Advancing the measurement of Critical Nutrition Literacy in adolescents

Desire Alice Naigaga

OsloMet Avhandling 2021 nr 43

OSLO METROPOLITAN UNIVERSITY
STORBYUNIVERSITETET

Advancing the measurement of Critical Nutrition Literacy in adolescents

Desire Alice Naigaga

Dissertation for the degree of philosophiae doctor (PhD)

Department of Nursing and Health Promotion
Faculty of Health Sciences
OsloMet – Oslo Metropolitan University

Autumn 2021

OsloMet Avhandling 2021 nr 43

ISSN 2535-471X (trykket)
ISSN 2535-5454 (online)

ISBN 978-82-8364-347-3 (trykket)
ISBN 978-82-8364-361-9 (online)

OsloMet – storbyuniversitetet
Universitetsbiblioteket
Skriftserien
St. Olavs plass 4,
0130 Oslo,
Telefon (47) 64 84 90 00

Postadresse:
Postboks 4, St. Olavs plass
0130 Oslo

Trykket hos Byråservice

Trykket på Scandia 2000 white, 80 gram på materiesider/200 gram på coveret

ABSTRACT

Aim: To evaluate the psychometric properties of scales measuring adolescents' proficiency to critically appraise nutrition information, their engagement in dietary behaviour and self-efficacy in the science topic 'Body and Health', and test a theoretical model linking these latent traits.

Methods: A sample of 1622 tenth graders at 58 randomly sampled schools in Norway voluntary responded to a questionnaire containing the three scales. The scale measuring adolescents' proficiency to critically appraise nutrition information was validated by applying the partial credit parameterization of the polytomous unidimensional Rasch model. The scale measuring self-efficacy in the science topic 'Body and Health' was validated by using the partial credit parameterization of the polytomous unidimensional Rasch model and confirmatory factor analysis. The theoretical model linking the latent traits was tested by using structural equation modelling. Structural equation modelling was preferred to regression models as this method properly handles latent traits measured with standard errors. The data was tested up against the appropriate Rasch model by using the statistical software package RUMM2030. The confirmatory factor model and the "structural equation modelling" model were estimated by using the statistical software package Lisrel 9. Owing to data at the ordinal measurement level, "diagonally weighted least square" estimation was applied. "Goodness of fit" indexes were also estimated by using robust maximum likelihood estimation, as published target values typically rely on this type of estimation.

Main results: The data collected sufficiently fit the polytomous unidimensional Rasch model. Confirmatory factor analysis displayed strong standardized factor loadings, which strengthened the idea of one underlying latent factor (unidimensional scale). The scales were slightly less than optimally targeted for the sample, as the distribution of person location estimates were somewhat shifted toward higher values than the distribution of item threshold locations. Overall, the scales were reliable and able to separate between students with different standing on the latent trait. The theoretical model was empirically strengthened, which means that the model implied variance-covariance matrix fairly well reproduced the observed variance-covariance matrix.

Conclusions: Valid and reliable scales for measuring adolescents' critical nutrition literacy were established, and self-efficacy in the health-related science topic explained variation in students' critical nutrition literacy. Hence, there is a relation between adolescents' self-perceived nutrition literacy and self-perceived health-related scientific literacy.

Keywords: adolescents, confirmatory factor analysis, critical nutrition literacy, diagonally weighted least squares, evaluation of information, engagement in dietary behaviour, maximum likelihood estimation, nutrition literacy, ordinal data, Rasch modelling, self-efficacy, structural equation modelling

SAMMENDRAG

Hensikt: Å evaluere de psykometriske egenskapene til måleskalaer som måler ungdoms evne til å kritisk vurdere ernæringsinformasjon (CNL-E), deres engasjement i kostholdsendrende atferd (CNLEng) og mestringsforventning i kjerneelementet «Kropp og helse», samt å teste en modell som beskriver sammenhenger mellom disse latente trekkene.

Metoder: Måleskalaene ble validert ved bruk av Rasch-modellering og konfirmerende faktoranalyse i et utvalg på 1622 elever på 10. årstrinn ved 58 tilfeldig utvalgte skoler, og den teoretiske modellen ble testet ved bruk av strukturell ligningsmodellering.

Hovedresultater: Dataene hadde tilstrekkelig tilpasning til polytom Rasch-modell (PCM). Samlet sett var de tre måleskalaene pålitelige, men de kunne vært bedre tilpasset elevenes dyktighet. De empiriske dataene styrket den teoretiske modellen.

Konklusjon: Valide og reliable måleskalaer for måling av ungdommers kritiske ernæringskompetanse ble etablert, og mestringsforventning, som er et helserelatert kjerneelement i naturfag, forklarte variasjon i kritisk ernæringskompetanse.

SUMMARY OF THE PAPERS

In paper I, I applied Rasch modelling to validate a five-item scale measuring perceived SEBH. Validation of the psychometric properties of the SEBH scale yielded a reliable scale with an overall acceptable fit to the partial credit model of the polytomous unidimensional Rasch model. However, the scale was not optimally targeted, and could benefit from including items with a wider range of item "difficulties", or more precisely, item thresholds.

In paper II, I assessed the psychometric properties of the newly developed CNL-E scale. Confirmatory factor analysis supported the analyses of dimensionality by confirming the underlying factor structure being measured by the items, which was consistent with the results of Rasch modelling. The items in this scale measured how well the respondents felt that they could evaluate nutrition information from different sources, and this aspect of "different sources" probably introduces, at a theoretical level, some multidimensionality into the scale. Using Rasch modelling revealed that the CNL-E scale had acceptable overall fit to the partial credit parameterization of the polytomous unidimensional Rasch model. Although the scale could have been better targeted at students' CNL levels, with more items targeting persons at the lower trait locations, the software estimated a high reliability coefficient. Even though there was a slight violation of the assumption of unidimensionality, multidimensional Rasch modelling and confirmatory factor analysis confirmed that the magnitude of multidimensionality was acceptable. In addition to providing an empirically sound measure of CNL-E, analyses implied that critical evaluation of nutrition information requires that persons are able to determine how reliable sources are, and transfer knowledge from other disciplines such as science, in order to judge the value of nutrition information.

In paper III, I tested the relationships between the latent traits reported in paper I and paper II by empirically testing a simple structural equation model. The focus was on covariance structure – not cause-effect or causal relations. The SEM model tests the theoretical assumptions that CNL-E and CNLEng are two distinct (cf. discriminant validity) but related aspects of CNL, and that both aspects are at least related to, if not are effects of, SEBH. Providing a theoretically and empirically founded inquiry into how the two aspects of CNL are associated, was important as research into this association is scarce; likewise, the association between self-efficacy and CNL is also inconclusive in different studies involving adolescents. I hypothesized that perceived self-efficacy *directly* influences adolescents' "critical evaluation of nutrition information" and that perceived self-efficacy *indirectly* influences adolescents' "critical evaluation of nutrition information" via the mediator "engagement in dietary behaviour". Owing to rating-scale items at the ordinal measurement level and a large sample size (N>1000), the SEM model was estimated using "diagonally weighted least squares" (DWLS) based on a polychoric covariance matrix. Evaluation of model fit supported the theoretically derived relationships between the three latent traits "critical evaluation of nutrition information" (CNL-E), "engagement in dietary behaviour" (CNLEng) and "perceived self-efficacy in Body and Health" (SEBH) in adolescents.

Table of contents

List of original papers

1. **Naigaga, D. A.,** Pettersen, K. S., Henjum, S. & Guttersrud, Ø. (2019). Assessing adolescent self-efficacy in 'body and health'-Exploring the psychometric properties of the SEBH scale. *Nordic Studies in Science Education, 15*(2), 145-158.
   DOI: **https://doi.org/10.5617/nordina.5913**
2. **Naigaga, D. A.,** Pettersen, K. S., Henjum, S., & Guttersrud, Ø. (2018). Assessing adolescents' perceived proficiency in critically evaluating nutrition information. International *Journal of Behavioral Nutrition and Physical Activity, 15*(1), 61.
   DOI: **https://doi.org/10.1186/s12966-018-0690-4**
3. **Naigaga, D. A.,** Pettersen, K. S., Henjum, S., & Guttersrud, Ø. (2022). Relating aspects of adolescents' critical nutrition literacy at the personal level. Nutrire 47, 1 (2022).
   DOI: **https://doi.org/10.1186/s41110-021-00149-1**

List of Figures

List of Tables

Abbreviations

Akaike Information criteria	AIC
Analysis of variance	ANOVA
Asymptotic distribution free	ADF
Comparative fit index	CFI
Confirmatory Factor Analysis	CFA
Critical evaluation of nutrition information	CNL-E
Critical Health Literacy	CHL
Critical Nutrition Literacy	CNL
Critical sample	CN
Degrees of Freedom	Df
Diagonally weighted least squares	DWLS
Differential item functioning	DIF
Distinct values	DV
Elaboration Likelihood model	ELM
Engagement in dietary behavior	CNLEng
Expected Parameter Change	EPC
Free parameters	FP
Front-Of-Package	FOP
Functional health literacy	FHL
Functional nutrition literacy	FNL
Food literacy	FL
Goodness of fit	GOF
Guideline daily amounts	GDA
Health literacy	HL
Interactive health literacy	IHL
Interactive nutrition literacy	INL
Item characteristic curve	ICC
Maximum Likelihood estimation	MLE
Media literacy	ML
Missing at random	MAR
Missing completely at random	MCAR

Modification Index	MI
Non-centrality parameter	NCP
Non-communicable diseases	NCDs
Non-normed fit index	NNFI
Nutrition literacy	NL
Organization for Economic Co-operation and Development	OECD
Partial credit model	PCM
Person separation index	PSI
Principal component	PC
Principal component analysis	PCA
Programme for International Student Assessment	PISA
Polytomous Unidimensional Rasch Model	PURM
Rasch Modelling	RM
Rasch Unidimensional Measurement Model	RUMM
Rating Scale Model	RSM
Root Mean Square Error of Approximation	RMSEA
Satorra-Bentler	SB
Self-efficacy in 'Body and Health'	SEBH
Scientific Literacy	SL
Standard Deviation	SD
Standardized root mean square residual	SRMR
Structural Equation Modelling	SEM
Socio cognitive theory	SCT
Tucker-Lewis Index	TLI
World Health Organization	WHO

Introduction/Background

The focus of this chapter is to present an overview of the main concepts addressed in the present study namely adolescents, nutrition literacy (NL), critical nutrition literacy (CNL) and self-efficacy. First off, I begin with a description of the population under study, which is adolescents. Following this, I describe what NL and CNL are, with emphasis on adolescents. I also highlight why it is important to focus on the CNL issues that concern this target group. Thirdly, I introduce how individual characteristics specifically self-efficacy beliefs might influence the CNL of adolescents. I then present the hypotheses and conclude the chapter with the research objectives of the study.

The World Health Organization (WHO) defines an adolescent as an individual between the ages of 10 and 19 years old (WHO, 2005, p.1). Adolescence marks the transition from dependence to independence in several aspects of one's life. One such area is in making decisions regarding their nutritional wellbeing. As adolescents assume more responsibility for their dietary choices, they also become more active in food-related communications, seeking out nutrition information from various sources including the media, peers, and social networks. However, because some of this information may be a result of a shared understanding of what 'good' nutrition is between the adolescents, fad-driven by culture, or merely the opinions of others, it is important that adolescents are able to identify what information is correct and relevant to their individual novel needs, from the 'jungle' of nutrition information available to them (Pettersen, 2005). Additionally, adolescents are highly impressionable and prone to social pressure, making it important that before adolescents can apply this information to take decisions, they have the critical skills required to interpret (understand), and evaluate (appraise) nutritional messages for any bias, misinformation and also establish the credibility of the sources of this information. Critical thinking refers to a type of thinking that involves careful consideration of evidence; it involves evaluation, critical appraisal, interpretation and the use of skills such as analysis, evaluation and inference (Profetto-McGrath, 2005).

The use of critical thinking skills like the aforementioned aligns with the concept of critical nutrition literacy (CNL), which is, beside functional nutrition literacy (FNL) and interactive nutrition literacy (INL), a main domain within the greater framework of nutrition literacy (NL). Being nutrition literate means that the adolescent has the capacity to access, interpret and appraise basic nutrition information and tools needed to take informed nutrition choices (Silk et al, 2008). There is ample evidence to support the importance of emphasizing nutrition literacy as part of the public health approach to nutrition promotion. Studies show that there is a direct and positive impact of NL on different aspects of one's diet including dietary quality and use of nutrition information; higher levels of NL are associated with better dietary quality, a higher likelihood to use available nutrition information while making food choices, and adherence to dietary patterns (Cha et al., 2014; Gibbs, 2016; Taylor et al., 2019; Zoellner et al., 2011). Moreover, given the strong association between nutrition and health outcomes, it is judicious to view NL as one of the 'drivers' of health promotion. From this perspective, identifying and addressing the NL needs of individuals is one avenue through which to channel efforts targeting the prevention of onset and exacerbation of diet-related conditions such as non-communicable diseases (NCDs) and obesity that are fast becoming a public health challenge. It is worth noting that whereas research in the field of NL has increased, adolescents remain an underserved target group. This is worrying given the significance of this life stage in shaping one's lifelong dietary practices and nutritional outcomes. For example, Lifshitz (2008) reports that up to 75%-80% of obese adolescents will remain obese even as adults. With findings like this, one might say that by focusing on promoting NL during adulthood, efforts appear reversed, instead they ought to focus on the NL needs of adolescents as this is the formative life stage for healthy dietary behaviour.

Schools play a pivotal role towards adolescents' development of skills that are associated with CNL through subjects in the curriculum like home economics in which they are taught practical skills such as food preparation skills, meal planning. Adolescents need these skills to use the nutrition information available to them from different sources while at school and away from school settings. Whereas various factors may influence how adolescents apply the information that they access from school and other sources, one noteworthy variable is their level of perceived self-efficacy. Self-efficacy influences what individuals do with the knowledge they have, how they approach tasks, behaviours and cognitive processes such as information processing. Within academic settings self-efficacy may refer to an individuals' convictions that they can successfully perform given academic tasks at designated levels (Britner & Pajares, 2006). 'A learner's convictions that they can successfully perform given academic tasks at designated levels' (Zimmerman, 2000; Schunk, 1991). Findings show that self-efficacy beliefs extend beyond the classroom and into daily dietary practices such as food selection (Massar & Malmberg, 2017).

Research shows that when adolescents encounter 'contradicting' or confusing nutrition information, some of them evaluate the quality of the nutrition information by looking for scientific cues and evidence (Barzilai & Zohar, 2012). The use of evidence-based information requires adolescents to understand how to use science knowledge as a yardstick to establish the value of science-related information such as nutrition information and the processes involved to create science knowledge. This requirement resonates with the concept of scientific literacy (SL), defined by the Organization for Economic Co-operation and Development (OECD) as “the ability to engage with science-related issues, with the ideas of science, as a reflective citizen” (OECD, 2013, p.7).

The need to address the CNL of adolescents

There is evidence that whereas adolescents have increased access to a multitude of nutrition information, they are often unable to use it to make dietary decisions that are beneficial for their overall nutritional wellbeing because they find nutrition information ‘confusing’. To address this, it is of utmost importance that researchers explore how to help adolescents navigate the ‘jungle’ of nutrition information so that they can correctly use the information take appropriate dietary actions. There is also need to explore and account for the influence of personal attributes such as self-efficacy on how adolescents appraise and apply nutrition information. One way to achieve this is by advancing measurement within the domain of CNL. This thesis exemplifies how to achieve some shifts in measurement studies by evaluating the psychometric properties the newly developed scales measuring the two aspects of CNL, namely, critical evaluation of nutrition information (CNL-E) and engagement in dietary behaviour (CNLEng) of tenth grade adolescents attending randomly sampled Norwegian schools.

Research hypothesis

The falsifiable hypothesis of the present study was that critical evaluation of nutrition information (CNL-E) is associated with engagement in dietary behaviour (CNLEng) and perceived self-efficacy in the science subject of ‘Body and Health’ (SEBH) taught in the tenth grade of Norwegian lower secondary school.

Specific objectives of the study

1. To examine the psychometric properties of a scale measuring SEBH (Paper I) and two newly developed CNL scales measuring the two aspects of CNL namely, CNL-E and CNLEng (Paper II).
2. To develop and test a theoretically derived structural equation model, which links CNL-E, CNLEng and SEBH at the personal level (Paper III).

Theoretical underpinnings of the study

This chapter is structured as follows: in the research philosophy of the study, I describe the research paradigm through which I viewed the study. Herein, I describe the theoretical perspective adopted, the ontological considerations made, the epistemology used, the data collection methods applied, data analysis methods applied, and the ethical considerations made. Following this, I describe the main theories that inspired the present study relating to the population under study (adolescents), the traits of interest namely, nutrition literacy (NL), critical nutrition literacy (CNL), self-efficacy (SE) and scientific literacy (SL). I conclude the chapter with a proposed theoretical model linking the main influences of CNL in adolescents based on existing literature and empirical evidence.

Research philosophy of the study

Research philosophy guides the selection and identification of the most appropriate approach to advance understanding of phenomena, as well as the factors that influence its occurrence, impact or perception. Researchers use research paradigms, which are 'worldview lenses' that guide researchers to help them to understand research findings and make sense of natural phenomena (Kivunja & Kuyini, 2017). Research paradigms contain the researcher's premises about the research that is, the theoretical perspective, ontology, epistemology, methodology, methods and axiology.

For the present study, I situated myself within the *scientific paradigm*, which is concerned with connecting the natural world to the scientific world through research (Reynolds et al., 2012). Accordingly, in the present study, I sought to advance understanding and achieve a broader grasp of the dimensions that comprise the phenomena of interest (CNL, SEBH) by linking them through measurement using scales that are comprised of items measuring the observable characteristics thought to be influenced by their respective latent traits. Within the scientific paradigm, I adopted the *post-positivist theoretical perspective* also known as 'empirical science', which generates knowledge that is considered observable and measurable (Abu-Alhaija, 2019; Assalahi, 2015). In the present study, we yielded empirical knowledge on how to measure the unobservable (latent) traits of CNL and SEBH using their respective scales (Abu-Alhaija, 2019).

Ontology in research reflects the philosophy or nature of reality (Kivunja & Kuyini, 2017). A key ontological issue for consideration in the present study was the principle of *reductionism.* Reductionism is an assumption that unobservable/latent phenomena exist and are capable of explaining the functioning of observable phenomena, making it possible to measure latent attributes (Kivunja & Kuyini, 2017; Ryan, 2006, p. 20). Situating this principle of reduction within the present study, we posited that these latent traits 'significantly explain' how the respondents respond to the items in their respective scales, and we were therefore able to develop and validate measures of CNL-E, CNLEng and SEBH.

Epistemologically, the main source of knowledge in the present study was empirical which is derived from perception of senses to the external world through observation and experimentation (Kivunja & Kiyuni, 2017). Epistemology is concerned with the theory of knowledge, particularly the nature of knowledge and justification of how we acquire knowledge (Kivunja & Kiyuni, 2017). Specifically, in the present study I sought out empirical evidence relating to adolescents' CNL and self-efficacy in a nutrition-related subject. From this epistemological standpoint, I was guided on what research methodology to adopt.

Methodology refers to the strategy or plan of action behind the choice and use of particular methods to gain valid knowledge about the phenomena of interest, linking the choice and use of methods to achieve the desired outcomes (Sobh & Perry, 2006). In respect to this, the present study employed a survey research approach, which provides a quantitative (numeric) description of trends, attitudes, or opinions of a population by studying a sample of that population (Creswell, 2014, p. 41). Additionally, using a cross-sectional design made it possible to collect data on CNL and SEBH at the same time, with the intent of generalizing findings from this sample to the population-all tenth graders in Norway. Since CNL-E, CNLEng and SEBH are unobservable phenomena, I used data collected using a self- administered electronic questionnaire survey system comprised of items that reflect these three underlying traits of interest. Random sampling of schools ensured that the selected sample was representative of adolescents aged 15-16 years tenth graders in schools across Norway. Furthermore, I also took measures to 'confine' the effect of any other underlying influences on the adolescents' responses to the scale items. These measures included checking the unidimensionality of the scales and the effect of person factors such as gender, age, socioeconomic status, ethnicity on adolescents' responses.

In addition, as per post-positivist guidelines, which seek to 'disconfirm' hypotheses instead of confirming hypotheses (Carpiano & Daley, 2006), the overall study hypothesis was falsifiable–'that CNL-E was associated with CNLEng and perceived SEBH'. The secondary hypotheses addressed the validation of the three scales used in the study. Data analysis involved the use of Rasch modelling (RM) and confirmatory factor analysis (CFA) for validation of the psychometric properties of the scales, and structural equation modelling (SEM) to relate the different traits under study.

Finally, the present study observed the four principles of axiology namely privacy, accuracy, property and accessibility. Axiology refers to the ethical principles and role of beliefs and values in conducting research (Kivunja & Kuyini, 2017; Mason, 1986). Accordingly, all participating schools and respondents were assigned anonymous identifier codes to ensure privacy. All participation was voluntary; the school principals gave consent of participation on behalf of the adolescents prior to data collection and they were informed that the data could be used in research. To ensure accuracy in the present study, data was checked for odd responses during data analysis. Lastly, as associate professor Øystein Guttersrud led the data collection, he granted access to use part of the data collected for this PhD project.

Theoretical underpinnings in the present study

Theory informs research and makes it possible to interpret empirical data. In social sciences, theory refers to an organized body of interrelated constructs (variables) and generalizations that systematically explains and predicts particular phenomena (Kerlinger, 1986, p. 9).

In this section, I present the theoretical positions of the main concepts of the present study namely, adolescent development, nutrition literacy, self-efficacy and scientific literacy. The theoretical underpinnings described below were selected based on their relevance to the study, and their association between the main concepts under study.

Adolescent development

Generally, theories of adolescent development argue for the two main schools of thought on human development; to what do we attribute the behaviour in adolescents, 'nature or nurture?' Different schools of thought shed light on either argument, ranging from theories that emphasize the biological perspective (nurture) to those that emphasize the environmental (nurture) (Steinberg, 2010, p. 13). For the present study, I leaned towards two organismic theories, *Piagetian' s theory of cognitive development* (Piaget, 1964; Huitt & Hummel, 2003) and the *information-processing perspective of adolescence* (Steinberg, 2010, p. 62). At the core of these theories is the notion that in addition to biological factors (such as the hormonal changes associated with adolescence) other contextual (environmental) factors interact with and may modify biological factors that influence behaviour in adolescents. I included these theories because they account for the influence of both internal and external factors on adolescents' behaviors such as engagement in dietary behaviour (CNLEng). More so, both these theories have a focus on cognition, which is central to the concept of critical evaluation of nutrition information (CNL-E), and even engagement in dietary behaviour (CNLEng).

Jean Piaget's theory of cognitive development suggests that cognitive development in children proceeds through a sequence of qualitatively distinct stages, marked by improvements in thinking about abstract concepts (Piaget, 1964). As per this theory, adolescence falls under the 'formal operational stage', beginning at the age 7-11years and lasting into adulthood. Herein, adolescence marks the transition from concrete to abstract thinking; individuals develop the ability to think about abstract concepts and logically test hypotheses (Huit & Hummel, 2003). Whereas a considerable amount of research into how young people think has emerged based on this theory, it is not without criticism. Antagonists of this theory claim there is no empirical evidence to support that cognitive development does in fact happen in a distinct stage-like fashion (Weiten, 1992). They argue that advanced reasoning capabilities develop gradually and continuously from childhood through adolescence and do not have distinct stages as Piaget suggested. Rather, they posit that these cognitive capabilities are indeed skills employed more often by older children (adolescents) than by younger ones (children) (Fischer & Bullock, 1984, p. 71). This discordance has paved the way for the emergence of another theoretical perspective that emphasizes how adolescents make meaning of information referred to as the information-processing view of adolescent cognitive development.

The information-processing view of adolescent cognitive development suggests that cognitive development is continuous, and as one's brain matures, cognitive processes that are actually components of information processing also advance in response to information stimuli, which may be audio-visual information that is encountered both actively and passively, and from one's social context (Kail & Ferrer, 2007). These cognitive processes include recognition, judging, remembering, and reasoning, among others. According to this perspective, cognitive improvements during adolescence result into improved concentration, memory skills for verbal and visual information, and increase speed in processing information (Hale et al., 1997). In addition, adolescents become more sensitive to social information, making them concerned about what others think they are thinking (Demetriou et al., 2002). This might explain why adolescents are prone to peer influence in decision-making. Whereas the present study did not have the opportunity to investigate whether these cognitive gains occur in distinct stages, this theory enriched our perspective on what cognitive advances one might expect to observe as adolescents interact with nutrition information such as the ability of adolescents to interpret abstract messages about nutrition they may encounter in the media.

These theoretical insights were relevant to the present study, because they provided a framework within which we can understand the cognitive processes that underlie critical evaluation of nutrition information in adolescents. Moreover, these 'skills' or 'cognitive advances' thought to occur during adolescence are central to the concept of nutrition literacy, specifically critical nutrition literacy.

Nutrition literacy

Theory development in the field of nutrition literacy is still in its stages of infancy. While different authors have proposed definitions of nutrition literacy, no known theory has been put forward. Nevertheless, recent efforts towards developing a theoretically founded model depicting nutrition literacy has been suggested by Pettersen (2019, p. 171). Although this theoretical model has not yet been empirically tested, it provides promising insight into how best to describe and operationalize nutrition literacy.

Pettersen's theoretical model and matrix of nutrition literacy draws on two empirically tested theoretical models in the field of health literacy (HL), that is, Nutbeam's HL model and Sørensen's integrated conceptual HL model (Nutbeam, 2000; Sørensen et al., 2012). Figure 1 shows Pettersen's proposed theoretical model and Figure 2 is the four-by-three matrix depicting the four competencies required to navigate nutrition literacy across three life cycle situations.

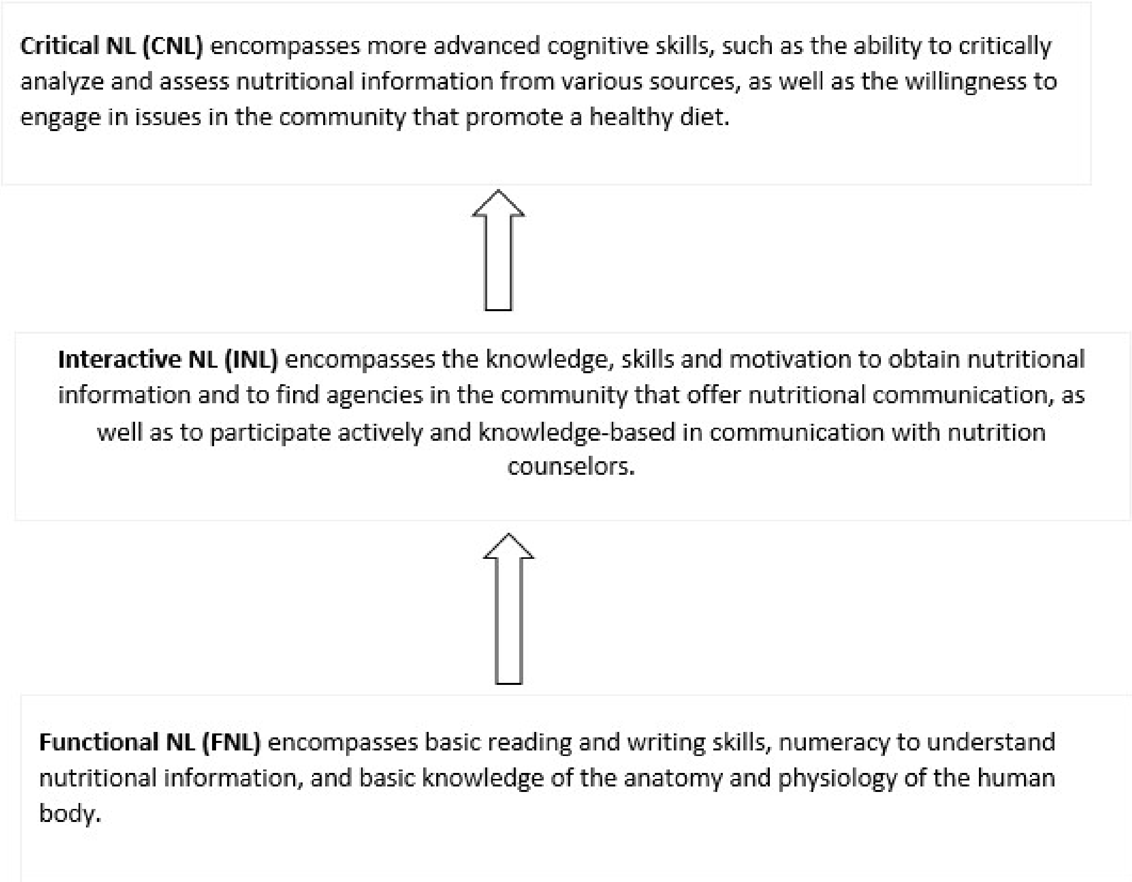

Figure 1: Pettersen's theoretical model of the three cumulative nutrition literacy competencies.

Model originally in Norwegian (Pettersen, 2019, p. 171), and translated to English by the author of this thesis.

Integrating the theoretical model, which was based on Nutbeam's model of health literacy, and the matrix shown in Figure 1 and 2 respectively, Pettersen (2019, p. 171) views nutrition literacy as an ongoing process that involves the use of four main cognitive skills (access, understand, appraise and apply) as described by Sørensen et al. (2019) to navigate the three main systems associated with nutrition (health care, disease prevention, health promotion). Some of these systems include nutritional care systems, marketing systems, and media environments.

		I	II	III	IV
		Access	**Understand**	**Appraise**	**Apply**
a)	**Health care**	Ability to access information on what to eat and not to eat in the event of illness	Ability to understand information about what to eat and not to eat during the event of illness	Ability to critically evaluate information about what to eat and not eat in the event of illness	Ability to make informed decisions about what to eat and not eat in the event of illness
b)	**Disease prevention**	Ability to access information on risk factors in the diet for the development of disease	Ability to understand information about risk factors in the diet for the development of disease	Ability to critically evaluate information on risk factors in the diet for the development of disease	Ability to make informed decisions to avoid risk factors in the diet for the development of disease
c)	**Health promotion**	Ability to access information about what is a healthy diet for you in your social and physical environment	Ability to understand information about what is a healthy diet for you in your social and physical environment	Ability to critically evaluate information about what is a healthy diet for you in your social and physical environment	Ability to make informed decisions about what is a healthy diet for you in your social and physical environment

Figure 2: The three-by-four matrix of nutrition literacy by Pettersen

The matrix in figure 2 by Pettersen (2019, p.167) was originally in Norwegian and translated to English by the author of this thesis. The building of the matrix in this figure was inspired by the empirically tested health literacy model by Sørensen et al. (2012). As far as I know, this is the first time a theoretical model of NL drawing upon the HL conceptual model by Sørensen et al. (2012) has been suggested. The wording in the cells is adjusted to depict nutritional issues and shows four individual cognitive skills (I-IV) reflecting nutrition literacy in three life cycle situations (a-c).

‘Access’ refers to an individual’s ability to seek out and obtain nutrition information (from the various nutrition-related systems) about nutritional wellbeing, prevention of malnutrition and promotion of nutrition. These systems include among others, media environments, food systems. The main skills employed herein is the ability to read and write, and basic numeracy skills (Velardo, 2015).

‘Understand’ refers to one’s ability to comprehend the nutrition information that one accesses to ensure nutritional wellbeing, prevention of malnutrition, and promo tion of nutrition. The cognitive skills are more advanced requiring the individuals to interact with agencies in the community that offer nutritional communication, for example, nutritional counsellors. Some of the cognitive skills involved include, inter-personal communication skills and inductive reasoning.

‘Appraise’ describes the ability to interpret, filter, judge and evaluate the nutrition information that individuals obtain to ensure nutrition wellbeing, prevent malnutrition and promote nutrition. The cognitive skills required herein are more advanced, requiring one to make comparisons of the information basing on a 'yardstick' and deductive reasoning.

'Apply' refers to the ability to communicate and use the nutrition information to take decisions to maintain and improve nutritional wellbeing for both individuals and their communities. Applying nutrition information requires individuals to comprehend and make meaning of the information obtained and translate it into solutions for their nutritional needs and those of their communities.

From the perspective of nutrition promotion, the goal of nutrition literacy is the practical application (use) of nutrition information, suggesting that the more 'transformational' skills that one has, the more nutrition literate they are. These transformational skills include among others, communication skills, social skills, information appraisal skills, computational skills, all of which enable one to derive meaning from nutrition information and apply that meaning in their noble situations. Advancing through the cumulative levels of nutrition literacy depicted in Figure 1 and across the lifecycle shown in Figure 2 requires that individuals are capable of making self-evaluations of nutrition information, how relevant it is to their specific nutrition needs, how able they are to form behaviors based on this information and thereby address their nutrition related issues.

Schwarzer and Renner (2000) posit that one of the important self-beliefs that individuals require to apply nutrition information and form beneficial behaviors that enhance their nutritional wellbeing is self-efficacy.

Self-efficacy

Bandura's theory of social cognitive theory (SCT) posits that adolescence is marked by different changes including changes in personal self-beliefs including self-efficacy (Stajkovic & Luthans, 1998). Self-efficacy refers to subjective judgments of one's capabilities to organize and execute courses of action to attain designated goals (Bandura, 1997). Self- efficacy beliefs are also personal expectations about what we believe we can do and are known to influence a number of behavioural processes (Bandura, 1997).

Within academic settings, self-efficacy refers to individuals' convictions that they can successfully perform given academic tasks at designated levels (Britner & Pajares, 2006). The expected output of science learning is that students are able to apply this knowledge away from school, and to their everyday life (Massar & Malmberg, 2017). This expectation is synonymous with the concept of 'scientific literacy' (SL). Research shows that one of the important factors that influences scientific literacy are the self-efficacy beliefs that students hold in their abilities to apply science knowledge in their everyday lives (Latifah et al., 2019). Herein, self-efficacy is utilized as a suitable way to determine student's perceptions of their capabilities to explore their scientific literacy-whether they comprehend the importance of skills and knowledge to be valued in their everyday life (Ait et al., 2015).

Scientific literacy

The Organization for Economic Cooperation (OECD) posits that understanding science and science-based technology is a crucial part of their 'preparedness for life' (OECD, 2013, p 3). Within the Programme for International Student Assessment (PISA) framework, scientific literacy (SL) refers to the knowledge *of* science and science-based technology (OECD, 2013, p 3); it thus follows that a scientifically literate adolescent should be able to apply science knowledge in their everyday life situations such as making healthy food choices.

Philosophers classify science knowledge into two broad categories namely, *declarative* (content) knowledge and *procedural* knowledge (Boshoff, 2014). Whereas these two types of knowledge are interrelated, they are distinct and make significant contributions towards one's scientific literacy. Declarative (content) knowledge also referred to as 'knowledge of what' and refers to information about concepts, ideas, and theories (Boshoff, 2014). In science, this includes knowledge of theories that explain the occurrence of phenomena in the natural world such as the occurrence of disease. Having this knowledge makes individuals more competent to explain scientifically the occurrence of phenomena. Conversely, procedural knowledge, also referred to as the knowledge of skills, concerns the 'know-how' of procedures by which scientists obtain factual knowledge (Boshoff, 2014). Scientifically literate individuals can use this information to conduct a critical review or appraisal of scientifically based claims.

The SL competencies described by PISA are the ability to identify scientific issues, explain phenomena scientifically, and use scientific evidence (OECD, 2013, p 5). Pettersen (2007) posits that CNL is part of SL as it requires applying similar competencies (understand, appraise, and apply) to nutrition information. Within the context of NL, we expect that individuals that are scientifically literate possess procedural knowledge, are thus able to critically appraise nutrition information (CNL-E), and apply that information in practical ways that address their nutritional needs and those of their communities (CNLEng).

In the present study, through the curriculum of science subjects such as ‘Body and Health’, we posit that adolescents gain declarative (factual) knowledge that they can use to appraise science-related information, such as nutrition information that they obtain from various sources such as media channels.

Influences of CNL in adolescents

Adolescents today obtain nutrition information from several sources including among others, media that is, both ‘new’ (digital) and ‘traditional’ (newspapers, radio, television etc.), social networks (peers, family, friends, schoolmates, teachers), qualified personnel (nutritionists, counsellors, doctors) and from school (as part of subject curriculums) (Coiro et al., 2015; Goodyear, et al., 2018, p 4; Paek et al., 2011). What's more, adolescents today are no longer just consumers of information as they are highly engaged in creating and sharing nutrition information. We can say that adolescents today find themselves cast in a ‘jungle of information’ of some sort. It is therefore not surprising that adolescents report being confused by all the nutrition information that they have access to (Goodyear et al., 2018, p 220). For this reason, it is important that adolescents are able to judge the quality and accuracy of nutrition information, and the reliability of sources from which they obtain this nutrition information. The skills that they need to achieve this are at the heart of CNL (Pettersen, 2019, p. 167).

The section that follows describes some of the key influences of CNL in adolescents, which were central to the development of the items in the scales applied in the present study (CNL-E, CNLEng, and SEBH). By exploring the factors that influence CNL and SEBH through measurement, the present study sought to give further insight into the cognitive skills adolescents apply when interacting with nutrition information from various sources.

Media use by adolescents

Adolescents today are living in the heyday of the ‘information age’. A study on the media consumption habits of adolescents in America showed that they spend more time interacting with media than they do in any other activity besides sleeping, an equivalent of more than 7 hours on average per day (Rideout et al., 2010, p 1).

Adolescents are particularly interested in the ‘newer’ type of media i.e. digital media, especially internet use. In fact, according to Brembeck and Johansson (2010), the internet has become a growing part of children’s foodscape. This has been fuelled largely by widespread access to the internet and ownership of internet-enabled mobile devices. For example, a study in Norway showed that 97% of young adolescents (9-16years old) have access to an internet- enabled mobile phone (Statistics Norway, 2016). The ownership of mobile phones from a very young age has fuelled ubiquitous media use by adolescents and it is therefore not surprising that the internet is the most popular medium of media in Norway among adolescents (Ní Bhroin & Rehder, 2018; Medietilsynet, 2016; Statistics Norway, 2016).

Adolescents today use the internet for a myriad of reasons, one of which is the communication about food and related aspects of nutritional wellbeing such as dieting and exercising (Chassiakos et al., 2016; Chau et al., 2018; Samoggia & Riedel, 2020; Seah & Koh, 2020). As adolescents share their views and those of others on food-related content on social media platforms such as blogs, Facebook, Twitter, Snapchat and TikTok they develop a shared understanding of healthy foods, unhealthy foods, ‘ideal’ diets and body image, among others. There is evidence that adolescents may model their dietary behaviour based on information obtained via different media channels (Buchanan, et al., 2018; Norman et al, 2016; Russell et al., 2019; Sidani et al., 2016; Smith et al., 2019). Studies on nutritional disorders in adolescents show a significant influence of media messaging on the occurrence of nutritional disorders like bulimia nervosa, obesity (Boyland & Whalen, 2015).

In addition, adolescents may encounter unsolicited food-related information while online via pop-up advertisements, promoted campaigns and sponsored content. Majority of these advertisements are by soft drink and fast food companies that capitalize on the increased level of independence in decision making and purchasing power attained during adolescence. A study on food advertising to adolescents online showed that global franchises including Nestle, Coca-Cola and Starbucks had some of the biggest expenditures on marketing on Facebook (Davis, 2018). Franchises like Coca-Cola (soft drink) and Dominos (pizza) have also been shown to have some of the best marketing strategies on social media. They employ marketing strategies that foster a high level of engagement by the adolescents often using hashtags, campaigns in which the consumers recreate advertisements including these products (Freeman et al., 2014; Laestadius & Wahl, 2017; Meyer et al., 2019; Montgomery & Chester, 2009; Potvin Kent et al., 2019; Vassaloo et al., 2018). By using these hashtags, adolescents unknowingly create more demand for these products and brands among their social circles.

Branding is a crucial part of food advertising and creates brand loyalty and increased demand for the products. During adolescence, brand awareness increases and brand loyalty becomes more important forming an important part of one's identity; loyalty to food brands at is established during this life stage often persists through adulthood influencing dietary practices like food selection and purchase selection (Roper & La Niece, 2009). Brand loyalty during adolescence is created both on the biological level and as a result of external exposure. At a biological level, exposure to images of food and food brands through advertising has been linked to cravings and reduced appetite control all of which are predominant during adolescence (Schienle et al., 2009). At external level, brand preference influences dietary practices like food selection and purchase behaviour and is shaped by factors such as media exposure to food advertisements and how 'prestigious' they perceive a food brand to be (Schienle et al., 2009).

Considering the influence of media on dietary behaviour of adolescents, it follows that if adolescents are to capably 'navigate the jungle' of nutrition information they need to be critically nutrition literate. This means that they must be equipped with the skills to evaluate the credibility of nutrition information, identify any self-serving interests of advertising in the media message and assess what information is relevant for their personal needs. Only then can they effectively apply nutrition information in their everyday lives.

Additionally, it is crucial that adolescents have strong self-belief in their ability to use this information in the correct way. In fact, evidence suggests that for competent functioning to be achieved, there needs to be harmony between self-efficacy, possessed skills and knowledge (Pajares, 2006, p. 4; Norman & Skinner, 2006). In this way, self-efficacy determines what adolescents do with the knowledge and skills that they possess.

Self-efficacy in adolescents

The role of self-efficacy in adolescents' lives extends beyond academic settings and may influence their choices and courses of action in situations in which they have options (Glasofer et al., 2013). One such area of life in which adolescents have multiple options and must make decisions is in their dietary choices. Accordingly, pertaining to food selection, we may view self-efficacy as the adolescent's confidence in their ability to use the skills and information they must make the best possible choice to achieve their desired nutritional outcomes. Studies show that there is a direct and positive association between self-efficacy and adopting and maintaining healthy dietary practices such as healthy food selection, increased intake of fruit and vegetables. The more efficacious the adolescent is, the greater the likelihood that they will engage and maintain positive dietary behaviors because they feel competent enough to make healthier food choices even when in a 'toxic food environment'(Anderson-Bill et al., 2011; Fitzgerald et al., 2013; Glasofer et al., 2013; Lubans et al., 2012; Rolling & Hong, 2016; Steele et al., 2011).

Self-efficacy influences how one uses one's cognitive resources and strategies; the more self-efficacious one is, the more cognitive strategies one applies to accomplish the tasks (Pajares, 2006, p. 5). What this means is if adolescents feel that they can apply the knowledge and skills that they have to achieve a given task, they will be motivated to persist in accomplishing that task. For example, studies on the use of nutrition labels during food purchase show that individuals that are confident in their ability to find and choose healthy foods using information provided on nutrition labels (high levels of self-efficacy), are more likely to make healthier food selections than their inefficacious counterparts are (Aboulnasr, 2013). In this context, we view self-efficacy as one's belief in one's ability to use the knowledge, resources or skills that one has to use the information on the nutrition label successfully during food selection.

The Elaboration Likelihood Model of Persuasion (ELM) provides a good theoretical framework within which to interpret and understand the influence of one's self-efficacy on the decision-making process. The ELM suggests that one's perceived ability to use information influences the extent of elaboration in information processing (cf. CNL-E in the present study). Accordingly, individuals with high levels of self-efficacy engage in detailed information processing; they consider more (*used quantitatively*) food product attributes that are nutrition-related, compare any claims to information and even compare different products, before selecting a product. On the other hand, individuals that are less self-efficacious do not engage in detailed information processing and consider few (often non-nutrition related) food product attributes such as taste and price when selecting food (Mai & Hoffmann, 2012). Situating this within the present study, we anticipated that adolescents with a high level of SEBH would exhibit a high level of CNL-E and CNLEng, as they feel competent enough to understand the information available, evaluate nutrition claims associated with the product, and in their ability to use the skills that they have to apply this information successfully when they encounter nutrition messages from various sources.

