous data this probability is 0·50. A polytomous item with an $m+1$ number of response categories has $m$ thresholds ($\tau_k$), where the index $k$ takes on values from 1 to $m$ and $x$ takes on values from 0 to $m+1$. The score $x$ indicates the number of $m$ thresholds a respondent has passed[37].

To treat the scales as linearly and directly related to the latent variable, where the succeeding response categories reflect successively more of the latent variable, we must examine whether the variables EDB and SE possess the properties of interval scales or are ordinal variables. If respondents use the rating scales in the questionnaire as expected, the observed succeeding thresholds should reflect successively more of the latent attitude and hence be ordered[38]. Disordered thresholds in the data violate the hypothesised ordering of response categories, meaning that respondents have not used the scales as expected. If so, the variables cannot be treated as interval variables[39].

The Fisher's likelihood ratio test available in RUMM might be used to assess the efficiency of the partial credit parameterisation as compared with the rating scale parameterisation of the PRM. The parameterisations are compared against each other for the same model specifications.

### Item discrimination, model fit, reliability and targeting

When an item, as part of a set of items, provides data that sufficiently fit a unidimensional Rasch model, the item provides an indication of attitude or ability along the latent variable. In Rasch analysis, this information is used to construct measures.

If the data do not fit the item characteristic curve – the theoretical expectation under the model – but rather approach a step function, the item is said to over-discriminate and the item might stratify the persons below and above a certain attitude estimate. If the data approach a constant function, the item is said to under-discriminate. Under-discriminating items tend to neither stratify nor measure. Strongly over- and under-discriminating items do not fit the Rasch model.

Fit residuals and item $\chi^2$ values are used to test how well the data fit the model[40]. Negative and positive item fit residuals indicate whether items over- or under-discriminate. Similarly, a person fit residual indicates how well a person's response pattern matches the expectation under the model[41,42].

Large $\chi^2$ indicate that persons with different attitudes do not 'agree on' item affective estimates, thus compromising the required property of invariance. To adjust $\chi^2$ probabilities for the number of significant tests performed, the probabilities are Bonferroni-adjusted[43] using RUMM2030[44].

Estimates of Cronbach's $\alpha$ and the person separation index (PSI) are used as indices of 'reliability'[45]. When the distribution of the items' threshold estimates matches the distribution of the persons' attitude estimates the instrument is well 'targeted'. Well-targeted instruments help reduce the measurement error.

## Method

### Frame of reference and data collection

One hundred randomly sampled public schools across Norway offering tenth grade were asked whether they could participate in a field-test trial for the 'national sample test' in science. The schools were contacted by regular mail on 21 November 2012, by email on 10 December 2012 and by telephone during the period 3–7 January 2013. As a result, 740 tenth-grade students with an age range from 14 to 15 years (48 % females and 9 % minority students) from twenty-seven public schools chose to take part in the voluntary field trial of the assessment

instruments. The number of participating schools was low as no incentive was offered and some schools experienced technical problems when enrolling their students in the test administration system.

Twenty-two out of the twenty-seven schools reported the number of students in the participating class. At these schools the number of students who actually responded to the achievement test and the questionnaire ranged from 67 to 100 % of the students, with an average of 81 %.

The field-trial data were collected during the period 16 January–15 February 2013. When logging on to the applied electronic assessment tool, each student was assigned to one out of four different electronic 'test booklets'. Each booklet contained science achievement test items and a student questionnaire that was completed at school within 90 min. Only one of these test booklets contained the EDB scale and the SE in science scale, and 178 students responded to this specific test booklet.

As the Scandinavian countries (Norway, Sweden and Denmark) have strong cultural and linguistic similarities, a student was defined as a majority student if at least one of his or her parents had been born in Scandinavia. Hence, a minority student in the present study is either an immigrant or a descendent of two immigrants (second generation).