Owing to the seemingly conflicting nutrition messages in the media, sources such as social networks, and the presence of nutrition claims on almost every product, adolescents today ought to engage in a detailed and rational decision-making process. To achieve this, adolescents need to have a wider scope of knowledge to which they can compare information obtained from various sources and the skills to apply this information. Enhancing scientific literacy provides a good opportunity to equip adolescents with this broader scope of knowledge and the skills they need to engage in a detailed process of information appraisal.

Scientific literacy in adolescents

In order for adolescents to make informed comparisons and evaluate nutrition information they obtain different sources, they need a 'yardstick'. Scientific knowledge provides a criterion against which they can judge nutrition information, as it is evidence-based and factual. Considering that nutrition is a science, comparisons based on scientific knowledge and evidence obtained through scientific processes are judicious (Smolin & Grosvenor, 2010). When adolescents apply science knowledge in real-life contexts such as when verifying the validity of nutrition claims, credibility of the messages conveyed in advertisements, we say that they are 'scientifically literate'. In this instance, we may define a scientifically literate adolescent as one who is able to apply science knowledge while making decisions related to their nutritional wellbeing. These actions may include for example food selection, establishing the credibility of sources of nutrition information and assessing how valid nutrition claims about food products are, among others.

Theoretical model linking selected influences of CNL in adolescents

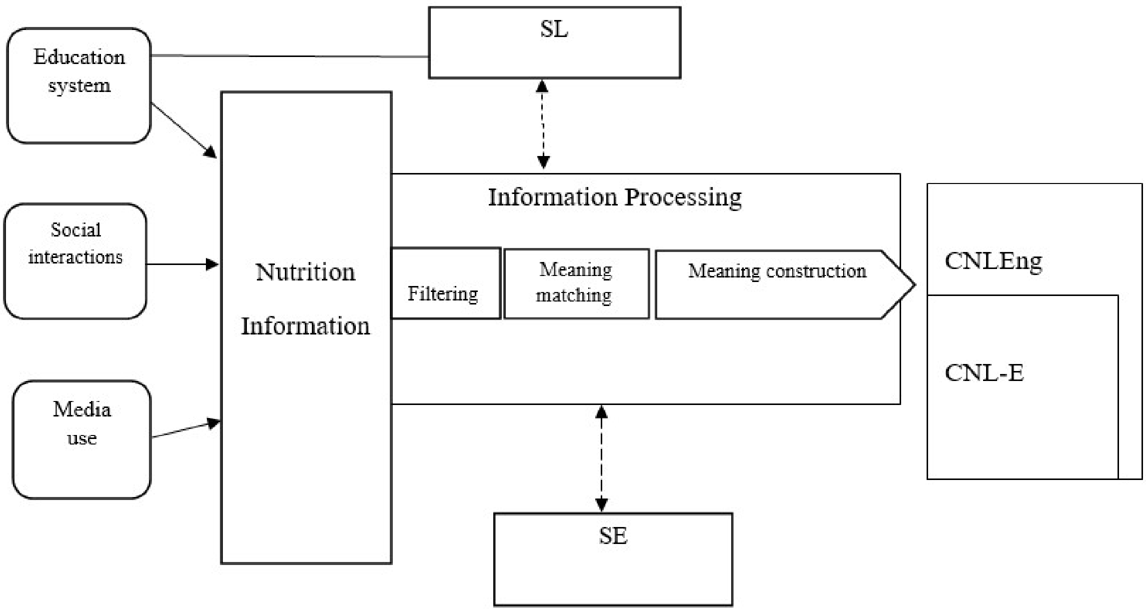

Figure 3: Theoretical model linking selected influences of CNL in adolescents made by the author of the thesis.

CNL: Critical nutrition literacy; CNLEng: Engagement in dietary behaviour; CNL-E: Critical evaluation of nutrition information; SL: Scientific literacy; SE: Self-efficacy.

Figure 3 is a graphic representation of how selected factors may influence CNL in adolescents. The solid arrows in the figure represent the direct influences of these selected factors on the different components of the theoretical model. The dashed arrows represent the more indirect influences of CNL in adolescents, representing attributes that affect how individuals interact with nutrition knowledge through the different phases of information processing.

Nutrition information: Adolescents get nutrition information from different sources including the education system, media, and through social interactions. These sources are some of the agents of the 'health socialization' process, defined as the processes through which young people acquire health-related orientations, skills, knowledge and attitudes, which in turn form their healthy lifestyles and behaviors (Paek et al., 2011). Situating this in the present study, health socialization refers to how adolescents acquire nutrition information via social interaction through media use and interpersonal socialization agents.

Media use: This serves a twofold purpose in this theoretical model. First as a source of nutrition information, and second as an avenue through which adolescents can hone the cognitive skills that they need to use this information. Adolescents obtain nutrition information from different types of media channels including traditional (analog, newspapers, radio, television) and digital sources (online, social media). Regarding skills, the continued use of (practice) media encourages improvement of skills relevant to critical use of media, such as evaluation, analysis and induction; in fact, media literacy interventions often aim at improving these skills through practice (Potter, 2004, p. 36).

Social networks: Adolescents also get nutrition information through interaction with their social networks. These networks may include family members, nutrition counsellors and peers. During adolescence, social relations are very important and influence many of their decisions. Findings indicate that peer level variables such as one's social environment are associated with dietary behavior in adolescents; social support from family and friends is important for healthy eating behavior in adolescents (Brown & Larson, 2009, p. 74; Wouters et al., 2010).

Education system: As adolescents advance through the education system, they study subjects related to nutrition such as biology, and food science, thereby gaining nutrition knowledge. Nutrition knowledge is classified as declarative and procedural. Declarative knowledge refers to the factual knowledge, theoretical knowledge, whereas procedural knowledge refers to the 'know-how' aspects of knowledge such as practical skills and abilities (Boshoff, 2014).

As part of the procedural nutrition knowledge, adolescents may acquire the 'how to' skills, which are the skills that they need to put nutrition information into practical use such as cooking. In addition to these, adolescents might also gain the skills they need to use media, such skills include how to evaluate information (CNL-E).

Furthermore, through the scientific curriculum adolescents gain the scientific knowledge that they can use as a yardstick against which to appraise the nutrition information that they get (CNL-E). Additionally, they may gain the competencies and skills required to evaluate information, interpret data and analyze evidence scientifically. They can then apply these skills to for example, assess how accurate nutrition information in the media is and identify any bias in nutrition claims. These skills are synonymous with the concept of scientific literacy (SL).

Self-efficacy: How young people evaluate information depends on among other factors, their perceptions of their abilities; more confident adolescents tend to seek out more 'advanced' sources of information such as digital, than those who don't feel competent to understand information (Gasser et al., 2012). During the process of information appraisal, self-perception beliefs influence what information the individual pays attention to, and what they do not (filtering). Self-efficacy has been shown to be an important determinant of consumer's information search determining the type of information-processing tasks that individuals engage when they encounter information (Mai & Hoffmann, 2012).

Information-processing: Potter (2004, p. 37) posits that there are three core components of the sequence of processing information. These are filtering, meaning matching and meaning-construction Figure 3 and that to advance along these tasks, individuals must apply different skills to the message encountered. First off, during *filtering*, when the adolescents encounter nutrition messages such as in media, they must decide which information to pay attention (filter in), and which to ignore. To do this, they use skills such as analysis which refers to breaking down the message into meaningful elements. The next task, *meaning matching* requires individuals to determine the meaning of these messages by matching it to previously learned meanings, the skills required for this include grouping, induction and deduction. Only after relating to this information, is one able to proceed to *meaning-construction,* a task which aims at creating meaning for oneself so that one can apply the information in novel situations. The key skills applied herein include, induction, deduction and evaluation (Potter, 2004, p. 36). It is crucial that adolescents are able to create meaning for themselves whenever they encounter nutrition messages, as this will help negate the influence of others (such as peers) on how they use nutrition information, ensuring that they correctly use nutrition information to solve their individual needs.

In the theoretical model shown in above, CNL should be an outcome of the interaction between the health socialization agents (media use, education system, and social interactions), individual attributes like self-efficacy, and skills applied to the nutrition information resulting from the information processing. In the figure CNL-E and CNLEng are the two different but related aspects of CNL. CNL-E is encased within CNLEng basing on the premise that the extent to which adolescents apply nutrition information as CNLEng depends on how well they understand this information during CNL-E. Moreover, findings show that adolescents may not correctly apply nutrition information they encounter because they find it confusing and difficult to understand (Babio et al., 2014).

In order to advance research on CNL in adolescents, it is important that studies develop and validate measures of CNL thereby lending empirical strength to the factors that influence CNL in this target group. The present study sought to contribute to the field of CNL this way.

Instruments measuring CNL in adolescents

Over the recent past, the field of nutrition literacy has undergone advancement through continued development of assessment instruments. A review by Yuen et al. (2018) on existing measures of nutrition literacy showed that there has been an increase in the number of instruments. Broadly, existing NL scales focus on the NL of individuals with for example, nutrition-related conditions (as in, Gibbs et al., 2018), the use of food labels (see for example, Ringland et al., 2016), and in primary care settings (see for example, Weiss et al., 2005). However, as more studies explore measurement of nutrition literacy, it has become apparent that compared to the functional and interactive domains of nutrition literacy, the domain of CNL remains largely underexplored (Velardo, 2015). Furthermore, few of these instruments are dedicated to measuring CNL on adolescent samples (Glasofer et al., 2013). Where available, instruments measure one aspect of CNL-CNL-E thereby presenting a limited scope of CNL. Nevertheless, these instruments give valuable insight into the different skills we expect adolescents to have and utilize when they interact with nutrition information from various sources. According to Guttersrud and Pettersen (2015) we can categorize CNL into two aspects namely CNL-E and CNLEng. In the section below, I present a brief overview of existing instruments measuring CNL on adolescent samples, categorized according to the two aspects of CNL.

Instruments measuring appraisal of nutrition information

Nutrition labelling on food products is one of the most cost-effective and commonly used tools for promoting healthy dietary practices (Viola et al., 2016). It is therefore no surprise that there is interest in whether individuals, particularly adolescents, understand this information on food labels (Cha et al., 2014; Dong, 2015; Haidar et al., 2017; Wang et al., 2016). This is judicious given the shift in purchasing power and decision making that happens during adolescence. During adolescence, individuals become more independent in making dietary choices. Also, of interest is how adolescents appraise nutrition information from different types of food labels. A study in Spain sought to compare how adolescents aged 14-16 years used two different front-of-package (FOP) labels to choose a diet that closely follows the nutritional recommendations (Babio et al., 2014). Both labelling systems contained the same information, guideline daily amounts (GDA), albeit presented differently. One system was color-coded to indicate varying quantities of the nutrient, whereas the other system used monochrome GDA. Following exposure to each system, the adolescents designed diets based on the respective systems, and reported on which of the systems was easier to use in selection of healthier food options. Study findings showed that adolescents found it easier to select healthier food options using the color coded-GDA system, compared to the monochrome- GDA labelling system (Babio et al., 2014). The authors suggested that adolescents find it easier to understand nutrition information on food labels that requires less cognitive capacity i.e. those with less 'text' and or figures.

Other studies have shown interest in the cognitive effort and specific skills that adolescents expend to interpret specific nutrition information available on food labels, such as the percent recommended daily allowance and percent daily values, as was the case in studies on Norwegian and Sri Lankan adolescents, respectively (Talagala & Arambepola, 2016; Wang et al., 2016). In both instances, adolescents had to have numeracy skills to compute the nutrient content of the snacks under comparison. Findings showed that in both studies, the adolescents' interpretation of this information was relatively poor, implying that the information might have been too complicated for them to use during snack selection. The authors concluded that there is a need for additional skills such as numeracy skills, to interpret nutrition information from food labels, 'complicates' the process of informat ion appraisal for adolescents.

Dietary informational resources are important factors for promoting nutrition literacy (Aihara & Minaj, 2011; Gibbs & Chapman-Novakofski, 2012). However, considering that adolescents have access to several sources of nutrition information, it is equally important that they are able to assess the credibility of these sources. This is because while access to diverse sources of nutrition information has increased, the assurances of the quality of information provided by these sources seems to be lagging behind (Ishikawa & Kiuchi, 2010). The authors' concern echoes the interest in equipping individuals, particularly children and adolescents, with the skills that they need to analyse and evaluate both the information and sources of information (Arke & Primack, 2009; Simovska et al., 2012). The present study sought to assess how adolescents evaluate the reliability of the sources of nutrition information using the CNL-E scale.

Instruments measuring engagement in dietary behaviour

Whereas researchers have an interest in dietary practices of adolescents, studies on the same are fewer than expected. Unlike appraisal of nutrition information, this aspect of CNL has fewer instruments reported in literature. To the extent of my knowledge, only one study by Guttersrud and Pettersen (2015) has developed, validated and reported on an instrument exclusively measuring the CNLEng of adolescents, which was developed and applied on a sample of Norwegian adolescents in 2013. To date, this scale has informed the measurement of CNLEng in adolescent samples (Joulaei et al., 2018). The scale, which was comprised of six items distributed on three levels-personal, social and global levels, sought to evaluate the extent to which adolescents engaged in dietary practices that improve their nutritional wellbeing. The items also established the adolescents' concern about the availability of healthy food options in local grocery stores, public places such as cafeterias and at global level. To date, no other known instrument probably exists that exclusively measures CNLEng on adolescent samples.

Differing from other instruments that are often applied in other studies, Guttersrud and Pettersen's (2015) scale was validated using Rasch analysis, which is a modern measurement approach to evaluating the psychometric properties of instruments used in quantitative studies. Validation of measurement instruments is important as it speaks to the quality of the instrument and the applicability of these instruments in other studies (Lai, 2013). In budding areas of research such as nutrition literacy, validating measurement instruments might result into increased use of these instruments in research. The present study exemplified how to validate adolescent-appropriate tools measuring the different aspects of CNL namely, CNL-E and CNLEng.

Evaluating the psychometric properties of measurement instruments

Often, the study of behavioral sciences involves assessing unobservable psychological quantities referred to as latent traits (Brzezińska, 2016). To achieve this, researchers use instruments comprised of items measuring observable behaviour thought to be influenced by the underlying latent trait. The study of instrument development and validation is referred to as psychometrics, which is a scientific discipline concerned with how observable variables can optimally represent latent traits and involves developing instruments to measure these traits (Borsboom & Molenaar, 2015, p. 418). Psychometric evaluation of assessment instruments involves the use of various statistical techniques to investigate these properties.

In the section that follows, I describe the analytical techniques applied to evaluate the psychometric properties of the three scales applied in the present study namely, *confirmatory factor analysis* (CFA) and *Rasch modelling (RM)*. I conclude the section with a comparison between these broad analytical frameworks to which these techniques (CFA and RM) belong, that is, *classical test theory* (CTT) and *item response theory* (IRT), respectively.

Confirmatory factor analysis

Confirmatory factor analysis (CFA) is a statistical technique belonging to the broader family of techniques of CTT. CFA is used to verify the factor structure of a set of observed (indicator) variables and seeks to confirm if the number of latent traits (factors) and the loadings of indicator variables on the respective factors conform to what is expected on the basis of theory (Suhr, 2006). Put another way, CFA tests the hypothesis that a relationship between the indicator variables and the underlying latent factor(s) exists and shows how well the indicator variables represent the latent factors to which they are related (Furr, 2013, p.351). Latent traits are the underlying theoretical concepts presumed to influence the effect/response to the observed variables, they are unobserved and are therefore 'operationalized' using specific indicator variables (Suhr, 2006). Conversely, indicators are observed and respondents' responses to these are influenced by the latent trait under investigation (Suhr, 2006). In CFA models, factors are represented by ellipses whereas indicators are represented by rectangles. The relationship patterns among observed and latent variables are represented by arrows. All relationships in confirmatory approaches should be informed by theory, empirical findings, or both and are specified *a priori*.

The objective of CFA is to solve a set of equations involving the indicator or item 'loadings' and the observed correlations between the indicators (cf. the elements in the observed sample correlation matrix $\boldsymbol{S}$) and obtain an estimate for each loading. We use these loadings to calculate a model-implied correlation matrix $\boldsymbol{\Sigma}$, and we assess to what degree the model-implied matrix recreates the observed correlation matrix $\boldsymbol{S}$. The set of equations has one unique solution only when *three* indicators operationalize each factor in the CFA. Then, the model-implied correlation matrix $\boldsymbol{\Sigma}$ is equal to the observed sample correlation matrix $\boldsymbol{S}$. This means that each element in $\boldsymbol{\Sigma}$ is equal to its counterpart in $\boldsymbol{S}$, there is no 'left-over' or residual, and $\boldsymbol{R}=\boldsymbol{S}-\boldsymbol{\Sigma}=\mathbf{0}$. In this case, all "goodness of fit" (GOF) indices take on their expected value and indicate good fit. When four or more indicators operationalize a factor in a CFA, the set of equations does not have one unique solution. Using maximum likelihood estimation (MLE) or some equivalent estimation method like diagonally weighted least squares (DWLS), we find the set of loadings that creates a model-implied correlation matrix $\boldsymbol{\Sigma}$ that is *as close* to the observed sample correlation matrix $\boldsymbol{S}$ as possible. We use GOF indices, like chi-square and SRMR to assess whether the $\boldsymbol{\Sigma}$ sufficiently recreates the $\boldsymbol{S}$. To solve the set of equations, the software 'needs' the observed correlation matrix $\boldsymbol{S}$ that is calculated from the empirical data. Either we may estimate and 'feed' this matrix into the software, or we can 'input' the data set or the covariance matrix and let the software calculate the correlation matrix based on the input.

SEM is a set of analytical techniques used to empirically test theory by studying the hypothesized relationships between factors, and CFA is used to analyze each "measurement model" or measurement scale in a SEM model. The aim of SEM is to determine whether the *a priori* specified theoretical model is consistent with the data collected (Lei & Wu, 2007).

Structural Equation Modelling

In the section that follows, I describe typical steps followed when conducting SEM namely, model specification, model identification, model estimation, evaluation of model fit and model modification (Teo et al., 2013).

Steps in conducting a SEM analysis

Model specification: involves determining all the relationships among the different variables; through specification, the analyst develops the measurement and structural components of the SEM. A measurement model (CFA) specifies which indicators or items that "load on" each factor. In the resultant measurement model, a loading (depicted by an arrow) between an indicator and a factor indicates that the factor or trait governs the responses to the item. The squared parameter or "loading" indicates the proportion of variance in the responses to the item that the factor "explains", and a loading close to zero indicates that the factor does not influence the responses to the item (Teo et al., 2013). The product of any two loadings is the correlation between the two respective items.

Using CFA, we estimate factor loadings (λ), unique variances (specific variance and error variance), and factor variance of the latent trait. Unique variance is the portion of the variance in the indicator that is not explained by the latent trait. If substantively justified, also specified in the measurement model are relationships among the specific variances of indicators. We may refer to this as error covariance.

Conversely, the structural model depicts the associations between the factors. Herein, the factors may be either exogenous or endogenous. An exogenous factor is one that does not depend on other factors in the model, whereas an endogenous factor is dependent on one or more factors in the model. Associations among factors are depicted as 'paths' and are specified to estimate the relationship between the factors. Following successful model specification, the next step is model identification.

Model identification: this step concerns establishing whether the sample covariance-variance matrix ***(S)*** contains enough information to obtain unique solutions/estimates for the free parameters (FP). We may first check whether the model fulfils the *'order condition'* for identifiability, which is a necessary (but not the only sufficient) requirement for model identification (Schumacker & Lomax, 2004, p. 58). For models to fulfil the *'order condition'*, the number of FP must be less than or equal to the number of distinct values (DV) in the observed correlation matrix (FP ≤ DV). The number of DV is computed as $[p\ (p + 1)]/2$, where p represents the number of indicators in the model, and the model degrees of freedom (df) = DV-FP. In an *'under-identified'* model, FP > DV, df is negative; as such there is not enough information in the variance-covariance matrix to obtain a solution. This may happen when too many factors are specified. However, should additional constraints be imposed, then such a model may become identified (Kline, 2011, p. 9). In a *'just-identified'* model, DV = FP, $df = 0$; there is enough information in the variance-covariance matrix to obtain one unique solution, i.e. estimate only one value for each FP (cf. CFA with three indicators for each factor). Lastly, in an *'over-identified'* model, DV > FP, *df* is positive, and it is possible to obtain different values for all FP depending on "in which order" we solve the equations. Using MLE the computer nevertheless suggests one specific value for each FP. The suggested or output values are those that make the model-implied covariance matrix as close as possible to the observed matrix. With 'over-identified' models GOF indices are estimated, and we can use these to decide whether the model-implied matrix is 'sufficiently close' to the observed matrix.

Model estimation: this step aims to generate numerical values of the parameters in the model (both free and constrained). Different estimation methods in SEM packages like LISREL utilise different fitting functions (mathematical algorithms) to measure how close the $\boldsymbol{S}$ is to the $\sum$**.** Of these, MLE is the more widely used available estimation method (Brown, 2006, p.21). MLE is an iterative procedure, meaning that the final parameter estimates are obtained through a numerical search process which repeatedly refines these estimates in an attempt to minimize the residual matrix $\boldsymbol{R} = \boldsymbol{S} - \sum$**.** When the program arrives at a set of parameter estimates that cannot be improved upon to further reduce the residual matrix, that the program convergences. At this point, the program has identified the 'optimal set of values for all the FP' to calculate the model-implied matrix. The use of MLE must meet some statistical assumptions, failure to fulfil these may result in biased estimates. A key underlying MLE is that the indicators are normally distributed continuous variables. Failure to meet this assumption might result in chi-square-based model fit indicator that may be biased, leading to potentially misleading conclusions from the empirical data (Kline, 2011). For example, while MLE yields relatively accurate parameter estimates with non-continuous data, bias in chi- square and standard errors increases with non-normality (Mindrila, 2010).

In instances when we have data that does not fulfil these conditions, for example ordinal data that is not normally distributed, we use asymptotic distribution-free (ADF) estimators that make no assumptions about the distribution of the data – where "asymptotic" refers to large sample size > 1000. This typically includes robust DWLS. A 'robust' MLE approach is less dependent on the assumption of multivariate normal distribution; herein the standard errors and chi-square test statistics are corrected to enhance the robustness of MLE against departures from normality (Li, 2016). For ordinal data, such as data collected using rating-scales, DWLS is a suitable estimation method because the assumption of normality is violated (Li, 2016). Whereas it makes no assumptions on the distribution of the observed variables, it holds an assumption on the distribution of the underlying latent variable – that the underlying latent trait is normally distributed (Mindrila, 2010).

During model estimation, we can also examine the indirect, direct and total effects of one factor on another. Direct effects are the influences unmediated by any other factor in the model, whereas indirect effects refer to the influence of latent independent variables on latent dependent variables as mediated by one or more intervening factors (Bollen & Stine, 1990): total effects = direct effects + indirect effects (Bollen & Stine, 1990). Using standardized solution, these parameters indicate that "one standard deviation change" in one factor implies, for example, "a half standard deviation change" in another factor. Following successful model estimation, the next step in SEM analysis is to check whether the specified model sufficiently re-creates the observed correlation matrix.

Evaluation of model fit: this step of SEM modelling refers to determining the degree to which the model-implied variance-covariance matrix **(∑)** fits the sample variance-covariance matrix ***S*** (Cortina et al., 2017). Fit indices gauge the 'closeness' of ***S*** and ∑ in different ways. GOF indices are classified as 'absolute', 'parsimony-adjusted' and 'incremental'. Absolute fit indices, also referred to as 'global fit indices', are the main fit indices used to assess overall model fit, that is, 'how well the covariances predicted from the parameter estimates reproduce the observed sample covariances' (Diamantopoulos & Siguaw, 2013, p. 87). Examples of absolute fit indices used in the present study are the chi-square test (χ^2) and the standardized root mean square residual (SRMR).

The chi-square statistic (χ^2) provides a 'test of perfect fit in which the null hypothesis is that the model fits the population data perfectly' (Brown, 2006, p. 84; Diamantopoulos & Siguaw, 2013, p. 83). A statistically insignificant χ^2 value calls for rejection of the null hypotheses, implying imperfect model fit and possible rejection of the model. The number of degrees of freedom *(df)* serves as a standard by which to judge whether the χ^2 is large or small. In model fit, the concept of 'degrees of freedom' refer to the number of independent pieces of information that were used (Cortina et al., 2017).

Despite its popular use, there are limitations associated with the use of the chi-square test statistic. Firstly, because the chi-square fit statistic is a significance test, it is sensitive to sample size. As the sample size increases (N>200), so does the sensitivity of the test, as such the χ^2 tends to indicate poor fit on larger sample sizes (Schumacker & Lomax, 2004, p. 86; McIntosh, 2007). Secondly, the chi-square fit test assumes an underlying multivariate normal distribution. Therefore, when used on non-normally distributed data, there is increased likelihood of rejecting an otherwise good fitting model (McIntosh, 2007). Thirdly, the chi- square statistic assumes that the model perfectly describes the observed covariances, something that is often implausible in practice, as all models are approximations of reality. For this reason, Diamantopoulos & Siguaw (2013, p. 85) suggest reporting test statistics that do not follow a χ^2 distribution, but rather follow a non-central χ^2 distribution with a non- centrality parameter (NCP), whose estimate is obtained by: $\chi^2 - df$. The NCP is the measure of the degree to which the null hypothesis is false (Schumacker & Lomax, 2004, p. 94). A related chi-square statistic used for model assessment is the *reduced chi-square test* = χ^2/df. For single model assessment, a reduced chi-square value $\chi^2/df > 1$ implies poor fit, whereas $\chi^2/df < 1$ implies over fit. For model comparison a χ^2/df close to 1 implies a better fit (Andrae et al., 2010).

In programs such as LISREL 9.30 applying DWLS yields an output that includes additional test statistics such as the Satorra-Bentler (SB) scaled chi-square statistic which corrects for the assumption of multivariate non-normality (Hu & Bentler, 1998; Satorra & Bentler, 2010). Significant SB-scaled values > 0.05 point to acceptable fit (Satorra & Bentler, 2001).

The standardized root mean square residual (SRMR) is another absolute fit index, which is a summary measure of the standardized fitted residuals in the residual matrix. The size of the standardized residuals tells us the extent to which the model's parameters underestimate or overestimate the relationship between the indicator variables. We interpret standardized residuals in terms of *z* scores, values ≥ 1.96 (which corresponds to a statistically significant *z* score at $p < .05$) or ≥ 2.56 (which corresponds to a statistically significant *z* score at $p < .01$). Large positive residual values indicate the presence of additional covariance between the indicators that is not accounted for by the model i.e., that the model is 'under- fitting'. Under-fit may be handled by adding indicators or adding correlation terms between the specific variances while 'over-fit' may be handled by removing indicators from the model.

As $R = S - \sum$, a negative element in R means that the corresponding element in $\sum$ is larger than the corresponding element in S. Then, the FP estimates from which we estimate the elements in $\sum$ produce "too large" elements (too large variances and covariances in $\sum$). We then say that our model "over-fit" the relationship between the two respective indicators. As the SRMR indicates the extent of error resulting from the estimation of the specified model, lower values are more favourable (Teo et al., 2013, p.

11). SRMR values can range between 0.0 and 1.0; the smaller the SRMR, the better the model fit. SRMR values < 0.05 suggest a well-fitting model, a value of 0.00 indicates 'perfect' fit; however, values as high as 0.08 also point to 'acceptable' fit (Hu & Bentler, 1998).

Parsimony-adjusted fit indices: these fit indices have an 'in-built correction' for model complexity, thereby favoring simpler models which have fewer estimated parameters. Although we tend to get better-fit indices when we let more parameters be estimated (thus increasing the complexity of the model), we cannot be certain whether the fit obtained was due to the correct model specification or because of freeing more parameters. These fit indices take into account the number of parameters required to achieve a given chi-square value and therefore automatically penalize any model with many parameters (Schumacker & Lomax, 2004, p. 90). Because root mean square error of approximation (RMSEA) and information criterion indices like Akaike information criteria (AIC) adjust for model complexity, they are considered as parsimony-adjusted fit indices.

RMSEA focuses on the discrepancy between the $\boldsymbol{S}$ and $\sum$ but per degree of freedom thereby considering model complexity. The RMSEA shows 'how well the model with unknown but optimally chosen parameter values would fit the $\sum$ if it were available' (Diamantopoulos & Siguaw, 2013, p. 85. At the core of the RMSEA statistic is parsimony- the fewer the number of parameters, the better the 'fit'. While authorities in the field suggest various cut-off values of 'acceptable' model fit, Hu and Bentler's cut-off value of RMSEA <0.06 [good model fit] is widely used in SEM literature (Hu & Bentler, 1999; Hooper et al., 2008). RMSEA values between 0.08 and 0.10 suggest mediocre fit, whereas values > 0.10 point to poor fit (Hu & Bentler, 1999). When reporting RMSEA, methodologists recommend reporting the associated confidence interval, which indicates the precision of the estimate. With confidence intervals, often set at the 90% threshold, the width of the confidence interval is informative about the precision in the estimate of the RMSEA. Lower values close to zero and upper values less than 0.08 suggest acceptable model fit.

Additionally, we may report on the closeness of fit (Cfit) associated with the RMSEA value. Cfit is the probability that RMSEA <.05, and this probability should be >.05. (Distefano, 2013, p. 258).

AIC is an information criterion which can be used with single models, it adjusts for sample size, and lower AIC values point to a more parsimonious model (Akaike, 1974). It is worth noting that to use information criterion such as AIC, the sample size must exceed the 'critical N statistic' (CN). CN is the sample size that the sample should be in order to accept the obtained fit (measured by χ^2) of a given model on a statistical basis (Bollen, 1990; Schumacker & Lomax, 2004, p. 41). Hoelter's CN is often used as the standard sample size and is computed as: CN = $\chi^2 + 2q$ where q = number of FP in the model (Teo et al., 2013, p.11).

Incremental (relative/comparative) fit indices: these indices measure the proportionate improvement in model fit by comparing the model being tested with a baseline (null) model, in which all indicators are uncorrelated (Diamantopoulos & Siguaw, 2013, p.87), i.e. all elements in $\boldsymbol{S}$ are 0 so $\boldsymbol{S} = 0$. Put another way, incremental indices check whether the model- implied correlation matrix is closer to the observed correlation matrix than a hypothetical correlation matrix that consist of only 0's. Examples of these indices applied in the present study include the non-normed fit index (NNFI) and the comparative fit index (CFI).

The non-normed fit index (NNFI), also referred to as the Tucker-Lewis index (TLI), assesses model fit by comparing the χ^2 of the specified model with that of the baseline model/ The TLI favours simple models and values >0.95 indicate good model fit (Schumacker & Lomax, 2004, p. 76).

The CFI indicates the relative lack of fit to the baseline model (the one in which all indicators are uncorrelated). Values range between 0.0 and 1.0, with values that are closer to 1.0 indicating good 'fit'-that the $\boldsymbol{\Sigma}$ and $\boldsymbol{S}$ is close. To ensure that mis-specified models are not erroneously accepted, a CFI value $\geq$ 0.95 implies good model fit (Hooper et al., 2008). Considering the number of fit indices that are available for model evaluation, it is advisable to report on at least one index from each of the three categories (absolute, parsimony-adjusted, comparative). Authors caution against relying on only the global fit indices of overall model fit, adding that in some instances, although global measures of fit might indicate a satisfactory data-model fit, certain estimates of parameters corresponding to the hypothesized relations may be non-significant (Diamantopoulos & Siguaw, 2000, p. 88). For these reasons, it is important that SEM analysts provide a detailed assessment of the measurement model and of the structural model in tandem with the overall model fit. It is judicious to begin with the evaluation of the measurement models because they are the 'building blocks' of the structural model. For this reason, we used RM and CFA to validate each of the three scales before putting them in a SEM.

Assessment of the measurement model

We may view each measurement model in a SEM model as a CFA model. Assessing fit of the measurement model entails determining the validity and reliability of the indicators used to represent the latent factor.

Evaluating validity and reliability of the measurement model

To evaluate the extent to which indicator variables measure what they are supposed to measure i.e. validity, and the extent to which they are consistent in measurement, analysts examine the factor loadings (λ) in the measurement model. Examining the factor loadings in the standardized solution indicates the correlation of the indicator variable with its respective factor. Standardized loadings >0.70 imply that the factor "explains" more than $0.70^2 \approx 0{,}50$ or more than 50 % of the variance in the responses to the indicator (Hair et al., 2010). Items with loadings <0.70 means more 'unique variance' and less variance common with the other items. Rasch modelling (RM) (see section below) is a different way to validate measurement scales. Using RM, we rely on 'probabilities' rather than 'item correlations'.

By examining the squared standardized loadings of the indicators (R^2), we are able to tell how reliable the indicator is, that is, the proportion of variance in an indicator that is explained by the underlying latent factor (Diamantopoulos & Siguaw, 2013, p. 90). High R^2 values denote high reliability for the indicator variable concerned. Following assessment of the reliability of measurement model, the next step is to assess the structural model.

Assessment of the structural model

Assessing the structural model aims at checking whether the theoretical relationships between the factors are supported by the data. To do this, we check for the validity and reliability of the structural model.

Evaluating validity and the reliability of the structural model

This involves checking the magnitude of the standardized parameter estimates (between the factors). Evaluating validity of the structural model provides information on the strength of these associations and can provide additional insights into the relative impact of exogenous factors on endogenous factors. Of relevance, too, is examining the direction of the

hypothesized relationships, as indicated by the signs (positive or negative). These signs should correspond with the theory underlying the hypothesized relationship.
Evaluating the reliability entails examining the squared standardized loadings of the indicator variables, these indicate the amount of variance in each endogenous factor accounted for by the exogenous factor that influences it. For example, in the present study, we interpret the R^2 values as the amount of variance in CNL-E and CNLEng accounted for by SEBH, and the variance in CNLEng that is accounted for by CNL-E as it is a mediating factor, i.e. is both an exogenous and endogenous factor. Furthermore, by examining the standardized parameter estimates for the paths between the factors, we can gauge the relative impact of the exogenous factors on the endogenous factors. Higher values indicate greater impact of the exogenous factor on the endogenous factor(s).

Following model evaluation, should model fit be less than adequate if necessary, analysts must then proceed to make changes to the model through re-specification during *model modification.* Modifying the model to improve fit may involve adding or deleting parameters, changing the parameters from fixed to free or vice versa. One way to examine misspecification and therefore correct it in a re-specified model is to examine the residual matrix R for large residuals (>2.58). Large positive residuals point to a model that underestimates the covariance between the indicator variables involved (under fitting). In this case, modification will require addition of paths achieved by freeing parameters fixed to zero in order to better account for that particular covariance between the two variables (Diamantopoulos & Siguaw, 2000, p. 108). Conversely, large negative residual values indicate that the model overestimates the covariance between the indicator variables (over fitting). Modification in this case will require the deletion of paths achievable by (fixing the relevant parameter that are associated with the covariance between the two variables to zero.

Another way is to examine the modification indices of the non-free parameters that is fixed at zero. A modification index (MI) provides an estimated value in which the model's chi-square test statistic would decrease if a (previously fixed) parameter were added to the model and freely estimated (Whittaker, 2012). Fixed parameters with large MI values (larger than a χ^2 critical value of 3.84, which corresponds with 1 *df* at an alpha level of 0.05) should be examined to check whether it is theoretically plausible to include them in the model and freely estimate them. However, it should be noted that MI is dependent on sample size and it is thus advisable to consider an MI in conjunction with its associated 'expected parameter change' (EPC) which is the approximate value of the new parameter if added to the model. Accordingly, large MIs with large EPCs suggest that that freeing the parameters might improve model fit. Some key considerations when using MIs include ensuring that the suggested changes are theoretically sound to avoid data-driven models that may not be generalizable across samples. Second is that parameter estimates that produce the largest improvement in fit be freed one at a time, this process is repeated until an adequate fit is achieved. Lastly, whenever possible, the re-specified model should be tested on a different sample to reduce type I error.

Rasch Analysis

Rasch analysis (RA) is the formal testing of ordinal data generated by items up against a Rasch model; testing whether the data fit the selected Rasch model by assessing whether the response pattern observed in the data corresponds to the theoretically expected pattern (Tennant & Conaghan, 2007). The Rasch model may be viewed a 1-parameter logistic model, belonging to the broader framework of item response theory (IRT). IRT is a statistical theory framework comprised of mathematical methods expressing the relationship between observed item responses and an underlying psychological trait or factors (Cappelleri et al., 2014). According to the Rasch model, there are two determinants of item response, namely respondent's standing on the latent trait i.e. person parameter *(*θ*)*, such as proficiency, and the item's location *(b)*, which may refer to difficulty. The probability that a person will affirm an item is a logistic function of the difference between θ and *b*. Rasch models are classified depending on the response structure used, such as *dichotomous* or *polytomous*.

For dichotomous response structures in which there are two response options such as yes/no or right/wrong, the item difficulty parameter *(b)* corresponds to the 'threshold' and represents the probabilistic midpoint between two adjacent response categories at which the probability of choosing one over the other is 50%.

Conversely, for polytomous response structures in which there are more than two response options, such as rating scales, we apply a polytomous Rasch model to represent the non-linear relationship between the respondent's location on the latent trait *(*θ*)* and the (conditional) probability of responding in a particular response category.

Use of polytomous response scales in research

The use of polytomously scored items in research is often preferred because they provide more information about the examinee using more response options (Dittrich et al., 2007). Polytomous response options extend the range over which the items measure the latent trait, and resultantly the range for which items provide information about the person locations (Salzberger, 2015, p. 380). Examples of commonly used variations of the Rasch model for polytomous response structures are Andrich's rating scale model (RSM) and Master's partial credit model (PCM) (Andrich, 1978; Masters, 1982).

The RSM is an extension of the dichotomous Rasch model applicable when all the items share the same response scale. The RSM assumes that the distance between each pair of ordinal response categories is equal across all items, resulting into equidistant thresholds across all the items (Andrich, 1978; Shea et al., 2009). Because all items share the same response scale structure, the RSM favours parsimony, which is important when comparing polytomous Rasch models.

The PCM is also an extension of the dichotomous Rasch model that makes no assumptions about the distances between ordinal response categories (Masters, 1982). The PCM permits that each item has its own unique response scale; creating a partial scale for each item introduces extra parameters by $(L-1)*(m-2)$ where L=number of items and m=number of response categories in the response scale, thereby increasing the number of parameters and making it a less parsimonious model (Sick, 2009, p. 7; Wright, 1998).

Assumptions of Rasch models

There are two basic assumptions of Rasch models namely, *unidimensionality* and *local independence.* As per the first assumption of unidimensionality, when data fits the selected Rasch model, the items summed together form a unidimensional scale (Tennant & Conaghan, 2007). When there is more than one latent trait captured by the items in the scale, this suggests violation of the assumption of unidimensionality, and we refer to this as trait dependence (Marais & Andrich, 2008). To investigate the presence of trait dependence, we may use the principal component analysis (PCA)/*t*-test protocol proposed by Smith (2002). The PCA examines the correlations between item residuals, where item residuals are defined as the difference between the observed and the model expected value, and isolates subsets of items that may reflect subscales present (Humphry, 2002). Based on two subsets of items, two sets of person locations are estimated for each respondent (dependent estimates) and, using dependent *t*-tests, we may estimate the proportion of respondents with significantly different person location estimates on the two subsets. If 5% or less of the dependent *t*-tests are significant, then unidimensionality is implied (Tennant & Conaghan, 2007). While values exceeding 5% suggest violation of the assumption of unidimensionality – that the two items subsets measure two different but related aspects of the latent trait. The presence of different aspects or subdomains may reflect the complexity of the latent variable of interest. There is a trade-off between unidimensionality and the validity of the measurement of a composite or 'multidimensional' trait (Marais & Andrich, 2008).

According to the assumption of local independence, the latent trait accounts for the entire association between the items. When items are linked in such a way that response to one item determines the response to a subsequent item, this means that the items share something more in common than the other items. Therefore, when two or more items asking for corresponding information or contain similar item content, they may display response dependence (Marais, 2013, p. 113). The effect of response dependence on measurement is varied, however, "it is well known that positive local response dependence inflates measures of reliability in Rasch models" (Marais, 2013, p. 115). For this reason, it cannot be ignored especially in validation studies such as the present. If the data fit the Rasch model, then residuals for any pair of items should be uncorrelated and close to zero. High item residual correlations may suggest that the pair of items has something more in common than all of the items have in common take together. Response dependence is typically checked using item residual correlations and estimated using Yen's Q3 statistic which represents the correlation between the two items after accounting for the person location estimates (Christensen et al., 2017). We compare each residual estimate up against the average residual correlation of the item set. When the condition of response independence holds, we anticipate the expected value of the Q3 statistic = -1/ (n-1), where n = number of items in the scale, a cut off of ±0.2 is used to identify response dependence. Accordingly, item pairs with residual correlation values that exceed the cut-off suggest possible violation of the assumption of response dependence.

Measurement theory underpinning Rasch analysis

The measurement theory underlying Rasch analysis fulfils the requirements of *fundamental measurement.* The Rasch model is a theoretical mathematical description of how fundamental measurement should operate with social or psychological attributes (Bond, 2015, p.275). Campbell (as cited by Tal, 2015) referred to “measurement procedures that satisfy the conditions of additivity ‘fundamental’ because they do not involve the measurement of any other magnitude). Rasch analysis operationalizes fundamental measurement by prescribing a structure essential to qualify a latent variable and relate observed scores (counts) and numbers thought to represent magnitudes of the latent traits of interest (Bond, 2004). Characteristics/principles of fundamental measurement include, invariance, specific objectivity, monotonicity and additivity/sufficient statistic (Wright, 1997).

The principle of *invariance* states that the values/measures attributed to variables by any measurement system should be independent of the measurement instrument used. Placing items and persons on the same linear scale provides a convenient framework for assessing a scale’s invariance meaning that item and person location estimates are independent of each other (Bond, 2004). This forms the basis for the requirement of invariance or stability. The validity of measures should derive from the principle of invariance of Rasch measures (Bond, 2004). Invariance regarding item parameters means that the item locations can be estimated independently of the distribution of the item’s location (Salzberger, 2013, p. 352). It thus follows that item estimates should remain invariant across analyses conducted in different settings, and that the items have to work invariantly across individuals and groups of individuals.

Invariance regarding person ability estimates requires that the difficulties of the items relative to person’s level of trait should remain stable across two substantially different subsamples. What this means is that all persons, regardless of their level of trait, taking the test must identify the same item as ‘easiest’ and the same item as ‘hardest’. When the principle of invariance is violated, it means that in addition to the latent trait, there is another underlying factor interacting with the persons/items causing the respondents at the same trait levels to experience the items differently. It also implies that the items function in a different way for different subgroups of respondents. This violation manifests as differential item functioning (DIF). In the presence of DIF, valid comparisons between subgroups of the respondents are not possible (Tennant & Conaghan, 2007).

The principle of *additivity* stems from the characteristic that in cases of fundamental measurement, the respondent’s total score is a *sufficient statistic*, meaning that the information needed to estimate the person’s location is contained in the total score. Likewise, the total score of an item, which is obtained by summing all the scores of different respondents, contains all the information for estimating the item location (Andersen, 1977). Therefore, all persons with the same total scores will get the same location estimate, irrespective of response pattern. As it fulfils this assumption, the Rasch model allows the responses (raw scores) from the items in the measurement instrument to be summated into the total score.

The principle of *monotonicity* implies that the probability of accurate response to an item increases with an increasing in persons' level of trait. When polytomous items meet this principle, the result is ordered response categories. A lower threshold category corresponds to a lower level of trait, while a higher threshold category corresponds to a higher level of trait (Mesbah, 2013, p. 242). When data fit the Rasch model, we can then ascribe meaning to the raw score (in terms of level of trait). However, if the response categories are wrongly scored, the item will misfit, and rescoring the item may resolve the misfit polytomous responses must be scored with successive integers starting at zero, meaning that the raw score is a count of all the thresholds the respondent has passed (Salzberger, 2010). Therefore, there is no need for 'half points' like 0.5 or 1.5. However, observed scores on response scales are, at best, ordinal.

The principle of *specific objectivity* means that the items and persons are independent of each other, that any person location estimate is independent of the specific items applied (Nielsen & Kreiner, 2013, p.318; Stenner 1994).

Rasch analysis provides a framework in which linear measures from counts of quantitatively ordered observations can be made (Salzberger, 2010). Only after this is *additivity* of scores from individual item scores into an overall score justified (Hagquist et al., 2009; Perline et al., 1979).

Steps in conducting a Rasch analysis

Tennant and Conaghan (2007) outline some key steps that one should follow and report on in any study in which one uses Rasch analysis for instrument development and or refinement. In brevity, these steps test how well the data sampled fits the selected Rasch model. In the section that follows, I describe each of these 'diagnostic' steps.

Selection of model of best fit: The goal of model selection is to find the most parsimonious model that adequately explains the observed item responses. To select the most suitable Rasch model for use, the number of response categories guides us. For a dichotomous number of response categories, we select the dichotomous model whereas for polytomous response patterns we use derivations of the polytomous model i.e. either the RSM or the PCM. As Rasch analysis favours the ideal of parsimony, the preferred model from the get-go is the RSM, which has fewer parameters than the PCM. Accordingly, the 'null' hypothesis is that the data fits the RSM derivation of the polytomous unidimensional Rasch model (PURM). For model selection, analysts often use the 'likelihood ratio test' (LRT) statistic, which provides a criterion for determining whether the addition of parameters creates significant improvement in fit of the model to a dataset. The LRT is useful when comparing nested models, in which the more complex model differs from the simpler model only by addition of parameters. When selecting a polytomous Rasch model for use, the RSM is nested within the PCM. The LRT is a statistical test calculated as the models' difference in -2LL (minus two times the log likelihood estimate) or 'deviance'. For large sample sizes the LRT has an approximately normally distributed chi-squared distribution. The associated degrees of freedom *(df)* for that distribution is equivalent to the difference between the number of parameters between the models, such as RSM and PCM: *df* = (number of parameters in the complex model - number of parameters in the simpler model). The larger the LRT value, the stronger the evidence that the more parsimonious model is a less than adequate fit to the data, compared to the more complex model. Associated p-values give insight into whether the more complex model fits the data significantly better than the simpler model. Significant values ($p<0.05$) indicate that we should reject the null hypothesis, that the more complex model provides a better fit to the data than the simpler model.

Ordering of thresholds: When dealing with polytomous data, such as that collected in the present study, thresholds partition the latent continuum of each item into adjacent intervals (Brown et al., 2015). With polytomous data, thresholds indicate the points on the latent scale where the conditional probability of two adjacent response categories is equal (Andrich et al., 1997, p. 61). Examining the ordering of thresholds checks whether the data reflects the increasing levels of trait across the response categories for each item. Although threshold estimates are essentially data-dependent, examining thresholds might give insight into the suitability of the response scale or format. Since the thresholds are located on the same continuum as the overall item location, the thresholds, and by implication the response categories they separate, are also supposed to represent more or less of the trait to be measured (Salzberger, 2015). Hence, the empirical ordering of the thresholds should match the proposed response format. Ordered thresholds imply ordered response categories, that the responses to the items are consistent with the underlying trait (Tennant & Conaghan, 2007). Ordered thresholds also indicate that the respondents used the response scale as intended. Disordered (reversed) thresholds do not imply that there is a problem with the response scale but rather suggest that the estimates of the thresholds defining the categories in the item are not ordered as require, an indicator that there are issues with the data arising from how the respondents used the response scale (Hagquist & Hellström, 2014; Salzberger, 2015). For example, when respondents with higher levels of the trait of interest do not consistently endorse higher response options for that specific item, their level of trait is not consistently reflected in progressively higher scores on the response scale options for specific items. In addition, when there are more response categories than the respondents can distinguish, we may encounter disordered thresholds. Disordered thresholds may also indicate that too few of the respondents have used a specific response category. To correct for disordered thresholds, we collapse adjacent response categories therefore reducing the number of response categories (Andrich et al., 1997). However, it is important that changes to the response format not be based on statistical evidence alone, engaging respondents through qualitative interviews might reveal avenues to more suitable response formats. When disordered thresholds occur, inspecting 'item category probability curves' might provide insight into possible causes for the disordering.

Category probability curves show the probability of each response category along the latent trait continuum and are determined by the response frequency in each response category (Wetzel & Carstensen, 2014). Accordingly, the more respondents that endorse a response category, the higher its category probability. For a well-fitting item, each response option has a highest probability of endorsement across the range of trait (Pallant & Tennant, 2007). It is worth noting that ordered thresholds mean that the threshold values must conform to the Guttman structure (Hagquist et al., 2009). This leads the next diagnostic step of 'item fit analysis'.

Assessing item fit to the Rasch model: The concept of 'item discrimination' refers to the degree with which an item separates individuals with higher person location estimates (trait level) from those with lower person estimates (trait level), and indicates the relevance of the item to the trait that is being measured. Rasch theory entails some degree of probabilistic uncertainty in responses to the items. Fit residuals are person-item differences between the observed data and theoretically expected values i.e. what is expected by the model for each person's response to each item (Marais, 2013, p. 120).

Item-fit statistics indicate how well items fit the Rasch model, and they can have 'good fit', be 'under-discriminating' or 'over-discriminating'. For acceptable item fit, the absolute value of z-fit residuals for individual items should be smaller than 2.56. When items *'over fit'* the selected Rasch model, it implies that they are too predictable and that there is little variation in responses. These items, identified by z-fit residual values below -2.56 imply that whereas these items separate persons according to the latent trait, they do so in a limited range, thereby providing little information to separate persons below or above that interval. While 'over fitting' items do not degrade the quality of measurement, they provide no unique information about the respondents (Sick, 2011). However, analysts should pay attention to these items because they inflate estimates of reliability because of the strengthened person separation (Wright et al., 1994).

Conversely, items that *'under fit'* the Rasch model have z-fit residual values exceeding 2.56 suggesting that the particular item does not separate sufficiently between persons with different standing on the latent trait.

As sample size highly influences chi-square statistic values, there is an increased risk of over rejecting items when we assess data-model fit using chi-square tests. Therefore, to assess how stable an item is, it is advisable to examine the individual item fit in amended sample sizes. Linacre (2002) suggests using sample sizes corresponding to 10, 20 and 30 persons per threshold. Items that fit the Rasch model in the sample size with the least number of persons per threshold (10) are the most informative and should be retained.

Although researchers rely on the chi-square sample size adjustment that is available in the Rasch Unidimensional Measurement Model (RUMM) program, its use is somewhat controversial. Alternatively, we may select five or more random 'amended' or smaller samples from the data set and create adjusted chi-square values. However, as chi-square testing involves 'more than one statistical test', analysts may adjust for the effect of multiple testing i.e. the significance of testing k number of items, by using a Bonferroni-adjusted individual item chi-square p-value. The Bonferroni-adjusted item chi-square p-value is $0.05/k$, where k is the number of statistical tests i.e. the number of items in the scale (Bland & Altman, 1995). That, alongside the z-fit residual estimate, gives insight into how well the item fits the Rasch model.

In addition to the fit statistics, we can also examine fit to the item characteristic curves (ICC) of the individual items. The observed proportion of responses for groups of respondents across the trait (class intervals) is plotted against the expected proportion ICC. Ideally, the observed values should fit perfectly with the theoretical curve (Hagquist et al., 2009). Items with good fit will show each of the group plots lying on the curve, those with plots that are steeper than the ICC indicate over discriminating items while those flatter than the curve point to under discriminating items (Pallant & Tennant, 2007). Following this, the next diagnostic step is assessment of person fit to the Rasch model.

Assessing person fit to the Rasch model: this gives insight into how well the respondents' response structure fit the expectations of the Rasch model. Person fit estimates indicate how well the observed response pattern matches the expected response pattern. A few respondents with bizarre response patterns may seriously affect fit at the item level (Pallant & Tennant, 2007). The expected response structure for the Rasch model is probabilistic, this means that the probability of endorsing an easier item has to be higher than that for endorsing a more difficult item. As with the item fit residuals, persons can have acceptable (good) fit, under fit or over fit the Rasch model. Acceptable z-fit residual values should exceed 2.56.

Persons that *'under fit'* the Rasch model have fit statistics values that are larger than +2.56. Such respondents may typically have responded correctly to 'easy' questions and wrongly to 'difficult' questions. Examining these respondents' responses might indicate lack of concentration, motivation, situational distraction, guessing and or cheating (Sjaastad, 2014). Conversely, persons that *'over fit'* the Rasch model have fit statistics values that are smaller than -2.56. This means that their response pattern is closer to the Guttman pattern than we would expect by chance. In cases of person misfit, analysts can remove these persons from the data set, if model fit improves in a validation study, then removal of these persons may be justified. Following assessment of person and item fit to the Rasch model, the next diagnostic step is to check for differential item functioning (DIF).

Checking for differential item functioning: DIF or item bias affects model fit because it violates the property of invariance and thereby compromises the ability to scale distinct groups onto a common metric (Reise & Waller, 2009). One way of checking for DIF is by applying statistical and graphical evidence i.e. analysis of variance (ANOVA) and ICCs. To check for DIF using ANOVA, the software program (such as RUMM2030 in this study) groups the respondents according to the trait levels, and then calculates the observed mean values for the different person factor levels, for example girls and boys (gender). The presence of significant values at p=0.05 (main effect) within the trait group levels suggest the presence of DIF. Investigating the ICC gives graphical evidence of the type of DIF that is, whether the differences in item response happen consistently along the trait (uniform DIF) or at different trait levels (non-uniform DIF). When an item displays uniform DIF, the issue may be remedied through splitting the item by person factor level and separating the item for each person factor level. However, in cases of non-uniform DIF, there is little that can be done to correct the issue and often item deletion is the preferred solution (Pallant & Tennant, 2007; Walker et al., 2001). After resolving the issues of threshold disordering, DIF, item and person fit, Tennant & Conaghan (2007) suggest that the next diagnostic step in conducting Rasch analysis is to check the targeting of the scale.

Checking the targeting of the scale: the Rasch model builds a hypothetical unidimensional line along which the items and persons are located according to their difficulty and level of trait measures. Targeting of a scale gives insight into how well the range of trait measured by the items addresses the range of trait in the sample. Ideal targeting which is when the severity of the items corresponds with the level of trait in the sample is important because poor targeting may cause imprecise estimates of the item parameters as well as the person parameters (i.e. high standard errors) (Hagquist et al., 2009). Less than optimal targeting compromises the reliability of the questionnaire making it difficult to differentiate between persons at different trait levels. Bad targeting also increases the likelihood of extreme scorers, which are situations in which respondents score in the highest or lowest categories across all items (Hagquist et al., 2009).

To check the targeting of the scale, analysts must compare the threshold distributions of the items. By default, in the RUMM2030 program, the mean item location is set at zero, we compare this against that of the person's mean location score. For a 'well targeted' measure, the distribution of the item thresholds is in the same region of the person locations depicting the trait levels of the respondents. A positive mean location value suggests that the sample as a whole is located at a higher trait level than the average difficulty of the scale, whereas negative mean location values suggest the opposite. Following this, the next diagnostic step is to examine the structure of the questionnaire.

Examining the structure of a questionnaire/test: RUMM2030 output includes a Wright map (person-item map) referred to as a person-threshold map. It shows the how person and item locations can be plotted on the same continuum along the axis (Ayele et al., 2014). The logit is the plot common unit of measurement; on the person-item map, the vertical scale line represents the unidimensional scale with interval properties. To the left of the vertical scale are the persons distributed according to level of trait whereas to the right is the item thresholds at the varying levels of the trait being measured. A well-structured questionnaire is one in which the order of questions reflects the increasing level of trait being measured. The person-item map allows us to compare the predicted order of item difficulty with the actual order of item difficulty in the data set. Examining the person-item map gives insight into construct validity by providing evidence that the instrument is measuring in a way that matches what a theory would predict. The next diagnostic step is evaluating the reliability of the questionnaire.

Evaluating the reliability of the questionnaire/test: reliability indices give an indication of how well the scale items separates and ranks the persons. Given unidimensionality, reliability indices measure the precision of the questionnaire/test. In RUMM2030, the person separation index (PSI) is an estimate of the internal consistency reliability. PSI values range between 0 and 1.00, PSI values exceeding 0.80 suggest that the scale items sufficiently separate the persons responding to the questionnaire (Tennant & Conaghan, 2007).

Evaluating the validity of the questionnaire: Rasch analysis is a powerful tool for evaluating scale validity. Threats to validity include construct underrepresentation and construct irrelevant variance (Ravand & Firoozi, 2016). Items that do not fit the Rasch model may suggest potential multidimensionality, item dependence, all of which affect the content aspect of construct validity. Examining the item map for gaps and overlaps between items also gives insight into the validity of scales tested against the Rasch model. The item hierarchy represents a spread of the items and gaps along the unidimensional continuum are indications of the construct under-representation suggesting that there may be big differences between the item difficulties meaning that the test does not measure the people with trait levels close to this part of the line.

Advancing research in budding fields such as nutrition literacy requires development of measurement studies. Having identified the benefits that CTT and IRT confer upon the process of instrument development and refinement, the present study sought to advance and thereby contribute, however small, towards development of measurement research in the field of nutrition literacy, specifically CNL.

Methodology applied in the study

As previously described in my theoretical perspective, I adopted the post-positivist theoretical perspective within the scientific paradigm of research. Without reiterating the details of the philosophical lens through which approached the present study, in the chapter that follows, I describe the methodological underpinnings that I undertook in the study, keeping in mind the beliefs of post-positivism. Herein I describe the methodological steps undertaken in the present study, beginning with the study design and sampling procedures used. I then describe the different measurement instruments (scales) applied in the present study, providing an overview of the process of instrument development. Following this, I describe the main data analysis procedures applied and how I handled missing data in the present study. I end the chapter detailing the ethical considerations made in the present study.

Conceptualization of the present study

The present study applied a cross-sectional study design, which is a study design that allows one to collect data from a representative sample from the population to make inferences about a population of interest at one point of time (Levin, 2006). In the subsection that follows, I describe the processes that I undertook addressing the different aspects of the study design considerations, as outlined by Hall beginning with a definition of the study population, the outcome variables of interest, classification of the independent variables, and formulating the hypotheses to be tested in the study (Hall, 2011, p. 172).

Definition of the study population

The study population in the present study was adolescents aged 15-16years enrolled in the tenth grade in schools in Norway. This population was selected because the tenth grade marks the end of compulsory education in Norway. By the end of this grade, it is assumed that the adolescents have acquired the skills they need for everyday living such as appraising nutrition information that they may encounter in the media, from peers and applying it in their dietary decisions.

Outcome variables of interest in the study population

The main outcome (dependent) variables of interest in the present study were CNL and SEBH. CNL was defined as the proficiency in critically analysing nutrition information and advice, as well as having the will to participate in emancipatory action to address barriers to good nutrition and influence the underlying determinants of health (Guttersrud et al., 2014). SEBH was defined as the subjective judgments of the adolescent's capabilities to organize and execute courses of action to attain designated goals in the specific topic of the Norwegian school subject 'Body and Health' which is taught through grades one to ten (Mullis et al., 2016).

Sample design employed in the present study

In order to select a representative sample from the study population, schools were sampled from a list of all lower secondary schools in Norway. From this list, 200 schools were randomly selected, and the school principals were contacted by email and telephone seeking consent to volunteer in the study on behalf of their students. As part of determining the sample design, the team scheduled data collection to coincide with the triennial PISA survey of 2015 citing convenience in terms of resource allocation, this was convenient for the participating schools. Therefore, the three scales (SEBH, CNL-E, and CNLEng) were included in the electronic survey system so that the participants responded to the instruments at once.

Calculation of sample size

As the main data analytical framework applied in the present study was RM, the criterion applied to determine sample size requirements for the present study followed the recommendation by Linacre who suggests having a minimum of ten (10) persons per threshold (Linacre, 2002). Accordingly, as all the scales used a six-item rating response scale, the minimum sample size required as per the six-item scale was calculated as follows:
N = (no. items *no. thresholds) *10

No. thresholds = (no. response categories - 1)

Hence, the minimum sample size (N) = (6*5)*10 = 300

The emergent sample size used (N = 1622) was well over five times the recommended sample size.

Scale design and use of rating scale items

In order to collect data about latent traits that are difficult to quantify such as CNL and SEBH, several authors suggest using questionnaires, which are groups of items designed to assess different aspects of one underlying latent trait (Artino et al., 2014).

Considerations followed in development of the scales

Artino and colleagues (2014) outline different steps that instrument developers should follow when developing questionnaires for use in academic research. The following considerations were made while developing the items in the three scales applied in the present study.

Firstly, in order to obtain a clear understanding of the definitions of the constructs under study, the developers of the items conducted a review of literature about the constructs of interest, specifically how these constructs have been operationalized in previous empirical studies, identifying which measurable variables are often used to operationalize these constructs, and the data analytical procedures used to develop and use the emergent measurement instruments. They then used this information and empirical evidence to develop the individual items for the three scales (CNL-E, CNLEng, and SEBH).

Secondly, because CNL is a narrowly defined concept and is still a budding area of research, the developers ensured that there were a minimum of six items for each of the two CNL scales. This followed) a recommendation by Gehlbach & Brinkworth (2011) to include more items that you might require, because of the likelihood of item deletion during instrument refinement.

Thirdly, development of the SEBH scale followed Bandura's guidelines for developing measures of self-efficacy (Bandura, 2006). Differing from existing measures of self-efficacy in the academic field, the SEBH scale was not 'general' and was therefore reflective of self-efficacy in the subject of 'Body and Health'. This lent validity to the instrument.

Use of a rating-scale response scale

The ease of use and adaptability for measuring different constructs makes Likert type response scales very popular (McCoach, Gable & Madura, 2013, p. 48). While several rating scales use five (5) alternative response options, the present study used six (6) response options because six-point rating scales have higher trends of discrimination and reliability, compared to five-point Likert-type response scales (Chomeya, 2010). For each of the three scales applied in the study, the extreme response categories of the response scales were anchored with a phrase (1) = 'strongly disagree' and (6) = 'strongly agree'. Whereas Krosnick (1999) cautions against labelling only the end points of the response options arguing that this leaves the meaning of the unlabelled options open to respondents' interpretation which may increase measurement error, this was not an issue for the present study as using Rasch analysis negates the dependence on the instrument (cf. *invariance*). Furthermore, during validation, checking for threshold ordering provided insight into how the respondents used the response options provided.

Scales applied in the present study

In the following section, I present the theoretical background that informed item development in each of the scales applied in the present study-the critical appraisal of nutrition information (CNL-E) scale, the engagement in dietary behaviour (CNLEng) scale and the perceived self- efficacy in the science subject of 'Body and Health' (SEBH) scale.

Development of the items in the CNL-E scale

In the study, CNL-E was defined as being proficient in critically analyzing nutrition information and advice as well as having the will to participate in emancipatory action to address barriers to good nutrition and influence the underlying determinants of health (Guttersrud et al., 2014) The respondents endorsed how easy they felt that it was for them to comprehend and interpret nutrition information from 'traditional' and 'new online' media sources, inter-personal interactions. In addition, some of the items sought to establish how respondents felt they could apply science knowledge to identify and falsify nutritional claims in media (Table 1).

Table 1: Wording of the items in the CNL-E scale

Item	Wording
1	Evaluate whether nutritional advice in the media (newspapers, magazines, television) is reliable?
2	Consider how reliable warnings about poor nutrition are, as warnings against malnutrition?
3	Consider whether information on websites for nutritional information is reliable?
4	Consider what it takes a scientific nutritional claim to be valid?
5	Evaluate nutritional advice in the media (newspapers, magazines, television) in a scientific way?

The response scale was anchored with the phrase "Nutrition" refers to the connection between diet and health. On a scale from "very difficult" to "very easy", how easy or difficult would you say it is to (1 = Very difficult, 6 = Very easy)

Theoretically, the five items in the CNL-E scale were generated basing on two frameworks namely, the integrated conceptual model of health literacy relating to the nutrition context and the PISA framework for assessing science literacy within the personal context (Sørensen et al., 2012; Thomson et al., 2013). From the integrated model of health literacy by Sørensen et al.(2012), the items reflected competences that are related to the process of 'understanding' and 'appraising' health-related information that is accessed from different sources (CNLE1-CNLE4). These sources were 'print' media including newspapers, magazines, 'traditional' sources such as television, and online media such as websites. Since critical evaluation of nutrition information requires judging information based on factual knowledge, the fifth item was also inspired by one of the three competencies within the PISA framework for assessing scientific literacy namely 'explaining phenomena scientifically' (Bybee, et al., 2009). This item reflected competencies associated with the application of science knowledge to identify and falsify nutritional claims in the media (CNLE5).

Development of the items in the CNLEng scale

The CNLEng scale applied in the present study is a modified version of the six-item 'Engagement in dietary behaviour' developed by Guttersrud & Pettersen (2015). The items in the scale explored the extent to which respondents were concerned about access to healthy foods at personal, social and global level and their level of commitment to ensuring that there is adequate access to healthy foods through their involvement in political action to address barriers to accessing healthy food. However, present study was restricted to the two items in the personal level of the CNLEng scale.

Table 2: Wording of items in the CNLEng personal subscale

Item	Wording
1	I am concerned about eating foods that provide the nutrients my body needs
2	I am concerned that there are healthy foods in the grocery shops that my family shops at

The items in the CNLEng scale were anchored with the phrase 'How much do you agree with the claim (1 = Disagree strongly, 6 = Agree strongly)'

Theoretically, the items included in the CNLEng scale were inspired by the multicomponent model of nutritious needs (Freedman, Blake & Liese, 2013). In the present study, only items at the personal level from Guttersrud and Pettersen's (2015) scale were included in the CNLEng scale. These two items explored the extent to which respondents were concerned about the quality of their diet in relation to their nutritional needs and the availability of healthy food options at the stores where their family shops. Herein, the domains of 'service delivery' and 'personal' informed the content for both items in the CNLEng subscale. 'Service delivery' relates to the various aspects of food stores, with particular reference to the dimension of 'foods sold' which is concerned with the quality and variety of foods sold in stores (CNLEng2) whereas the second domain, 'personal' includes factors that enhance or diminish access to nutritious foods (CNLEng1). Within these domains, three dimensions were of interest namely, 'health status', which refers to the aspects of individual and family health that influence access to and selection of nutritious foods (Freedman et al., 2013). The second dimension 'food knowledge' refers to the importance of food and nutrition knowledge for accessing nutritious food. Third was 'food preferences' which refers to how respondents' food preferences determine where and what nutritious foods they purchase (CNLEng2).

Development of the items in the SEBH scale

The SEBH-scale applied in the present study is a modified version of a 'Self-efficacy in science' scale reported by Guttersrud & Pettersen (2015). In the present study, the items were reworded to reflect self-efficacy in the specific subject of 'Body and Health' and rated on a six-point response scale. In addition, a new item '*I am confident that I can apply the knowledge that I have in Body and Health in new and unfamiliar situations'* was added (SEBH5). This item reflects the aspect of *adaptability*, which is the transferability of self-efficacy beliefs to novel and changing situations and deeper learning – the mastering of core academic content at high levels (Martin et al., 2013; Pellegrino & Hilton, 2012, p. 16). This item replaced an original item, which was shown to be weak and a poor fit to the scale (Guttersrud & Pettersen, 2015).

Table 3: Wording of items in the SEBH scale

Item	Wording
1	I am confident that if I have to learn something very thoroughly in Body and Health, I will be able to
2	I am confident that I can do an excellent job with difficult tasks in Body and Health
3	I am confident that I can do very well in tests in Body and Health.
4	I am confident that I can understand difficult learning material in Body and Health
5	I am confident that I can apply the knowledge that I have in Body and Health in new and unfamiliar situations.

The items in the SEBH scale were anchored with the phrase 'Body and Health' are a main subject area in science. How much do you agree with the claim (1 = Disagree strongly, 6 = Agree strongly)

Items in the SEBH scale were inspired by measures of learning confidence namely, the index of perceived self-efficacy and the index of control expectation as applied in PISA 2000 survey. Except for one item (SEBH1), item wording followed the guidelines for measuring general self-efficacy, thus all the items were phrased in terms of *can do* rather than *will do,* to reflect the judgment of capability (Bandura, 2006). However, this discrepancy may be attributed to difference in sentence structure after translation from Norwegian to English. In addition, the items reflected gradations of challenge on tasks in the subject of 'Body and Health', and in so doing, they sufficiently reflected the measurable aspects of self-efficacy; namely *level*, *generality* and *strength* (Bandura, 2006, p. 7).

Person factors assessed in the present study

In addition to the items in the three scales, the students reported on five person (demographic) factors dichotomized into two levels namely, *gender, age, cultural background, language predominantly spoken at home* and *number of books owned at home*.

Gender was reported as either male or female; *age* at the time of the survey was either 5 or 16 years old. Students also reported their birthplace and their parents' as 'Norway', 'Denmark or Sweden' or 'elsewhere'. These variables were then re-coded into a new variable *'cultural background'* with the levels 'majority' (if at least the student or one of the parents were born in a Scandinavian country) or 'minority' (if at least the student or one of the parents were born elsewhere).

Language predominantly spoken at home was reported as 'Norwegian', 'Danish or Swedish' or 'other'. This was also re-coded to two levels to which the student responses were scored as either being 'Scandinavian' (if student reported Norwegian/Danish/Swedish) or 'other' (if student reported other language). Re-coding the levels for linguistic and cultural background is justified by the similarities in culture among Scandinavian countries (Holmberg & Platzack, 2005). For instance the term 'Scandinavian languages' refers to the generally mutually understandable languages of the three continental Scandinavian countries i.e. Norway, Sweden and Denmark (Holmberg & Platzack, 2005, p.421).

Students' socioeconomic status (SES) was indicated by the number of books owned at home. In order to help improve the response accuracy, a picture of how books (clustered in groups of ten) ranging from 10 through to 200 was included. This variable was also re-coded to form two levels to which responses were either 'less than 100' or 'more than 100'. Using the number of books owned at home as an indicator of SES is justified by findings in research on SES and family resources that show a strong correlation between children's reading skills with the home literacy environment, particularly the number of books owned (Foster et al., 2005). Findings show that children from poor households often have less access to learning materials including books, computers and skill-building lessons to create a positive literacy environment (Aikens & Barbarin, 2008; Bradley et al., 2001; Bergen et al., 2017; Orr, 2003). In addition, an evaluation of compulsory schooling in Sweden reported that when assessing SES in a heterogeneous student sample, such as the one in the present study, the number of books in a home gives a clearer difference between children from different backgrounds than other variables (Brese & Mirazchiyski, 2010). Furthermore, commonly used indicators such as parental income, ownership of certain home possessions are open to bias related to the age of the child, whether the child knows how much the parents earn and the mere fact that the value of home possessions is not universal (Orr, 2003).

Data analysis applied in the present study

The main aim of data analysis was to refine the measurement instruments applied in the present study through validation. To achieve this, the present study applied two main analytical theoretical frameworks, CTT and RM. In the section that follows, I describe in detail the specific statistical procedures and software programs used, presenting them chronologically corresponding to the manuscripts (I-III) emerging from the study.

Breakdown of data analyses

In data analyses (I), I applied RM using RUMM2030 and ConQuest4 statistical packages to check the properties of the scales and individual items (Andrich & Luo, 2003; Yang, & Gustafsson, 2004). In analysis (II), in addition to RM, I used CFA using Lisrel 9.3 to test the dimensionality of the factor structure for the CNL-E scale (II). This was done to exemplify the use of both IRT and CTT statistical methods. Furthermore, it set the 'stage' for analysis in paper III. In analysis (III), in order to test a theoretical model linking perceived SEBH and CNL at a personal level, I applied SEM using Lisrel 9.30, applying robust DWLS and MLE procedures.

Analysis (I, II): Validating the CNL-E, CNLEng and SEBH scales applying Rasch analysis

Table 4 shows the steps in Rasch analysis that were applied in the validation of CNL-E, CNLEng and SEBH scales with the corresponding cut-off values/references and interpretations considered in the study. The present study used Rasch analysis to test the following propositions:

H1) All three scales have an acceptable overall fit to the rating scale model (RSM) of the polytomous unidimensional RM (PURM) and are each comprised of locally independent items, representing a well-targeted and reliable measurement scales

H2) Each of the items in the three scales had ordered response categories, was devoid of item bias (DIF) and had an acceptable individual fit to the RSM

Table 4: Types of Rasch analyses applied with reference values

Analyses	Test	Reference value	Interpretation
Derivation of PURM used	Likelihood ratio test (LRT)	p <0.05	Significant values suggest that the less parsimonious model provides a better fit to the sampled data
Data-model fit	Item-trait interaction chi-square	Large x^2 values, with significant *p*-value	Poor model fit Violation of the property of invariance
Dimensionality	PCA/ *t*-test procedure	<5% significant *t*-tests	The test items measure a single dominant underlying trait
	Fractal indices (r, c, A)	Low values of r, A High values of c	Possibility of multidimensionality, sub dimensions The items measure more than one underlying trait
Response dependence	Yen's $Q3$ statistic Compare $Q3$ statistic against the average residual correlations	$Q3$ values > ± 0.2 average residual correlation	Violation of the assumption of local dependency
Targeting	Compare distribution of person mean location with item mean location	Logit values close to 0	Ideal targeting
Reliability	Person separation index (PSI)	>0.85 at individual level >0.65 at group level	Sufficient reliability
Ordering of response categories	Examine category structure	Ordered response categories	The proposed response format is consistent with the underlying metric estimate of the underlying trait
Person-fit	Examine person-fit residuals	Values exceeding -2.56 Values exceeding +2.56	'Over-fitting' 'Under-fitting'
Individual item-fit analysis	Examine item-fit residuals	Values exceeding -2.56 Values exceeding +2.56	Poor item fit 'Over-discriminating' 'Under-discriminating'
	Mean item x^2 statistics with associated *p*- values (based on random samples)	Associated *p*-values <0.05	Item misfit
Differential item functioning (DIF)	Analysis of variance (ANOVA) Examine ICCs across different person factor levels (groups)	High values of F-values with probability <5% Bonferonni-adjusted values Consistent systematic difference in responses of one group over the other Non-systematic difference in responses	Presence of DIF Uniform DIF Non-uniform DIF

Analysis II - Refining the revised CNL-E scale applying CFA

In addition to Rasch analysis, the present study applied CFA to confirm the dimensionality of the latent factor (CNL-E). The *a priori* specified model (M1) was a one-factor measurement model in which all five indicator variables operationalized the single latent factor. The CFA specified ten (10) FP for estimation; factor loadings (4), unique variances (5), and factor variance (1). The factor loading for CNLE2 was set as the reference variable (λ=0.86). The resultant *a priori* specified model was 'over-identified' with 5*df* [DV-FP; (DV: counted as [k (k+1)/2, in which k=5 (no. items); *df* =15-10]. Because the data had a non-continuous distribution, in addition to MLE, the present study also used the DWLS estimation procedure. To evaluate fit of the *a priori* specified model, I used the reference values for the GOF indices shown in Table 5.

Table 5: GOF indices applied with reference cut-off values

	Absolute GOF indices			**Parsimony-adjusted GOF indices**		**Incremental GOF indices**	
	SB-scaled χ^2	Reduced chi-square χ^2/df	SRMR	RMSEA (90% CI)	Cfit	CFI	NNFI
Target value	*p*>.05	<3	<.05	<.06 (<.05; <.08)	>.05	>.95	>.95

SRMR = Standardized Root Mean Square Residual; Cfit = Closeness of Fit; CFI = Comparative Fit Index; NNFI = Non-Normed Fit Index; *df* = degrees of freedom

Analysis III - Linking CNL and SEBH at a personal level applying SEM

In the section that follows, I present the theory underlying the five steps followed during the development of the SEM linking CNL-E, CNLEng and SEBH at a personal level. I begin by presenting the theory underpinning the model specification of the SEM.

Self-efficacy is an important determinant of human behaviour, particularly what individuals do with the information, knowledge and skills that they acquire (Schwarzer, & Renner, 2000). Relatedly concerning nutritional outcomes, high levels of perceived self-efficacy have been noted to boost young adults' confidence to engage in positive 'healthful' dietary practices, because they feel competent enough to successfully accomplish the tasks (Aboulnasr, 2013; Adams et al., 2015; Guttersrud & Pettersen, 2015). Self-efficacy is one of the determinants of information processing (evaluation) as people make choices about food (AbuSabha & Achterberg, 1997). Findings show that individuals with high levels of perceived self-efficacy are expected to engage in an information processing that is more detailed comparing nutrition claims with factual knowledge because they feel competent eno ugh to understand the information. SEBH was the latent independent (exogenous) variable whereas CNL-E and CNLEng were latent dependent (endogenous) variables. Following this, the study held the following proposition related to the association between CNL-E, CNLEng and SEBH:

> **H1):** Adolescents' perceived self-efficacy in a nutrition-related subject explains a "significant" portion of the variability in the CNL variables (CNL-E and CNLEng).

Secondly, theory and empirical evidence suggest that individuals that are concerned about their health will engage in positive 'healthful' practices in order to sustain or enhance their state of well-being (Mai & Hoffmann, 2012). Relating to nutritional wellbeing, such consumers whilst making healthy food choices will engage in a more extensive information processing than those that are less concerned about their nutritional wellbeing. This includes for example, relying on attributes that they consider important such as the nutrient value of different food options, instead of peripheral attributes such as cost (Mai & Hoffmann, 2012). Therefore, it was also propositioned that in the present study:

> **H2):** Adolescents' engagement in dietary behaviour might affect their proficiency in critically evaluating nutrition information.

We referred to CNLEng as a *mediator* i.e., an endogenous variable that also serves as an independent variable. Accordingly, using the SEM approach, the study sought to test the following proposition:

> **H3):** "Critical nutrition literacy" (CNL), which is comprised of the proficiency to evaluate nutrition information (CNL-E) and to engage with nutrition-related issues (CNLEng) is related to the adolescents' self-efficacy exemplified in the present study as self-efficacy in the main subject area "Body and Health" (SEBH).

Ethical considerations in the present study

Prior to study implementation, the study proposal was presented to the Norwegian Ethics Committee, and Øystein Guttersrud[1] received immediate feedback that this study was not notifiable, since the respondents' answers to the questionnaire did not involve sensitive and personally identifiable information (Appendix II, III). Following this, through the Norwegian Centre for Science Education (Naturfagsenteret) at University Of Oslo sought consent from school principals to participate in the study. The school principals gave consent on behalf of the students in their tenth grade. All data was anonymized by removing all identifying information by assigning code values for the name of the school, name of respondent. Additionally, the students were informed that the data provided could be used for further research.

[1] Øystein is one of the co-supervisors of the PhD candidate

Study results

This chapter presents the outcomes of the investigation into the CNL of adolescents in Norway attending the tenth grade aged 15-16 years old and sets the stage for the discussion about CNL in adolescents. First, I present the overall study findings, the summary of papers reported in the three manuscripts emerging from the study. Next, I report on the sample characteristics as characterized by the person factors, missing data analysis. I conclude the chapter by presenting the study results according to the validation of the scales and linking CNL and SEBH at personal level.

Summary of overall study findings

Overall, empirical evidence partially refuted the hypothesis that high levels of CNL-E are associated with high levels of CNLEng and perceived SEBH. Specifically, the conclusion that high levels of CNL-E are associated with high levels of CNLEng was held plausible in the present study. However, the presence of empirical evidence suggests that the findings refuted the hypothesis that high levels of CNL are associated with high levels of perceived SEBH.

Summary of findings reported in the manuscripts

Regarding the secondary hypotheses about the individual scales, empirical evidence partially refuted the hypotheses regarding fit to the more parsimonious RSM and individual item characteristics. Instead, findings showed that all scales fit the less parsimonious PCM, were less than optimally targeted, but they were reliable measures of their respective traits. At item level, not all items displayed acceptable fit the RSM, had ordered response categories or were DIF-free.

In **Paper (I),** applying Rasch analysis, I evaluated the psychometric properties of the five-item SEBH scale. Empirical findings supported the viability of using the SEBH-5 scale for measuring adolescents' perceived SEBH.

Paper (II) detailed the analysis of assessing the psychometric properties of the five-item CNL-E scale applying both Rasch analysis and CFA. Empirical evidence showed that responses to the items in the CNL-E scale were influenced by one main underlying trait, CNL-E.

Paper (III) 'wrapped up' how CNL and SEBH relate at personal level by applying SEM. The paper reported a resultant theoretically derived and empirically sound model that confirmed the hypothesized association between CNL and SEBH-that high levels of SEBH of adolescents aged 15-16years old.

Sample characteristics according to person factors

The sample was representative of all the five regions of Norway i.e. Nord-Norge, Trøndelag, Vestlandet, Østlandet and Sørlandet. Of the 200 schools contacted to participate in the study, approximately 30% (n=58) of these responded positively and were included in the final sample. This rather low response rate is not an issue as findings indicate that higher response rates are not necessary for sample representativeness (Fincham, 2008). Moreover, 60.3% of the schools in the final sample were from the largest part of Norway (i.e. Oslo and Akershus), the most populated region of Norway (Kristiansen, 2015).

In terms of person factors, the final sample was relatively balanced in terms of gender with slightly more boys than girls, and majority of the respondents on the lower range of the age group (15years old). However, the sample was not very representative of the cultural background; less than 25% of the respondents were classified as 'minority'. This distribution reflects the low proportion of 'minorities' in Norway; 15.6% of population were either immigrants or born to non-Scandinavian parents (Kristiansen, 2015). Table 6 shows a summary of the sample characteristics by the person factors investigated in the study.

Table 6: Summary of sample characteristics by demographic factors

Person factors (PF)		Gender		Age		Cultural background		Language spoken at home		No. books owned at	
PF.1	n	Girl	792	15yr	1027	Minority	190	Other	204	Less than 100	913
	%		48.8		65		12.3%		13.1		58.3
PF.2	n	Boy	830	16yr	553	Majority	1355	Scandinavian	1350	More than 100	654
	%		51.2		35		87.7%		86.9		41.7
Missing	n	0		42		77		68		55	
	%	0		2.6		4.7		4.2		3.4	
Total	n	1622		1580		1545		1584		1567	
	%	100		97.4		95.3		95.8		96.6	

Missing data analysis in the present study

In total, 230 respondents had one or more missing responses to all items in the three scales translating into 4.95% of values missing from the data set. Missing item-response data ranged from a low 3.5% for the SEBH scale (SEBH 1, n = 56) to a high of 7.1% for the CNL-E scale (CNLE5, n = 115).

Psychometric evaluation of the SEBH scale (Paper I)

This section presents results of validating the SEBH scale using Rasch analysis. First, I describe the sample scale response characteristics, extreme scorers and missing data analysis. Next, I present results of Rasch analysis at overall level and individual item level for the SEBH scale. I conclude this subsection with a summary of the psychometric properties of the validated SEBH-5 scale.

SEBH scale response characteristics

Of the 1622 students in the study sample, 1568 (97%) responded to the SEBH scale. Of the 1493 valid scores, 75 (approx. 5%) were extreme scorers i.e. respondents scoring the lowest value (1, n=138) or highest value (6, n = 136) across all items. There were 367 (approx. 23.4%) missing responses to all items with the highest number of missing responses (n=69) on SEBH5 and the least (n=56) on SEBH1.

Psychometric properties of the SEBH scale using Rasch analysis

In the section below, I present the findings from the investigation of whether the observed responses to the items in the SEBH-5 scale as per the steps followed in Rasch analysis. First off, I report on model selection, then onto overall fit analysis and finally individual item fit analysis.

Concerning model selection, a significant LRT statistic (LRT χ^2 (p=.000019) indicated that it was more appropriate to use the PCM which is less parsimonious than the RSM.

Overall fit analysis

An examination of the overall fit to the PCM yielded a large and significant total item-trait chi-square statistic ($\chi2$ = 122.859; p ($\chi2$) = 0.000) suggesting an ideal fit to the PURM.

Investigation of unidimensionality using the PCA/t-test procedure indicated acceptable dimensionality as 6.2% of the t-tests were significant. Subset analysis revealed that the two subsets created according to whether the items loaded positively or negatively on the principal component (PC1) factor shared a large common variance, had high latent correlation and a small specific variance (A=0.90, r=0.90, c=0.17). This indicated that responses to all the items in the SEBH-5 scale were influenced by one main underlying trait-SEBH.

A mean person location estimate of 1.4 indicated that the SEBH-5 scale was less than optimally targeted for the sample as fewer respondents located at the lower trait levels at which the easy to endorse items are located. This explains the presence of disordered thresholds in the easiest item (SEBH1). Person-fit residuals showed that 20 and 89 respondents had fit residuals above and below the +/-2.56 cut-off value respectively. Removal of these person responses with most unlikely responses i.e. fit residuals greater than +2.56 did not improve the overall fit, so they were retained in the sample.

Overall, the SEBH-5 scale proved to be a reliable estimate of SEBH; PSI values of 0.88 for the original and complete data set, and Cronbach's alpha estimates of 0.88 (excluding extremes) and 0.92 including extremes.

Investigation of DIF showed that all five items were locally independent, displaying no DIF on the amended sample sizes of N=250, N=500 and N=750. These amended sample sizes were computed Andrich's (2011) rule of thumb corresponding to 10, 20 and 30 persons per threshold i.e. 25 thresholds (5 items with 5 thresholds).

All five items were locally independent as the item residual correlation values did not exceed their average item residual by 0.2 or more. All item-item correlations were less than 0.02, which was the average item residual.

Psychometric evaluation of the CNL-E scale (Paper II)

In the following section, I present the findings from the evaluation of the CNL-E scale using Rasch analysis and CFA. As in the previous section, I describe the sample scale response characteristics; extreme scorers, missing data analysis, the results of Rasch analysis at overall level and individual item level for the CNL-E scale. Lastly, I conclude this section presenting the results of the CFA for the 5-item CNL-E scale.

CNL-E scale response characteristics

Approximately 95.2% of the students in the sample (n=1544), responded to the CNL-E scale. Of these, 65 students (approx. 4.2%) were extreme scorers meaning that they responded either 1 to all items or 6 to all items. There were 488 (approx. 32%) missing responses to all six items with the highest frequency (*n*=109) recorded for CNLE4 and the lowest (*n*=80) for CNLE1. A significant value of Little's test for data missing completely at random (MCAR) test (*p*=.003) implied that the missing data were not MCAR, but possibly missing at random (MAR).