### The engagement in dietary behaviour and self-efficacy scales, the achievement test, the socio-economic status indicator and the items asking for the students' cultural and linguistic background

All items in the EDB and the SE scales are reported in Table 1. The EDB scale, consisting of six items, is a revised version of the engagement in dietary habits scale reported by the authors[1]. Items 68 and 69 are at the personal level, items 70 and 71 are at the social level, and items 72 and 73 are at the global level. The SE in science scale, consisting of five items, is based on the SE in science scale and the control expectation scale applied in PISA[19]. Six-point rating scales, with the extreme response categories anchored with the phrases 'strongly disagree' (=1) and 'strongly agree' (=6), were applied for all items in the EDB and the SE scales.

The achievement test in the same field-test booklet as the EDB scale and the SE scale consisted of fifty-nine items, of which two were open-ended. The items were distributed across the competence aims in the science curriculum after grade ten and across the described cognitive categories.

An item asking for the number of books at home taken from the PIRLS (Progress in International Reading Literacy Study) student questionnaire[46] was applied as an indicator for SES. The categories for number of books at home were 0–10, 11–25, 26–100, 101–200 and >200 books. To help students decide the number of books, pictures of how ten, twenty-five, 100 and 200 books might look in shelves were provided.

**Table 1** The wording of the items in the engagement in dietary behaviour (EDB) scale and the self-efficacy (SE) in science scale (originally stated in Norwegian)

| Item | Context | Item phrasing; item 68–73 (EDB scale) and item 74–78 (SE scale) |
|---|---|---|
| 68 | Personal | I am concerned about eating healthy foods |
| 69 | Personal | I am concerned that there is a wide selection of healthy foods in the grocery stores I shop at |
| 70 | Social | I am concerned that most people in this country can afford to buy and eat healthy foods |
| 71 | Social | I am concerned that the cafeterias and vending machines at Norwegian schools and workplaces offer healthy foods |
| 72 | Global | I engage myself politically to ensure that the world's population will have good access to healthy foods |
| 73 | Global | I require that rich countries commit themselves to ensure that populations in poor countries have enough healthy food |
| 74 | – | I am confident that if I want to learn science properly, I am able to do so |
| 75 | – | I am quite sure that I can do an excellent job on science achievement tests |
| 76 | – | I am quite sure that I understand even the hardest subject matter in science |
| 77 | – | I am confident that I can do an excellent job in solving difficult tasks in science |
| 78 | – | I will do better in science than most in my class |

Student's cultural background was obtained from an item asking for the students' and the parents' birth place. The three categories for birth place were: (i) Norway; (ii) Sweden or Denmark; and (iii) 'Other'. The categories (i) and (ii) were merged into one category. The students also reported linguistic background: the language spoken at home most of the time. The two categories for linguistic background were: (i) Norwegian, Swedish or Danish; and (ii) 'Other'. Students' gender was available in the applied electronic national assessment tool.

## Results

### Differential item functioning in the engagement in dietary behaviour and the self-efficacy data

No item showed DIF associated with the person factor gender, but this finding might be a result of the rather few respondents in the sample. However, this implies that the items and the variable defining groups (gender) are conditionally independent given the person estimate corresponding to the total scale score (attitude). There were too few minority students in the sample to draw any meaningful conclusions regarding DIF associated with cultural and linguistic background.

### Response violations of local independence in the engagement in dietary behaviour data – disordered thresholds observed in a dependent subsequent item