Psychometric properties of the CNL-E-5 scale using Rasch analysis

The initial test of data fit to the two derivations of the PURM yielded a significant LRT statistic (LRT χ^2 (p=0.536). Based on this finding, I proceeded to assess data fit against the less parsimonious PCM, as it described the data significantly better than the more restrictive RSM. I then present the findings of the overall and individual item properties of the 5-item CNL-E scale to the PCM.

Overall fit of data to the Rasch model

The total item-trait chi-square (χ^2) had a non-significant associated p-value of 0.536. A closer look at the person and item residual means and standard deviation (SD) values shows that these mean and SD values were close to the cut-off values of 0 and 1 respectively. This implies an acceptable fit between the persons and the PCM.

Unidimensionality of the CNL-E scale

Overall, the assumption of unidimensionality was upheld. Following the PCA/t-test procedure, 5.7% of the t-tests of difference between the person location tests were significant. As this value is slightly over the recommended cut-off value (5%), we conclude that the responses to the five items in the CNL-E scale are all influenced by one single dominant trait - CNL-E. We further investigated the dimensionality of the CNL-E-scale applying ConQuest4 statistical program. To achieve this, we compared a unidimensional (1-Dim) model (with all 5 items) with a multidimensional (2-Dim) model comprised of two 'subscales' based on how the five items loaded on the PC1 factor (CNLE1 - CNLE3 (negative), CNLE4 - CNLE5 (positive). Results showed high correlation values (0.984) between the two 'subscales', implying that in the 2-Dim model, the items in the two subscales measured very closely related traits. Overall, the data fit the 1-Dim model better than the 2-Dim model further providing evidence of unidimensionality.

Reliability and targeting of the CNL-E scale

As the interpretation of PSI and Cronbach's alpha is the same, I discuss the results in view of the Cronbach's alpha. The Cronbach's alpha and PSI values (0.896 and 0.878) respectively, were well within the acceptable range of the recommend cut-off values which indicated that the CNL-E scale is a reliable measure of the trait of interest (CNL-E). Furthermore, as Cronbach's alpha depends on the dimensionality of the trait measured by the items in the scale, it provides confirmation of unidimensionality, when items capture more than one trait, reliability is compromised.

Examination of the distribution of the person-item thresholds revealed that there was a skewed distribution towards the higher locations of the trait, meaning that the sample was located at a higher trait level (0.42 logits) than the average difficulty of the five-item CNL-E scale. In addition, there were persons located at the lower locations whose trait levels were not captured by the items in the scale. This indicates that in order to improve targeting, the five-item CNL-E scale could benefit from the inclusion of more difficult items and easier items.

Examination of the item map revealed that CNL-E scale left 80 respondents beyond the scope of measurement by five items in the scale (Figure 4). Specifically, on the 'upper' end of the vertical unidimensional scale, these respondents were located at trait locations beyond 4.10 logits, whereas on the 'lower' end, respondents located at locations beyond -4.00 logits were outside the scope of item difficulty range.

Furthermore, the item structure was shown to be relatively poor in the CNL-E scale as shown by the ordering of the items according to difficulty in the item-map; item structure did not match the chronological order in the predicted scale i.e. beginning with the easiest to most difficult items. In the item-map, CNLE1 was the most difficult item, while the remaining items were similar in difficulty level (CNLE2 - CNLE4). In addition, CNLE5 was located between the easiest and most difficult items. This order did not match the chronological order in the scale.

Ordering of thresholds in the CNL-E scale

All items in the CNL-E scale displayed ordered thresholds. This result suggests that the respondents used the response categories as intended. Ordered thresholds also imply that CNL-E the responses to the items in the CNL-E scale were consistent with the underlying metric estimate of the CNL-E trait.

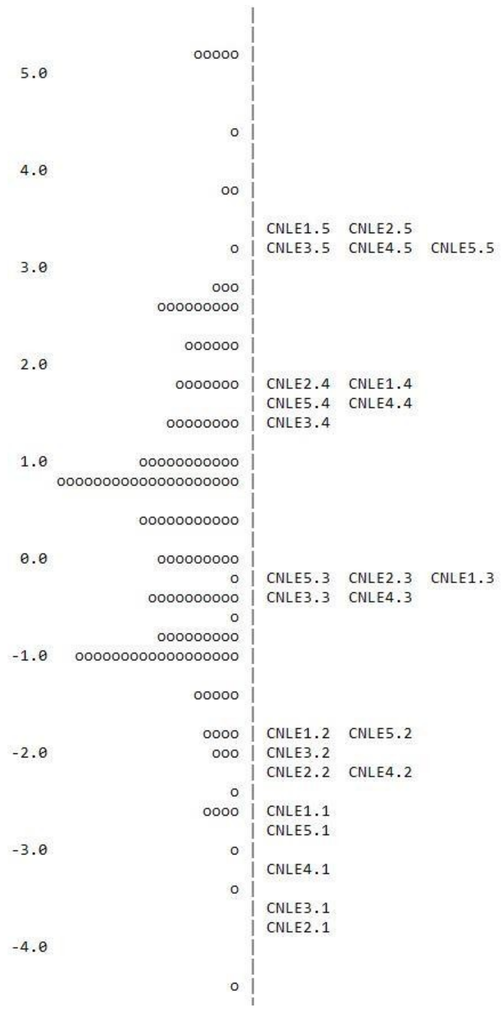

Figure 4: Item map for the CNL-E scale

◦ represents 10 persons

Person fit of the CNL-E scale to the PCM

A total of 357 persons 'over fit' the Rasch model with fit residuals larger than -2.56. This can be explained by the less than optimal targeting whereby there were fewer more difficult items in the scale. Because of this, the respondents at these trait locations had less items that were closer to their level of trait and therefore found the scale items too easy. The same explanation holds for the persons located at the lower trait levels, because there were few items at the lower levels of the trait. These persons would find that the 'easiest' items in the scale were too difficult for them.

Individual item fit analysis in the CNL-E scale

All items in the CNL-E scale had fit residual estimates within the acceptable range of +/- 2.56 and associated non-significant *p*-values. When checked against the amended sample sizes of N=350, 700 and 1050 corresponding to 10, 20 and 30 persons per threshold, the item-fit residual estimates were still within the acceptable range. This means that even at the lowest number of persons per threshold (10), these items provide enough information to sufficiently discriminate between persons at different trait levels.

Test for local independence in the CNL-E scale

None of the item pairs in the CNL-E scale displayed response dependence. Item correlation values were less than the cut-off value i.e. lower than ± 0.2 of the average item residual correlation value of the two items. CNL-E. This suggests that in each item pair, the responses to one of these items was not influenced by the response to the other item.

Psychometric properties of the CNL-E scale using CFA

Empirical findings from examining the underlying dimensionality of the 5-item CNL-E scale applying CFA showed that the items in the scale measured a single underlying factor (CNL- E).

Assessment of fit to the measurement model

High factor loadings (completely standardized solutions) ranging between 0.79 and 0.86 imply that the five items in the CNL-E scale represented measurable aspects that are valid indicators of the latent trait. CNLE2 provided the most information about the latent trait because it had the highest factor loading, whereas CNLE5 had the lowest factor loading meaning that it provided the least information about the trait. A qualitative look at the item wording provides insight into these findings. For example, CNLE2 states *'On a scale from "very difficult" to "very easy", how easy or difficult would you say it is to consider how reliable warnings about poor nutrition are, as warnings against malnutrition?'* CNLE2 seeks information about how easy the adolescents find it to consider (*cf. think through*) how reliable warnings against poor nutrition malnutrition are. This item most likely provides the most information about the trait because assessing the reliability of information requires the person to judge the information and the source of information as well (Larsen & Martey, 2013; Dorey & McCool, 2009; Meyer et al., 2012; Rieh & Danielson, 2007).

On the other hand, item CNLE5 *'evaluate nutritional advice in the media (newspapers, magazines, television) in a scientific way?'* provided the least information about the trait, probably because it sought information on more than the trait of interest (CNL-E). Qualitative inspection of the item wording showed that this item CNLE5 also sought information about media use, scientific evaluation of information. We posit that the more constructs that an item seeks out, the less focus the respondents will pay to the main trait of interest.

Evaluation of model fit to the GOF indices

The *a priori* specified model (M1) was a one-factor measurement model in which latent trait (CNL-E) was operationalized by five indicator variables (CNLE1-CNLE5). Ten FP were estimated applying robust DWLS and ML estimation procedures and the model was 'over- identified' with 50 degrees of freedom (*df*). Evaluation of model fit for M1 showed that the absolute and incremental fit GOF indices supported the hypothesis of a well-fitting model. Examination of the residual matrix indicated that all standardized residual values were well below the acceptable +2.58 (values ranged from -1.668 - +1.627).

Modification indices suggested including correlation between CNLE4 '*consider what it takes a scientific nutritional claim to be valid'* and CNLE5 '*evaluate nutritional advice in the media (newspapers, magazines, television) in a scientific way*?' (MI =111.86, Expected change = 0.309). This modification is justified as both items seek to establish how competent the students feel about their ability to evaluate nutritional information in scientific way. Model modification by adding this association between CNLE4 and CNLE5 in the modified model (M2) revealed an improvement in the fit against all the GOF indices.

Relating aspects of CNL and SEBH at personal level using SEM (Paper III)

Empirical evidence supported the development of a simple yet theoretically derived model that shows how CNL-E, CNLEng and SEBH link at personal level. In the section that follows, I present the findings of linking CNL-E, CNLEng and SEBH using a SEM framework. Without reiterating, the results reported in paper III, I present a summary of results from evaluating model fit to the GOF indices in Table 7.

Table 7: Overall fit of the *apriori* specified and post hoc modified model to ML estimation

Model	**Absolute GOF indices**			**Parsimony-adjusted**		**Incremental GOF indices**	
	SB-scaled χ^2	Reduced chi-square χ^2/df	SRMR	RMSEA (90% CI)	Cfit	CFI	NNFI
Model M1 (*df* =	164.54 ***p*=0.000**	**3.23**	0.03	0.07 (**0.06;** 0.07)	**0.00**	0.99	0.99
Model M2 (*df* =	158.765 ***p*=0.000**	**3.11**	0.03	0.07 (0.06; 0.07)	**0.00**	0.98	0.97
Target value	*p*>.05	<3	<.05	<.06 (<.05; <.08)	>.05	>.95	>.95

Estimation using Maximum Likelihood Estimation (MLE); M1 = *a priori* specified model, M2 = posthoc modified model; SB = Satorra-Bentler scaled chi-square, SRMR = standardized root-mean-square residual; Root-mean-square-error of approximation = RMSEA; close fit = Cfit; comparative fit index = CFI; non-normed fit index = NNFI.

Discussion of findings

This chapter presents the discussion of the study findings based on my interpretations and existent literature. It is structured as follows: a statement on the overall findings in relation to the primary and secondary hypotheses presented in each of the three papers, followed by a posthoc discussion for findings from each of the three articles on which this dissertation is based. What follows this, is a critical analysis of the study-its strengths and weaknesses. I then conclude this section with suggestions of implications for future research in the field of CNL especially among adolescents.

Overall, empirical evidence supported the main study hypothesis that the CNL of Norwegian adolescents aged 15-16years old is related to their perceived level of SEBH.

Empirical findings did not support the secondary hypotheses that data from the three scales (CNL-E, CNLEng, and SEBH) fit the more parsimonious PCM of the PURM. Instead, data from two of the three scales-CNL-E and SEBH, fit the less parsimonious PCM instead of the RSM. This finding is explained by the fact that the PCM derivation has more parameters than the RSM and thereby provides more information than the RSM. Each of the items has its own unique rating scale structure, and responses that indicate some level of knowledge towards the item still receive a partial credit towards a correct response (Wright, 1998).

Philosophically, one might explain the preferred fit to the PCM in terms of the post-positivist ontological stance that 'while reality exists, it can never be fully apprehended, but can only be approximated' (Guba, 1990, p22). Additionally, as post-positivist methodology aims at capturing as much of the reality as possible this lends strength of fit to the less parsimonious PCM (Denzin & Lincoln, 2005, p. 203). Therefore, from this viewpoint, we might conclude that partial credit awarded towards responses that indicate a level of the respondent's ability (about the traits) yields more information about the person's trait, compared to if the more restrictive RSM were to be used, in which case incorrect responses, even though indicative of a certain level of trait, would not receive any credit.

Discussion of findings in Paper I: Assessing adolescent 'self-efficacy in body and health'- Exploring the psychometric properties of the SEBH scale.

Empirical findings showed that overall, the data from the SEBH scale did not fit the more parsimonious RSM, but rather fit the less parsimonious PCM of the PURM. At individual item level, all five items in the SEBH scale were locally independent and represented a well- targeted and reliable measurement scale of SEBH.

Only one item (SEBH1) displayed disordered thresholds, indicating the possibility that this item may have been unsuitably worded addressing an additional trait that is related to SEBH (Salzberger, 2015). Qualitative inspection of the item wording *'I am confident that if I have to learn something very thoroughly in Body and Health, I will be able to manage it'* points to the existence of two related constructs reflected in the item, namely, self-efficacy and self-confidence. In item SEBH1, the first part of the item reflects self-efficacy as the respondent's belief that they can learn something thoroughly in 'Body and Health' whereas the second part reflects self-confidence as the belief in the degree of certainty in managing to correctly learn the content in the subject of 'Body and Health'. As these two constructs that are often interchanged in literature, it is likely that the respondents misinterpreted the item (Cramer et al., 2009). Merkle and Zandt (2006) define self-confidence as the degree of certainty about a perception or outcome. This differs from self-efficacy, which is an affirmation of one's perceived ability and strength of belief (Bandura, 1977).

The less than optimal targeting of the SEBH scale can be attributed to the positively skewed location of persons at a higher trait than the average that the item location. What this means is that respondents at a higher trait location (high levels of SEBH) find the questions very easy to endorse, implying that the SEBH scale could benefit from the addition of more difficult items to capture the respondents with high levels of SEBH.

The presence of DIF between some items in the SEBH scale at the original sample size and gender highlight the well-documented gender-based differences in academic self- efficacy were aligned with studies from previous studies. For example, we might explain the observed DIF in items SEBH1 and SEBH5 by studies that show girls are better at controlling their efforts and interest in learning situations perceived as challenging (Peklaj & Pečjak, 2002). For students, new and unfamiliar situations and new material, reflected in the item wording, might be challenging. Boys were more likely to endorse item SEBH4 compared to the girls, a finding that can be attributed to response bias between boys and girls (Noddings, 1996). Studies indicate that boys are more likely to express confidence in skills that they do not have or express over confidence in those that they have, whereas girls are more modest in their responses to self-rated judgments (Pajares, 2002; Wigfield & Guthrie, 1997).

The absence of DIF between items in the SEBH scale and socioeconomic status at original sample and amended sample sizes presents an interesting refute to studies that have shown significant association between academic self-efficacy and SES. While studies have reported a positive and significant correlation between family SES and a child's self-efficacy, the present study showed the opposite. Findings that a positive home literacy environment described in part by the number of books owned, is correlated with children's initial reading competence, do not seem supported by the findings in the present study (Aikens & Barbarin, 2008; Han et al., 2014). While children belonging to a family having a high SES are most likely exposed to a wider range of learning material, and thus feel more confident to tackle tasks in the subject of Body and Health, empirical evidence from this study shows that a student's level of SES and home literacy environment do not influence their self-beliefs in how they perceive their abilities to perform on tasks in the science subject of Body and Health.

Discussion of findings reported in Paper II: Assessing adolescents' perceived proficiency in critically evaluating nutrition information.

Empirical evidence only partially supported the primary hypothesis that the CNL-E scale had an acceptable overall fit to the RSM and that it represented a well-targeted and reliable measurement scale. At item-level, the findings indicated that the CNL-E scale was comprised of locally independent items, was less than optimally targeted but represented a reliable measurement of CNL-E.

Findings from the individual item-analysis in the CNL-E scale revealed that the most difficult items in the scale as indicated by their location on the item map (CNLE1, CNLE5), were concerned with how adolescents evaluated the reliability of nutritional messages in 'traditional' sources of media i.e. newspapers, television, magazines. This finding is in line with other studies that show adolescents are not as well versed with using traditional media sources for information, preferring newer sources like the internet, social media (La Ferle et al., 2000; Twenge et al., 2019). It is therefore not surprising that they find the items related to using traditional media sources the most difficult. For example a study in Sweden on teenagers' ability to determine the credibility of news from various sources found that although the respondents reported being good at searching for and evaluating information online, the majority (88%) struggled to separate news from advertisements in a common digital newspaper (Nygren & Guath, 2019). This may also be attributed to the way that information is packaged in traditional sources like newspapers; mixing and sharing of information from established media sources like newspapers with social media makes it difficult to assess the credibility of information (Fletcher & Park, 2017). This finding highlights the need to equip adolescents with the skills that they need to identify and establish the credibility of nutritional information in traditional sources like newspapers and magazines, especially as they migrate to digital platforms.

The less than ideal targeting observed with the CNL-E scale implied that the sample was located at a slightly higher level of trait (CNL-E) than the average of the items in the scale. This suggests that the CNL-E scale could benefit from including more items that reflect more difficult aspects of CNL-E. The shortage of items measuring CNL is an acknowledged challenge in nutrition literacy measurement studies. As experienced by others, developing items that reflect the competencies associated with CNL is demanding, requiring careful wording, often resulting into measurement instruments that contain very few items (Krause et al., 2016; Doustmohammadian et al., 2017). This may be attributed to the lack of theory and conceptualization of CNL compared to FNL and INL, thereby making it challenging to develop items measuring CNL.

The items in the CNL-E scale were shown to measure two related aspects namely, reliability of nutrition information and scientific evaluation of nutrition information. This finding highlights the need to emphasize how adolescents' asses the reliability of nutrition and application of scientific knowledge as a yardstick for assessing nutrition information. Evaluating reliability of messages such as warnings is closely related to credibility of messages. Reliability is one of the concepts associated with credibility, defined as the believability of information that rests largely on the trustworthiness and expertise of the information source or message, as interpreted by the information receiver (Rieh & Danielson, 2007). Reliability refers to the extent to which the recipients can trust the message or the channel (Roberts, 2010). When establishing the reliability of nutrition information, adolescents must look out for cues such as authorities behind the information (warning) of poor nutrition, when the information was published, and details of what exactly comprises of poor nutrition. It is particularly important to equip adolescents with the skills they need to establish the reliability of media messages about nutrition because media have been perceived as a less reliable source of nutrition information, partly because they provide biased, inconsistent or confusing information (Dorey & McCool, 2009; Larsen & Martey, 2013). Situating items CNLE1, CNLE2 and CNLE3 within the context of media credibility, we realize that CNLE1 and CNLE3 reflect 'source credibility' whereas CNLE2 reflects 'message credibility'.

Inspection of results of the item-fit analysis of the two 'source credibility' items reveals that item CNLE3 *'consider whether information on websites for nutritional information is reliable?'* was easier to endorse than item CNLE1 *'evaluate whether nutritional advice in the media (newspapers, magazines, television) is reliable?'*. This finding suggests that adolescents may find it easier to establish the reliability of nutrition information obtained from online sources than they do information from traditional sources. This finding finds support in previous studies that have shown that individuals trust information sources with which they are familiar (Flanagin & Metzger, 2000). Considering how better versed adolescents today are with using the internet for nutrition information than they are using traditional media sources, it is not surprising that the respondents would report finding item CNLE3 easier to endorse than CNLE1.

Qualitative investigation of the item wording of CNLE2 shows that the item sought out how easily respondents found it to establish how reliable warning messages about poor nutrition were in the media. 'Message credibility' refers to an 'individual's judgment of the reliability of the content of communication' (Appelman & Sunder, 2016), and has been shown to be closely associated with 'source credibility'. In fact, in the present study, at the original sample size, we noted the presence of response dependence between CNLE1 and CNLE2, suggesting that how a respondent evaluates credibility of the source may influence how they evaluate the credibility of the message. What this means is that for adolescents to ably evaluate the reliability of nutrition information, it is paramount that they can identify any self- seeking interests, bias and accuracy of nutrition information by the source and in the message, respectively. To achieve this, it is important that adolescents have evidence-driven knowledge against which to compare nutrition messages and sources. Being scientifically literate provides this. It is therefore not surprising that these aspects of reliability and scientific inquiry are related.

Theoretically, Potter's model of ML serves as a good starting point from which we can explore the relevance of media skills in evaluation of nutrition information. As per his model, media content is one of the knowledge structures upon which the model rests. The more knowledge structures that one has access to, the more context one has within which to analyze new media messages.

In addition, within his model, Potter identifies tools (skills) that are necessary for one to interact with and use of information that they may encounter; these skills include among others, *analysis* and *evaluation.* Other authors cite critical thinking skills as one of the elements that are central to one's use of media information (Silverblatt et al., 2014, p. 4). Part of critical thinking involves scrutinizing messages, applying checks for consistency in the message and the sources of the message i.e. establishing the media credibility. Because adolescents obtain nutrition information from both traditional and social media channels, it is crucial that they are equipped with the skills necessary to evaluate the reliability of the medium from which they get the information (e.g. television, social media platforms) and also the assess the validity of the source of the message (the authority behind the message). Previous studies have shown that adolescents trusted nutrition information found on government websites, newspapers more than that elsewhere; however, in the advent and rise of social media use, this is no longer the case as more institutions are increasingly using social media channels and 'non-professionals' such as celebrities to disseminate and endorse nutrition messages. Therefore, adolescents need to know what to look out for in messages they encounter, and then make informed choices about these messages.

The second aspect that two of the items in the CNL-E scale reflected is the use of scientific inquiry to establish the validity of nutritional claims and messages in the media. The role of science knowledge in establishing the credibility of nutrition information and sources of nutrition information in the media cannot be underestimated. Having a yardstick against which to compare the nutrition messages in the media is especially important for adolescents' CNL-E because they seem to lack criterion against which to gauge the level of truthfulness and consistency of the messages (Dudley et al., 2018, p. 147; Moons et al., 2009). The use of scientific knowledge and evidence-based inquiry provides this. A recent study by Wiblom et al. (2020) on how students engage in the critical examination of nutritional science in news media showed that respondents were able to scrutinize nutrition claims in media by drawing on experiences of scientific knowledge and investigations. Against these findings, we conclude that it is imperative that efforts targeting improvement of CNL-E of adolescents incorporate and emphasize the application of skills associated with scientific literacy.

Psychometric properties of the CNLEng scale-personal subscale

Both items showed DIF associated with only one of the person factors (gender). Specifically, at all location estimates, girls had a higher probability than boys did of endorsing both items '*I am concerned about eating foods that provide the nutrients my* *body needs'* and *'I am concerned that there are healthy foods in the grocery shops that my family shops at'.*

Gender differences in the adoption and execution of healthful dietary behaviors are well documented; showing that females make more healthful choices than males do and express more ethical concerns than males do (Beardsworth et al., 2002). In Norway, findingsshow that Norwegian women have higher levels of health knowledge than men have and are more likely to change their dietary choices in line with dietary recommendations (Fagerli & Wandel, 1999). Similar findings are seen in self-reported studies involving adolescents - in spite of possessing the same level of food knowledge, girls were more likely than boys to make more healthful shopping choices like choose low-fat milk during shopping because they are generally more concerned with their body image (Dong, 2015).

Discussion of findings reported in Paper III: Relating aspects of adolescents' critical nutrition literacy at the personal level.

The present study yielded a rather simple yet theoretically derived and empirically sound model exploring and confirming a link between CNL and self-efficacy at a personal level on an adolescent sample.

The finding that CNL was associated with perceived SEBH in adolescents is consistent with previous findings that evaluated aspects of CNL on adolescent samples such as use of food labels, dietary behaviour, evaluation of nutrition information (Dong, 2015; Haidar et al., 2017; Wang et al., 2016; Cha et al., 2014). This means that adolescents that are confident in their ability to learn and apply science knowledge will most likely be able to use it when interacting with nutrition information-interpreting information and advice and making dietary choices.

Adolescents' proficiency to engage in dietary behaviour (CNLEng) was strongly and directly related to their self-efficacy in the subject of 'Body and Health' (SEBH), a finding that is consistent with some empirical studies (Guttersrud & Pettersen, 2015; Dong, 2015). In a study on Norwegian adolescent's engagement in dietary behaviour, findings showed that self-efficacy in science explained up to approximately 7% of the variance observed in the respondents' scores on a self-reported measure of engagement in dietary behaviour (Guttersrud & Pettersen, 2015). As nutrition is a science, it is tempting to suggest that self- efficacy in a science related subject may have an influence on adolescents' dietary practices. Although measures of self-efficacy are domain-specific, self-efficacy beliefs are not limited to specific tasks, meaning that transference of self-efficacy beliefs across similar domains is possible (Bandura, 1997; Artino et al., 2014). Therefore, in the present study, I posit that adolescents' SEBH might be transferred to situations in which they have to interact with or engage in dietary practices such as selecting healthy food options. Accordingly, adolescents who feel confident in their ability to apply science knowledge in new situations would be highly efficacious in engaging in dietary habits.

Implications of results

The following section describes the implications of findings from my present PhD study, I categorize these into two, as theoretical and practical implications. *Theoretical implications* refer to the implications that the present study that can inspire efforts to address any existing shortcomings in the field of CNL. *Practical implications* are as the name suggests, practical, and refer to the ways in which researchers can apply the direct findings from the present study.

Theoretical implications

The present study yielded two psychometrically sound measures of CNL (CNL-E, CNLEng), thereby addressing the well-documented need for instruments measuring nutrition literacy within the 'critical' domain. Several authors note that developing items measuring CNL is challenging, owing to the 'abstract' nature of the domain. By focusing on two established aspects of the domain, and employing theoretical and empirical approaches, the present study proves that CNL is not too abstract a concept and can indeed be operationalized and measured in adolescent samples. However, by yielding two short, direct scales applicable for use on adolescents, a target group that in deed requires as little 'abstractness' as possible, the present study exemplified that it is possible to develop measures of CNL that are not abstract or very difficult for respondents to understand.

Secondly, the present study explored a previously underexplored association between two subdomains of CNL (CNL-E, CNLEng), and in so doing also widened the scope of CNL aspects that are measured in this specific population. This is important, as existing instruments have often focused on use of food labels in this sample and on adolescents in clinical settings, among others.

Thirdly, this study exemplified how the use of modern measurement approaches to instrument validation, yields psychometrically superior instruments measuring CNL. In a field like CNL in which there are very few valid and reliable scales, this is a step towards developing a single measure of CNL.

The study also contributes to the ongoing debate on which literacy should be emphasized for health promotion (Martin et al., 2013). A recent review showed that despite perceived similarities in content, health, food, nutrition and media literacy conceptualize the relationship between health and education differently (Martin et al., 2013). Findings from the present study highlight the association of nutrition literacy and health promotion, specifically engagement in dietary behaviour that promotes overall wellbeing. Also shown is how media literacy can and should be emphasized among adolescents in order to ensure the correct use of nutrition information for overall health promotion. These findings and associations can go a long way in informing researchers, educators and health and nutrition practitioners in interventions targeting adolescents.

Practical implications

Evaluation of psychometric properties using Rasch analysis means that the emergent CNL scales from the present study are reliable and valid measures of CNL. Furthermore, these scales are readily available for use without the need for re-validation on different samples, a process that is often costly. Furthermore, the items in the CNL scales may be used to build item banks, and in so doing, researchers in the field of nutrition literacy might develop more CNL measures that are comprehensive.

Weaknesses of the study

As measurement of CNL is still in its budding stages, there was limited research with which to compare the findings in the present study. The lack of studies in similar contexts and similar studies means that the findings in the present study are somewhat novel. This however creates an opportunity for furthering research in this field by providing results with which to compare findings.

In testing the structural equation model, re-testing the modified model on the same sample increased the risk of errors due to sampling. The study would have benefitted from testing the modified model on a different sample.

Strengths of the study

The use of theory in all aspects of the study lent support to the empirical evidence and interpretation of findings. The use of Rasch analysis as a validation approach means that all the instruments are invariant measures of their respective traits (CNLEng, CNL-E and SEBH). This implies that they can be used on other samples. This is especially important since there are very few existing measures of CNL on adolescent samples.

The study applied both classical test theory and modern approaches to measurement validation - Rasch modelling, this yielded psychometrically sound instruments that can be used on varied samples without the need for re-validation.

Conclusion

The present study exemplifies the possibility to advance CNL research through measurement by implementing three main shifts; widening scope of CNL aspects, using modern approaches to measurement validation, and focusing more on underexplored populations such as adolescents. In summary, critical appraisal does play a significant role in the extent to which persons use nutrition information; policy makers should therefore work towards equipping adolescents with the critical appraisal skills they need to utilize information from various sources.

Suggestions for future research

Perhaps the most crucial suggestion for future research is to develop an encompassing theory of nutrition literacy. This will inspire more research on nutritional literacy and CNL as a whole. Methodologically, is to apply Rasch analysis during the initial development of CNL instruments, not just for validation of instruments. This approach will save researchers resources like time, funds and contribute towards the development of more comprehensive instruments and refinement of instruments.

Additionally, by expanding on the existing instruments, by including more items in each of the scales that are more difficult, and easier. This could be achieved by conducting qualitative interviews during scale development. This will increase the range over which the scales measure the respective CNL traits. Another way of expanding on these instruments is by applying the scales on other samples, this will give insight into the applicability of the validated scales on different samples thereby lending strength to their reliability.

References

Aboulnasr, K. (2013). The role of self-efficacy, response efficacy, and nutrition knowledge in consumers' utilization of nutrition labels. *Marketing Management, 23*(1), 1-13.

Abu-Alhaija, A. S. (2019). From Epistemology to Structural Equation Modeling: An Essential Guide in Understanding the Principles of Research Philosophy in Selecting the Appropriate Methodology. *Australian Journal of Basic and Applied Sciences, 13*(9), 122-128. 10.22587/ajbas.2019.13.9.12

AbuSabha, R., & Achterberg, C. (1997). Review of self-efficacy and locus of control for nutrition-and health-related behavior. *Journal of the American Dietetic Association, 97*(10), 1122-1132. Hoff10.1016/S0002-8223(97)00273-3

Adams, R., Wu, M., & Wilson, M. (2015). *ACER ConQuest: generalized item response modelling software [Computer software] (Vol. 4).* Camberwell: Australian Council for Educational Research.

Aihara, Y., & Minai, J. (2011). Barriers and catalysts of nutrition literacy among elderly Japanese people. *Health Promotion International, 26*(4), 421-431. https://doi.org/10.1093/heapro/dar005

Aikens, N., & Barbarin, O. (2008). Socioeconomic differences in reading trajectories: The contribution of family, neighborhood, and school contexts. *Journal of Educational Psychology, 100*(2), 235. https://psycnet.apa.org/doi/10.1037/0022-0663.100.2.235

Ait, K., Rannikmäe, M., Soobard, R., Reiska, P., & Holbrook, J. (2015). Students' self- efficacy and values based on a 21st century vision of scientific literacy–A pilot study. *Procedia-Social and Behavioral Sciences, 177*, 491-495.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716-723. 10.1109/TAC.1974.1100705

Andersen, EB. (1977). Sufficient statistics and latent trait models. *Psychometrika,* 42, 69–81. https://doi.org/10.1007/BF02293746

Anderson-Bill, Eileen, S., Winett, R. A., & Wojcik, J. R. (2011). Social cognitive determinants of nutrition and physical activity among web-health users enrolling in an online intervention: the influence of social support, self-efficacy, outcome expectations, and self-regulation. *Journal of medical Internet Research, 13*(1), e28. https://doi.org/10.2196/jmir.1551

Andrae, R., Schulze-Hartung, T., & Melchior, P. (2010). *Dos and don'ts of reduced chi- squared.* https://arxiv.org/abs/1012.3754

Andrich D, De Jong, J., & Sheridan, B. (1997). Diagnostic opportunities with the Rasch model for ordered response categories. In J. Rost & R. Langerheine (Eds.), *Applications of latent trait and latent class models in the social sciences* (pp. 59-70). Münster.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561-573. https://doi.org/10.1007/BF02293814

Andrich, D. (2011). Rating scales and Rasch measurement. *Expert Review of Pharmacoeconomics & Outcomes Research, 11*(5), 571-585. https://doi.org/10.1586/erp.11.59

Andrich, D., & Luo, G. (2003). Conditional Pairwise Estimation in the Rasch Model for Ordered Response Categories using Principal Components. *Journal of Applied Measurement, 4(*3), 205-221.

Appelman, A., & Sundar, S. S. (2016). Measuring message credibility: Construction and validation of an exclusive scale. *Journalism & Mass Communication Quarterly*, *93*(1), 59-79. https://doi.org/10.1177%2F1077699015606057

Arke, E. T., & Primack, B. A. (2009). Quantifying media literacy: Development, reliability, and validity of a new measure. *Educational media international*, *46*(1), 53-65. https://doi.org/10.1080/09523980902780958

Artino Jr, A. R., La Rochelle, J. S., Dezee, K. J., & Gehlbach, H. (2014). Developing questionnaires for educational research: AMEE Guide No. 87. *Medical Teacher, 36*(6), 463-474. https://doi.org/10.3109/0142159X.2014.889814

Assalahi, H. (2015). The philosophical foundations of educational research: A beginner’s guide. *American Journal of Educational Research, 3*(3), 312-317. https://doi.org/10.12691/education-3-3-10

Ayele, D. G., Zewotir, T., & Mwambi, H. (2014). Using Rasch modeling to re-evaluate rapid malaria diagnosis test analyses. *International Journal of Environmental Research and Public Health, 11*(7), 6681-6691. https://doi.org/10.3390/ijerph110706681

Babio, N., Vicent, P., López, L., Benito, A., Basulto, J., & Salas-Salvadó, J. (2014).

Adolescents’ ability to select healthy food using two different front-of-pack food labels: a cross-over study. *Public Health Nutrition, 17*(6), 1403-1409. https://doi.org/10.1017/S1368980013001274

Bandura, A. (1977). Self-efficacy: Toward a Unify in Theory of Behavioral change. *Psychological Review, 84*(2). https://psycnet.apa.org/doi/10.1037/0033-295X.84.2.191

Bandura, A. (1997). The anatomy of stages of change. *American Journal of Health Promotion, 12*(1), 8-10. https://doi.org/10.4278/0890-1171-12.1.8

Bandura, A. (2006). Guide for constructing self-efficacy scales. In T. Urdan & F. Pajares (Eds.) *Self-efficacy beliefs of adolescents* (pp. 307-337). Connecticut, USA: Information Age Publishing.

Barzilai, S., & Zohar, A. (2012). Epistemic thinking in action: Evaluating and integrating online sources. *Cognition and Instruction, 30*(1), 39-85. https://doi.org/10.1080/07370008.2011.636495

Beardsworth, A., Bryman, A., Keil, T., Goode, J., Haslam, C., & Lancashire, E. (2002). Women, men and food: the significance of gender for nutritional attitudes and choices. *British Food Journal, British Food Journal, 107*, 470-491. https://doi.org/10.1108/00070700210418767

Bergen, E., Zuijen, T., Bishop, D., & de Jong, P. (2017). Why are home literacy environment and children's reading skills associated? What parental skills reveal. *Reading Research Quarterly, 52*(2), 147-160. https://doi.org/10.1002/rrq.160

Bland, J. M., & Altman, D. G. (1995). Multiple significance tests: the Bonferroni method. *BMJ, 310*(6973), 170. https://doi.org/10.1136/bmj.310.6973.170

Bollen, K. A. (1990). Overall fit in covariance structure models: Two types of sample size effects. *Psychological Bulletin, 107*(2), 256. https://psycnet.apa.org/doi/10.1037/0033-2909.107.2.256

Bollen, K. A., & Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological Methodology, 20,* 115–140. 10.2307/271028

Bond, T. (2004). Validity and assessment: a Rasch measurement perspective. *Metodología de las Ciencias del Comportamiento, 5*(2), 179-194.

Borsboom D, & Molenaar, D. (2015). Psychometrics. In J. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 418-422.). Elsevier.

Boshoff, N. (2014). Types of knowledge in science-based practices. J*ournal of Science Communication, 13*(3), A06.

Boyland, E. J., & Whalen, R. (2015). Food advertising to children and its effects on diet: review of recent prevalence and impact data. *Pediatric Diabetes, 16*(5), 331-337. https://doi.org/10.1111/pedi.12278

Bradley, R., Corwyn, R., McAdoo, H., & García Coll, C. (2001). The home environments of children in the United States part I: Variations by age, ethnicity, and poverty status. *Child Development, 72*(6), 1844-1867. https://doi.org/10.1111/1467-8624.t01-1-00382

Brembeck, H., & Johansson, B. (2010). Foodscapes and children's bodies. *Culture Unbound, 2*(5), 707-818. 10.3384/cu.2000.1525.10242797

Brese, F., & Mirazchiyski, P. (2010). Measuring students' family background in large-scale education studies. Paper presented at *The 4th IEA International Research Conference*, Gothenburg, Sweden.

Britner, S. L., & Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. Journal of Research in Science Teaching: *The Official Journal of the National Association for Research in Science Teaching, 43*(5), 485-499. https://doi.org/10.1002/tea.20131

Brown, B., & Larson, J. (2009). Peer relationships in adolescence. In R. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (2nd ed., pp. 74-104): John Wiley & Sons, Inc.

Brown, C., Templin, J., & Cohen, A. (2015). Comparing the two-and three-parameter logistic models via likelihood ratio tests: A commonly misunderstood problem. *Applied Psychological Measurement, 39*(5), 335-348. https://doi.org/10.1177%2F0146621614563326

Brown, T. (2006). *Confirmatory Factor Analysis for Applied Research: Methodology in the social sciences* (1st ed.). New York: The Guilford Press.

Brzezińska, J. (2016). Latent variable modelling and item response theory analyses in marketing research. *Folia Oeconomica Stetinensia, 16*(2), 163-174. https://doi.org/10.1515/foli-2016-0032

Buchanan, L., Kelly, B., Yeatman, H., & Kariippanon, K. (2018). The effects of digital marketing of unhealthy commodities on young people: a systematic review. *Nutrients, 10*(2), 148. https://doi.org/10.3390/nu10020148

Bybee, R. W. (2009). Program for International Student Assessment (PISA) 2006 and Scientific Literacy: A Perspective for Science Education Leaders. *Science Educator*, *18*(2), 1-13.

Bybee, R., McCrae, B., & Laurie, R. (2009). PISA 2006: An assessment of scientific literacy. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *46*(8), 865-883. https://doi.org/10.1002/tea.20333

Cappelleri, J. C., Lundy, J. J., & Hays, R. D. (2014). Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient - reported outcomes measures. *Clinical Therapeutics, 36*(5), 648-662. https://doi.org/10.1016/j.clinthera.2014.04.006

Carpiano, R. M., & Daley, D. M. (2006). A guide and glossary on postpositivist theory building for population health. *Journal of Epidemiology & Community Health, 60*(7), 564-570. 10.1136/jech.2004.031534

Cha, E., Kim, K. H., Lerner, H. M., Dawkins, C. R., Bello, M. K., Umpierrez, G., & Dunbar, S. B. (2014). Health literacy, self-efficacy, food label use, and diet in young adults. *American journal of health behavior*, *38*(3), 331-339. https://doi.org/10.5993/AJHB.38.3.2

Chassiakos, Y. L. R., Radesky, J., Christakis, D., Moreno, M. A., & Cross, C. (2016). Children and adolescents and digital media. Pediatrics, 138(5), e20162593. https://doi.org/10.1542/peds.2016-2593

Chau, M. M., Burgermaster, M., & Mamykina, L. (2018). The use of social media in nutrition interventions for adolescents and young adult – A systematic review. *International Journal of Medical Informatics, 120*, 77-91. https://doi.org/10.1016/j.ijmedinf.2018.10.001

Chomeya, R. (2010). Quality of psychology test between Likert scale 5 and 6 points. *Journal of Social Sciences, 6*(3), 399-403.

Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical values for Yen's Q3: Identification of local dependence in the Rasch model using residual correlations. *Applied Psychological Measurement, 41*(3), 178-194. https://doi.org/10.1177/014662161667750

Cohen, L., Manion, L., & Morrison, K. (2013). Research methods in education: Routledge.

Coiro, J., Coscarelli, C., Maykel, C., & Forzani, E. (2015). Investigating criteria that seventh graders use to evaluate the quality of online information. *Journal of Adolescent & Adult Literacy, 59*(3), 287-297. https://doi.org/10.1002/jaal.448

Cortina, J. M., Green, J. P., Keeler, K. R., & Vandenberg, R. J. (2017). Degrees of freedom in SEM: Are we testing the models that we claim to test? *Organizational Research Methods, 20*(3), 350-378. https://doi.org/10.1177%2F1094428116676345

Cramer, R., Neal, T., & Brodsky, S. (2009). Self-efficacy and confidence: Theoretical distinctions and implications for trial consultation. *Consulting Psychology Journal: Practice and Research, 61*(4), 319–334. https://psycnet.apa.org/doi/10.1037/a0017310

Creswell J. W. (2014). The selection of a research approach. In J. W. Creswell (Ed.), *Research design: Qualitative, quantitative, and mixed methods approaches* (pp. 3-24). Thousand Oaks, California: SAGE Publications.

Davis, B. (2018). *30 brands with excellent social media strategies.*https://econsultancy.com/30-brands-with-excellent-social-media-strategies/

Demetriou, A., Christou, C., Spanoudis, G. & Platsidou, M. (2002) The development of mental processing: Efficiency, working memory, and thinking. *Monographs of the Society of Research in Child Development 67*, Serial No. 268. https://psycnet.apa.org/doi/10.1111/1540-5834.671173

Denzin, N., & Lincoln, Y. (Ed.). (2005). Paradigmatic controversies, contradictions, and emerging confluences. In *Handbook of qualitative research* (3rd ed., pp. 191–215). Sage Publications Ltd.

Diamantopoulos, A., & Siguaw, J. (2013). *Introducing LISREL: A guide for the uninitiated.* Thousand Oaks, California: SAGE Publications.

Distefano, C., Mindrila, D., & Monrad, D. (2013). Investigating factorial invariance of teacher climate factors across school organizational levels. In M. Khine (Ed.), *Application of structural equation modeling in educational research and practice* (pp.257-275). Rotterdam: The Netherlands: Sense Publishers.

Dittrich, R., Francis, B., Hatzinger, R., & Katzenbeisser, W. (2007). A paired comparison approach for the analysis of sets of Likert-scale responses. *Statistical Modelling, 7*(1), 3-28. 10.1177/1471082X0600700102

Dong, Z. (2015). How to persuade adolescents to use nutrition labels: effects of health consciousness, argument quality, and source credibility. *Asian Journal of Communication, 25*(1), 84-101. https://doi.org/10.1080/01292986.2014.989241

Dorey, E., & McCool, J. (2009). The role of the media in influencing children's nutritional perceptions. *Qualitative Health Research*, *19*(5), 645-654. https://doi.org/10.1177%2F1049732309334104

Doustmohammadian, A., Omidvar, N., Keshavarz-Mohammadi, N., Abdollahi, M., Amini, M., & Eini-Zinab, H. (2017). Developing and validating a scale to measure Food and Nutrition Literacy (FNLIT) in elementary school children in Iran. *PLoS One*, *12*(6), e0179196. https://doi.org/10.1371/journal.pone.0179196

Dudley, D., Van Bergen, P., McMaugh, A., & Mackenzie, E. (2018). The role of social media in developing young people’s health literacy. In V. Goodyear & K. Armour (Eds.), *Young People, Social Media and Health*. (pp. 147-161). Routledge.

Fagerli, R., & Wandel, M. (1999). Gender differences in opinions and practices with regard to a" healthy diet". *Appetite, 32*(3), 170-190. https://doi.org/10.1006/appe.1998.0188

Fincham, J. (2008). Response rates and responsiveness for surveys, standards, and the Journal. *American Journal of Pharmaceutical Education, 72*(2), 43. https://dx.doi.org/10.5688%2Faj720243

Fischer, K. W. & Bullock, D. (1984).). *Cognitive Development in School-Age Children: Conclusions and New Directions.* In: Collins W.A. Collins (Ed). Development during Middle Childhood: The Years from Six to Twelve. (pp. 70-146). National Academies Press (US).

Fitzgerald, A., Heary, C., Kelly, C., Nixon, E., & Shevlin, M. (2013). Self-efficacy for healthy eating and peer support for unhealthy eating are associated with adolescents' food intake patterns. *Appetite, 63*, 48-58. https://doi.org/10.1016/j.appet.2012.12.011

Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of Internet information credibility. *Journalism & Mass Communication Quarterly, 77*(3), 515-540. https://doi.org/10.1177%2F107769900007700304

Fletcher, R., & Park, S. (2017). The impact of trust in the news media on online news consumption and participation. *Digital Journalism, 5*(10), 1281-1299. http://doi.org/10.1080/21670811.2017.1279979

Foster, M., Lambert, R., Abbott-Shim, M., McCarty, F., & Franze, S. (2005). A model of home learning environment and social risk factors in relation to children's emergent literacy and social outcomes. *Early Childhood Research Quarterly, 20*(1), 13-36. https://doi.org/10.1016/j.ecresq.2005.01.006

Freedman, D. A., Blake, C. E., & Liese, A. D. (2013). Developing a multicomponent model of nutritious food access and related implications for community and policy practice. *Journal of Community Practice, 21*(4), 379-409. https://doi.org/10.1080/10705422.2013.842197

Freeman, B., Kelly, B., Baur, L., Chapman, K., Chapman, S., Gill, T., & King, L. (2014). Digital junk: food and beverage marketing on Facebook. *American Journal of Public Health, 104*(12), e56-e64.

Furr, R. M., & Bacharach, V. R. (2013). Confirmatory Factor Analysis. In M. Furr (Ed.), *Psychometrics: An Introduction.* London: SAGE Publications.

Gasser, U., Cortesi, S., Malik, M. M., & Lee, A. (2012). Youth and digital media: From credibility to information quality. *Berkman Center Research Publication,* (2012-1). https://dx.doi.org/10.2139/ssrn.2005272

Gehlbach, H., & Brinkworth, M. E. (2011). Measure twice, cut down error: A process for enhancing the validity of survey scales. *Review of General Psychology, 15*(4), 380-387. https://doi.org/10.1037%2Fa0025704

Gibbs, H. D., Ellerbeck, E. F., Gajewski, B., Zhang, C., & Sullivan, D. K. (2018). The nutrition literacy assessment instrument is a valid and reliable measure of nutrition literacy in adults with chronic disease. *Journal of Nutrition Education and Behavior*, *50*(3), 247-257. https://doi.org/10.1016/j.jneb.2017.10.008

Gibbs, H. D., Ellerbeck, E. F., Befort, C., Gajewski, B., Kennett, A. R., Yu, Q., ... & Sullivan, D. K. (2016). Measuring nutrition literacy in breast cancer patients: development of a novel instrument. *Journal of Cancer Education*, *31*(3), 493-499. https://doi.org/10.1007/s13187-015-0851-y

Gibbs, H., & Chapman-Novakofski, K. (2012). A review of health literacy and its relationship to nutrition education. *Topics in Clinical Nutrition*, *27*(4), 325-333. 10.1097/TIN.0b013e31826f8dc5

Glasofer, D. R., Haaga, D. A., Hannallah, L., Field, S. E., Kozlosky, M., Reynolds, J., Tanofsky-Kraff, M. (2013). Self-efficacy beliefs and eating behavior in adolescent girls at-risk for excess weight gain and binge eating disorder. *International Journal of Eating Disorders, 46*(7), 663-668. https://doi.org/10.1002/eat.22160

Goodyear, V., Armour, K., & Wood, H. (2018). *The impact of social media on young people's health and wellbeing: Evidence, guidelines and actions*. University of Birmingham: Routledge.

Guba, E. (1990). The paradigm dialog. Paper presented at *The Alternative Paradigms Conference*, Indiana University, School of Education.

Guttersrud, Ø., & Pettersen, K. S. (2015). Young adolescents' engagement in dietary behavior-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. *Public Health Nutrition, 18*(14), 2565-2574. https://doi.org/10.1017/S1368980014003152

Guttersrud, Ø., Dalane, J. Ø., & Pettersen, S. (2014). Improving measurement in nutrition literacy research using Rasch modelling: examining construct validity of stage-specific 'critical nutrition literacy' scales. *Public health nutrition*, *17*(4), 877-883. 10.1017/S1368980013000530

Hagell, P. (2014). Testing rating scale unidimensionality using the principal component analysis (PCA)/t-test protocol with the Rasch model: the primacy of theory over statistics. *Open Journal of Statistics, 4*(6), 456-465. http://dx.doi.org/10.4236/ojs.2014.46044

Hagquist, C., & Hellström, L. (2014). The psychometric properties of the Early Development Instrument: a Rasch analysis based on Swedish pilot data. *Social indicators research*, *117*(1), 301-317. https://doi.org/10.1007/s11205-013-0344-5

Hagquist, C., Bruce, M., & Gustavsson, J. P. (2009). Using the Rasch model in nursing research: an introduction and illustrative example. *International Journal of Nursing Studies, 46*(3), 380-393. https://doi.org/10.1016/j.ijnurstu.2008.10.007

Haidar, A., Carey, F. R., Ranjit, N., Archer, N., & Hoelscher, D. (2017). Self-reported use of nutrition labels to make food choices is associated with healthier dietary behaviours in adolescents. *Public Health Nutrition, 20*(13), 2329-2339. https://doi.org/10.1017/S1368980017001252

Hair, J., Black, W., Babin, B., & Anderson, R. (2010). *Multivariate data analysis.* New Jersey: Pearson.

Hale, S., Bronik, M. D., & Fry, A. F. (1997). Verbal and spatial working memory in school- age children: developmental differences in susceptibility to interference. *Developmental Psychology, 33*(2), 364. https://doi.org/10.1037//0012-1649.33.2.364

Hall, J. (2011). Cross-sectional survey design. In P. Lavrakas (Ed.), *Encyclopedia of survey research methods* (pp. 172 - 173). Thousand Oaks: Sage Publications Inc.

Han, J., Chu, X., Song, H., & Li, Y. (2014). Social capital, socioeconomic status and self- efficacy. *Applied Economics and Finance, 2*(1), 1-10. 10.11114/aef.v2i1.607

Hobart, J., & Cano, S. (2009). Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technology Assessment*, *13*, 1-200. 10.3310/hta13120

Holmberg, A., & Platzack, C. (2005). The Scandinavian Languages. In G. Cinque & R. Kayne (Eds.), *The Oxford Handbook of Comparative Syntax* (pp. 420-458). Oxford University Press.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6*(1), 53-60.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424. https://psycnet.apa.org/doi/10.1037/1082-989X.3.4.424

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: a Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Huitt, W., & Hummel, J. (2003). Piaget's theory of cognitive development. *Educational Psychology Interactive, 3*(2), 1-5.

Humphry, S. (2002). Residuals and rating scales. *Rasch Measurement Transaction*, 16(1), 866. https://www.rasch.org/rmt/rmt161j.htm

Ishikawa, H., & Kiuchi, T. (2010). Health literacy and health communication. *BioPsychoSocial Medicine, 4*(1), 18. https://doi.org/10.1186/1751-0759-4-18

Joulaei, H., Keshani, P., & Kaveh, M. H. (2018). Nutrition literacy as a determinant for diet quality amongst young adolescents: A cross sectional study. *Progress in Nutrition, 20*, 455-464. 10.23751/pn.v20i3.6705

Kail, R. V., & Ferrer, E. (2007). Processing speed in childhood and adolescence: Longitudinal models for examining developmental change. *Child Development, 78*(6), 1760-1770. https://doi.org/10.1111/j.1467-8624.2007.01088.x

Kerlinger, F.N. (1986). *Foundations of behavioral research* (3rd ed.). Holt, Rinehart, & Winston.

Kivunja, C., & Kuyini, A. B. (2017). Understanding and applying research paradigms in educational contexts. *International Journal of Higher Education, 6*(5), 26-41.

Kline, R. (2011). *Principles and Practice of Structural Equation Modeling* (3rd ed.). New York: Guilford Press.

Krause, C., Sommerhalder, K., & Beer-Borst, S. (2016). Nutrition-specific health literacy: development and testing of a multi-dimensional questionnaire. *Ernahrungs Umschau*, *63*(11), 214-220. 10.4455/eu.2016.046

Kristiansen, J. E. (2015). *This is Norway 2015: What the figures say*. http://www.ssb.no/en/norge/

Krosnick, J. (1999). Survey research. *Annual Review of Psychology, 50*(1), 537-567. https://doi.org/10.1146/annurev.psych.50.1.537

La Ferle, C., Edwards, S. M., & Lee, W. N. (2000). Teens' use of traditional media and the Internet. *Journal of Advertising Research*, *40*(3), 55-65. 10.2501/JAR-40-3-55-65

Laestadius, L. I., & Wahl, M. M. (2017). Mobilizing social media users to become advertisers: Corporate hashtag campaigns as a public health concern. *Digital Health, 3,* 1-12. https://doi.org/10.1177%2F2055207617710802

Lai, P. (2013). Validating instruments of measure: Is it really necessary? *Malaysian Family Physician: The Official Journal of the Academy of Family Physicians of Malaysia*, *8*(1), 2.

Larsen, J. N., & Martey, R. M. (2013). Adolescents seeking nutrition information: motivations, sources and the role of the internet Learning tools and teaching approaches through ICT advancements. *International Journal of Information and Communication Technology Education, 7*(3), 74-85. 10.4018/jicte.2011070107

Latifah, S., Susilowati, N., Khoiriyah, K., & Rahayu, R. (2019). *Self-Efficacy: Its Correlation to the Scientific-Literacy of Prospective Physics Teacher.* Paper presented at the Journal of Physics: Conference Series.

Lei, P. W., & Wu, Q. (2007). Introduction to structural equation modeling: Issues and practical considerations. *Educational Measurement: Issues and Practice, 26*(3), 33-43. https://doi.org/10.1111/j.1745-3992.2007.00099.x

Levin, K. A. (2006). Study design III: Cross-sectional studies. *Evidence-Based Dentistry, 7*(1), 24-25. https://doi.org/10.1038/sj.ebd.6400375

Li, C. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936-949. https://doi.org/10.3758/s13428-015-0619-7

Lifshitz, F. (2008). Obesity in children. *Journal of Clinical Research in Pediatric Endocrinology, 1*(2), 53. https://dx.doi.org/10.4008%2Fjcrpe.v1i2.35

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*(1), 85-106. https://pubmed.ncbi.nlm.nih.gov/11997586/

Lubans, D. R., Plotnikoff, R. C., Morgan, P. J., Dewar, D., Costigan, S., & Collins, C. E. (2012). Explaining dietary intake in adolescent girls from disadvantaged secondary schools. A test of Social Cognitive Theory. *Appetite, 58*(2), 517-524. https://doi.org/10.1016/j.appet.2011.12.012

Mai, R., & Hoffmann, S. (2012). Taste lovers versus nutrition fact seekers: how health consciousness and self-efficacy determine the way consumers choose food products. *Journal of Consumer Behaviour, 11*(4), 316-328. https://doi.org/10.1002/cb.1390

Marais, I. (2013). Local dependence. In K. Christensen, S. Kreiner, & M. Mesbah (Eds.), *Rasch models in health* (pp. 111-130). John Wiley & Sons Incorporated.

Marais, I., & Andrich, D. (2008). Formalizing dimension and response violations of local independence in the unidimensional Rasch model. *Journal of Applied Measurement*, *9*(3), 200-215. https://pubmed.ncbi.nlm.nih.gov/18753691/

Martin, A., Nejad, H., Colmar, S., & Liem, G. (2013). Adaptability: How students' responses to uncertainty and novelty predict their academic and non-academic outcomes. *Journal of Educational Psychology, 105*(3), 728-746. https://psycnet.apa.org/doi/10.1037/a0032794

Mason, R.O. (1986). Four ethical issues of the information age. *MIS Quarterly, 10,* 5-12. 10.2307/248873

Massar, K., & Malmberg, R. (2017). Exploring the transfer of self-efficacy: academic self- efficacy predicts exercise and nutrition self-efficacy. *International Journal of Social Psychology*, *32*(1), 108-135. https://doi.org/10.1080/02134748.2016.1248026

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149-174. http://dx.doi.org/10.1007/BF02296272

McCoach D.B., Gable, R.K., &, Madura, J.P. (2013). *Instrument development in the affective domain: School and corporate applications*. Springer.

McIntosh, C. N. (2007). Rethinking fit assessment in structural equation modelling: A commentary and elaboration on Barrett (2007). *Personality and Individual Differences, 42*(5), 859-867. 10.1016/j.paid.2006.09.020

Medietilsynet (2016). *Children and media 2016—facts about children and youths (9–16) use and experience of media.* Fredrikstad: Medietilsynet.

Merkle, E., & Zandt, T. (2006). An application of the Poisson Race Model to confidence calibration. *Journal of Experimental Social Psychology, 135*, 391–408. https://psycnet.apa.org/doi/10.1037/0096-3445.135.3.391

Mesbah, M. (2013). From measurement to analysis. In K. Christensen, S. Kreiner, & M. Mesbah (Eds.), *Rasch models in health* (pp. 35-256). London: England: Wiley.

Meyer, M., Adkins, V., Yuan, N., Weeks, H. M., Chang, Y.-J., & Radesky, J. (2019). Advertising in young children's apps: A content analysis. *Journal of Developmental & Behavioral Pediatrics, 40(*1), 32-39. 10.1097/DBP.0000000000000622

Meyer, S. B., Ward, P. R., & Jiwa, M. (2012). Does prognosis and socioeconomic status impact on trust in physicians? Interviews with patients with coronary disease in South Australia. *BMJ Open, 2*(5). http://dx.doi.org/10.1136/bmjopen-2012-001389

Mindrila, D. (2010). Maximum likelihood (ML) and diagonally weighted least squares (DWLS) estimation procedures: A comparison of estimation bias with ordinal and multivariate non-normal data. *International Journal of Digital Society, 1*(1), 60-66. 10.20533/ijds.2040.2570.2010.0010

Montgomery, K. C., & Chester, J. (2009). Interactive food and beverage marketing: targeting adolescents in the digital age. *Journal of Adolescent Health, 45*(3), 18-29. https://doi.org/10.1016/j.jadohealth.2009.04.006

Moons, W., Mackie, D., & Garcia-Marques, T. (2009). The impact of repetition-induced familiarity on agreement with weak and strong arguments. *Journal of Personality and Social Psychology, 96*(1), 32. https://psycnet.apa.org/doi/10.1037/a0013461

Mullis, I. V. S., Martin, M. O., Goh, S., & Cotter, K. (2016). TIMSS 2015 Encyclopedia: The Science curriculum in primary and lower secondary grades. http://timssandpirls.bc.edu/timss2015/encyclopedia/countries/norway/the-science-curriculum-in-primary-and-lower-secondary-grades/

Ní Bhroin, N., & Rehder, M. (2018). *Digital natives or naïve experts? Exploring how Norwegian children (aged 9-15) understand the Internet*. http://www.lse.ac.uk/media- and-communications/research/research-projects/eu-kids-online/news

Nielsen, T., & Kreiner, S. (2013). Improving items that do not fit the Rasch model. In K. Christensen, S. Kreiner, & M. Mesbah (Eds.), *Rasch models in health* (pp. 317-334). John Wiley & Sons Incorporated.

Noddings, N. (1996). Current directions in self research: Self-concept, self-efficacy, and possible selves. Paper presented at *The American Educational Research Association*, New York.

Norman, C. D., & Skinner, H. A. (2006). eHealth literacy: essential skills for consumer health in a networked world. *Journal of Medical Internet Research, 8*(2), e9. https://doi.org/10.2196/jmir.8.2.e9

Norman, J., Kelly, B., Boyland, E., & McMahon, A. T. (2016). The impact of marketing and advertising on food behaviors: evaluating the evidence for a causal relationship. *Current Nutrition Reports, 5*(3), 139-149. https://doi.org/10.1007/s13668-016-0166-6

Nutbeam, D. (2000). Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. *Health Promotion International, 15*(3), 259-267. https://doi.org/10.1093/heapro/15.3.259

Nygren, T., & Guath, M. (2019). Swedish teenagers' difficulties and abilities to determine digital news credibility. *Nordicom Review, 40*(1), 23-42. 10.2478/nor-2019-0002

Organization for Economic Co-operation and Development [OECD]. (2013). *PISA 2015 Draft Science Framework.* Retrieved from: https://read.oecdilibrary.org/education/pisa-2015-assessment-and-analytical-framework/pisa-2015-science-framework_9789264281820-3-en#page1

Organization for Economic Co-operation and Development. (2000). Measuring student knowledge and skills: the PISA 2000 assessment of reading, mathematical and scientific literacy. http://hdl.voced.edu.au/10707/64880

Orr, A. (2003). Black-white differences in achievement: The importance of wealth. *Sociology of Education, 76*(4), 281-304. 10.2307/1519867

Paek, H.-J., Reber, B. H., & Lariscy, R. W. (2011). Roles of interpersonal and media socialization agents in adolescent self-reported health literacy: a health socialization perspective. *Health Education Research, 26*(1), 131-149. https://doi.org/10.1093/her/cyq082

Pajares, F. (2002). Gender and perceived self-efficacy in self-regulated learning. *Theory into Practice,41*(2), 116-125. https://www.tandfonline.com/action/showCitFormats?10.1207/s15430421tip4102_8

Pajares, F. (2006). Self-efficacy during childhood and adolescence. In Urdan T & P. F (Eds.), *Self-efficacy beliefs of adolescents* (pp. 339-367). Information Age Publishing.

Pallant, J. F., & Tennant, A. (2007). An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology, 46*(1), 1-18. https://doi.org/10.1348/014466506x96931

Peklaj, C., & Pečjak, S. (2002). Differences in students' self-regulated learning. Studia *Psychologica, 44*(1), 29.

Pellegrino, J., & Hilton, M. (2012). *Education for Life and Work: Developing transferable knowledge and skills in the 21st century.* The National Academies Press.

Perline, R., Wright, B. D., & Wainer, H. (1979). The Rasch model as additive conjoint measurement. *Applied Psychological Measurement, 3*(2), 237-255. https://doi.org/10.1177%2F014662167900300213

Pettersen K.S. (2005). Norwegian health journalists' ability to report on health research: A concern to science education? *NorDiNa, 1*(1), 5-16. https://doi.org/10.5617/nordina.462

Pettersen, K. (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* [Doctoral dissertation, University of Oslo].

Pettersen, K. (2019). Health literacy og Nutrition literacy. In L. Dagrun, E. Torheim, & N. Øverby (Eds.), Samfunnsernæring (pp. 167-174). Oslo: Universitetsforlaget.

Piaget, J. (1964). Cognitive development in children: Piaget. *Journal of research in science teaching*, *2*(3), 176-186.

Potter, W. J. (2004). *Theory of media literacy: A cognitive approach.* Sage Publications. http://dx.doi.org/10.4135/9781483328881

Potter, W. J. (2010). The state of media literacy. *Journal of Broadcasting & Electronic Media, 54*(4), 675-696. https://doi.org/10.1080/08838151.2011.521462

Potvin Kent, M., Pauzé, E., Roy, E. A., de Billy, N., & Czoli, C. (2019). Children and adolescents' exposure to food and beverage marketing in social media apps. *Pediatric Obesity, 14*(6), e12508. https://doi.org/10.1111/ijpo.12508

Profetto-McGrath J. (2005). Critical thinking and evidence-based practice. *Journal of Professional Nursing, 21(*6), 364-371. https://doi.org/10.1016/j.profnurs.2005.10.002

Programme for International Student Assessment [PISA] & Organization for Economic Co - operation and Development [OECD]. (2006). Assessing scientific, reading and mathematical literacy: A framework for PISA 2006. OECD. 10.1787/9789264026407- en

Programme for International Student Assessment [PISA]. (2006). Assessing scientific, reading and mathematical literacy: A framework for PISA 2006. Paris: Publications de l'OCDE.

Ravand, H., & Firoozi, T. (2016). Examining construct validity of the master's UEE using the Rasch model and the six aspects of the Messick's framework. *International Journal of Language Testing*, *6*(1), 1-18.

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology,* *5*, 27-48. https://doi.org/10.1146/annurev.clinpsy.032408.153553

Reynolds, K., Turner, J., Branscombe, N., Mavor, K., Bizumic, B., & Scotland, J. (2012). Exploring the philosophical underpinnings of research: Relating ontology and epistemology to the methodology and methods of the scientific, ınterpretive, and critical research paradigms. *English Language Teaching, 5*(9), 9-16. 10.5539/elt.v5n9p9

Rideout, V., Foehr, U., & Roberts, D. (2010). *Generation M2: Media in the lives of 8- to 18- year-olds.* Menlo Park: CA: Henry J. Kaiser Family Foundation.

Ringland, E., Gifford, J., Denyer, G., & Thai, D. (2016). Evaluation of an electronic tool to assess food label literacy in adult Australians: A pilot study. Nutrition & Dietetics, 73(5), 482-489. 10.1111/1747-0080.12271

Rieh, S. Y., & Danielson, D. R. (2007). Credibility: A multidisciplinary framework. *Annual Review of Information Science and Technology, 41*(1). https://doi.org/10.1002/aris.2007.1440410114

Roberts, C. (2010). Correlations among variables in message and messenger credibility scales. *American Behavioral Scientist, 54*(1), 43-56. https://doi.org/10.1177%2F0002764210376310

Rolling, T. E., & Hong, M. (2016). The effect of social cognitive theory-based interventions on dietary behavior within children. *Journal of Food and Nutrition Research, 4*(5), 1-9. http://dx.doi.org/10.15226/jnhfs.2016.00179

Roper, S., & La Niece, C. (2009). The importance of brands in the lunch-box choices of low-income British school children. *Journal of Consumer Behaviour: an International Research Review, 8*(2-3), 84-99. https://doi.org/10.1002/cb.275

Russell, S. J., Croker, H., & Viner, R. M. (2019). The effect of screen advertising on children's dietary intake: A systematic review and meta-analysis. *Obesity Reviews, 20(*4), 554-568. https://doi.org/10.1111/obr.12812

Ryan, A. B. (2006). Post-positivist approaches to research. In Antonesa M. (Ed.), *Researching and Writing your Thesis: a guide for postgraduate students* (pp. 12-26). Southampton: University of Southampton.

Salzberger, T. (2010). Does the Rasch model convert an ordinal scale into an interval scale? *Rasch Measurement Transaction, 24*(2), 1273-1275.

Salzberger, T. (2013). Reporting a Rasch Analysis. In K. Christensen, S. Kreiner, & M. Mesbah (Eds.), *Rasch models in health* (pp. 347-361). London: England: Wiley.

Salzberger, T. (2015). The validity of polytomous items in the Rasch model-The role of statistical evidence of the threshold order. *Psychological Test and Assessment Modeling, 57*(3), 377-395.

Samoggia, A., & Riedel, B. (2020). Assessment of nutrition-focused mobile apps' influence on consumers' healthy food behaviour and nutrition knowledge. *Food Research International, 128,* 108766. https://doi.org/10.1177%2F1356336X20930741

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*(4), 507-514. https://doi.org/10.1007/BF02296192

Schienle, A., Schäfer, A., Hermann, A., & Vaitl, D. (2009). Binge-eating disorder: reward sensitivity and brain activation to images of food. *Biological Psychiatry, 65*(8), 654-661. https://doi.org/10.1016/j.biopsych.2008.09.028

Schumacker, R. E., & Lomax, R. G. (2004). Developing structural equation models. Part I. In R.E. Schumacker & R.G. Lomax (Eds.), *A Beginner's Guide to Structural Equation Modeling* (4th ed., pp. 194-208). Psychology Press.

Schumacker, R., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling.*(3rd ed.). New York: Routledge.

Schunk, D. H. (1991). Self-efficacy and academic motivation. *Educational Psychologist, 26*(3-4), 207-231. https://doi.org/10.1080/00461520.1991.9653133

Schwarzer, R., & Renner, B. (2000). Social-cognitive predictors of health behavior: action self-efficacy and coping self-efficacy. *Health Psychology, 19*(5), 487. https://psycnet.apa.org/doi/10.1037/0278-6133.19.5.487

Schwarzer, R., & Renner, B. (2009). *Health-specific self-efficacy scales.* Berlin, Germany.

Seah, M. L. C., & Koh, K. T. (2020). The efficacy of using mobile applications in changing adolescent girls' physical activity behaviour during weekends. *European Physical Education Review*, 1356336X20930741. https://doi.org/10.1177%2F1356336X20930741

Shea, T. L., Tennant, A., & Pallant, J. F. (2009). Rasch model analysis of the Depression, Anxiety and Stress Scales (DASS). *BMC Psychiatry, 9*(1), 21. https://doi.org/10.1186/1471-244X-9-21

Sick, J. (2009, January). Rasch measurement in Language education Part 3: The family of Rasch models. Shiken: JALT Testing & Evaluation SIG Newsletter, 13(1), 4-10.

Sick, J. (2011, March). Rasch measurement in Langauge education Part 6: Rasch measurement and factor analysis. Shiken: JALT Testing & Evaluation SIG Newsletter, 15(1), 15-17.

Sidani, J. E., Shensa, A., Hoffman, B., Hanmer, J., & Primack, B. A. (2016). The association between social media use and eating concerns among US young adults. *Journal of the Academy of Nutrition and Dietetics, 116*(9), 1465-1472. https://doi.org/10.1016/j.jand.2016.03.021

Silk, K. J., Sherry, J., Winn, B., Keesecker, N., Horodynski, M. A., & Sayir, A. (2008). Increasing nutrition literacy: testing the effectiveness of print, web site, and game modalities. *Journal of Nutrition Education Behavior*, *40*(1), 3-10. https://doi.org/10.1016/j.jneb.2007.08.012

Silverblatt, A., Smith, A., Miller, D., Smith, J., & Brown, N. (2001). *Media literacy: Keys to interpreting media messages* (2nd ed.). Praeger.

Simovska, V., Paakkari, L., & Paakkari, O. (2012). Health literacy as a learning outcome in schools. *Health Education, 112*(2), 133-152. https://doi.org/10.1108/09654281211203411

Sjaastad, J. (2014). Enhancing measurement in science and mathematics education through Rasch analysis: Rationale and properties. *NorDiNa, 10*(2), 212-230. https://doi.org/10.5617/nordina.662

Smith Jr, E. V. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement, 3*(2), 205-231.

Smith, R., Kelly, B., Yeatman, H., & Boyland, E. (2019). Food marketing influences children's attitudes, preferences and consumption: A systematic critical review. *Nutrients, 11*(4), 875. https://doi.org/10.3390/nu11040875

Smolin, L. A., & Grosvenor, M. B. (2010). *Nutrition: Science and applications:* Hoboken, NJ: Wiley.

Sobh, R., & Perry, C. (2006). Research design and data analysis in realism research. *European Journal of Marketing.* https://doi.org/10.1108/03090560610702777

Sørensen, K., Van den Broucke, S., Fullam, J., Doyle, G., Pelikan, J., Slonska, Z., & Brand, H. (2012). Health literacy and public health: a systematic review and integration of definitions and models. *BMC Public Health, 12*(1), 80. https://doi.org/10.1186/1471-2458-12-80

Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta- analysis. *Psychological bulletin*, *124*(2), 240. https://psycnet.apa.org/doi/10.1037/0033-2909.124.2.240

Statistics Norway. (2016). *Media usage in Norway.*https://www.statista.com/topics/4258/media-usage-in-norway/

Steele, M. M., Daratha, K. B., Bindler, R. C., & Power, T. G. (2011). The relationship between self-efficacy for behaviors that promote healthy weight and clinical indicators of adiposity in a sample of early adolescents. *Health Education & Behavior, 38*(6), 596-602. https://doi.org/10.1177%2F1090198110387514

Steinberg L. (2010). *Adolescence* (9th ed.). New York: McGraw-Hill.

Stenner, A. (1994). Specific objectivity-local and general. *Rasch Measurement Transactions*, *8*(3), 374.

Suhr, D. (2006). Exploratory or Confirmatory Factor Analysis? Paper presented at the *Proceedings of the 31st Annual SAS Users Group International Conference*, Cary, NC.

Tal, E. (2015, June 15). Measurement in Science. The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/archives/fall2020/entries/measurement-science/

Talagala, I. A., & Arambepola, C. (2016). Use of food labels by adolescents to make healthier choices on snacks: a cross-sectional study from Sri Lanka. *BMC Public Health, 16*(1), 739. https://doi.org/10.1186/s12889-016-3422-1

Taylor, M. K., Sullivan, D. K., Ellerbeck, E. F., Gajewski, B. J., & Gibbs, H. D. (2019). Nutrition literacy predicts adherence to healthy/unhealthy diet patterns in adults with a nutrition-related chronic condition. *Public Health Nutrition*, *22*(12), 2157. https://dx.doi.org/10.1017%2FS1368980019001289

Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Care & Research, 57*(8), 1358-1362. https://doi.org/10.1002/art.23108

Teo T, Tsai LT, & Yang, C. (2013). Applying structural equation modeling (SEM) in educational research. In M. Khine (Ed.), *Application of structural equation modeling in educational research and practice* (pp. 3-21). Rotterdam: The Netherlands: Sense Publishers.

Thomson, S., Hillman, K., & De Bortoli, L. (2013). *A teacher's guide to PISA reading literacy.* https://research.acer.edu.au/ozpisa/11/ (Original work published August 2013).

Twenge, J. M., Martin, G. N., & Spitzberg, B. H. (2019). Trends in US Adolescents' media use, 1976–2016: The rise of digital media, the decline of TV, and the (near) demise of print. *Psychology of Popular Media Culture*, *8*(4), 329. https://psycnet.apa.org/doi/10.1037/ppm0000203

Vassallo, A. J., Kelly, B., Zhang, L., Wang, Z., Young, S., & Freeman, B. (2018). Junk food marketing on Instagram: content analysis. *JMIR Public Health and Surveillance, 4*(2), e54. https://doi.org/10.2196/publichealth.9594

Velardo, S. (2015). The nuances of health literacy, nutrition literacy, and food literacy. *Journal of Nutrition Education and Behavior, 47*(4), 385-389. e381. https://doi.org/10.1016/j.jneb.2015.04.328

Viola, G. C. V., Bianchi, F., Croce, E., & Ceretti, E. (2016). Are food labels effective as a means of health prevention? *Journal of Public Health Research, 5*(3), 768. https://dx.doi.org/10.4081%2Fjphr.2016.768

Walker, C. M., Beretvas, S. N., & Ackerman, T. (2001). An examination of conditioning variables used in computer adaptive testing for DIF analyses. *Applied Measurement in Education, 14*(1), 3-16. https://doi.org/10.1207/S15324818AME1401_02

Wang, Q., Oostindjer, M., Amdam, G. V., & Egelandsdal, B. (2016). Snacks with nutrition labels: tastiness perception, healthiness perception, and willingness to pay by Norwegian adolescents. *Journal of Nutrition Education and Behavior, 48*(2), 104-111. e101. https://doi.org/10.1016/j.jneb.2015.09.003

Weiss, B. D., Mays, M. Z., Martz, W., Castro, K. M., DeWalt, D. A., Pignone, M. P., Hale, F. A. (2005). Quick assessment of literacy in primary care: The Newest Vital Sign. Annals of Family Medicine, 3(6), 514-522. 10.1370/afm.405

Weiten, W. (1992). *Psychology Themes and Variations.* California: Wadsworth: Inc.

Wetzel, E., & Carstensen, C. H. (2014). Reversed thresholds in partial credit models: A reason for collapsing categories? *Assessment, 21*(6), 765-774. https://doi.org/10.1177%2F1073191114530775

Whittaker, T. A. (2012). Using the modification index and standardized expected parameter change for model modification. *The Journal of Experimental Education, 80*(1), 26-44. https://doi.org/10.1080/00220973.2010.531299

Wiblom, J., Andrée, M., & Rundgren, C.-J. (2020). Navigating alarming media messages about nutrition and health. *Science & Education, 29*(1), 75-100. https://doi.org/10.1007/s11191-019-00099-1

Wigfield, A., & Guthrie, J. (1997). Relations of children's motivation for reading to the amount and breadth or their reading. *Journal of Educational Psychology, 89*(3), 420.

World Health Organization. (2005). *Nutrition in adolescence: issues and challenges for the health sector: issues in adolescent health and development*. World Health Organization.

Wouters, E. J., Larsen, J. K., Kremers, S. P., Dagnelie, P. C., & Geenen, R. (2010). Peer influence on snacking behavior in adolescence. *Appetite, 55*(1), 11-17. https://doi.org/10.1016/j.appet.2010.03.002

Wright, B. (1997). Fundamental Measurement. *Rasch Measurement Transaction, 11*(2), 558. Wright, B. (1998). Rating scale model (RSM) or partial credit model (PCM). *Rasch Measurement Transactions, 12*(3), 641-642. https://www.rasch.org/rmt/rmt1231.htm

Wright, B., Linacre, J., Gustafsson, J., & Martin-Loff, P. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*(3), 370.

Yang, Y., & Gustafsson, J. (2004). Measuring socioeconomic status at individual and collective levels. *Educational Research and Evaluation, 10*(3), 259-288. https://doi.org/10.1076/edre.10.3.259.30268

Yuen, E. Y., Thomson, M., & Gardiner, H. (2018). Measuring nutrition and food literacy in adults: A systematic review and appraisal of existing measurement tools. *HLRP: Health Literacy Research and Practice*, *2*(3), e134-e160. https://doi.org/10.3928/24748307-20180625-01

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology, 25*(1), 82-91. https://doi.org/10.1006/ceps.1999.1016

Zoellner, J., You, W., Connell, C., Smith-Ray, R. L., Allen, K., Tucker, K. L., . . . Estabrooks, P. (2011). Health literacy is associated with healthy eating index scores and sugar- sweetened beverage intake: findings from the rural Lower Mississippi Delta. *Journal of the American Dietetic Association, 111*(7), 1012-1020. https://doi.org/10.1016/j.jada.2011.04.01

Appendix I: Scales applied in the study (in Norwegian)

SEBH scale

"Kropp og helse" er et hovedområde i læreplanen i naturfag.

Hvor enig er du i påstanden? (1 = Svært uenig, 6 = Svært enig)

	1	2	3	4	5	6
Jeg er trygg på at hvis jeg må lære noe i kropp og helse veldig godt, kan jeg klare det						
Jeg er trygg på at jeg kan gjøre en utmerket jobb på vanskelige oppgaver i kropp og helse						
Jeg er trygg på at jeg kan gjøre en utmerket jobb på prøver i kropp og helse						
Jeg er trygg på at jeg kan forstå vanskelig lærestoff om kropp og helse						
Jeg er trygg på at jeg kan bruke kompetansen jeg har i kropp og helse, i nye og ukjente situasjoner						
Jeg har mer kunnskap om kropp og helse enn de fleste andre i klassen						

CNLEng scale (personal', 'social', global') levels included

Hvor enig er du i påstanden? (1 = Svært uenig, 6 = Svært enig)

	1	2	3	4	5	6
Jeg er opptatt av å spise mat som gir de næringsstoffene kroppen min trenger						
Jeg er opptatt av at det er sunne matvarer i de matbutikkene familien min handler i						
Jeg er opptatt av at befolkningen i Norge skal ha råd til å kjøpe og spise sunne matvarer						
Jeg er svært opptatt av at kantiner og automater på norske skoler og arbeidsplasser tilbyr sunne matvarer						
Jeg engasjerer meg politisk for at verdens befolkning skal ha god tilgang til sunne matvarer						
Jeg stiller krav om at rike land forplikter seg til å sørge for at befolkninger i fattige land har nok sunn mat						

CNL-E scale

"Ernæring" handler om sammenhengen mellom kosthold og helse.

På en skala fra "veldig vanskelig" til "veldig lett", hvor lett eller vanskelig vil du si det er å (1 = Veldig vanskelig, 6 = Veldig lett)

	1	2	3	4	5	6
vurdere om ernæringsråd i media (aviser, ukeblader, fjernsyn) er pålitelige						
vurdere hvor pålitelige advarsler om dårlig ernæring er, som advarsler mot feilernæring						
vurdere om informasjon på ernæringssider på Internett er pålitelig						
vurdere hva som skal til for at en vitenskapelig påstand om ernæring er holdbar						
vurdere om ernæringsrådene du får av lærere på skolen, er pålitelige						
vurdere om ernæringsråd gitt gjennom sosiale medier (Facebook, Twitter) er pålitelige						
vurdere ernæringsråd i media (aviser, ukeblader, fjernsyn) på en vitenskapelig måte						
vurdere om påstander om ernæring er vitenskapelige						
vurdere effekten av ernæringsråd fra helsepersonell, som for eksempel leger						

Appendix II: NCD Ethical approval to conduct study

Svar Svar til alle Videresend

Elizabeth Blomstervik <Elizabeth.Blomstervik@nsd.no> Kjell Sverre Pettersen 10.C

RE: Utfylt første side i Meldeskjema for Desire Naigaga sin phd og hennes brukte spørreskjema

Du svarte på meldingen 10.02.2021 16:38.

Hei, igjen

Har du referansenummer til det nye meldeskjemaet du opprettet?

Dersom det aldri ble behandlet personopplysninger i prosjektet til phd-studenten, så trenger det ikke å meldes til oss nå. Det er ikke lenger aktuelt at man skal melde inn et prosjekt for så å motta vurdering at det ikke er meldepliktig. Det eneste som avgjør om man skal melde et prosjekt til NSD, er om det blir behandlet personopplysninger,dette vil også inkludere en behandling av avidentifiserte personopplysninger når det finnes en koblingsnøkkel. Det er kun bruk av helt anonymiserte personopplysninger som det ikke er behov for å melde til NSD.

Vennlig hilsen
Elizabeth Blomstervik
rådgiver | Seksjon for personverntjenester
NSD – Norsk senter for forskningsdata AS
t: 55 58 32 62

Appendix III: Confirmation of use of anonymized data in study

NATURFAGSENTERET
NASJONALT SENTER FOR NATURFAG I OPPLÆRINGEN

Naturfagsenteret
Postboks 1106 Blindern
0317 Oslo
Telefon: 22 85 53 37

E-post: post@naturfagsenteret.no
Internett: www.naturfagsenteret.no

Karakterstøttende prøve i naturfag – bekreftelse på sletting av personopplysninger

Naturfagsenteret viser til avtale om utvikling av karakterstøttende prøver i naturfag for perioden 1. oktober 2015 – 31. desember 2019 og brev datert 23. november 2016 angående avslutning av avtalen.

Naturfagsenteret bekrefter herved at personopplysninger har blitt slettet regelmessig.

Med hilsen

Merethe Frøyland

Merethe Frøyland
Leder

Appendix IV: Papers I-III

Paper I

Naigaga, D. A., Pettersen, K. S., Henjum, S. & Guttersrud, Ø. (2019). Assessing adolescent self-efficacy in 'body and health'-Exploring the psychometric properties of the SEBH scale. *Nordic Studies in Science Education*, *15*(2), 145-158.

DOI: **https://doi.org/10.5617/nordina.5913**

Naigaga Desire Alice is a PhD candidate studying Health Sciences, with a background in Public Health Nutrition (MSc) and Human Nutrition and Dietetics (BSc). She has a keen interest in health literacy and nutrition literacy of adolescents in both developed and developing economies.

Kjell Sverre Pettersen is a professor in health communication, with experience in health science teaching, supervising students, and researching on topics related to health literacy and nutrition literacy.

Sigrun Henjum is an associate professor with extensive research experience within the field of nutritional epidemiology, particularly dietary assessment, maternal and infant nutrition and research on environmental determinants of dietary behaviour among adolescents.

Øystein Guttersrud is an associate professor in science education. He has experience in physics/physics education, quantitative measurement, and supervision of students. He is currently working on different evaluation projects in education.

DESIRE ALICE NAIGAGA (corresponding author)
PhD student, Oslo Metropolitan University, Norway
dnaiga@oslomet.no

KJELL SVERRE PETTERSEN
Professor, Oslo Metropolitan University, Norway
spetters@oslomet.no

SIGRUN HENJUM
Associate professor, Oslo Metropolitan University, Norway
shenjum@oslomet.no

ØYSTEIN GUTTERSRUD
Associate professor, University of Oslo, Norway
oystein.guttersrud@naturfagsenteret.no

Assessing adolescent 'self-efficacy in body and health'- Exploring the psychometric properties of the SEBH scale

Abstract
Self-efficacy beliefs are significant predictors of achievement in education. However, majority of existing self-efficacy measures are rather 'general' and assess aggregated perceptions of students' proficiencies within broad academic disciplines. Applying Rasch analysis, the present study explored the psychometric properties of the five-item 'self-efficacy in body and health' (SEBH) scale as administered to more than 1600 tenth-graders aged 15-16years in Norway. Based on our sample, the SEBH

scale stood out as well targeted and reliable with acceptable overall fit to the partial credit parameterization of the polytomous unidimensional Rasch model. Except for a slightly reversed threshold in item 1, which could be explained by few persons located at low levels of self-efficacy, the locally independent items had ordered response categories and functioned in the same way for the different levels of relevant person factors. Adapting this scale to different fields of education would contribute to development of more specific measures of perceived capability.

Introduction

Does it matter whether you have the will and belief that you can? Of course it does! Self-efficacy signifies a person's belief that he or she is able to execute successfully the behaviours required to produce a specific outcome. Thus, self-efficacy is the person's belief in his or her capability to control and execute actions in spite of potential obstacles. A person's perceived self-efficacy has a direct influence on the choice of activities and settings, and the stronger the perceived self-efficacy, the more active the efforts to cope with the task at hand (Bandura, 1977). Therefore, self-efficacy affects individuals' decisions concerning the effort and endurance they will put into a task. In general, higher self-efficacy is linked with greater effort, perseverance and resilience (Van Dinther, Dochy, & Segers, 2011; Zeegers, 2004).

In school achievement, self-efficacy refers to an individual's belief in his or her ability to successfully accomplish academic tasks or to achieve academic goals (Schunk, 1991). Scales measuring academic self-efficacy evaluate the extent to which students perceive they can accomplish established academic goals (Marsh, Hau, Artelt, Baumert & Peschar, 2006; Pastorelli et al, 2001). However, according to Bong and Skaalvik (2003), majority of the existing academic self-efficacy measures are 'wide-ranging', aiming at school proficiency in general, thus making them more reflective of 'academic self-concept'. Self-efficacy is a specific view of one's capacities in a given domain and it follows that efficient self-efficacy measures be tailored to the particular domain of interest (Bandura, 2006).

Studies have found that adolescents do display autonomy associated with health and self-efficacy (Schunk & Meece, 2006; Taylor, Adelman & Kaser-Boyd (1984). Today, adolescents are exposed to diverse sources of health and nutrition information that shape their lifestyle attitudes (Gray, Klein, Noyce, Sesselberg & Cantrill, 2005; Neuhauser, Rothschild & Rodríguez, 2007). Notable among these sources are schools, through the health topics offered (Brown, Tuefel & Birch, 2007).