The x-axis on Fig. 1 reports the person attitude levels on the EDB scale and the y-axis indicates probability. The six
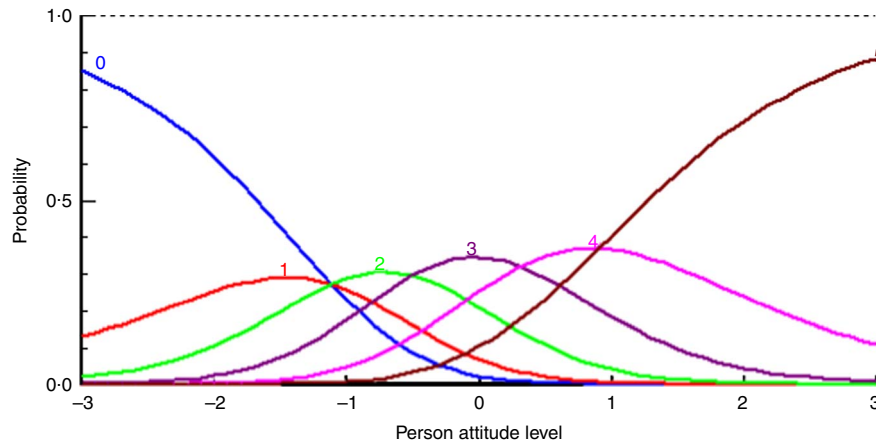
**Fig. 1** (colour online) The response category probability curves for item 69 of the engagement in dietary behaviour (EDB) scale. The curves marked 0–5 illustrate the probability of ticking off in each of the six response categories on the rating scale applied in the questionnaire as a function of the estimated attitude levels on the EDB scale. The line - - - - - is an asymptote (probability equals 1, i.e. 100 %)

curves marked 0–5 in Fig. 1 illustrate the probability of ticking off in each of the six response categories on the rating scale applied in the questionnaire as a function of the estimated attitude levels on the EDB scale, i.e. engagement in dietary habit. Figure 1 indicates that item 69 had disordered thresholds as the category curve marked '1' is not the most likely for any attitude level, and this was indeed considered a problem. Further, item 69 was dependent on item 68 and a subtest was created to absorb the dependency between items 68 and 69. The resulting super-item had disordered thresholds (not reported), but these were not considered a problem and were not rescored.

### Dimension violations of local independence in the engagement in dietary behaviour and the self-efficacy data

The correlation coefficient between the residual of each item and the first principal component was checked for both the EDB and the SE items respectively. Applying the equating tests procedure in RUMM2030, the *t*-test procedures indicated no problematic multidimensionality in any scale. No further subtest analyses were performed.

### Item discrimination, item fit and person fit in the engagement in dietary behaviour and the self-efficacy data
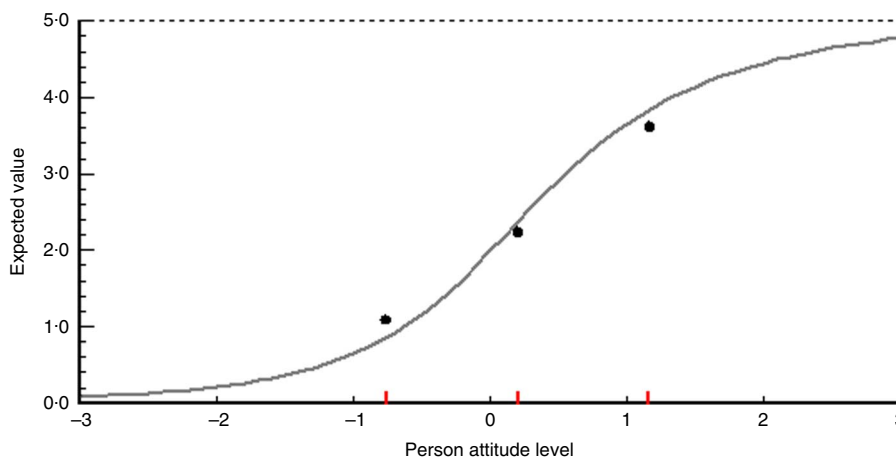
The *x*-axis on Fig. 2 indicates person attitude level and the *y*-axis indicates expected value (0–5) on the six-point rating scale applied in the questionnaire. The mean person attitude level of each of three class intervals is marked on the *x*-axis. The observed mean response category value for each class interval is plotted in the diagram (circles) and compared with the expected values described by the theoretical graphical representation of the PRM (curve).

When measured against the model, the analysis in Fig. 2 reveals that persons with low attitude levels on average tick off in response categories higher on the scale than expected when they respond to item 69. Likewise, persons with high attitude levels on average tick off in response categories lower on the scale than expected. Hence, item 69 is not able to discriminate as strongly as expected between persons with low and high attitude on the EDB scale.

Table 2 refers to scale, item location (i.e. item affective level), standard error, *z*-fit residual, degrees of freedom, $\chi^2$ value, $\chi^2$ probability, whether the item had disordered thresholds or was dependent on other items, and action taken to solve any problem. According to the item fit residual statistic (see bold value in Table 2) and the observed values' fit to the PRM (Fig. 2), item 73 was slightly under-discriminating. The item's fit to the PRM improved when the subtest of items 68 and 69 was created. The fit also improved in an additional analysis where item 69 actually was discarded (analysis not reported). Individual person fit residuals showed that twelve and twenty-three students had a *z*-fit residual outside the range $z = \pm 2.5$ on the EDB and the SE scale, respectively.

### Comparing the parameterisations of the polytomous Rasch model using likelihood ratio test and $\chi^2$ statistics

The likelihood ratio test was used to determine the best-fitting parameterisation. The likelihood values for the EDB scale were −1290·16 for the partial credit mode and −1294·40 for the rating mode. The likelihood ratio test $\chi^2$ statistic based on these two values was 8·48 and the probability that these two likelihood values would occur by chance alone, based on the 14 df, was 86 %.

**Fig. 2** (colour online) The item characteristic curve for item 73 of the engagement in dietary behaviour (EDB) scale, plotting the expected value of response on the six-point rating scale applied in the questionnaire *v.* person attitude level on the EDB scale. The line - - - - - is an asymptote. The mean person attitude level of each of three class intervals is marked on the *x*-axis. The observed mean response category value for each class interval (●) is compared with the expected values described by the theoretical graphical representation of the polytomous Rasch model (———)

**Table 2** Initial analysis of the engagement in dietary behaviour (EDB) scale and the self-efficacy (SE) in science scale applying the partial credit parameterisation of the polytomous Rasch model

| Item | Scale | Loc | SE | Res | df | $\chi^2$ | df | $P(\chi^2)$ | Disord | Dep | Action |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 68 | EDB | −0·26 | 0·08 | 0·16 | 131·1 | 4·8 | 2·0 | 0·092 | | | |
| 69 | EDB | −0·30 | 0·07 | −1·14 | 130·3 | 2·0 | 2·0 | 0·370 | x | 68 | Subtest with 68 |
| 70 | EDB | −0·39 | 0·07 | −0·93 | 131·1 | 2·5 | 2·0 | 0·285 | | | |
| 71 | EDB | −0·20 | 0·07 | −0·58 | 131·1 | 8·2 | 2·0 | 0·017 | | | |
| 72 | EDB | 0·79 | 0·07 | 0·89 | 131·1 | 0·3 | 2·0 | 0·855 | | | |
| 73 | EDB | 0·36 | 0·07 | **2·68** | 131·1 | 2·0 | 2·0 | 0·373 | | | |
| 74 | SE | −0·75 | 0·10 | 1·77 | 121·8 | 1·0 | 2·0 | 0·604 | | | |
| 75 | SE | −0·77 | 0·11 | −0·86 | 121·8 | 3·6 | 2·0 | 0·162 | | | |
| 76 | SE | 0·29 | 0·10 | −1·71 | 120·3 | 4·5 | 2·0 | 0·107 | | | |
| 77 | SE | 0·23 | 0·11 | −1·89 | 121·0 | 2·5 | 2·0 | 0·292 | | | |
| 78 | SE | 1·00 | 0·11 | 2·47 | 98·1 | 11·0 | 2·0 | 0·004 | | | |

Data presented are item number, scale, item location (i.e. item affective level), standard error, *z*-fit residual, degrees of freedom, $\chi^2$ value, $\chi^2$ probability, whether the item had disordered thresholds or was dependent on other items, and action taken to solve any problem.

The corresponding values for the SE in science scale resulted in a probability of 52 %.