In health literacy research, studies have shown that self-efficacy is a significant predictor of health behaviour and achievement (Gutiérrez-Doña, Lippke, Renner, Kwon, & Schwarzer, 2009; Rayane & Achterberg, 1997; Schwarzer, 2008). As a result, various health-specific self-efficacy assessment tools have been developed (see, for example, Schwarzer & Renner, 2009; Young, Mills, Woolmore, Hawkins, & Tennant, 2012).

In spite of the advantages that item response theory (IRT) models, and Rasch models in particular, have over classical test theory (CTT), few health-related and health literacy scales have been evaluated using IRT and Rasch models (see, for example, Davidson, Keating & Eyres, 2004; Escobar et al., 2015; Huang et al. 2018; Nguyen, Paasche-Orlow, Kim, Han and Chan, 2015). One such advantage is concerned with the assumption of item-sample independence, which is strongly emphasized in IRT and Rasch models. While violations of local independence in IRT and Rasch models, and 'error correlations' in confirmatory factor models (CFM), might refer to similar 'problems' in the data, there is no direct link between the probabilistic IRT and Rasch models and the correlation-based CFM. Unlike descriptive IRT-models, the family of prescriptive Rasch models satisfy the requirements of fundamental measurement (Andrich, 1988).

To fill in the gaps identified, there is a need for a measurement scale for the evaluation of how adolescents perceive their proficiency in accomplishing specific academic tasks within health, which meets the assumptions and satisfies the requirements of fundamental measurement. To exemplify this in the field of science education, the current study focuses on the subject area of 'body and health' in the Norwegian compulsory school science curriculum. This subject area; which focuses on the structure of our bodies, how the body is affected by nutrition and lifestyle and how the body changes over time; will play a vital part in the new and forthcoming interdisciplinary school topic 'public health and wellbeing' (KD, 2016). A self-efficacy in 'body and health' scale might be efficient for evaluating the proficiency with which adolescents perceive they can apply that knowledge to solve complex problems in new and unfamiliar contexts and adopting critical thinking skills associated with 'deeper learning' (Paakkari, L. & Paakkari, O., 2012; Pellegrino & Hilton 2012; KD, 2016).

The main objective of the current study is therefore to, applying Rasch-analysis, validate a five-item measurement scale tailored towards assessing adolescent self-efficacy in 'body and health' at the end of compulsory school (tenth grade). We will test the following hypotheses:

> H1) The 'self-efficacy in body and health' (SEBH) scale has acceptable overall fit to the rating scale parameterization of the polytomous unidimensional Rasch model, consists of locally independent items, and represents a well-targeted and reliable measurement scale.
>
> H2) Each SEBH-scale item has ordered response categories, is functioning in the same way for the different levels of relevant person factors, and shows acceptable fit to the rating scale parameterization of the polytomous unidimensional Rasch model.

Our first hypothesis is concerned with the overall SEBH-scale psychometric properties, while our second hypothesis refers to the psychometric properties at the individual item level. With the goal of estimating as few parameters as possible (parsimony rule), we hypothesized a unidimensional scale with items sharing the same set of thresholds.

Method - Sample

A sample of 200 Norwegian lower secondary schools was randomly selected, and the school principals were contacted by email and telephone seeking consent to volunteer. Fifty-eight schools (30%) accepted the invitation. From April to May 2015, 1622 students in the tenth grade (47% girls) responded by using an electronic assessment tool.

The substantive theory of the SEBH latent variable

The SEBH-scale is a revised and further developed version of a self-efficacy scale reported by Guttersrud & Pettersen (2015), which was based on self-efficacy measures in science and the control expectation scale applied in PISA (Organization for Economic Co-operation and Development [OECD], 2001). The items were reworded to reflect competencies within 'body and health', with one additional item (Table 1): *'I am confident that I can apply the knowledge that I have in Body and Health in new and unfamiliar situations'*. This item reflects aspects of adaptability–the transferability of self-efficacy beliefs to novel and changing situations (Martin, Nejad, Colmar, & Liem, 2013; Pellegrino & Hilton, 2012) and deeper learning – the mastering of core academic content at high levels (Pellegrino & Hilton, 2012).

Table 1. The wording of the items in the self-efficacy in body and health (SEBH) scale (originally stated in Norwegian). A six-point rating scale with the extreme response categories anchored with a phrase 1 = 'strongly disagree' and 6 = 'strongly agree' was used.

Item	Item wording
1	I am confident that if I have to learn something very thoroughly in Body and Health, I will be able to manage it.
2	I am confident that I can do an excellent job with difficult tasks in Body and Health
3	I am confident that I can do very well in tests in Body and Health.
4	I am confident that I can understand difficult learning material in Body and Health
5	I am confident that I can apply the knowledge that I have in Body and Health in new and unfamiliar situations.

Person factor levels and data processing

Students reported the following five person factors (with levels indicated in parentheses); gender (male/female); age at the time of the survey (15 or 16 years old); language predominately spoken at home (Norwegian, Danish/ Swedish (i.e., Scandinavian languages) or 'other'); student's, mother's and father's place of birth (Norway, Denmark/ Sweden or 'other'); and the number of books at home (five categories). A picture showing how different numbers of books might appear like on shelves was included to improve validity or 'response accuracy'.

The variables for birthplace were re-coded into a new variable named 'cultural background' with the levels 'majority' (if at least the student or one of the parents were born in any of the Scandinavian countries i.e., Norway, Denmark or Sweden) and 'minority'. This classification is valid as countries within Scandinavia share strong cultural and linguistic similarities. The five levels of 'number of books at home' were merged into the categories 'less than 100 books' and '100 or more books'. These two levels reflected the largest difference in SEBH-scale score (cf. DIF analysis). The number of books was used as an indicator of socioeconomic status (SES), as research on SES and family resources shows that children's initial reading competency is correlated with the home literacy environment and number of books owned; with children from poor households often having less access to learning materials, including books, computers and skill-building lessons to create a positive literacy environment (Aikens & Barbarin, 2008; Bergen, Zuijen, Bishop, & Jong, 2016; Bradley, Corwyn, McAdoo, & García Coll, 2001; Orr, 2003). As a consequence, research indicates that children from low-SES households develop academic skills slower than children from higher SES groups (Morgan, Farkas, Hillemeier, & Maczuga, 2009).

SEBH-scale response characteristics

A six-point rating scale with the extreme response categories anchored with a phrase 1 = 'strongly disagree' and 6 = 'strongly agree' was applied for all SEBH-scale items. Out of the 1622 student responses there were 1568 valid responses: There were 166 extreme scorers of which 12 students attained the lowest possible raw score on the SEBH-items responded to and 154 students attained the highest possible score (ceiling effect) on the items responded to. There were a total of 36 missing responses to the five items, with item 5 having the highest number of these (15) and item 1 having the least (2). We have no evidence weakening the hypothesis stating that 'data are missing completely at random' (MCAR; Allison, 2001).

The unidimensional Rasch model–a rationale for the methodological decisions

The prescriptive Rasch models estimate the probability of endorsing an item based on the difference between the person location (proficiency or attitude) and item location (difficulty or affective level)

(Rasch, 1960; Shaw, 1991). Person and item location estimates refer to the point estimate of a person's or an item's location on the latent trait scale, respectively (Harris, 1989). In the current study, person location refers to an individual's self-reported perceived proficiency in body and health. The different threshold locations reflect the locations at which the probability of a response in two adjacent categories is equal. For example, a dichotomously scored item has one threshold, and the threshold location refers to the location at which the probability of a response in the two adjacent categories is 0.5. In this paper, we applied RUMM 2030 for all analyses (Andrich, Lyne, Sheridan, & Lou, 2010). RUMM uses pairwise maximum likelihood estimation (PLME) and Warm's mean weighted likelihood estimation (WLE) for estimating item location estimates and person locations respectively (Katsikatsou, Moustaki, Yang-Wallentin, & Joreskog, 2012; Warm, 1989).

The concept 'item discrimination' refers to the degree with which an item separates individuals with higher person location estimates from those with lower location estimates. An under-discriminating item differentiates weaker between such respondent groups than the RM expects, given the item location.

Using the Rasch Model (RM), raw scores at the ordinal level (presumes 'ordered response categories' otherwise nominal) are transformed into interval implying additivity (Andrich, 1989; Perline, Wright & Wainer, 1979; Salzberger, 2010). Fit to Rasch models implies that the property of *invariance* holds meaning that the item-trait relationships are stable for the different person locations along the latent trait scale (Andrich, 1988). Rasch models satisfy *specific objectivity* which refers to the requirement of item-person independence; any person location estimate must be independent of the specific measurement device or items applied (Stenner, 1994). As the raw scores contain all the information needed to estimate Rasch models parameters i.e., item and person locations, the raw score is a *sufficient statistic* for Rasch models (Andersen, 1977).

While both Rasch models and other IRT models assume locally independent data–unidimensional and statistically independent data, only the family of Rasch models ensure additivity, invariance, specific objectivity and sufficiency as described above. Therefore, we applied prescriptive Rasch models and not descriptive IRT models in this study.

Overall model fit

The parameters of the rating scale parameterization (RSM; Andrich, 1978) of the RM are a subset of the parameters of the partial credit parameterization (PCM; Masters, 1982) of the RM, so the RSM is nested in the PCM. We compare data-model fit for nested models using likelihood ratio test (LRT). The LRT test statistic – the change in deviance (D) – is asymptotically χ^2 distributed (i.e., for large samples) with degrees of freedom (df) equal to the difference in model estimated parameters (Wilks, 1938, p. 62). A 'significant' χ^2 value implies rejecting the 'null hypothesis' stating that the less complex and nested model, describing the data using fewer threshold estimates, is preferred (cf. hypothesis 1). Compared to RSM, the df of PCM is larger and the PCM therefore usually accounts better for the observed data.

Individual item and person fit

To account for our somewhat large sample size (N = 1622), we drew five random samples of 250, 500 and 750 persons from the SPSS file storing the data – a total of fifteen samples. These sample sizes correspond to 10, 20 and 30 persons per thresholds (Andrich, 2010). We estimated individual item χ^2 and overall χ^2 for each sample, and we reported the mean values. To account for the significance testing of k individual items, we Bonferroni-adjusted the individual item χ^2 p-values by the number of χ^2 tests performed: $0.05/k = 0.01$ (see Bland & Altman, 1995).

Person z-fit shows how well a person's response pattern conforms to the 'Guttmann structure' (Andrich, 1978). The difference in difficulty of the items caused by dependence is reported as a z-fit residual statitic at a conservative 1% level of significance (z = 2.56), a positive z-fit >2.56 indicates an unexpected response pattern (Andrich & Kreiner, 2010).

Local independence–response independency and unidimensionality

Once we have extracted the Rasch factor–the unidimensional underlying latent trait "self-efficacy", we assume there are no further patterns in the residuals (Wright, 1996). This assumption is tested by checking for response dependency and multidimensionality. *Response dependency* implies that items are linked in such a way that the responses to one item influence the responses to other items, and we identify this phenomenon by inspecting the item residual correlation matrix. The commonly used conservative item residual correlation of < 0.30, has recently come under criticism for being too conservative. Therefore, Yen (1984) proposed exploring local dependence based on comparing the item residual correlation values up against the average item residual correlation with values 0.2 above the average item residual as displaying dependency.

Unidimensionality means that only one latent trait - self-efficacy - explains all the covariances between the items (cf. partial correlations). A combined principal component analysis (PCA) of residuals and paired *t*-tests procedure is applied to check for unidimensionality (Hagell, 2014). If approximately 5% or less of the dependent *t*-tests comparing respondents' location estimates on two distinct subscales are significant, then unidimensionality is assumed (Smith Jr, 2002; Tennant & Pallant, 2006).

Furthermore, by creating a 'sub*test* structure' for a pair of item subsets identified, we can estimate fractal indices (*r*, *c* and *A*) specific to the 'sub*test* structure'. The index *A* describes the amount of common variance among the two subsets or subscales identified, *c* identifies the magnitude of unique subscale variance, and *r* is the correlation between the two subsets (RUMM, 2009). High values for both *A* and *r,* and a low value for *c*, might therefore indicate an approximately unidimensional scale (Andrich, 2016; Andrich, 2015).

Targeting, reliability, ordering of response categories and differential item functioning

In a well-targeted scale, the distribution of the person estimates matches the distribution of the item threshold estimates centred at 0.0 logits. Poor targeting might increase the risk of unordered response categories and disordered thresholds, large standard errors, extreme person scores, and therefore deflated reliability indices and poor information at certain locations along the latent trait scale.

The internal consistency reliability of the latent trait measurement scale is reported as Person Separation Index (PSI), which is analogous to Cronbach's alpha, and indicates the capacity to separate persons with higher location estimates from those with lower location estimates on the latent trait (Andrich, 1982). Different criteria are suggested for PSI, with values >0.70, >0.80 and >0.90 indicating 'acceptable', 'good' and 'excellent' reliability respectively (Duncan, Bode, Lai & Perera, 2003). Often 0.7 is used as the minimum value for group and 0.85 as the minimum value for assessments at the individual item level (Cronbach, 1951).

Differential item functioning (DIF) or 'within-item bias' might occur when different 'levels' or 'groups' of a person factor, such as males and females, at equivalent levels of the underlying construct have different probabilities of endorsing an item (Holland & Wainer, 1993; Walker, Beretvas, & Ackerman, 2001). When persons belonging to a particular 'level' show a consistent systematic difference in their responses to an item, uniform DIF is implied. In cases where the differences vary across levels of the attribute between the person factor groups, non-uniform DIF is indicated. Items that display non-uniform DIF are discarded from the instrument.

A procedure in RUMM2030, allows for the resolution of uniform DIF by resolving the item into multiple items, one for each group levels and comparing the estimates of the item parameters from the different 'levels'.

RESULTS

We found that the SEBH items did not share the same set of threshold difficulties. A significant likelihood ratio test statistic LRT χ^2 (p =0.000019; *df* = 11) indicated that the PCM (partial credit parameterization) of the polytomous unidimensional Rasch model described the data 'significantly' better than the RSM (rating scale parameterisation).

In Table 2, we report the overall adjusted mean χ^2 value for each of the amended sample sizes estimated from five random samples reflecting 10, 20 and 30 individuals per scale threshold, as $\chi 2$ is a sample size dependent fit statistic. The PCM of the polytomous unidimensional Rasch model was applied.

Table 2. Overall mean χ^2 fit statistics for the SEBH scale using amended sample sizes.

df (estimated parameters)	Amend sample size (N)	χ^2	p(χ^2)	Scale thresholds
45	750	71.09	0.015	25
35	500	53.45	0.159	
15	250	20.81	0.197	

To sum up, Table 2 indicates that hypothesis 1 (the SEBH data is sufficiently described by RSM) is not fully supported.

Individual person residuals showed that 20 and 89 students had z-fit above/below the cut-off criterion of +/-2.56, respectively (Andrich & Kreiner, 2010). Concerns were raised about values above the +2.56 threshold, as these indicate response patterns that are unlikely i.e., deviate significantly from the Guttmann pattern given the self-efficacy score sum. However, removing these few responses did not significantly change any fit parameter estimates.
The assumption of a locally independent scale holds for the SEBH-scale as no response dependence between any pair of items was observed, and only 6.2% of paired *t*-tests were significant. The *t*-test structure was based on two subsets of items empirically indicated by the PCA of residuals procedure (the easily endorsable items 1–3 (subscale 1) versus items 4 and 5 (subscale 2), see Table 3). A sub-*set* analysis indicated that these two subscales measured strongly related latent traits (high subscale common variance A = 0.90, subscale correlation r = 0.97 and low subscale unique variance c = 0.17).

Table 3. Individual mean item χ^2 fit statistics for the SEBH scale using the amended sample sizes.

Item	Loc	SE	χ^2 (N=750)	p(χ^2)	χ^2 (N=500)	p(χ^2)	χ^2 (N=250)	p(χ^2)
1	-0.61	0.04	15.82	0.13	11.24	0.19	5.11	0.19
2	0.16	0.04	17.09	0.07	9.63	0.27	5.36	0.17
3	-0.03	0.04	10.48	0.33	8.12	0.40	3.04	0.44
4	0.26	0.04	6.27	0.70	8.03	0.44	2.54	0.51
5	0.22	0.04	21.43	0.02	16.43	0.08	4.76	0.26

Note. The location estimates with the standard errors are based on the full sample. Each χ^2 value is the mean value estimated from five random samples of sample sizes corresponding to 10, 20 and 30 persons per thresholds respectively (N = 250, 500, 750).

For all random sample sizes of 250 and 500, all the chi-square values were insignificant ($p(\chi^2)$). For the random samples of 750, the chi-square value for item 5 was significant in two of the five random samples.

When centering the average item location at 0.0 logits, the resulting average person proficiency was at 1.4 logits, pointing to a scale that could have been better targeted. The positively skewed distribution of person self-efficacy estimates deviates somewhat from the locations at which the items measure most efficiently.

The above results suggest that the SEBH-scale is a rather *valid* measure of self-efficacy in tenth graders. Sufficiently high reliability indices indicated a *reliable* measure (PSI = 0.88 for original and complete data sets and Cronbach's alpha = 0.88 (excluding extremes) and 0.92 (including extremes) for the complete data set where the 36 respondents with missing data for one or more SEBH-items were discarded). Hence, the SEBH-scale is an accurate and precise measure of self-efficacy.
Moving from the overall analyses to the single item level, the slightly disordered response categories observed for item 1 (Figure 1) is explained by the somewhat poorly targeted SEBH-scale (Figure 2). The curves in Figure 1 show the probability of endorsing each of the six response categories (1 = 'strongly disagree' and 6 = 'strongly agree') versus person location. The second category does not function as intended. The dotted line is the upper limit asymptote, where probability equals 100%.

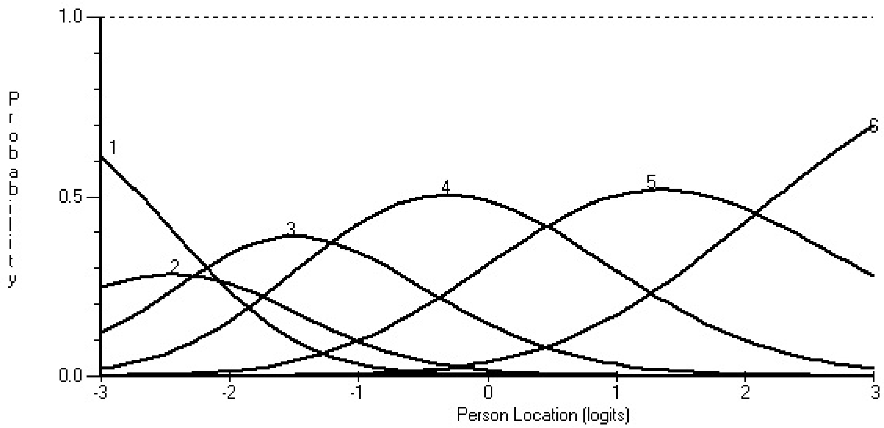

Figure 1. Category probability curves for item 1.

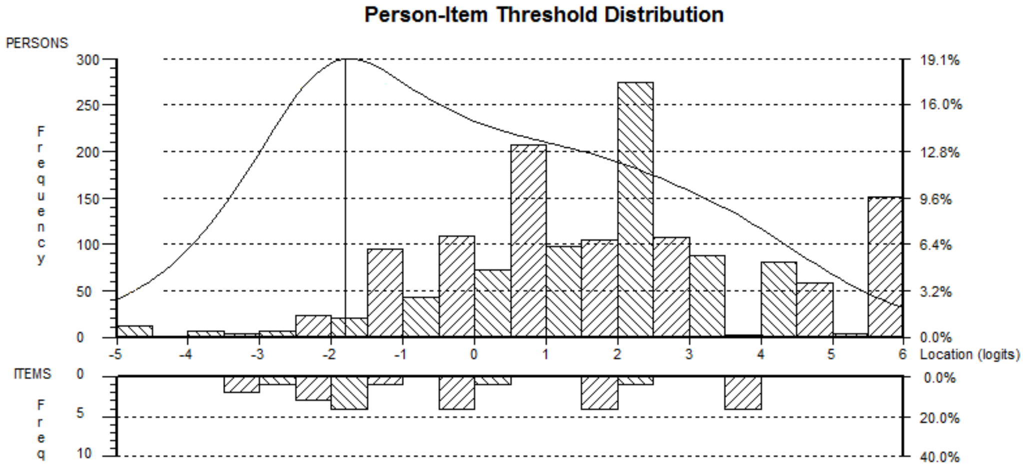

Figure 2. Histogram showing the distributions of person and item threshold locations including Fisher's information function (curve) for the SEBH scale

A skewed distribution toward higher locations of self-efficacy indicates that the the items could have been better targeted at the sample. This skewed distribution leaves few persons located at the lower end of the continuum – the trait locations at which the easily endorsable item 1 (item location at -0.61 according to Table 3) has its lower thresholds. We therefore interpret the SEBH-scale raw score as a sufficient statistic at the ordinal level.

Finally, we investigated DIF using the amended sample sizes based on the rule of thumb of 10, 20 and 30 persons per threshold (Andrich, 2011), with a total of 25 thresholds (5 items with 5 thresholds). No DIF was observed for any person factor (gender, age, cultural background, language at home and books at home) using the amended sample sizes of $N = 250$, $N = 500$ and $N = 750$ based on a total of 25 thresholds (5 items with 5 thresholds)

Discussion

Empirical data partially support our two composite hypotheses. The first hypothesis was strengthened except for a deviation from our ideal of parsimonity: The partial credit parameterization (PCM), estimating one set of threshold parameters for each item, described the data better than the less complex rating scale parameterization (RSM) estimating one set of threshold difficulties common for all items.

Furthermore, the targeting of the SEBH-scale was not optimal with few items at higher locations. The lack of items providing information at higher levels of the latent trait is a well-known problem in health-literacy measurement (Nguyen et al., 2015). One of few exceptions is the 'Claim Evaluation Tools' developed by the Informed Health Choices group. The second hypothesis was strengthened except for slightly disordered thresholds observed for item 1. The disordering of response categories for item 1 has a simple explanation: The distribution of person estimates is skewed toward higher locations thereby locating few persons at the lower end of the continuum–the locations at which we find the lower threshold parameters for item 1.

Since the SEBH-scale built on a self-efficacy scale published by Guttersrud & Pettersen (2015), the scale seems to easily translate to different fields of education improving the generalizability and external validity of our findings. We interpret this as a serious strength of our study. A limitation to our study is the low school participating rate (58 out of 200 or 30%). This might result in responses from students enrolled in classes taught by above average motivated and enthusiastic teachers–teachers more likely to see the benefits of external assessment resources like the one we developed. This possible difference between the target sample and the accessed sample might explain the high mean self-efficacy estimate in our sample, which again could cause the skewed distribution of self-efficacy person location estimates and the disordering observed for item 1.

Conclusions

The present paper provides insights into an issue that seems to have passed health literacy research by: the application of Rasch analysis to evaluate the psychometric properties of measurement scales. By fitting the Rasch model, our findings indicate that the SEBH-scale meets the assumptions and satisfies the requirements for fundamental measurement.

The SEBH-scale presented in the study exemplifies that Rasch analysis is a powerful tool for evaluating construct validity of measurement instruments. This is indicated by the absence of construct-irrelevant variance, as all five items fit the Rasch model, implying that the items don't capture unrelated constructs that affect responses in a manner irrelevant to the construct. On the other hand, by meeting the assumption of unidimensionality albeit with the presence of strongly correlated sub-dimensions, the SEBH-scale points to the absence of construct underrepresentation-another threat to construct validity, in which the assessment is too narrow and fails to capture different facets and sub dimensions of the construct.

Furthermore, the total score on the SEBH-scale can be viewed as one of several possible sets of indicators of the construct-perceived self-efficacy in a science subject. An important recommendation is to include more items in the SEBH-scale in order to improve the preciseness with which the abilities of persons that fall between successive items along the hypothesized unidimensional continuum are measured.

The positive effect of perceived self-efficacy on management of diseases is well documented, developing and validating equivalent measures for 'non-sick' individuals particularly adolescents in different domains, as exemplified in the present study, will go a long way in providing measurement tools to inform, design successful health literacy polies and interventions within public health and education.

Acknowledgements

The authors would like to thank the students who responded and the extraordinary support from school heads and teachers who selflessly facilitated the data collection process and the reviewers who provided constructive comments and suggestions.

Disclosure statements

Conflict of interests: None.
Funding: This research received no specific grant for the work done.
Ethics of human subject participation: Ethics approval was not required.

References

Aikens, N. L. & Barbarin, O. (2008). Socioeconomic differences in reading trajectories: the contribution of family, neighborhood and school contexts. *Journal of Educational Psychology*, *100*(2), 235-251. doi:10.1037/0022-0663.100.2.235

Allison, P. D. (2001). Missing Data: Sage University Papers Series on Quantitative Applications in the Social Sciences (7-136). *Thousand Oaks*, CA.

Andersen, E. B. (1977). Sufficient statistics and latent trait models. *Psychometrika*, *42*(1), 69-81. doi:http://dx.doi.org/10.1007/BF02293746

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561-573. doi:http://dx.doi.org/10.1007/BF02293814

Andrich, D. (1982). An index of person separation in latent trait theory, the traditional KR. 20 index, and the Guttman scale response pattern. *Education Research and Perspectives*, *9*(1), 95-104. Retrieved from http://www.rasch.org/erp7.htm

Andrich, D. (1988). Rasch Models for Measurement. Beverly Hills: Sage Publications, Inc.

Andrich, D. (1989). Distinctions between assumptions and requirements in measurement in the social sciences. In J.A. Keats, R. Taft, R.A. Heath, & S.A. Lovibond (Eds.) *Mathematical and Theoretical Systems* (Vol.4, pp. 7-16). North Holland: Elsevier Science Publishers BV.

Andrich, D. & Kreiner, S. (2010). Quantifying response dependence between two dichotomous items using the Rasch model. *Applied Psychological Measurement*, *34*(3). 181 - 192. doi: https://doi.org/10.1177/0146621609360202

Andrich, D., Lyne, A., Sheridan, B., & Luo, G. (2010). RUMM2030, Computer software and manual. Perth, Australia: RUMM Laboratory Pty Ltd.

Andrich, D. (2011). *EDUC5606: Advanced course in Rasch Measurement Theory [lecture handout].* Faculty of Education. The University of Western Australia, Australia.

Andrich, D. (2015). Components of Variance of Scales with a Subscale Structure Using Two Calculations of Coefficient alpha. (*Pensamiento Educativo*). *Journal of Latin American Educational Research*, *52*(2), 6-33.

Andrich, D. (2016). Components of variance of scales with a Bifactor subscale structure from two calculations of α. *Educational Measurement: Issues and Practice, 35*(4), 25–30. doi:10.1111/emip.12132

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. P*sychological Review, 84*(2), 191-215. doi: http://dx.doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1997). Self-efficacy: The exercise of self-control. New York: W.H. Freeman and Company.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.) *Self-efficacy beliefs of adolescents* (Vol 5, pp. 307-337). Greenwich, CT: Information Age Publishing.

Bergen, E., Zuijen, T., Bishop, D., & Jong, P. F. (2017). Why are home literacy environment and children's reading skills associated? What parental skills reveal. *Reading Research Quarterly, 52*(2), 147-160. doi:10.1002/rrq.160

Bland J.M., & Altman, D.G. (1995). Multiple significance tests: the Bonferroni method. *British Medical Journal, 310,* 170. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2548561/pdf/bmj00576-0038.pdf

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational psychology review*, *15*(1), 1- 40. doi:10.1023/A:1021302408382

Bradley, R. H., Corwyn, R. F., McAdoo, H. P., & García Coll, C. (2001). The home environments of children in the United States part I: Variations by age, ethnicity, and poverty status. *Child Development*, *72*(6)*,* 1844-1867. doi:10.1111/1467-8624.t01-1-00382

Brown, S. L., Teufel, J. A., & Birch, D. A. (2007). Early adolescents' perceptions of health and health literacy. *Journal of School Health*, *77*(1), 7-15. doi:10.1111/j.1746-1561.2007.00156.x

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334. doi:10.1007/BF02310555

Davidson, M., Keating, J. L., & Eyres, S. (2004). A low back-specific version of the SF-36 Physical Functioning scale. *Spine, 29*(5), 586-594. doi:10.1097/01.BRS.0000103346.38557.73

Duncan, P. W., Bode, R, K., Lai, S. M., Perera, S. (2003). Rasch analysis of a new stroke-specific outcome scale: the stroke impact scale. *Archives of Physical Medicine and Rehabilitation, 84*(7), 950-963. doi:https://doi.org/10.1016/S0003-9993(03)00035-2

Dunn, O. J. (1959). Estimation of the medians for dependent variables. *Annals of Mathematical Statistics*, *30*(1), 192-197. doi:10.1214/aoms/1177706374

Escobar, A., del Mar Trujillo-Martín, M., Rueda, A., Pérez-Ruiz, E., Avis, N. E., & Bilbao, A. (2015). Cross-cultural adaptation, reliability and validity of the Spanish version of the Quality of Life in Adult Cancer Survivors (QLACS) questionnaire: application in a sample of short-term survivors. *Health and Quality of Life Outcomes*, *13*(1), 182-195. doi: http://doi.org/10.1186/s12955-015-0378-2

Gray, N. J., Klein, J. D., Noyce, P. R., Sesselberg, T. S., & Cantrill, J. A. (2005). The Internet: a window on adolescent health literacy. *Journal of adolescent health*, *37*(3), 243.e1-243.e7. doi:10.1016/j.jadohealth.2004.08.023

Gutiérrez-Doña, B., Lippke, S., Renner, B., Kwon, S., & Schwarzer, R. (2009). Self-Efficacy and Planning Predict Dietary Behaviors in Costa Rican and South Korean Women: Two Moderated Mediation Analyses. *Applied Psychology: Health and Well-Being*, *1*(1), 91-104. doi:10.1111/j.1758-0854.2009.01004.x

Guttersrud, Ø., & Pettersen, K. S. (2015). 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. *Public Health Nutrition, 18*(14), 2565-2574. doi: https://doi.org/10.1017/S1368980014003152

Harris, D. (1989). Comparison of 1-, 2-, and 3-Parameter IRT Models. *Educational Measurement: Issues and Practice*, *8*(1), 35-41. doi: 10.1111/j.1745-3992.1989.tb00313.x

Hofer, B. K., & Pintrich, P. R. (1997). The Development of Epistemological Theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, *67*(1), 88-140. doi:10.3102/00346543067001088

Holland, P., & Wainer, H. (1993). *Differential Item Functioning*. Hillsdale (NJ): Lawrence Erlbaum Associates, Publishers.

Huang, Y.-J., Chen, C.-T., Gong-Hong, L., Wu, T.-Y., Chen, S.-S., Lin, L.-F., ... Hsieh, C.-L. (2018). Evaluating the European Health Literacy Survey Questionnaire in patients with stroke: A latent trait analysis using Rasch modeling. *The Patient - Patient-Centered Outcomes Research*. *11*(1), 83-96. doi:10.1007/s40271-017-0267-3

Katsikatsou, M., Moustaki, I., Yang-Wallentin, F., & Jöreskog, K. G. (2012). Pairwise likelihood estimation for factor analysis models with ordinal data. *Computational Statistics & Data Analysis*, *56*(12), 4243-4258. Retrieved from https://www.diva-portal.org/smash/get/diva2:458843/FULLTEXT01.pdf

KD (2016). *Fag – Fordypning – Forståelse. En fornyelse av Kunnskapsløftet* [No English title available]. The Office of the Prime Minister. Retrieved from https://www.regjeringen.no/contentassets/e8e1f41732ca4a64b003fca213ae663b/no/pdfs/stm201520160028000dddpdfs.pdf

Linacre, J. M. (1998). Detecting multidimensionality: which residual data-type works best? *Journal of outcome measurement, 2*(3), 266-283. Retrieved from https://pdfs.semanticscholar.org/d970/94395b593b123607ba2fc40376edd5130f51.pdf#page=99

Marsh, H. W., Hau, K. T., Artelt, C., Baumert, J, & Peschar, J. L. (2006). OECD's Brief self-report measure of educational psychology's most useful affective constructs: cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing*, *6*(4), 311-360. Retrieved from http://dx.doi.org/10.1207/s15327574ijt0604_1

Martin, A. J., Nejad, H. G., Colmar, S., & Liem, G. A. D. (2013). Adaptability: How students' responses to uncertainty and novelty predict their academic and non-academic outcomes. *Journal of Educational Psychology, 105*(3), 728. doi:10.1037/a0032794

Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174. doi: 10.1007/BF02296272

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2009). Risk factors for learning-related behavior problems at 24 months of age: Population-based estimates. *Journal of Abnormal Child Psychology, 37*(3), 401-413. doi:10.1007/s10802-008-9279-8

Neuhauser, L., Rothschild, R., & Rodríguez, F. M. (2007). MyPyramid. gov: assessment of literacy, cultural and linguistic factors in the USDA food pyramid web site. *Journal of nutrition education and behavior*, *39*(4), 219-225. doi:10.1016/j.jneb.2007.03.005

Nguyen T.H., Paasche-Orlow M.K., Kim M.T., Han H.-R. & Chan K.S. (2015) Modern measurement approaches to health literacy scale development and refinement: Overview, current uses, and next steps. *Journal of Health Communication 20* (sup2), 112-115. doi: 10.1080/10810730.2015.1073408

Organization for Economic Co-operation and Development, Programme for International Student Assessment (2001). Knowledge and skills for life. Paris: OECD.

Paakkari, L., & Paakkari, O. (2012). Health literacy as a learning outcome in schools. *Health Education*, *112*(2), 133-152. doi: https://doi.org/10.1108/09654281211203411

Pastorelli, C., Caprara, G. V., Barbaranelli, C., Rola, J., Rozsa, S., & Bandura, A. (2001). The structure of children's perceived self-efficacy: A cross-national study. *European Journal of Psychological Assessment*. *17*(2). doi: 10.1027//1015-5759.17.2.87

Pellegrino, J. W., & Hilton, M. (2012). Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century. Washington, DC: The National Academies Press.

Perline, R., Wright, B. D., & Wainer, H. (1979). The Rasch model as additive conjoint measurement. *Applied Psychological Measurement, 3*(2), 237-255. doi:http://dx.doi.org/10.1177/014662167900300213

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests* (Expanded edition 1983 ed.). Copenhagen: Danish Institute for Educational Research.

Rayane, A., & Achterberg, C. (1997). Review of self-efficacy and locus of control for nutrition and health-related behavior. *Journal of the academy of nutrition and dietetics*. *97*(10), 1122-1132. doi: 10.1016/S0002-8223(97)00273-3

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

Salzberger, T. (2010). Does the Rasch model convert an ordinal scale into an interval scale? *Rasch Measurement Transactions 24*(2) 1273-1275. Retrieved from https://www.rasch.org/rmt/rmt242a.htm

Schunk, D. H. (1991). Self-efficacy and academic motivation. *Educational Psychologist, 26*(3-4), 207-231. doi:10.1207/s15326985ep2603&4_2

Schunk, D. H., & Meece, J. L. (2006). Self-efficacy development in adolescence. In F. Pajares & T. C. Urdan (Eds.) *Self-efficacy beliefs of adolescents* (pp. 72-96). Greenwich, Connecticut: Information Age Publishing.

Schwarzer, R. (2008). Modeling health behavior change: How to predict and modify the adoption and maintenance of health behaviors. *Applied Psychology*, *57*(1), 1-29. doi:10.1111/j.1464-0597.2007.00325.x

Schwarzer, R., & Renner, B. (2009). Health-specific self-efficacy scales. Retrieved from http://www.academia.edu/2391923/Health-Specific_Self-Efficacy_Scales

Shaw, F. (1991). Descriptive IRT vs. prescriptive Rasch. *Rasch Measurement*, 5(1), 131. Retrieved from http://www.rasch.org/rmt/rmt51f.htm

Smith Jr., E.V. (2002) Detecting and Evaluating the Impact of Multidimensionality Using Item Fit Statistics and Principal Component Analysis of Residuals. *Journal of Applied Measurement*, 3, 205-231. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/12011501?dopt=Abstract

Stenner, A. J. (1994). Specific objectivity-local and general. *Rasch Measurement Transactions*, *8*(3), 374. Retrieved from http://www.rasch.org/rmt/rmt83e.htm

Taylor, L., Adelman, H. S., & Kaser-Boyd, N. (1984). Attitudes toward involving minors in decisions. *Professional Psychology: Research and Practice*, *15*(3), 436-49. Retrieved from http://psycnet.apa.org/fulltext/1984-25481-001.pdf

Tennant, A. and Pallant, J. (2006) Unidimensionality Matters! (A Tale of two Smiths?). *Rasch Measurement Transactions*, *20*(4), 1048-1051. Retrieved from https://rasch.org/rmt/rmt201.pdf

Tsai, C.-C., Ho, H. N. J., Liang, J.-C., & Lin, H.-M. (2011). Scientific epistemic beliefs, conceptions of learning science and self-efficacy of learning science among high school students. *Learning and Instruction*, *21*(6), 757-769. doi:10.1016/j.learninstruc.2011.05.002

Walker, C.M., Beretvas, S.N. & Ackerman, T. (2001). An Examination of Conditioning Variables Used in Computer Adaptive Testing for DIF Analyses. *Applied Measurement in Education*, *14*(1), 3-16. Retrieved from http://www.learntechlib.org/p/93401

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, *54*(3), 427–450. doi:https://doi.org/10.1007/BF02294627

Wright, B. D. (1996). Local dependency, correlations and principal components. *Rasch Measurement Transactions*, *10*(3), 509–511. Retrieved from https://www.rasch.org/rmt/rmt103b.htm

Young, C. A., Mills, R. J., Woolmore, J., Hawkins, C. P., & Tennant, A. (2012). The unidimensional self-efficacy scale for MS (USE-MS): developing a patient based and patient reported outcome. *Multiple Sclerosis Journal*, *18*(9), 1326-1333. doi:10.1177/1352458512436592

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology, 25*(1), 82-91. doi:10.1006/ceps.1999.1016

Paper II

Naigaga, D. A., Pettersen, K. S., Henjum, S., & Guttersrud, Ø. (2018). Assessing adolescents' perceived proficiency in critically evaluating nutrition information. International *Journal of Behavioral Nutrition and Physical Activity, 15*(1), 61.

DOI: **https://doi.org/10.1186/s12966-018-0690-4**

Naigaga *et al. International Journal of Behavioral Nutrition and Physical Activity* (2018) 15:61
https://doi.org/10.1186/s12966-018-0690-4

International Journal of Behavioral Nutrition and Physical Activity

RESEARCH Open Access

Assessing adolescents' perceived proficiency in critically evaluating nutrition information

Desire Alice Naigaga[1*], Kjell Sverre Pettersen[1], Sigrun Henjum[1] and Øystein Guttersrud[2]

Abstract

Background: Over the recent past, there has been an increase in nutrition information available to adolescents from various sources, which resulted into confusion and misinterpretation of the dietary advice. Results from international assessment frameworks such as PISA and TIMMS reflect the need for adolescents to critically appraise health information. While a number of scales measuring the critical health literacy of individuals exist; very few of these are devoted to critical nutrition literacy. More so, these scales target individuals with an advanced level of nutrition education, often gaging their proficiency in information appraisal in relation to principles of evidence-based medical research. The purpose of the present study was to examine the psychometric properties of a newly developed critical nutrition literacy scale (CNL-E) measuring adolescents' perceived proficiency in 'critically evaluating nutrition information from various sources'.

Methods: During spring 2015, more than 1600 tenth graders aged 15–16 years from approximately 60 schools in Norway responded to the five-item questionnaire using an electronic survey system. Applying Rasch analysis approach, we examined the psychometric properties of the CNL-E scale employing the RUMM2030 statistical package. To further investigate the dimensionality of the scale and test the underlying structure, we applied multidimensional Rasch modelling using the ConQuest 4 software and confirmatory factor analysis (CFA) using the Lisrel 9.30 software.

Results: In our sample, the CNL-E stood out as a valid, reliable and well-targeted scale with good overall fit to the partial credit parameterization of the polytomous unidimensional Rasch model (PCM). All the items were sufficiently statistically independent, had ordered response categories and showed acceptable individual fit to the PCM. No item displayed within-item bias or differential item functioning (DIF).

Conclusions: From the observed CNL-E sum score, it is possible to draw plausible conclusions about how individuals critically evaluate nutrition information. Efforts to improve communication of nutrition information could benefit from applying validated measures such as the CNL-E scale. The CNL-E scale provides insight into how individuals without an advanced level of nutrition education, such as adolescents, determine the validity and reliability of nutrition information from various sources.

Keywords: Critical nutrition literacy, Scientific literacy, Media literacy, Adolescents, Rasch analysis, Rasch modelling, Confirmatory factor analysis

* Correspondence: dnaiga@oslomet.no
[1]Department of Nursing and Health Promotion, Faculty of Health Sciences, OsloMet - Oslo Metropolitan University, Kunnskapsveien 55, 2007, KC228 Kjeller, Norway
Full list of author information is available at the end of the article

Background

'*For good health, increase the intake of proteins and lower the intake of fat. No, wait a second- increase the fat intake and lower the carbohydrates*'. To critically interpret such seemingly 'contradicting' nutrition information and dietary advice from different sources, a high level of nutrition literacy is needed.

Dietary habits acquired during early adolescence are often life-long and have a strong impact on one's future health. Adolescents today are exposed to a vast amount of nutrition information from various sources including traditional print media such as newspapers and magazines; online media like websites, blogs, social media platforms like Facebook; advertisements on television, radios; from health experts like dietitians, doctors and from social interaction with family and peers. While increased access to nutrition information is a welcome progression in efforts to advance nutrition promotion strategies, with it has emerged an increase in the 'confusion' associated with having too much information, a characteristic of 'information overload' [1]. This points to the concept of 'filter failure' indicating that the strategies for deciding which information is relevant have not evolved at the same pace as the means for producing the information [2]. In addition, the assurances about the quality of information provided by these sources seems to be lagging behind [3]. These concerns have fueled the interest in exploring how individuals appraise nutrition information obtained from various sources, prior to making nutrition-related decisions.

Appraisal skills encompass the ability to interpret, filter, judge and evaluate health information obtained [4]. In the field of nutrition, these skills are associated with nutrition literacy, specifically critical nutrition literacy (CNL). Broadly, nutrition literacy refers to an individual's capacity to access, process and understand nutrition information needed to make appropriate decisions regarding one's nutrition [5–7]. In the domain of critical nutrition literacy, information appraisal skills are emphasized in one's ability to evaluate the quality of the nutrition information and advice received [8]. Methodological advancement in the field of nutrition literacy has yielded a number of assessment instruments used to measure the skills and competencies associated with nutrition literacy in both clinical and non-clinical settings [9–14]. However, only a few of these measures specifically target the domain of 'critical' nutrition literacy [7, 15, 16]. Furthermore, these measures of critical nutrition literacy have predominantly focused on how individuals with an advanced level of education appraise information and nutrition claims based on principles of evidence-based research [16]. This focus seems to overlook how individuals without an advanced level of nutrition education, appraise and contextualize nutrition information in the media.Therefore, the aim of the present study was to examine the psychometric properties of a newly developed CNL scale measuring perceived proficiency in evaluating nutrition information from various sources, targeting individuals without an advanced level of nutrition education. Owing to the evaluation aspect of CNL, we refer to the scale as the critical nutrition literacy – evaluation (CNL-E) scale. We translated our aim into the following three hypotheses:

H1) The CNL-E scale has acceptable overall fit to the restricted rating scale parametrization of the polytomous unidimensional Rasch model (PCM), consists of locally independent items, and represents a well targeted and reliable measurement scale.
H2) Each item in the CNL-E scale has ordered response categories, displays no within-item bias or differential item functioning (DIF), and has acceptable individual fit to the PCM.
H3) Using confirmatory factor analysis, the CNL-E scale has acceptable factorial validity and discriminant validity.