Table 3 refers to total item $\chi^2$, df, $\chi^2$ probability, the PSI, mean *z*-fit residual, mean person location (i.e. attitude level) and SD. Based on the $\chi^2$ statistic in Table 3, the rating scale parameterisation might provide the best fit for the data from the SE in science scale. The scales' item fit residual mean and SD deviated slightly from their expected values, i.e. 0 and 1, as their values were 0·21 (1·11) and 0·31 (2·67), respectively (Table 3).

### Reliability estimates and the targeting of the engagement in dietary behaviour and the self-efficacy scales

Cronbach's $\alpha$ coefficient was estimated using the statistical software package IBM SPSS Statistics 20. The $\alpha$ coefficients for the EDB scale data and the SE in science scale data were 0·86 and 0·92, respectively. The PSI were 0·79 for the EDB scale and 0·90 for the SE in science scale (Table 3). The average person location values were 0·08 for the EDB scale and 0·59 for the SE in science scale (Table 3). Except for item 71, the EDB items at the global level had a higher affective level than the items at the social level, and the items at the social level had a higher affective level than the items at the personal level.

### Reliability and targeting of the achievement test in science

In the test booklet under consideration, five of the fifty-nine achievement test items, two of which were open-ended and one was scored polytomously (ordered score values), were discarded. One of the items was discarded due to technical issues in the electronic testing system and four items were discarded as they under-discriminated. The fifty-four remaining achievement test items had acceptable fit to the Rasch model and constituted a well-targeted (mean person

**Table 3** Summary statistics for the engagement in dietary behaviour (EDB) scale and the self-efficacy (SE) in science scale

| Scale | Model | $x^2$ | df | $P(x^2)$ | PSI | z | SD | Loc | SD | Disordered or subtest |
|-------|-------|-------|----|----------|-----|----|----|-----|----|----------------------|
| EDB | Partial | 19·7 | 12 | 0·07 | 0·84 | 0·18 | 1·43 | 0·15 | 1·18 | 69 |
| EDB | Rating | 19·7 | 12 | 0·07 | 0·84 | 0·06* | 1·56 | 0·14 | 1·18 | |
| *EDB* | *Partial* | *23·1* | *10* | *0·01* | *0·81* | *0·24* | *1·16* | *0·07* | *1·03* | *Subtest (items 68 and 69)* |
| SE | Partial | 22·5 | 10 | 0·01 | 0·90 | 0·05* | 2·03 | 0·61 | 2·21 | |
| *SE* | *Rating* | *12·8* | *10* | *0·24* | *0·90* | *0·31** | *2·67* | *0·59* | *2·16* | |

Data presented are scale, model parameterisation, total item $x^2$, degrees of freedom, $x^2$ probability, person separation index, mean $z$-fit residual and its standard deviation, mean person location (i.e. attitude level) and its standard deviation, whether the item had disordered thresholds or was dependent on other items, and any subtest structure created.
Negative estimates are marked *. Analyses marked in italic are considered in Table 4.

location $= -0.221$) and sufficiently reliable ($\alpha = 0.87$ and PSI $= 0.87$) cluster of achievement test items measuring ability in science.

### *Exploring the relationships between the variables*

Table 4 shows the Pearson correlation coefficients between the estimated attitude levels on the EDB scale applying the partial credit parameterisation of the PRM after creating a subtest structure of items 68 and 69 (the analysis is marked in italics in Table 3), the estimated attitude levels on the SE in science scale applying the rating parameterisation of the PRM (the analysis is marked in italics in Table 3) and the ability in science as measured by the achievement test. The point biserial coefficients between these scales and gender (1 = girl and 2 = boy) and the Spearman $\rho$ between these scales and SES, as measured by the number of books at home, are also reported in Table 4. All of the bivariate correlations above 0·20 in Table 4 were statistically significantly different from zero at the 1 % level.

Table 4 shows that the estimated attitude levels on the EDB scale were positively correlated with the estimated attitude levels on the SE in science scale, that the estimated attitude levels on the EDB scale were negatively correlated with gender (i.e. in favour of girls) and that the estimated attitude levels on the EDB scale's correlation with SES (number of books at home) was close to zero. Further, SES was positively correlated with the estimated attitude levels on the SE in science scale and with scientific literacy – the ability in science as measured by the achievement test. On average, boys did not report higher SE in science or higher SES than girls (not reported in Table 4).

### Discussion and conclusions

From a conceptual point of view, the EDB scale has a structure like that of 'multiple domains' consisting of the three contextual 'levels' referred to as personal, social and global. These levels are equally weighted in the entire scale. If we discard item 69 (reversed thresholds) from the personal level, that aspect is under-represented and we are left with a conceptually unbalanced scale. In a purely unidimensional instrument, omitting an item would probably not have played an important role. The fact that

**Table 4** The correlation matrix (the estimated correlation coefficients, $\rho$, with significance levels, $P$) for the variables engagement in dietary behaviour (EDB), ability in science (Ability), self-efficacy (SE) in science, socio-economic status (SES) and gender

| | EDB | | Ability | | SE | |
|---|---|---|---|---|---|---|
| | $\rho$ | $P$ | $\rho$ | $P$ | $\rho$ | $P$ |
| Gender | −0·26 | 0·00 | 0·04 | 0·64 | 0·07 | 0·36 |
| SES | 0·04 | 0·63 | 0·26 | 0·00 | 0·25 | 0·00 |
| SE | 0·26 | 0·00 | 0·43 | 0·00 | – | – |
| Ability | −0·08 | 0·33 | – | – | – | – |

SES, socio-economic status (the number of books at home).

the fit of item 73 improved when item 69 was omitted supports this idea. The underlying composite latent variable changes somewhat and becomes more dominated by the social and the global perspectives. Hence, item 73 reflecting the global perspective has a better fit to the model.

There is a trade-off between a conceptually balanced scale and the model fit. By retaining item 69, we manage to retain the construct and keep as much information about the person's attitude levels as possible. The subtest structure helps absorb the dependency and avoid violating the requirement of local dependence. Hence, retaining item 69 can be defended from both a conceptual point of view and a methodological perspective. The observed disordering in the super-item of item 68 and item 69 is viewed a symptom of the extra dependency of those items and is not considered a problem.

Except for item 71, the EDB scale seems to be 'stage specific' with the items measuring global level at the highest affective level and the items measuring personal level at the lowest level. Further validations of the EDB construct are needed and we suggest that item 68 be modified to make it less broad so as to avoid the observed redundancy in the data provided by items 68 and 69.

No item showed DIF related to gender, but the sample contained too few participants to draw a robust conclusion. In addition, there were too few minority students in the sample to conclude anything about DIF associated with cultural and linguistic background.

The hypothesis of unidimensionality and the requirement of local independence hold for both the EDB scale

and the SE in science scale after creating the subtest consisting of items 68 and 69. We might conclude that our two scales represent interval variables and hence construct measurements. This assumption is crucial in order to investigate relationships between the scales and the person factors.

Based on the likelihood ratio test, we concluded that the partial credit parameterisation does not contain more information about the data than does the rating parameterisation for either the EDB scale or the SE in science scale. The $\chi^2$ statistic indicated that the data from the SE in science scale had a somewhat better fit to the rating parameterisation. The partial credit parameterisation was applied for the EDB scale and the rating parameterisation was applied for the SE in science scale.

The analyses indicate that the scale in focus of our study – the EDB scale – had excellent targeting, sufficient fit to the PRM and acceptable reliability at the group level. The SE in science scale was well targeted, had sufficient fit to the PRM and acceptable reliability. As the rating scale parameterisation provided a good fit for the data from both scales, we can conclude that the distances between the thresholds were fairly equal across the items within each scale.

SES (the number of books at home) seems to predict SE in science and ability in science. The number of books at home explained approximately 6–7 % of the variance in both SE in science and ability in science. As SE explained 18 % of the variance in ability in science, self-reported expectations about success are clearly useful predictors for achievement. However, the relationships reported do not justify SES as an explicit predictor for tenth-grade students' engagement in dietary behaviour at the personal, social and global level.