It follows from empirical support of the above hypotheses that reasonable claims about adolescents' critical evaluation of nutrition information from various sources, that go beyond the observed CNL-E scale score sums, are plausible.

Method

Frame of reference

We randomly selected 200 schools from a list of lower secondary schools in Norway and the respective school principals were contacted by email and telephone seeking consent to volunteer in the study. From the 58 schools that accepted to participate, we collected data during the period of April to May 2015 by use of an electronic survey system from 1622 students aged 15–16 years.

The substantive theory of the CNL-E latent variable

Basing on Nutbeam's tripartite model of health literacy [17], nutrition literacy is categorized into three cumulative levels referred to as *functional nutrition literacy* (FNL), *interactive nutrition literacy* (INL) and *critical nutrition literacy* (CNL). FNL is concerned with basic writing and reading skills that are required to access information about nutrition. INL is comprised of the interpersonal communication and cognitive skills which enable individuals to translate and apply information in their daily lives with the aim of improving their overall nutritional status. Thirdly, CNL is concerned with higher level cognitive and social skills that enable individuals to critically appraise nutrition information and advice, as well as engage in actions that are aimed at addressing the barriers to good nutrition at individual and group levels [15, 16, 18]. The critical dimension of health literacy (CHL), which is akin to CNL, is also conceptualized as 'judgement skills'- the ability to judge information based on factual knowledge necessary to deal with novel situations [19]. With respect to nutrition literacy, factual or declarative knowledge is characterized by an awareness of the facts and processes that pertain a certain nutrition benefit or condition [15]. Therefore, individuals that are 'critically nutrition-literate' are expected to meaningfuly interprete and skillfully establish how reliable, valid and credible nutrition information and dietary advice is, by comparing this information to established nutrition facts (factual knowledge).

This aspect of judging information against established factual knowledge is reflective of scientific literacy (SL), which is the capacity to apply factual scientific knowledge to identify scientific issues, explain scientific phenomena and to draw evidence-based conclusions in order to inform decisions in personal, social and global contexts [20]. Scientific knowledge in different contexts for example in the field of nutrition, provides the criteria against which information is judged. Additionally, scientific literacy is concerned with the skills that enable individuals to assess the trustworthiness (validity) of information and their willingness to participate in science-related issues, with the ideas of science as constructive, concerned and reflective citizens [20]. Therefore, it thus follows that critical nutrition literacy (CNL) is part of scientific literacy (SL) as CNL involves the application of nutrition knowledge to explain, evaluate and interpret nutrition information basing on scientific factual knowledge about nutrition and involves emancipatory action to address barriers to good nutrition [21].

Owing to the vast amount of nutrition information that is available from various sources, it is important that individuals are 'media literate'. Media literacy encompasses the competencies and skills that enable persons to access, analyze, evaluate and produce communication in a variety of forms [22]. A proposed theory of media literacy suggests that in order for individuals to become more media literate, they must possess the capacity to comprehend information, that is to find meaning in information hosted by the media (meaning matching) and the capacity to transform information from the media and create meaning for oneself (meaning construction) [23, 24]. Central to both these inter-twined capacities of media literacy are 'evaluation or appraisal skills', which are suggested as one of the most relevant critical thinking skills required for the effective appraisal of messages in the media [23]. The process of critical thinking involves skillfully analyzing and synthesizing information as a guide to action and is a core component of health literacy. This is especially important if individuals are to create meaningful links between health information obtained from numerous sources in the media [25].

From the above, it is evident that evaluation skills are a crucial link between CNL, ML and SL as they enable persons to adequately identify nutrition claims, assess the consistency of nutrition information in the media and establish the validity of the underlying messages through comparison with established scientific knowledge, thereby informing their action towards overcoming barriers to good nutrition.

The CNL-E items

The five-item CNL-E scale shown in Table 1, uses a six-point response scale anchored with the phrase 'on a scale from 'very difficult' to 'very easy', how easy or difficult would you say it is to (1 = Very difficult, 6 = Very easy)' The phrase is adapted from the European Health Literacy Survey Questionnaire (HLS-EU-Q47) [4]. The items were generated basing on competencies related to the process of understanding and appraising health-related information, as reflected in the integrated model of health literacy; and the category of 'scientific enquiry', according to the PISA framework for assessing scientific literacy [4, 26]. The sources of information were categorized into 'traditional' sources covering television, print sources such as newspapers, magazines and 'online' sources such as websites. Items 1–3 assessed the extent to which respondents felt that they could trust the nutrition information from different sources. These items explored how competent the respondents were in comprehending and interpreting nutrition information in order to maintain adequate nutritional status and prevent malnutrition. Items 4 and 5 assessed the proficiency with which respondents felt they could establish the falsifiability of nutrition claims by judging the information against basic knowledge about nutrition (facts). As the 10th grade marks the end of compulsory education in Norway, students in the 12th grade have acquired basic knowledge about nutrition (factual) and are expected to ably apply these facts while making decisions about nutrition [27].

Person factors and data properties

In addition to the CNL-E scale, students reported on the following person factors; gender as either male or female, language predominantly spoken at home as Norwegian, Danish/Swedish (Scandinavian languages) or 'other', and their mother's, father's and own place of birth as Norway, Denmark/Sweden or 'other'. A dummy variable, 'cultural background' was created with the person factor levels '*majority*' if at least either the student or one of the parents were born in a Scandinavian country and '*minority*' if elsewhere. The levels of linguistic and cultural background are justified by the similarities among the Scandinavian

Table 1 Wording of the CNL-E scale items (originally stated in Norwegian)

Item	Item wording
1	evaluate whether nutritional advice in the media (newspapers, magazines, television) is reliable?
2	consider how reliable warnings about poor nutrition are, as warnings against malnutrition?
3	consider whether information on websites for nutritional information is reliable?
4	consider what it takes a scientific nutritional claim to be valid?
5	evaluate nutritional advice in the media (newspapers, magazines, television) in a scientific way?

Note: The six-point response scale was anchored with the phrase 'on a scale from 'very difficult' to 'very easy', how easy or difficult would you say it is to (1 = Very difficult, 6 = Very easy)'

countries. Lastly, as an indicator of socioeconomic status (SES), the students reported how many books they could access at home [28]. The number of books in the home was used as an indicator of SES as research on SES and family resources shows that the aspects of the home literacy environment, such as 'opportunity'-which includes the number of books in a home; are strongly correlated with children's reading skills [29, 30]. Additionally, when measuring SES at student level in heterogonous groups, the number of books shows clearer differences between children from different backgrounds [31]. In order to help improve the response accuracy, a picture of how different numbers of books might appear on a bookshelf, in five groups of 10 through to 200 might look like, was included. A dichotomous variable with the levels 'less than 100 books' and '100 or more books' was thereby defined.

CNL-E scale response characteristics

Of the 1622 students in the sample, 78 did not respond to any of the CNL-E items (invalid records) and 137 students (less than 10%) had one or more missing responses. Item 1 had the lowest number of missing responses (80) and item 4 had the highest number of missing responses (109). There were 75 extreme scorers in the data set; 28 of whom responded "1" to all five items and 45 of whom responded "6" to all five items. With 78 invalid records and 75 extreme scorers, there were 1469 students with valid scores available for analyses.

Validating measurement models approach 1: Rasch analysis (RA) – Testing the empirical data up against the theoretical requirements of fundamental measurement

A measurement model describes how responses to a set of items (observed variables) reflect a unidimensional latent trait (unobserved variable), such as 'critically evaluating nutrition information. Theoretically, Rasch models fulfill the assumptions and requirements of fundamental measurement such as unidimensionality, equal item distribution, specific objectivity and additivity [32–39]. Rasch analysis makes it possible to assess the psychometric properties of new and existing scales, by assessing whether the response patterns in the data fit the expectations of Rasch models [40]. Based on the prescriptive Rasch models, the distance between the item location (difficulty) and person location (proficiency) defines the expected probability of a certain response [41]. The polytomous unidimensional Rasch model (PCM) assumes two parameterizations, the 'unrestricted' partial credit parameterization (PCM) [42] or the 'restricted' rating scale parameterization (RSM) [43], where the latter is nested within the first. Data-model fit of 'competing' nested models is compared applying likelihood ratio tests (LRT) [44]. The LRT test statistic is the difference or change in deviance, which is the asymptotically chi-square (χ^2) distributed statistic with degrees of freedom (*df*) equal to the difference in number of estimated parameters. A large and significant χ^2 value indicates that the null hypothesis, which states that the less complex nested model or parameterization describes the data better than the more complex model or parameterization; should be rejected. If we compare models before and after discarding and adding items, we no longer have nested models and apply measures such as the Akaike Information criteria (AIC) [43–45].

Fit to the Rasch model

Comparing the 'theoretically or model expected' probabilities of responses to the 'empirically observed' portions, yields a formal 'chi-square test of goodness-of-fit'. The concept 'item discrimination' indicates how well an item is capable of discriminating or separating between individuals with higher person location estimates along the latent trait from those with lower estimates. An 'under-discriminating' item is indicative of a weaker distinction between such respondents than what is expected by the RM and is indicated by a large and nonsignificant item chi-square value ($p(\chi^2) < 5\%$) as compared to the χ^2 distribution on that degrees of freedom. To account for statistical misfit that might arise owing to chance, we adjust the significance level by the number of χ^2 tests applied, applying the Bonferroni adjustment [39].

Ordering of response categories and differential item functioning

When respondents use the rating scales as intended, ordered thresholds reflect the increasing levels of severity across each response category [40]. The different threshold locations reflect the locations at which the probability of a response in two adjacent categories is equal.Within-item bias is examined by checking for the presence of differential item functioning (DIF). DIF is indicated when individuals with the same standing on the latent trait belonging to different categories of a person factor (such as gender) have different probabilities of endorsing an item [41, 46].

Targeting and reliability

In a well-targeted scale, the distribution of the person estimates matches the distribution of the item threshold estimates, where either the person or item estimates are centered at 0.0 logits. Poor targeting increases the risk of extreme scores and unordered response categories. The Person Separation Index (PSI), which is analogous to Cronbach's alpha, indicates how precise the measurement is, given unidimensionality. Values greater than 0.70 suggest better internal consistency reliability [42].

The assumption of local independence

The Rasch models assume locally independent items – i.e., all covariance between the items is attributed to the latent trait variable or 'Rasch factor'. Violation of local independence is reflected as either *multidimensionality* implying that more than one latent variable influences the responses, or *response dependence* where subsets of items share further similarities than those accounted for by the latent trait variable.). Response dependence is indicated by significant correlation between item model residuals (>.30) [43, 44].

To empirically check for unidimensionality, a combined principal components analysis of residuals (PCA) and paired *t*-tests procedure is available in RUMM. If 5% or less of the *t*-tests are significant, then the proportion of instances in which two item subsets yield "significantly" different person location estimates is small enough to retain the hypothesis of unidimensionality [46–49]. Additionally, subtest structures based on theoretical or empirical assumptions of subsets of items might be formed to investigate violations of local independence [50].

Furthermore, additional tests of dimensionality can be carried out by applying multidimensional Rasch modelling in the ConQuest [51] program and confirmatory factor analysis (CFA) using Lisrel [52].

Validating measurement models approach 2: Confirmatory factor analysis (CFA) – Covariance characteristics defining latent traits

A structural equation model involves *measurement models* that define latent variables and a *structural model* to indicate how the latent variables are related [53].

Model specification

A confirmatory factor model is based on theoretical assumptions, that the observed variables represent the latent variable accurately, albeit with a unique variance (error). The model in Fig. 1 demonstrates that the latent variable "critical nutrition literacy evaluation" (CNL-E) is measured by the observed variables CNL1, CNL2, CNL3, CNL4 and CNL5, taking into account the unique variance associated with each of the observed variables

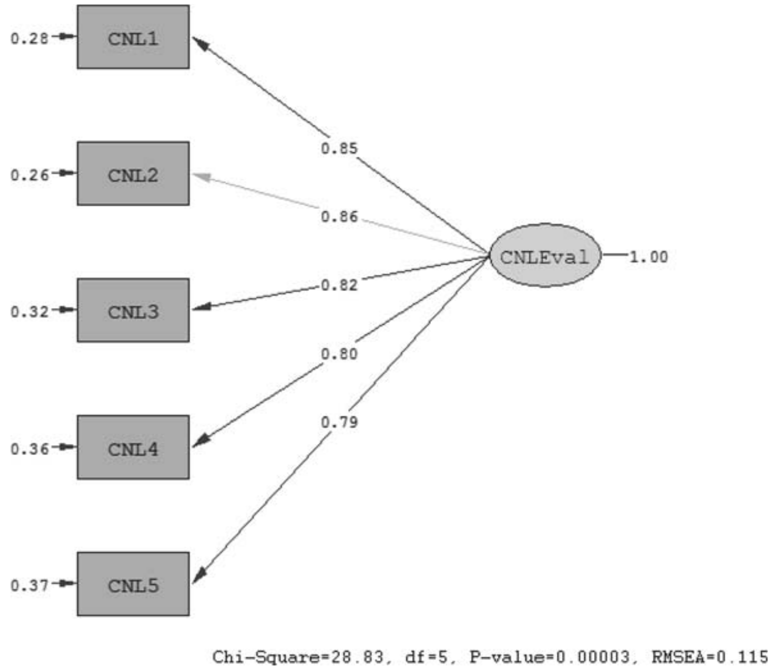

Fig. 1 The highest factor loading (CNL2) was fixed to 1.0 (completely standardized solution). The rectangular boxes represent the observed variables (CNL1-CNL5), while the oval shape represents the unobserved latent factor, CNL-E. The single arrows pointing towards the observed variables indicate the specific variance for each of the five variables

CNL1-CNL5. Formally, the model in Fig. 1 is a hypothesized a priori 1-factor confirmatory factor model (M1) testing the hypothesized relationship between the latent variable and the observed variables i.e.; whether the responses to the questionnaire items measure the latent variable. The factor structure for the hypothesized a priori 1-factor confirmatory factor in Fig. 2 for a post hoc modified model (M2) is conceptually diagrammed, based on modification indices suggested by Lisrel between items CNL1 and CNL4 and CNL2 and CNL5. In both figures, the variance of the latent variable (CNL-E) is fixed to 1.0 (completely standardized solutions).

Model identification

Refers to deciding on whether a single value for each unknown model parameter, referred to as a free parameter (FP) is obtainable from the observed data. Because latent variables are unobservable, they have no scale of their own; therefore their origin and unit of measurement can be assumed by defining the unit of the latent variable in relation to reference variable-an observed variable whose coefficient (factor loading) is fixed; and or by fixing the variance of the latent variable to 1.0, thereby assuming that it is a standardized variable.

As the observed variance-covariance matrix ***(S)*** is an "unstructured" and symmetric matrix, it contains $k(k+1)/2$ unique or distinct values (DV). The single factor CFA measurement model in Fig. 1 has $k = 5$ items so there are 15 DV in $\boldsymbol{S}$. The number of free parameters (FP) to be estimated for the model in Fig. 1 were counted as following (see Table 5): 4 factor loadings (with 1 other factor loading (CNL2) fixed to 1), 5 item unique variances, 0 item unique covariances[1] (or correlations in standardized solution) and 1 latent variable variance – a total of 10 free parameters. For a model to be identified, FP must be less than or equal to DV. Therefore, the model in Fig. 1 is "over-identified" with degrees of freedom $df = \text{DV-FP} = 5$. This means it is possible to obtain a single value for each unknown FP from the observed data and still have degrees of freedom available for estimating data-model fit.

Model estimation

Refers to the estimation of data-model fit. Using diagonally weighted least squares (DWLS) and or maximum likelihood (ML) estimation, the FP are estimated with the aim of obtaining a model-based (implied) variance-covariance matrix ($\boldsymbol{\Sigma}$) with elements as close as possible to the elements in $\boldsymbol{S}$ [53, 54]. The objective of the estimation is therefore to achieve a residual variance-covariance matrix $\boldsymbol{S}\text{-}\boldsymbol{\Sigma}$ where the elements are as small as possible. Because ML assumes multivariate normality, in cases of non-continous distribution, such as with observations made on ordinal variables like those obtained from rating scales as in the current study; asymptomatic distribution-free (ADF) estimators are used [55]. These include the diagonally weighted least squares (DWLS), which make no assumptions about the

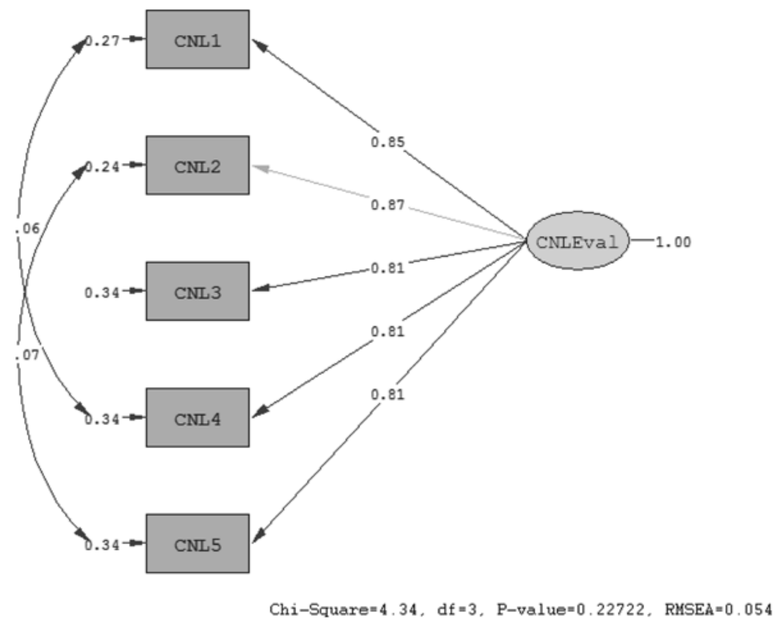

Fig. 2 The double-headed arrows represent the uniqueness correlations between CNL1 & CNL4, and CNL2 & CNL5. The highest factor loading (CNL2) is fixed to 1.0 (completely standardized solution). The rectangular boxes represent the observed variables (CNL1-CNL5), while the oval shape represents the unobserved latent factor, CNL-E. The single arrows pointing towards the observed measured variables indicate the uniqueness, a composite of specific variance and measurement error specific to each of the five variables

distribution of the observed variables. DWLS minimizes the chance over-estimating chi-square fit values and underestimating standard errors [53].

Model evaluation

Refers to evaluating the discrepancy between $\boldsymbol{S}$ and $\boldsymbol{\Sigma}$. Absolute fit indices such as the chi-square (χ^2) and standardized root-mean-square residual (SRMR), are used to achieve this. A statistically significant chi-square value implies imperfect model fit and points to rejection of the model [55], therefore it is advisable to report other fit indices as they provide different information about model fit, providing more conservative and reliable evaluation of the fit to the model. The "parsimony correction indices", such as the root mean square error of approximation (RMSEA) and "close" fit (Cfit), evaluate the discrepancy between $\boldsymbol{S}$ and $\boldsymbol{\Sigma}$ while penalizing complex models with many parameters [53–55]. Using ML or DWLS estimation along with the asymptotic covariance matrix, LISREL implements the mean-adjusted Santorra-Bentler scaled χ^2 to adjust for non-normality. Incremental fit indices, such as the comparative fit index (CFI) and the non-normed fit index (NNFI) or Tucker-Lewis index (TLI), assess absolute or parsimonious fit relative to a baseline model hypothesizing no relationships among the variables. The latter indices are therefore rather liberal, with values greater than 0.95 for the CFI, NNFI (TLI), indicative of an acceptable model-data fit [54]. Values <.05, .05- < .08, .08–.10 imply good fit, reasonable fit (.05- < .08) and mediocre fit (.08–.10), while values >.10 indicate poor fit. An associated fit index is the C-fit value, which is a test of the closeness of fit when RMSEA < 0.05. Values greater than 0.05 indicate a good model fit [54–56].

Lastly, the critical sample (CN) statistic; which shows the size that a sample should reach in order to accept the fit of a given model on a statistical basis. Values > 200 indicate that the model is an accurate representation of the data [57].

Model modification

Refers to adding or removing items and/ or paths to obtain better data-model fit – that there are alternative models predicting the observed variables better. Modification indices > 3.84 indicate which previously fixed parameters should be set free (added) in order to improve model fit maximally [53]. However, this should only be done if the modifications fit the underlying theory. It is advisable that where possible, researchers test the resultant post-hoc model on a different sample as adjusting models after initial testing increases the chances of capitalizing on sampling error, in that 'idiosyncratic characteristics of the sample may influence the modifications performed' [58]. Furthermore, because model modifications generally result into better fitting models, there is a risk of having more data-driven than theory-driven models which are not generalizable across samples [59]. Therefore, it is important to justify any model modifications on empirical and/or conceptual grounds such as item content and violations of local independence [60].

Results

Rasch analysis

This section begins with a discussion of the dimensionality of the data, fit to the Rasch model, the individual item response dependency (local independence) and finally elaborates on the validity of the theoretically derived model using confirmatory factor analysis, for the latent variable 'critically evaluating nutrition information from various sources'. All analyses run smoothly.

Using ConQuest, the data was fitted to the partial credit parameterization (PCM) (deviance =15,544, number of estimated parameters = 21) and to the rating scale parameterization (RSM) (deviance = 19,004, number of estimated parameters = 10) of the unidimensional polytomous Rasch model. Comparing these nested parameterizations yielded a significant LRT chi-square statistic χ^2 ($\Delta df = 11$, $N = 1469$, $p < 0.05$, critical value = 19.68, implying that the PCM describe the data significantly better than the RSM. The change in deviance is asymptotically χ^2 distributed (see Step 1 in Fig. 3).

Applying the principal component analysis in RUMM, the subset of items CNL4 and CNL5 loaded positively on the first principal component, while the subset of items CNL1, CNL2 and CNL3 loaded negatively on that component. These two item subsets might therefore tap into two different aspects of the overall underlying trait 'critically evaluating nutrition information and therefore possibly define two subscales which might rank individuals differently. Regarding local independence, none of the residual correlations between any pairs of items in the scale excceded 0.3, implying that there was no significant response dependence between the items. However the presence of large negative residual correlations less than − 0.3, pointed to the presence of possible underlying dimensions (multidimensionality). Furthermore, a PCA of item residuals yielded two item sets comprised of items CNL1, CNL2 and CNL3 and items CNL4 and CNL5, respectively. Subsequently paired *t*-tests implied a sufficiently unidimensional scale as approximately 5% of the paired t-tests were significant. Using ConQuest, the data was fitted to the partial credit parametrization of the 2-dimensional polytomous Rasch model (deviance = 19,072, number of estimated parameters =31), where the two item subsets defined the two dimensions (see Step 2 in Fig. 1), and to the partial credit parametrization of the unidimensional polytomous Rasch model (deviance =15,544, number of

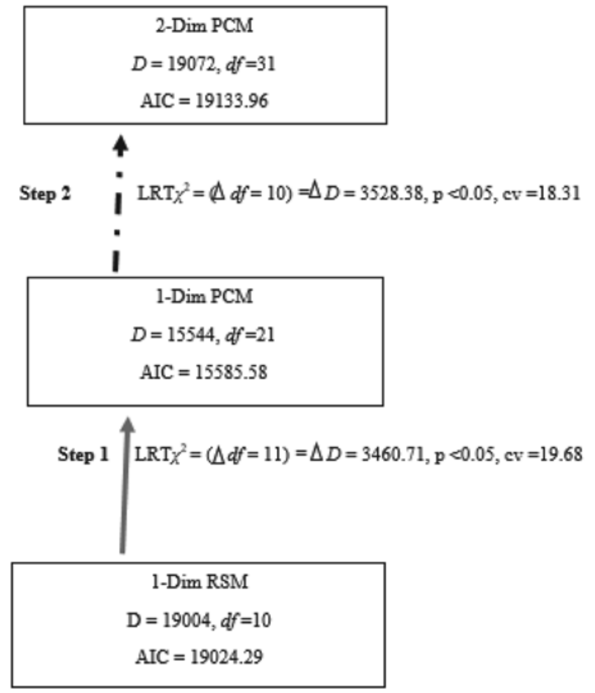

Fig. 3 Step 1 tests the null hypothesis that the restricted rating scale parameterization (RSM) describes the data as well as the more complex partial credit parameterization (PCM) of the polytomous unidimensional Rasch model (PCM). Step 2 investigates dimensionality, comparing fit to the PCM of the 1-dimensional scale and a 2-dimensional scale. In the 2-dimensional scale, the items are categorized into two 'sub-dimensions' i.e., items 1, 2, 3 and items 4 and 5, based on qualitative interpretation of item content, confirmed by the PCA/*t*-test procedure in RUMM2030. The dashed arrow indicates that the 1-dim PCM is the preferred parameterization

estimated parameters = 21). Comparing these nested models yielded a non-significant LRT chi-square statistic χ^2 ($\Delta df = 10$, $N = 1469$) $p < 0.01$, critical value = 23.2 (Step 2 in Fig. 1) pointing to better data-model fit for the the unidimensional than the multidimensional Rasch-model.

At the individual item level, all five CNL-scale items meet the model expectations and fit well to the partial credit parametrization of the unidimensional polytomous Rasch model (PCM)as shown in Table 2 (chi-square values for N = 1469).

All of the items in the CNL-E scale displayed ordered response categories, implying that the CNL-E raw score produces data at the ordinal level and can be transformed to interval using Rasch-modelling.

No item displayed within-item bias (DIF) across the different person factor levels for the available person factors (age, gender, socio-economic status/books at home, linguistic background/language spoken at home and cultural background/place of birth).

With the average item location centered at 0.0 logits, the mean person location at 0.42 logits suggests a sufficiently well-targeted scale, meaning that the items in the CNL-E scale sufficiently captured the range of the latent trait within the sample. From RUMM, using the weighted maximum likelihood estimator (WMLE) high reliability indices (PSI = 0.88 for the original data set (with missing values) and Cronbach's alpha = 0.90 for the complete data set without missing values), indicate that the CNL-E scale was a reliable measure in our sample. Likewise, estimation of the multidimensional model in ConQuest applying marginal maximum likelihood estimation (MLE), showed that both subscales in the 2-dimensional scale had MLE person separation reliability coefficients larger than 0.70 i.e.; dim1 (0.754), dim 2(70.21). Therefore, the CNL-E scale seems to measure "critical evaluation of nutrition information from various sources" sufficiently reliably in our sample.

Confirmatory factor analysis of the underlying latent structure

Using Lisrel, we conducted a confirmatory factor analysis in which we specified the a priori one-factor (CNL-E) measurement model with five observed variables (items CNL1 – CNL5). Applying robust maximum likelihood estimation, the goodness-of-fit SRMR index was well below its target value (Table 3) and the Satorra-Bentler χ^2 value was insignificant possibly owing to large sample size ($N = 1469$). The CFI and NNFI were both very

Table 2 Individual item fit statistics for the critical nutrition literacy-evaluation (CNL-E) scale (pairwise maximum likelihood estimation using RUMM)

Item	Loc	SE	Thresholds					FitRes	χ^2 ($N = 1395$)	$p(\chi^2)$
3	− 0.171	0.04	− 3.3	− 1.8	− 0.1	1.7	3.5	−0.48	4.5	0.88
4	−0.099	0.04	−3.1	−2.0	−0.2	1.9	3.4	1.40	6.1	0.73
2	−0.076	0.04	−3.7	−2.0	−0.0	2.0	3.7	−2.14	12.4	0.19
5	0.097	0.04	−2.8	−1.8	−0.2	1.6	3.3	1.24	9.9	0.36
1	0.248	0.04	−2.7	−2.0	−0.3	1.7	3.2	−1.72	9.6	0.39

Note: Items sorted by location order, applying the partial credit parameterization (PCM) of the polytomous unidimensional Rasch model (PCM) was applied, $df = 9$ Estimates shown are from unidimensional Rasch analysis using RUMM2030 software showing item location point estimate (*Loc*) with standard error (*SE*). The $p(\chi^2)$ reports the probability of observing the χ^2 value on the given degrees of freedom (*df*) where $df = G$-1 and *G* is the number of proficiency groups applied in the formal test of fit. The *thresholds* indicate the location of the items along the latent continuum, covering the latent trait from approximately −3.7 to + 3.7 logits. The rounded average of the five thresholds are approximately − 3, − 2, − 0, 2 and 3 logits respectively

Table 3 Model evaluation by goodness-of-fit indices (GOFI) for the a priori specified measurement model M1 in Fig. 2 and the re-specified and data-driven post hoc modified measurement model M2 in Fig. 3 (robust maximum likelihood estimation using the statistical package LISREL)

Model (M) and GOFI goodness-of-fit target value	Absolute GOFI		Parsimony-adjusted GOFI		Incremental GOFI	
	SB scaled χ^2 with p-value	SRMR	RMSEA (90% CI)	CFit	CFI	NNFI
M1 ($df = 5$, $N = 1485$)	28.83, $\mathbf{p = .0000}$	0.021	**0.115 (0.096; 0.135)**	**0.000**	0.995	0.990
M2 ($df = 3$, $N = 1485$)	4.34, $p = .2272$	0.009	0.054 (0.030; 0.081)	0.358	1.000	0.999
target value	$p > .05$	< .05	< .06 (< .05; < .08)	> .05	> .95	> .95

Note: M2 is the more restricted and nested model obtained from M1 by the addition of the covariance between the uniqueness variance components of items CNL1 and CNL4, and items CNL2 and CNL5

df = degrees of freedom, *N* = effective sample size (list wise deletion). Goodness-of-fit indices (GOFI) are classified as absolute, parsimony-adjusted and incremental: SB scaled χ^2 = Satorra-Bentler scaled chi-square, SRMR = Standardized Root Mean Square Residual, RMSEA = Root Mean Square Error of Approximation, CFit = p-value for test of Close Fit (i.e., the probability that RMSEA < 0.05), CFI = Comparative Fit Index, NNFI = Non-Normed Fit Index = TLI = Tucker & Lewis fit index. Bold values imply mediocre to poor data-model fit (the SB scaled χ^2 *p*-value for M1 is insignificant owing to large sample size)

high and clearly above their respective target value. As the RMSEA was above .06 the CFit was below .05. Therefore, the absolute and the incremental fit indices, as opposed to the parsimony-adjusted indices, strengthened the hypothesis of a well-fitting measurement model. The post hoc model (model M2 in Table 3), in which we added the uniqueness covariance between items CNL2 & CNL5 and between CNL1 & CNL4, as suggested by the modification indices in Lisrel; showed a significant improvement in all six fit indices as seen in Table 3. The modifications were purely data-driven and might therefore capitalize on sampling error. It is warranted that future studies define both models a priori and test whether model 2 (M2) is preferred to model 1 (M1).

Furthermore, investigation of the standardized residual matrix pointed to improved 'local fit' in the post hoc modified model, as indicated by a rise in the standardized residual values following the addition of parameters between the error variances as suggested by the modification indices (Table 4).

Investigation of the parameter estimates of both the a-priori model (M1) and post hoc modified model are shown in Tables 5 and 6 respectively. The better fitting model-the post hoc modified model, shown in Table 6, show that all the factor loadings exceeded 0.71 and all unique variances were below 0.50, an indicator that the latent trait under study largely explained the variance in the responses to the observed variables. Taken together, the five variables measured accounted for approximately 70% of the variance in the latent factor, as indicated by a mean R^2 value of 0.69. Both the a priori specified and post hoc modified models were over-identified as the difference between the number of distinct values (15) and the number of free parameters (10 and 12, respectively) were larger than 0 ($df = 15–10 = 5$ and $df = 15–12 = 3$, respectively). However, since the better fitting model- the post hoc model (M2) was tested on the same sample, there is a possibility of 'capitalizing on sampling error'.

Discussion

Empirical data from the Rasch-modelling approach supports hypotheses H1 and H2 with one exception; the less complex parameterization of the polytomous unidimensional Rasch model (PCM) described the CNL-E scale data 'significantly' better than the more restricted rating scale parameterization (RSM) did. This means that the PCM contained more information about the data as it estimated one set of threshold parameters for each item, unlike the RSM, which estimates one set of step difficulties common for all items. We therefore offer the following post hoc explanation; that the four thresholds are not equal in size across the five items and there is need to estimate one set of threshold parameters for each item. And while using the same sample to evaluate fit of post hoc model modifications is not advised, we were not able to obtain another sample on which to test the modified model (M2). However we further justify these modifications based on the high negative residual correlations observed in Rasch analysis, indicative of items from different dimensions in the latent structure of the variable CNL-E.

Furthermore, qualitative interpretation of item content and categorization of the items into the subsets identified based on PCA residuals confirmed the substantive theory of the underlying latent trait (CNL-E); that critical evaluation of nutrition information from various sources requires skills that are well recognized and central to 'media literacy' and 'scientific literacy' [24, 61]. 'Media literacy' is concerned with skills pertaining to the ability to assess the consistency (reliability) of information while 'scientific literacy' is concerned with the skills that enable individuals to assess the the trustworthiness of information (validity of information) [21].

Higher item order on the latent continuum suggests that items addressing skills related to assessing the validity of information appear to provide more information about the latent trait (CNL-E) than those concerned with assessing the reliability of information. This supposition finds support in Potter's cognitive model of media literacy [23],

Table 4 Standardized residual matrices for the critical nutrition literacy evaluation (CNL-E) measurement models

Variable	CNL1	CNL2	CNL3	CNL4	CNL5
Original a priori model:					
CNL1					
CNL2	0.380				
CNL3	0.658		0.000		
CNL4	−1.668	1.627	−0.862	0.000	
CNL5	0.673	−1.380		1.390	
Modified *posthoc* model:					
Variable	CNL1	CNL2	CNL3	CNL4	CNL5
CNL1	0.000				
CNL2		0.000			
CNL3	0.959		0.000		
CNL4	0.000		−1.024	0.000	
CNL5	−0.341	0.000	0.875	0.215	0.000

Note: All standardized residuals of the a priori and post hoc modified models are within the accepted range of ≤ +/− 1.96. The largest values (−1.668, − 1.380, 1.390) indicate that the a priori model does not account very well for the correlations between CNL1 and CNL4, CNL2 and CNL5, and CNL4 and CNL5 respectively. Adding parameters between the error covariances of CNL1 and CNL4, and CNL2 and CNL5 in the post hoc modified model results into a decrease in the residual values, indicating better fit

Table 5 Model identification and model estimation for the a priori measurement model in Fig. 2 (applying robust DWLS and ML using the statistical package LISREL)

Model Identification		Unstandardized solution				Completely standardized solution	
		DWLS		ML		DWLS	ML
FP	Observed variables	Estimate	(SE)	Estimate	(SE)	Estimate	Estimate
1	CNL1 factor loading	.981	(.024)	.984	(.020)	.856	.847
	CNL2 factor loading	1.000*		1.000*		.872	.861
2	CNL3 factor loading	.932	(.020)	.958	(.020)	.813	.824
3	CNL4 factor loading	.932	(.023)	.928	(.021)	.813	.799
4	CNL5 factor loading	.931	(.021)	.923	(.022)	.812	.794
5	CNL1 unique variance	.267		.282		.267	.282
6	CNL2 unique variance	.239		.259		.239	.259
7	CNL3 unique variance	.339		.320		.339	.320
8	CNL4 unique variance	.339		.362		.339	.362
9	CNL5 unique variance	.340		.369		.340	.369
	Latent variable						
12	CNL-Eval variance**	.742	(.023)	.741	(.024)	1.000	1.000

Note. CNL1 - CNL5 are the observed variables, CNL-Eval is the latent variable. FP = Free parameter (counting the number of free parameters to be estimated with reference to the unstandardized soultion), DWLS = Diagonally Weighted Least Squares estimation, ML = Maximum Likelihood estimation, SE = Standard Error, Factor loading = the proportion of the total variance that an item shares with the other items ie., is common to the items (a variance component accounted for by the latent variable in the model), Unique variance = the proportion of the total variance that is unique to an item (a variance component not accounted for by the latent variable model in the model i.e., the *un*modelled variance component). Additional correlation was specified between the error covariances of CNL1 and CNL4 and CNL2 and CNL5

#) Lisrel reports unique variance components as 1-R^2 for both the standardized and the unstandardized solutions, where R^2 is the squared standardized factor loading when the item only load on one factor

*) Factor loading constrained to 1 owing to item being used as reference or marker variable to resolve the origin and unit of measurement problem

**) The variance of the latent variable is the "covariance with itself" in the *un*standardized solution and the "correlation with itself" in the standardized solution. The latter is always 1

Table 6 Model identification and model estimation for the post hoc modified measurement model in Fig. 3 (applying robust DWLS and ML using the statistical package LISREL)

Model Identification		Unstandardized solution				Completely standardized solution	
		DWLS		ML		DWLS	ML
FP	Observed variables	Estimate	(SE)	Estimate	(SE)	Estimate	Estimate
1	CNL1 factor loading	.981	(.024)	.976	(.024)	.856	.852
	CNL2 factor loading	1.000*		1.000*		.872	.873
2	CNL3 factor loading	.932	(.020)	.930	(.020)	.813	.812
3	CNL4 factor loading	.932	(.023)	.933	(.023)	.813	.814
4	CNL5 factor loading	.931	(.021)	.931	(.021)	.812	.813
5	CNL1 unique variance	.267		.274		.267	.274
6	CNL2 unique variance	.239		.238		.239	.238
7	CNL3 unique variance	.339		.341		.339	.341
8	CNL4 unique variance	.339		.337		.339	.337
9	CNL5 unique variance	.340		.339		.340	.339
10	CNL1,CNL4 uniqueness relationship**	– 0.061	(.021)	– 0.059	(.021)	–.061	–.059
11	CNL2,CNL5 uniqueness relationship**	– 0.066	(.018)	–0.067	(.018)	–.066	–.067
	Latent variable						
12	CNL-Eval variance***	.761	.024	.762	.024	1.000	1.000

Note. CNL1 - CNL5 are the observed variables, CNL-Eval is the latent variable. FP = Free parameter (counting the number of free parameters to be estimated with reference to the unstandardized soultion), DWLS = Diagonally Weighted Least Squares estimation, ML = Maximum Likelihood estimation, SE = Standard Error, Factor loading = the proportion of the total variance that an item shares with the other items ie., is common to the items (a variance component accounted for by the latent variable in the model), Unique variance = the proportion of the total variance that is unique to an item (a variance component not accounted for by the latent variable model in the model i.e., the *un*modelled variance component). Additional correlation was specified between the error covariances of CNL1 and CNL4 and CNL2 and CNL5

#) Lisrel reports unique variance components as 1-R^2 for both the standardized and the unstandardized solutions, where R^2 is the squared standardized factor loading when the item only load on one factor

*) Factor loading constrained to 1 owing to item being used as reference or marker variable to resolve the origin and unit of measurement problem

**) The relationship refers to the covariance (in the *un*standardized solution) and the correlation (in the standardized solution) between the uniquene variance components of the repective observed variables. These relationships are data-driven re-specifications of M1

***) The variance of the latent variable is the "covariance with itself" in the *un*standardized solution and the "correlation with itself" in the standardized solution. The latter is always 1

in which he describes the advancement in skills associated with the different levels of information-processing, starting with *filtering* of messages, analogous to assessing the reliability of information and sources; through to *meaning-matching*, *meaning-making* and finally *meaning-construction*. The latter steps, which point to advanced information-processing, require individuals to refer to previously learned knowledge in order to determine the meaning of a message and thereby create their own meaning that is relevant for them. Similarly, it can be anticipated that the ability to assess validity of nutrition information, requiring the ability to effectively interprete and use scientific knowledge as a criterion to appraise nutrition information from various sources like the media; reflects an advanced level of CNL evaluation.

Limitations of the study

While it is recommended to evaluate fit of post hoc model modifications on a different sample in order to minimize the chances of capitalizing on sampling error, we were unable to obtain another sample on which to test the post hoc modified model. Therefore we recommend that the CNL-E scale is applied on different age-groups and populations in order give better insight into the validity of the modified model.

In the current study, the number of books owned at home was used as an indicator of family SES; while it is an appropriate indicator in studies involving young children and adolescents, it is rather outdated. With the widespread use of digital learning platforms including e-books, a better suited indicator of family SES could be the number of computers or e-readers that they have access to at home.

The sources of nutrition information that were captured by the items were limited to 'traditional' media and online media sources. Other information sources such as dietitians, peers, family could have been included, as it is equally important to establish the credibility of this information. Furthermore, rewording the items to remove complex jargon terms such as 'claims' might benefit the respondents who might not be familiar with the term.

Conclusion

A significant *theoretical* outcome of our study is that we managed to overcome a well-known challenge of nutrition literacy measurement; the lack of a clear theoretical basis and thereby poorly founded methodological advancement [61].

An important practical outcome of this study was that we were able to develop a set of short non-abstract user-friendly test items assessing how individuals with a basic level of nutrition education (12th grade) evaluate nutrition information obtained from various sources. This is of significance as existing measures of critical evaluation of nutrition information are comprised of items which require the subjects to have an advanced knowledge about evidence-based medicine.

Furthermore, while it is recognized that measuring critical health literacy is demanding, requiring careful consideration of wording and context [61, 62]; the current study shows that by focusing on established aspects of the critical dimension of nutrition literacy such as 'evaluation of nutrition information and advice', it is possible to operationalize and measure nutrition literact at the 'more advanced' level (critical domain). Additionally, this study reveals the potential benefits of critical thinking skills in effective evaluation of nutrition information from various sources. By emphasizing the skillful analyzing, translating and application of established scientific knowledge across different disciplines like nutrition; individuals will be better equipped to identify potentially harmful nutrition claims, thereby lessening the 'confusion' caused by the seemingly contradicting nutrition information from various sources.

Lastly, as the field of nutrition literacy advances, applying instruments such as the valid, accurate and precise CNL-E scale presented in this paper, that may be beneficial in evaluating the impact of interventions and programs that are primarily focused on nutrition education.

Endnotes

[1]No item unique correlations were specified a priori based on prior research in different samples. The FP increases to 12 and *df* decreases to 3 after specifying 2 unique correlations post hoc (model in Fig. 3).

Abbreviations

AIC: Akaike Information criteria; CFA: Confirmatory factor analysis; CFI: Comparative fit index; CNL: Critical nutrition literacy; CNL-E: Critical Nutrition Literacy-Evaluation; *df*: Degrees of freedom; DIF: Differential item functioning; DWLS: Diagonally weighted least squares; FNL: Functional nutrition literacy; ICC: Item characteristic curve; INL: Interactive nutrition literacy; LRT: Likelihood ratio test; ML: Maximum likelihood; NNFI: Non-normed fit index; PCM: Partial credit parameterization of the Rasch model; PCMRSM: Rating scale parametrization of the Rasch model; PSI: Person Separation Index; RMSEA: Root mean square error of approximation; SB: Satorra-Bentler; SRMR: Standardized root mean square residual

Acknowledgements

The authors would like to thank the students who responded and the extraordinary support from school principals and teachers who selflessly facilitated the data collection process and the reviewers who provided constructive comments and suggestions. We also would like to thank Lena Victoria Nordheim for contributing to the development of the CNL-E scale items.

Availability of data and materials

The data that support the findings of this study are available from The Norwegian Directory for Education but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of The Norwegian Directory for Education.