On average, girls seem to attain higher engagement in dietary behaviour than boys. There is no sign that boys on average report either higher SES or SE in science than girls. Gender and SE in science each explained approximately 6–7 % of the variance in engagement in dietary behaviour.

Given the limited explanatory power of the variables considered, further studies should consider other demographic factors that might play a role when specifying and identifying a structural model for a multiple regression analysis (structural equation modelling analysis). Effects of parents' education on children's dietary behaviours, the home nutrition environment and students' own nutrition literacy might influence students' responses to the EDB scale. Level of physical activity might influence individuals' nutrition literacy and thereby their EDB level. It could also be interesting to study whether being on certain diets or suffering from illnesses influencing food intake have certain impacts on individuals' EDB level. People who 'often' use the Internet to search for health-related issues might, on average, have different attitudes associated with nutrition than others. Political engagement, such as being a member of a political party, might in general influence people's engagement in a variety of health-related issues. In other samples of respondents, one could possibly study the effects of parenthood and how smoking and the use of alcohol influence responses to the EDB scale.

The Rasch analyses imply that the scales measuring engagement in dietary behaviour at the personal, social and global levels and SE in science both construct measures. The study of relationships between the variables implied that girls and those students who expected to perform well in science reported higher levels of engagement in dietary behaviour than other groups of students. Our study indicates that students' engagement in the dietary behaviour aspect of CNL seems to be associated with students' SE in science but not their actual ability in science. Surprisingly, SES did not predict tenth-grade students' engagement in dietary behaviour at the personal, social and global levels. These conclusions build on high-quality data from students at randomly sampled schools.

More quantitative research applying diverse, valid and reliable measures of the different aspects of CNL, SE, SES and proficiency in health and nutrition is needed to validate our conclusions and understand how background factors influence individuals' CNL.

## Acknowledgements

## References

1. Guttersrud Ø, Østerholt Dalane J & Pettersen KS (2014) Improving measurement in nutrition literacy research using Rasch modelling: examining construct validity of stage-specific 'critical nutrition literacy' scales. *Public Health Nutr* **17**, 877–883.
2. Nielsen-Bohlman L, Panzer AM & Kindig DA (editors) (2004) *Health Literacy: A Prescription to End Confusion.* Washington, DC: National Academies Press.
3. American Medical Association (1999) Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, Health Literacy: Report of the Council on Scientific Affairs. *JAMA* **281**, 552–557.
4. 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.

5.  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.

6.  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.

7.  Pettersen KS (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. PhD Thesis, University of Oslo.

8.  Sykes S, Wills J, Rowlands G *et al.* (2013) Understanding critical health literacy: a concept analysis. *BMC Public Health* **13**, 150.

9.  Zimmerman BJ (2000) Attaining self-regulation: a social cognitive perspective. In *Handbook of Self-regulation*, pp. 13–40 [M Boekaerts, PR Pintrich and M Zeidner, editors]. San Diego, CA: Academic Press.

10. Bandura A (1977) Self-efficacy: toward a unifying theory of behavioral change. *Psychol Rev* **84**, 191–215.

11. Bandura A (1988) Organizational application of social cognitive theory. *Aust J Manage* **13**, 275–302.

12. Schwarzer R & Fuchs R (1996) Self-efficacy and health behaviours. In *Predicting Health Behaviour: Research and Practice with Social Cognition Models*, pp. 163–196 [M Conner & P Norman, editors]. Buckinghamshire: Open University Press.

13. Schwarzer R (2008) Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. *Appl Psychol* **57**, 1–29.

14. Gutiérrez-Doña B, Lippke S, Renner B *et al.* (2009) How self-efficacy and planning predict dietary behaviors in Costa Rican and South Korean women: a moderated mediation analysis. *Appl Psychol Health Well-Being* **1**, 91–104.

15. Norwegian Directorate for Education and Training (2006) *The Knowledge Promotion Curricula*. Oslo: NDET.

16. Adler NE & Ostrove JM (2006) Socioeconomic status and health: what we know and what we don't. *Ann N Y Acad Sci* **896**, 3–5.

17. Ostrove JM & Adler NE (1998) The relationship of socioeconomic status, labor force participation, and health among men and women. *J Health Psychol* **3**, 451–463.

18. Winkleby MA, Jatulis DE, Frank E *et al.* (1992) Socioeconomic status and health: how education, income, and occupation contribute to risk factors for cardiovascular disease. *Am J Public Health* **82**, 816–820.

19. Organisation for Economic Co-operation and Development, Programme for International Student Assessment (2001) *Knowledge and Skills for Life*. Paris: OECD.

20. Ammermueller A & Pischke J-S (2009) Peer effects in European primary schools: evidence from PIRLS. *J Labour Econ* **27**, 315–348.

21. Andrich D (1988) *Rasch Models for Measurement*. Beverly Hills, CA: SAGE Publications.

22. Rasch G (1960) *Probabilistic Models for Some Intelligence and Achievement Tests*. Copenhagen: Danish Institute for Educational Research. Expanded edition 1983. Chicago, IL: MESA Press.

23. Andrich D (1978) A rating scale formulation for ordered response categories. *Psychometrika* **43**, 561–573.

24. Smith EV (2002) Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas* **3**, 205–231.

25. Linacre JM (1998) Detecting multidimensionality: which residual data-type works best? *J Outcome Meas* **2**, 266–283.

26. Andrich D & Hagquist C (2001) *Taking Account of Differential Item Functioning through Principals on Equating. Research Report* no. 12. Perth, WA: Social Measurement Laboratory, Murdoch University.

27. 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.

28. 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 Meas* **10**, 247–265.

29. RUMM (2009) *Interpreting RUMM2030 Part IV Multidimensionality and Subtests in RUMM*, 1st ed. Perth, WA: RUMM Laboratory Pty Ltd.

30. Marais I & Andrich D (2008) Formalising dimension and response violations of local independence in the unidimensional Rasch model. *J Appl Meas* **9**, 200–215.

31. Smith EV (2005) Effect of item redundancy on Rasch item and person estimates. *J Appl Meas* **6**, 147–163.

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. Andrich D & Kreiner S (2010) Quantifying response dependence between two dichotomous items using the Rasch model. *Appl Psychol Meas* **34**, 181–192.

34. Marais I & Andrich D (2008) Effects of varying magnitude and patterns of response: dependence in the unidimensional Rasch model. *J Appl Meas* **9**, 105–124.

35. 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.

36. Wright BD & Masters GN (1982) *Rating Scale Analysis: Rasch Measurement*. Chicago, IL: MESA Press.

37. 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.

38. Andrich D (1995) Models for measurement, precision and the non-dichotomization of graded responses. *Psychometrika* **60**, 7–26.

39. Singh J (2004) Tackling problems with item response theory: principles, characteristics and assessment, with an illustrative example. *J Bus Res* **57**, 184–208.

40. Smith RM & Plackner C (2009) The family approach to assessing fit in Rasch measurement. *J Appl Meas* **10**, 424–437.

41. Andrich D (1985) An elaboration of Guttman scaling with Rasch models for measurement. In *Sociological Methodology*, pp. 33–80 [N Brandon-Tuma, editor]. San Francisco, CA: Jossey-Bass.

42. 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.

43. Bland JM & Altman DG (1995) Multiple significance tests: the Bonferroni method. *BMJ* **310**, 170.

44. RUMM (2009) *Extending the RUMM2030 Analysis*, 7th ed. Perth, WA: RUMM Laboratory Pty Ltd.

45. Traub RE & Rowley GL (1981) Understanding reliability. *Educ Meas Issues Pract* **10**, 37–45.

46. International Association for the Evaluation of Educational Achievement (2005) Progress in International Reading Literacy Study PRLS 2006 Student Questionnaire. http://timssandpirls.bc.edu/PDF/P06_StudentQuestionnaire.pdf (accessed December 2012).