Authors' contributions

DAN conducted the statistical anaysis and drafted the manuscript. KSP read through and offered guidance on the contents of the manuscript. SH read through and contributed towards the manuscript. KSP developed the items that were included in the questionnaire applied in the study. ØG conceived of the study, participated in the collection of data, statistical analysis and drafting of the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

No ethics approval was required. Data analyzed was collected as part of a field test trial of the national science test.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details

[1]Department of Nursing and Health Promotion, Faculty of Health Sciences, OsloMet - Oslo Metropolitan University, Kunnskapsveien 55, 2007, KC228 Kjeller, Norway. [2]Norwegian Centre for Science Education, University of Oslo, Naturfagsenteret, Postboks 1106, Blindern, 0317 Oslo, Norway.

Received: 3 January 2018 Accepted: 11 June 2018
Published online: 28 June 2018

References

1. Yang CC, Chen H, Honga K. Visualization of large category map for internet browsing. Decis Support Syst. 2003;35:89–102.
2. Shirky C. It's not information overload. It's filter failure. Mas Context. 2014. http://mascontext.com/pdf/MAS_Context_Issue07_INFORMATION.pdf. Accessed 16 July 2017.
3. Ishikawa H, Kiuchi T. Health literacy and health communication. Biopsychosoc Med. 2010. https://doi.org/10.1186/1751-0759-4-18.
4. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, Brand H. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health. 2012. https://doi.org/10.1186/1471-2458-12-80.
5. Neuhauser L, Rothschild R, Rodríguez FM. MyPyramid.gov: assessment of literacy, cultural and linguistic factors in the USDA food pyramid web site. J Nutr Educ Behav. 2007;39:219–25.
6. Silk KJ, Sherry J, Winn B, Keesecker N, Horodynski MA, Sayir A. Increasing nutrition literacy: testing the effectiveness of print, web site, and game modalities. J Nutr Educ Behav. 2008;40:3–10.
7. Zoellner J, Connell C, Bounds W, Crook L, Yadrick K. Nutrition literacy status and preferred nutrition communication channels among adults in the lower Mississippi Delta. Prev Chronic Dis. 2009;6:A128.
8. Guttersrud Ø, Pettersen KS. 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. Public Health Nutr. 2015. https://doi.org/10.1017/S1368980014003152.
9. Diamond JJ. Development of a reliable and construct valid measure of nutritional literacy in adults. Nutrition. 2007. https://doi.org/10.1186/1475-2891-6-5.
10. Krause C, Sommerhalder K, Beer-Borst S, Abel T. Just a subtle difference? Findings from a systematic review on definitions of nutrition literacy and food literacy. Health Promot Int. 2016:daw084.
11. Rothman RL, Housam R, Weiss H, Davis D, Gregory R, Gebretsadik T, Elasy TA. Patient understanding of food labels: the role of literacy and numeracy. Am J Prev Med. 2006;31:391–8.
12. TenHave TR, Van Horn B, Kumanyika S, Askov E, Matthews Y, Adams-Campbell LL. Literacy assessment in a cardiovascular nutrition education setting. Patient Educ Couns. 1997;31:139–50.
13. Weiss BD, Mays MZ, Martz W, Castro KM, DeWalt DA, Pignone MP, Hale FA. Quick assessment of literacy in primary care: the newest vital sign. Ann Fam Med. 2005;3:514–22.
14. Carbone ET, Zoellner JM. Nutrition and health literacy: a systematic review to inform nutrition research and practice. J Acad Nutr. 2012;112:254–65.
15. Velardo S. The nuances of health literacy, nutrition literacy, and food literacy. J Nutr Educ Behav. 2015;47:385–9.
16. Guttersrud Ø, Dalane JØ, Pettersen S. Improving measurement in nutrition literacy research using Rasch modelling: examining construct validity of stage-specific 'critical nutrition literacy' scales. Public Health Nutr. 2014;17:877–83.
17. Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. Health Promot Int. 2000;15:259–67.
18. Berman M, Lavisso-Mourey R. Obesity prevention in the information age. J Am Med Assoc. 2008;300:433–5.
19. Schulz P, Nakamoto K. Emerging themes in health literacy. Stud Commun Sci. 2005;5:1–10.
20. Organisation for Economic Co-operation and Development (OECD). Assessing scientific, reading and mathematical literacy: A framework for PISA 2006. OECD, Paris: 2006.
21. Pettersen K.S. Health claims and scientific knowledge. A study of how students of Health Sciences, their teachers and newspaper journalists relate to health claims in society. University of Oslo, Oslo: 2007.
22. Aufderheide P, Firestone C. Media Literacy: A report of the national leadership conference on media Literacy Queenstown MD: Aspen Institute; 1993.
23. Potter WJ. Theory of media literacy: A cognitive approach. London: Sage Publications Inc; 2004.
24. Frisch AL, Camerini L, Diviani N, Schulz PJ. Defining and measuring health literacy: how can we profit from other literacy domains? Health Promot Int. 2012;27:117–26.
25. Paakkari L, Paakkari O. Health literacy as a learning outcome in schools. Health Educ. 2012;112:133–52.
26. Thomson S, Hillman K, De Bortoli L. A teacher's guide to PISA scientific literacy. ACER press, Australia: 2013.
27. Ministry of Education and Research, Norway: The Education System. https://www.regjeringen.no/en/topics/education/school/the-norwegian-education-system/id445118/. Accessed 17 Apr 2018.
28. Aikens NL, Barbarin O. Socioeconomic differences in reading trajectories: the contribution of family, neighborhood, and school contexts. J Educ Psychol. 2008;100. https://doi.org/10.1037/0022-0663.100.2.235.
29. Foster MA, Lambert R, Abbott-Shim M, McCarty F, Franze S. A model of home learning environment and social risk factors in relation to children's emergent literacy and social outcomes. Early Child Res Q. 2005;20:13–36.
30. Bergen E, Zuijen T, Bishop D, Jong PF. Why are home literacy environment and children's reading skills associated? What parental skills reveal. Read Res Q. 2017;52:147–60.
31. Skolverket. Nationella utvärderingen av grundskolan 2003 [The National evaluation of the compulsory school]. Stockholm: Skolverket; 2004.
32. Marais I, Andrich D. Formalizing dimension and response violations of local independence in the unidimensional Rasch model. J Appl Mes. 2008;9:200–15.
33. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied and what should one look for in a Rasch paper? Arthritis Care Res. 2007;57:1358–62.
34. Adams R, Wu M. Multidimensional models.In ConQuest tutorial. 2010. Retrieved from https://www.acer.org/files/Conquest-Tutorial-7-MultidimensionalModels.pdf.
35. Rasch G. An individualistic approach to item analysis. In: Lazarsfeld PF, Henry NW, editors. Readings in mathematical social science. Chicago: science research associates; 1966. p. 89–108.
36. Andersen EB. Sufficient statistics and latent trait models. Psychometrika. 1977;42:69–81.
37. Salzberger T. Does the Rasch model convert an ordinal scale into an interval scale. RMT. 2010;24:1273–5.
38. Andrich D. Rasch models for measurement 1988. Beverly Hills: Sage.
39. Perline R, Wright BD, Wainer H. The Rasch model as additive conjoint measurement. Appl Psych Meas. 1979;3:237–55.
40. Rasch G. Probabilistic models for some intelligence and attainment tests. 1983 ed. Copenhagen: Danish institute for Educ Res; 1960.
41. Dunn OJ. Estimation of the medians for dependent variables. Ann Math Stat. 1959. https://doi.org/10.1214/aoms/1177706.
42. Walker CM, Beretvas SN, Ackerman T. An examination of conditioning variables used in computer adaptive testing for DIF analyses. Appl Meas Educ. 2001;14:3–16.
43. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951; https://doi.org/10.1007/BF02310555.
44. RUMM Laboratory Pty Ltd, Interpreting RUMM2030: part III estimation and statistical techniques, 2009.
45. Andrich D. An index of person separation in latent trait theory, the traditional KR. 20 index, and the Guttman scale response pattern. ERP. 1982; 9:95–104.
46. RUMM Laboratory Pty Ltd, Differential item function (DIF) analyses in extending the RUMM2030 analysis, November 2009.
47. RUMM Laboratory Pty Ltd, Quantifying response dependence in extending the RUMM2030 analyses, November, 2009.
48. Wright B. Local dependency, correlations and principal components. RMT. 1996;10:509–11.
49. Hagell P. Testing rating scale Unidimensionality using the principal component analysis (PCA)/t-test protocol with the Rasch model: the primacy of theory over statistics. OJS. 2014;4:456–65.
50. Smith EV Jr. Understanding Rasch measurement: detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. JAM. 2002;3:205–31.
51. ACER ConQuest command reference, July 2015.
52. Joreskog K, Sorbom D. LISREL (Version 9.3). 2009; Chicago: Scientific Software: Inc.
53. Teo T, Tsai LT, Yang CC. Applying structural equation modeling (SEM) in educational research. In: Myint SK, editor. Application of structural equation modelling in educational research and practice. Sense: Publishers; 2013. p. 3–21.
54. Diamantopoulos A, Siguaw JA. Introducing LISREL. A guide for the uninitiated. London: SAGE publications; 2000.
55. Bandalos DL, Sara JF. Factor analysis: Exploratory and confirmatory. In: Mueller RO, Hancock GR, editors. The Reviewer's guide to quantitative methods in the social sciences. Routledge; 2010.
56. Browne MW, Cudek R. Alternative ways of assessing model fit. In: Bollenand KA, Long JS, editors. Testing structural equation models. Newbury Park: SAGE; 1993.
57. Hoelter JW. The analysis of covariance structure: goodness of fit indices. Sociological methods and Research. 1983;11:325–44.
58. MacCallum R.C. 1995. Model specification: procedures, strategies and related issues. In: Hole RH, editor. Struct Equ Model: Concepts, issues and applications pp. 16–36. Thousand Oaks: Sage; 1995.
59. Martens MP. The use of structural equation modelling in counselling psychology research. Couns Psychol. 2005;33:269–98.
60. MacCallum RC, Roznowski M, Necowitz LB. Model modifications in covariance structure analysis: the problem of capitalization on chance. Psychol Bull. 1992;111:490–504.
61. Steckelberg A, Hülfenhaus C, Jürgen K, Jürgen R. Mühlhauser. How to measure critical health competences: development and validation of the critical health competence test (CHC test). Adv in health. Sci Educ. 2009;14:11–22.
62. Krause C, Sommerhalder K, Beer-Borst S. Nutrition-specific health literacy: development and testing of a multi-dimensional questionnaire. Ernahrungs-Umschau. 2016;63:214–20.

Paper III

Naigaga, D. A., Pettersen, K. S., Henjum, S., & Guttersrud, Ø. (2022). Relating aspects of adolescents' critical nutrition literacy at the personal level. Nutrire 47, 1 (2022).

DOI: **https://doi.org/10.1186/s41110-021-00149-1**

Nutrire (2022) 47:1
https://doi.org/10.1186/s41110-021-00149-1

RESEARCH

Relating aspects of adolescents' critical nutrition literacy at the personal level

Desire Alice Naigaga[1] · Kjell Sverre Pettersen[1] · Sigrun Henjum[1] · Øystein Guttersrud[2,3]

Received: 28 June 2021 / Accepted: 8 November 2021

Abstract
Efforts targeting adolescents' dietary behaviour have often focused on improving their access to nutrition information; however, adolescents report finding nutrition information difficult to understand. Exploring adolescents' critical nutrition literacy might provide insight into how best to improve their use of available nutrition information.

Purpose The purpose of this article is to explore how the two aspects of the critical nutrition literacy - critical evaluation of nutrition information' and 'engagement in dietary behaviour' are linked at personal level. Additionally, the study sought to establish the association between critical nutrition literacy and self-efficacy in nutrition related subjects.

Methods Applying a cross-sectional study design, the study sampled 1622 adolescents aged 15-16years, enrolled in 58 secondary schools in Norway. The adolescents responded to scales measuring self-efficacy and CNL. Using Lisrel 9.30, the study evaluated a structural equation model linking CNL and SEBH.

Results The study yielded a simple yet theoretically sound model depicting the link between CNL and self-efficacy.

Conclusion Efforts promoting adolescents' nutrition literacy might benefit from increasing their self-efficacy in nutrition-related subjects.

Keywords Adolescents · Critical nutrition literacy · Self-efficacy · Structural equation modelling

Health literacy (HL) has fast become an area of interest within the broader scope of public health. Defined as the cognitive and social skills which determine the motivation and ability of individuals to gain access to, understand and use health information in ways which promote and maintain good health, HL has been identified as one of the building blocks of health and a significant influence of health outcomes [1]. The consequences of low health literacy are varied and include among others, low responsiveness to available health services, poor self-management of disease, and low participation of communities in population health programs, among others [2]. HL is context-specific, taking different forms within the field of health; one such important domain is nutrition literacy (NL) defined as 'the capacity to obtain, process and understand nutrition information needed to make appropriate decisions regarding one's health' [3, 4]. There are three domains of NL namely, functional nutrition literacy (FNL), interactive nutrition literacy (INL) and critical nutrition literacy (CNL) [5, 6]. FNL refers to the basic writing and reading skills that are required to access information about nutrition, while INL is comprised of the interpersonal communication and cognitive skills which enable individuals to translate and apply information in their daily lives with the aim of improving their overall nutritional status. CNL refers to proficiency in critically analysing nutrition information and advice, alongside increased awareness and engaging in action to address barriers to sufficient nutrition at personal, social and global levels [6, 7]. At the individual level, CNL might be assessed by the two aspects, 'critical evaluation of nutrition information' (CNL-E) and 'engagement in dietary behaviour' (CNLEng) [8].

During adolescence individuals develop their dietary behaviours. It is therefore plausible that improving NL during adolescence might increase their chances of developing

✉ Øystein Guttersrud
oystein.guttersrud@naturfagsenteret.no

1 Department of Nursing and Health Promotion, Faculty of Health Sciences, OsloMet- storbyuniversitetet, Kunnskapsveien 55, 2007, KC231 Kjeller, Norway

2 Norwegian Centre for Science Education, University of Oslo, Blindern, Norway

3 Naturfagsenteret, Postboks 1106, 0317 Blindern, Oslo, Norway

Published online: 23 December 2021

healthy dietary behaviours and prevent health risks during adulthood. Studies show that adolescents generally find nutrition information difficult to understand, they inadequately interpret nutrition information and are unable to establish the credibility of the sources of this information [9–11]. Therefore, it is not surprising that in spite of having increased access to nutrition information, adolescents rarely use this information properly when making dietary choices [12, 13]. Other studies indicate that individuals engage in a more detailed process of information-appraisal in order to sustain or enhance their state of well-being [14]. This might suggest that in order to improve the NL of adolescents, it is important to provide nutrition information that they can comprehend. By exploring the 'critical' dimension of nutrition literacy, stakeholders might be better informed about how adolescents understand nutrition information, what cues they use to interpret the information, and can therefore tailor the information accordingly. Presently, there are only a few scales for assessing NL, and even fewer for assessing CNL [11, 15]. Moreover, existing CNL instruments have mainly been validated using classical test theory (CTT) techniques. While CTT has long-standing benefits, researchers are adopting the use of modern measurement validation approaches like Rasch modelling which yields psychometrically defendable scales [16].

Self-efficacy refers to the judgements that one holds of one's capabilities to organize and execute actions to achieve designated goals [17]. Accordingly, an individual's self-efficacy perceptions determine one's behaviour such as participation in activities that require the use of the knowledge and skills attained. Self-efficacy influences the choices that adolescents make when faced with options and how they use their cognitive resources and strategies. For example, when making dietary choices, studies show that adolescents that are confident in their ability to apply the information that they have to make dietary-related choices (high self-efficacy) are more likely to make healthier food choices based on detailed comprehension of the nutrition-related cues [9]. Relatedly, studies show that adolescents' self-efficacy in science subjects is associated with their engagement in dietary behaviour (CNLEng) [8]. This finding is judicious as nutrition is a science [18]. Therefore, in the present study we anticipated that the adolescents' self-efficacy in the science subject topic of 'body and health'—one of the main five subject topics in the broader subject of 'nature science', 'body and health' focuses on the structure of the human body, and the impact of lifestyle on an individual's physical and mental health. One of the key elements within 'Body and Health' is nutrition. Herein, the subject topic is concerned with how an individual's lifestyle and health specifically relating to nutrition, diet, dietary patterns and eating disorders will influence how the adolescents comprehend the nutrition information that they encounter (CNL-E) and how they apply this information to achieve their dietary goals (CNLEng).

Methods

We randomly selected 200 schools from the list of lower secondary schools in Norway and contacted the respective school principals via email and telephone, seeking consent to volunteer in the study. Of these, 58 schools (approx. 30%) accepted and were included in the study. During the period of April to May 2015, we collected data from 1622 tenth grade students aged 15–16 years who responded using an electronic survey system.

Analyses

We used three scales measuring each of the three traits of interest to the study, namely, SEBH, CNL-E and CNLEng. Following this, we were able to explore the associations among these traits. The study applied the five-item CNL-E scale to measure the adolescents' perceived proficiency in evaluating nutrition information from various sources [19]. The scale uses a six-point rating scale and captures skills required for evaluating the 'consistency' and 'trustworthiness' of nutrition information. To measure adolescents' engagement in dietary behaviour, we used the two items of the 'engagement in dietary behaviour (EDB) scale' that measures adolescents' engagement in dietary behaviour at the *personal* level [20]. These items relate to how concerned the respondents are about eating healthy foods and having a variety of healthy foods available to them [8]. In order to measure the adolescents' perceived self-efficacy in mastering the health content taught in the science subject topic of 'body and health' in the Norwegian science curriculum, we used the five-item SEBH scale [20].

As we measured the three latent traits (SEBH, CNL-E and CNLEng) using twelve six-point rating scale items we treated all items as categorical variables at the ordinal level. Using the structural equation modelling (SEM) framework to test the hypothesized model, we therefore applied the "diagonally weighted least square" (DWLS) estimator–an asymptotically distribution free estimator available in the Lisrel 9.30 software package. "Asymptotically" refer to "large sample size" $N > 1000$. We followed the steps in conducting a SEM analysis. In the section that follows, we describe the rationale behind the SEM models shown in Fig. 1 and 2.

The model in Fig. 1 shows the measurement models of each of the three latent variables (CNL-E, CNLEng and SEBH) and depicts the hypothesized relationships between them when they are allowed to freely covary.

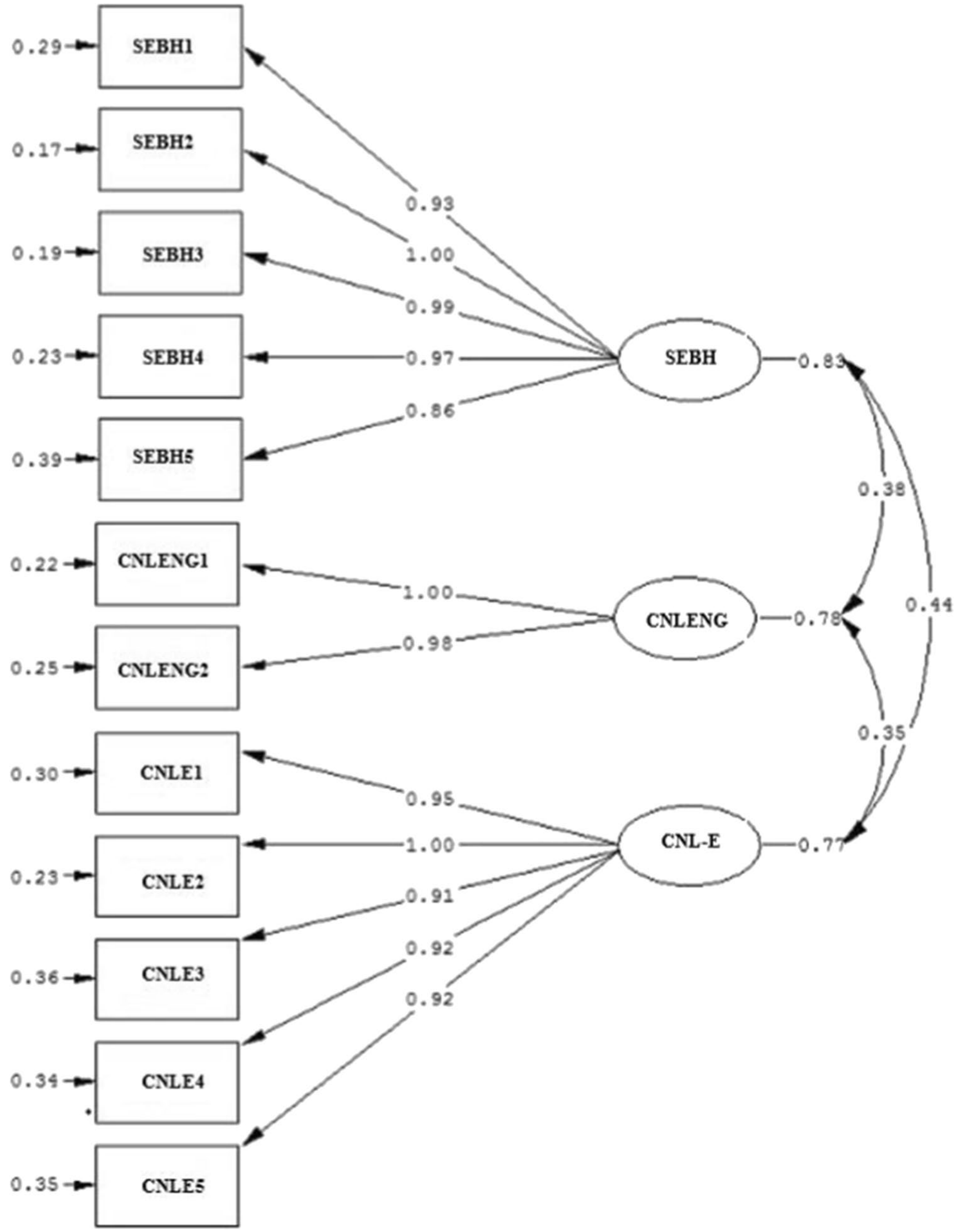

Fig. 1 Structural equation model linking CNL and SEBH using DWLS estimation in which the latent variables are free to covary

In Fig. 2, we specify that adolescents' perceived SEBH explains a portion of the variability in CNL-E and CNLEng. Herein, SEBH serves as an independent latent variable, and CNL-E and CNLEng are dependent latent variables. In the model shown in Fig. 2, we modelled the association between SEBH and CNLEng as a direct effect, the association between SEBH and CNL-E as a direct effect, and an indirect effect where CNL-E facilitates or "mediates" the relationship between SEBH and CNLEng; herein, CNL-E is both an independent variable and a mediator variable.

The SEM model shown in Fig. 2 has $p(p + 1)/2 = 12(12 + 1)/2 = 78$ distinct values (DV), where $p = 12$ is the number of items or indicators in the measurement models. We identified the 27 free parameters (FP) to be estimated (Table 1).

Therefore, the specified model is 'over-identified' with DV-FP = 78 – 27 = 51 degrees of freedom (df). As DV > FP the *order condition* is fulfilled and there are degrees of freedom available to estimate "goodness-of-fit" indexes (GOFI). As all indicators were categorical rating scale items, we estimated all FP by applying the DWLS estimator. As "target values" for GOFI refer to simulation studies using maximum likelihood (ML) estimation, and we want to compare our GOFI up against these target values, we estimated the model in Fig. 2 by also applying ML estimator. GOFI depict the degree to which the model implied variance-covariance

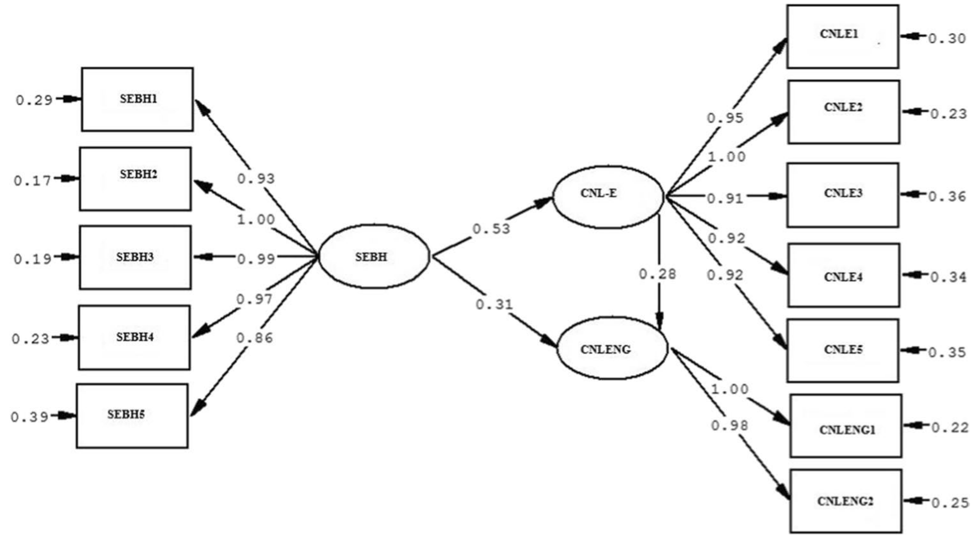

Fig. 2 Structural equation model linking CNL and SEBH using DWLS estimation

matrix ($\boldsymbol{\Sigma}$), which is based on our model in Fig. 2, is able to re-create the "actual" or observed variance-covariance matrix ($\boldsymbol{S}$), which we estimate from the observed empirical data. The smaller the elements in the residual matrix are ($\boldsymbol{R} = \boldsymbol{S} - \boldsymbol{\Sigma}$), the better the $\boldsymbol{\Sigma}$ re-creates the $\boldsymbol{S}$, the more the GOFI estimates approach their target values, and hence the better the "fit". There are three categories of GOFI namely, absolute, parsimony-adjusted and comparative GOFI.

Examples of absolute GOFI considered in the present study include the Satorra-Bentler (SB) scaled chi-square test (χ^2), which is robust to non-normality (we do not assume that categorical indicators based on rating scales are normally distributed), the reduced chi-square (χ^2/df), and the standardized root-mean-square residual (SRMR). SB (χ^2) values > 0.05 point to a good model fit; (χ^2/df) values < 3 suggest good fit; SRMR values < 0.05 suggest a well-fitting model, and value of 0.00 indicates 'perfect' fit; however, values as high as 0.08 also point to 'acceptable' model fit [21]. As other GOFI are derived using the chi-squared test, they also become robust to non-normality. Although SB chi-square values are not MLE-based, they may be comparable to the target values based on MLE [22].

The parsimony-adjusted GOFI considered in the present study is the root mean square error of approximation (RMSEA) with its associated "close" fit (Cfit) value at a 90% confidence interval. RMSEA values < 0.06 suggest good fit, values 0.08–0.10 suggest mediocre fit, while values > 0.10 point to poor fit; C-Fit values < 0.05 suggest acceptable fit [14]. Comparative fit GOFI applied in the present study were the comparative fit index (CFI) and the non-normed fit index (NNFI)–also known as the Tucker-Lewis index (TLI), CFI, and NNFI values ≥ 0.95 suggest good fit [21].

In addition to assessing overall model fit, we evaluated the fit of the three measurement models for each factor (Fig. 1). To do this, we examined the factor loadings, each factor loading should exceed 0.70 i.e., $0.71^2 = 0.50$ meaning that the latent variable explains at least 50% of the common variance in each item or indicator. As all scales were validated using RM prior to the SEM analysis, we expect that all factor loadings exceed 0.70 by a good margin [19, 20]. We also evaluated the SEM model using local fit indices, where insignificant residual matrix (R) elements (values exceeding 2.56) may indicate substantial specification and prediction error.

Results

As all DWLS-based *standardized* factor loadings exceed 0.78, we may conclude that all items are valid indicators for their respective latent factors i.e., the respective latent factor "governs the responses to the items" and can explain more than 0.78^2 or at least 60% of the variance in the responses to the items. Further, this means that less than 40% of the variance in any of our indicators is "unique variance", that is, "specific variance" caused by latent traits not included in our model or "error variance" caused by measurement error. All DWLS-based unique variances are smaller than 0.40. The smaller the uniqueness, the more of the variance that is common with the other items (communality).

The standardized residual matrix indicated that all but four elements had z-values smaller than − 2.56 or z-values larger than + 2.56, suggesting that there are small differences between the elements of the sample variance-covariance matrix $\boldsymbol{S}$ and the model implied variance-covariance matrix $\boldsymbol{\Sigma}$. This indicates that the specified SEM-model describes the patterns in the observed data quite well. One may reduce the size of these standardized residuals by defining "correlated error terms", that is, allowing items' specific variances to correlate (i.e., we state that items have

Table 1 Parameters estimated in the hypothesized model in Fig. 1

Observed variable	Model identification	Model estimation			
		Unstandardized solution		Completely standardized solution	
	Free parameter	ML	DWLS	ML	DWLS
Factor loadings		estimate (SE)	estimate (SE)	estimate	estimate
CNLE1	1	.98 (.02)	.95 (.02)	.85	.84
CNLE2		*1.000	*1.000	.87	.88
CNLE3	2	.94 (.02)	.91 (.02)	.82	.80
CNLE4	3	.93 (.02)	.92 (.02)	.80	.81
CNLE5	4	.92 (.02)	.92 (.02)	.80	.81
CNLENG1		*1.000	*1.000	.88	.88
CNLENG2	5	.99 (.04)	.98 (.04)	.87	.87
SEBH1	6	.92 (.01)	.93 (.02)	.85	.84
SEBH2		*1.000	*1.000	.92	.91
SEBH3	7	.99 (.01)	.99 (.01)	.91	.90
SEBH4	8	.95 (.02)	.97 (.02)	.87	.88
SEBH5	9	.83 (.02)	.86 (.02)	.76	.78
Unique variances (sum of specific variance and error variance)					
CNLE1	10	.28 (.04)	.30 (.06)	.28	.30
CNLE2	11	.25 (.04)	.23 (.06)	.25	.23
CNLE3	12	.33 (.04)	.36 (.06)	.33	.36
CNLE4	13	.36 (.04)	.34 (.06)	.36	.34
CNLE5	14	.37 (.04)	.35 (.06)	.37	.35
CNLENG1	15	.23 (.04)	.22 (.06)	.23	.22
CNLENG2	16	.24 (.05)	.25 (.06)	.24	.25
SEBH1	17	.28 (.03)	.29 (.06)	.28	.29
SEBH2	18	.16 (.03)	.17 (.06)	.16	.17
SEBH3	19	.17 (.03)	.19 (.06)	.17	.19
SEBH4	20	.25 (.03)	.23 (.06)	.25	.23
SEBH5	21	.42 (.04)	.39 (.06)	.42	.39
Latent factor associations (structural coefficients)					
SEBH-CNL-E	22	.52 (.03)	.53 (.03)	.54	.55
SEBH-CNLENG	23	.31 (.03)	.31 (.03)	.32	.32
CNL-E-CNLENG	24	.29 (.04)	.28 (.04)	.28	.28
Prediction residual of latent dependent factors					
CNL-E	25	53 (.03)	.54 (.03)	.70	.70
CNLENG	26	.56 (.03)	.56 (.03)	.72	.72
Variance of latent independent variable					
SEBH	27	.84 (.02)	.83 (.02)	1.000	1.000

DWLS = diagonally weighted least squares, *ML* = robust maximum likelihood estimation

*The factor loading of the variable is fixed to 1 on the independent latent variable. All other observed variables for that latent variable are interpreted in relation to the unit of measurement for this *reference variable*

"common variance" that refer to factors or constructs not being part of our model). However, such post hoc model modifications may be sample-dependent due to some bias in the specific sample.

Evaluating the DWLS-based standardized structural coefficients, we found that SEBH acts as a substantial "predictor" of students' CNL-E (standardized total effect = standardized direct effect = .552) and of students' CNLEng (standardized total effect = standardized direct effect + standardized indirect effect = .319 + (.552 × .281) = .319 + .155 = .475). Table 2 reports the GOFI for the models depicted in Figs. 1 and 2 based on ML estimation.

An inspection of the GOFI between models depicted in Figs. 1 and 2 shows that the model arising from specification

Table 2 Model evaluation by goodness-of-fit (GOF) indexes based on ML estimation

Model	Absolute GOF			Parsimony-adjusted GOF		Incremental GOF	
	SB-scaled χ^2	Reduced chi-square χ^2/df	SRMR	RMSEA (90% CI)	Cfit	CFI	NNFI
Model in Fig. 1 (df = 51, N = 1453)	164.543 $\boldsymbol{p = 0.000}$	**3.226**	0.025	0.067 (**0.061;** 0.073)	**0.000**	0.991	0.989
Model in Fig. 2 (df = 51, N = 1453)	158.765 $\boldsymbol{p = 0.000}$	**3.113**	0.027	0.065 (0.059; 0.071)	**0.000**	0.977	0.970
Target value	p > .05	< 3	< .05	< .06 (< .05; < .08)	> .05	> .95	> .95

SRMR = standardized root mean square residual, *Cfit* = closeness of fit, *CFI* = comparative fit index, *NNFI* = non-normed fit index, df = degrees of freedom, N = effective sample size, defined as the number of cases with responses on all 12 items/indicators

Model-fit values in bold deviate from the target values in the literature

in which the latent variables were associated based on theory (in Fig. 2) had better fit than the model in which the latent variables were free to covary (Fig. 1). Therefore, I conclude that the specified SEM model depicted in Fig. 2 sufficiently describes the observed structure of the sample data.

Discussion of findings

Empirical findings supported the hypothesis that self-efficacy in the science subject 'body and health' (SEBH) was associated with the two aspects of critical nutrition literacy (CNLEng, CNL-E). This significant positive association is similar to findings from a study conducted on young adolescents in Norway in which students that expected to perform well on the science test reported higher levels of engagement in dietary behaviours than their counterparts [8].

Similarly, consumer research shows that for individuals who are concerned about their health, the extent to which they engage in actions that promote their health depends on their 'nutrition self-efficacy' [14]. 'Nutrition self-efficacy' refers to a person's belief in his or her ability to overcome the barriers that are associated with healthy eating and is often associated with healthy dietary behaviour [14].

The extent to which young adolescents undertake positive dietary behaviours depends on their perceptions of competency to accomplish the task (self-efficacy) and understanding of the information relating to the task. Similarly, Mai and Hoffmann [14] suggest that self-efficacy which influences the extent of elaboration in information processing, determines food decision strategies. While there is no obvious directional association, findings in the present study support this notion, as shown by the stronger relationship between self-efficacy in 'Body and Health' (SEBH) and critical 'evaluation of nutrition information' (CNL-E) in comparison to that between SEBH and CNLEng, and CNL-E and CNLEng. It is for this reason that studies exploring the level of engagement in positive practices, such as using nutrition labels during shopping, suggest a two-tiered approach to increasing adolescents' use of nutrition labels: through enhancing adolescents' confidence in understanding nutrition labels and simplifying the information on the nutrition labels [23, 24].

Compared to self-efficacy, there are fewer instruments for measuring CNL; the present study showed that health-related self-efficacy in a science subject topic is related to CNL. Thus, in the absence of instruments specifically measuring CNL, it may be possible to use existing measures of self-efficacy during screening to forecast the adolescents' possible outcomes of nutrition interventions and improve the efficacy of nutrition interventions targeting adolescents.

Findings from the present study suggest that the extent to which adolescents are engaged in participating in dietary-related practices (CNLEng) may influence their food consumption decisions such as the use of available information and knowledge for the development of the skills required to execute positive dietary practices. This result finds support in a previous study in which children that closely participated in practical food preparation reported an increase in the consumption of vegetables, an example of positive dietary practice [23].

Whereas the present study showed a significant direct effect of SEBH on the two aspects of CNL, this result differs from previous studies in which engagement in household food tasks contributed to increased self-efficacy [23]. They argue that perceived self-efficacy is greater when individuals have practical experience with the necessary skills for completion. This exhibits the interconnected nature of psychosocial attributes and the skills associated with the critical domain of nutrition literacy, a notion that is consistent with Nutbeam's description of the skills associated with the critical level of health literacy namely higher-level cognitive and interactive social skills [5]. Therefore, when planning for and evaluating the outcome of health or nutrition programs, it will be beneficial to consider psychosocial attributes such as self-efficacy in related disciplines. In addition, developing nutrition-related science topics such as 'body and health' could benefit from taking into consideration how students understand the information therein, and what this could mean for their application of this knowledge in their daily life.

Conclusion

Evidence presented in this paper highlights the need to incorporate self-efficacy interventions in nutrition-related interventions targeting adolescents.

Implications and contributions

This study gives insight into the relations of psychosocial attributes (self-efficacy) and critical nutrition literacy in adolescents. These findings are particularly important for informing policy makers on how to develop tailored and targeted nutrition information, for adolescents' health and nutrition-related curricula and interventions addressing critical nutrition literacy needs of adolescents within the larger scope of media use and educational settings.

Abbreviations CNL-E: Critical Evaluation of Nutrition information; CNLEng: Engagement in dietary behaviour; DWLS: Diagonally weighted least squares; GOFI: Goodness-Of-Fit Indexes; CTT : Classical test theory; CNL: Critical nutrition literacy; DV: Distinct values; FP: Free parameters; FNL: Functional nutrition literacy; HL: Health literacy; INL: Interactive nutrition literacy; ML: Maximum likelihood; NNFI: Non-normed fit index; NL: Nutrition literacy; RM: Rasch modelling; RMSEA: Root mean square error of approximation; SB: Satorra-Bentler; SEBH: Self-efficacy in 'body and health'; SEM: Structural equation modelling; TLI: Tucker-Lewis Index

Acknowledgements The authors thank the students that participated and the school principals and teachers who facilitated the data collection. We would also like to thank Lena Victoria Nordheim for contributing to the development of the CNL-E scale items.

Authors' contributions Desire Alice Naigaga conducted the statistical analysis and drafted the manuscript. Kjell Sverre Pettersen offered guidance on the contents of the manuscript and read through the manuscript. Sigrun Henjum read through and contributed towards the manuscript. Kjell Sverre Pettersen developed the items that were included in the scales applied in the study. Øystein Guttersrud participated in the collection of data, statistical analysis, and drafting of the manuscript. All authors read and approved the final manuscript.

Funding Open access funding provided by University of Oslo (incl Oslo University Hospital). The study did not receive any special grant/funding.

Availability of data and materials The data supporting the findings of this study is available from the Norwegian Directory for Education but restrictions apply to the availability of this data, and so are not publicly available. Data is however available from the authors upon reasonable request and with permission from the Norwegian Directory for Education.

Code availability Not applicable

Declarations

Ethics approval This study was approved by the Norwegian Centre for Research (NSD).

Consent to participate All data were anonymized by assigning code values for the name of the school and respondent. The students were also informed that the data provided could be used for further research.

Consent for publication: All authors have consented to the publication of the manuscript.

Competing interests All authors declare that they have no competing interests.

References

1. Nutbeam D. The evolving concept of health literacy. Soc Sci Med. 2008;67:2072–8.
2. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, Brand H. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health. 2012;12:80. https://doi.org/10.1186/1471-2458-12-80.
3. Kickbusch I. Health literacy: engaging in a political debate. Int J Public Health. 2009;54:131–2. https://doi.org/10.1007/s00038-009-7073-1.
4. Sillk KJ, Sherry J, Winn B, Keesecker N, Horodynski MA, Sayir A. Increasing nutrition literacy: testing the effectiveness of print, web site, and game modalities. JNEB. 2008;40:3–10. https://doi.org/10.1016/j.jneb.2007.08.012.
5. Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. Health Promot. 2000;15:259–67. https://doi.org/10.1093/heapro/15.3.259.
6. Velardo S. The nuances of health literacy, nutrition literacy, and food literacy. JNEB. 2015;47:385-9.e1. https://doi.org/10.1016/j.jneb.2015.04.328.
7. Guttersrud Ø, Dalane JØ, Pettersen S. Improving measurement in nutrition literacy research using Rasch modelling: examining construct validity of stage-specific 'critical nutrition literacy' scales. Public Health Nutr. 2014;17:877–83. https://doi.org/10.1017/s1368980013000530.
8. Guttersrud Ø, Pettersen KS. 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. Public Health Nutr. 2015;18:2565–74. https://doi.org/10.1017/S1368980014003152.
9. Dong Z. How to persuade adolescents to use nutrition labels: effects of health consciousness, argument quality, and source credibility. Asian J Commun. 2015;25:84–101. https://doi.org/10.1080/01292986.2014.989241.
10. Talagala IA, Arambepola C. Use of food labels by adolescents to make healthier choices on snacks: a cross-sectional study from Sri Lanka. BMC Public Health. 2016;16:739. https://doi.org/10.1186/s12889-016-3422-1.
11. Velardo S, Drummond M. Emphasizing the child in child health literacy research. J Child Health Care. 2016;1:5–13. https://doi.org/10.1177/1367493516643423.
12. Saha S, Vemula SR, Mendu VVR, Gavaravarapu SM. Knowledge and practices of using food label information among adolescents attending schools in Kolkata, India. JNEB. 2013;45:773–9. https://doi.org/10.1016/j.jneb.2013.07.011.
13. Wojcicki JM, Heyman MB. Adolescent nutritional awareness and use of food labels: Results from the national nutrition health and examination survey. BMC Pediatr. 2012;12:55. https://doi.org/10.1186/1471-2431-12-55.
14. Mai R, Hoffmann S. Taste lovers versus nutrition fact seekers: how health consciousness and self-efficacy determine the way consumers choose food products. J Cust Behav. 2012;11:316–28. https://doi.org/10.1002/cb.1390.
15. Krause C, Sommerhalder K, Beer-Borst S. Nutrition-specific health literacy: development and testing of a multi-dimensional questionnaire. Ernahrungs Umschau. 2016;63:214–20. https://doi.org/10.4455/eu.2016.046.
16. Cappelleri JC, Lundy JJ, Hays RD. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. Clin Ther. 2014;36:648–62. https://doi.org/10.1016/j.clinthera.2014.04.006.
17. Bandura A. Human agency in social cognitive theory. Am Psychol. 1989;44(9):1175–84. https://doi.org/10.1037/0003-066X.44.9.1175.
18. U.S. national Library of Medicine. Joint Collection Development Policy: Human Nutrition and Food. 1998. Available on https://wayback.archive-it.org/org-350/20190227225406/. https://www.nlm.nih.gov/pubs/cd_hum.nut.html. Accessed 15 Sept 2020.
19. Naigaga DA, Pettersen KS, Henjum S, Guttersrud Ø. Assessing adolescents' perceived proficiency in critically evaluating nutrition information. IJBNPA. 2018;15:1–13. https://doi.org/10.1186/s12966-018-0690-4.
20. Naigaga DA, Pettersen KS, Henjum S, Guttersrud Ø. Assessing adolescent self-efficacy in Body and Health - Exploring the psychometric properties of the SEBH scale. NORDINA. 2019;15:145–58. https://doi.org/10.5617/nordina.5913.
21. Hu L, Bentler PM. Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Modeling. 1999;6:1–55. https://doi.org/10.1080/10705519909540118.
22. Bryant FB, Satorra A. Principles and practice of scaled difference chi-square testing. Struct Equ Model. 2012;19(3):372–98.
23. Caraher M, Seeley A, Wu M, Lloyd S. When chefs adopt a school? An evaluation of a cooking intervention in English primary schools. Appetite. 2013;62:50–9. https://doi.org/10.1016/j.appet.2012.11.007.
24. Chu YL, Farmer A, Fung C, Kuhle S, Storey KE, Veugelers PJ. Involvement in home meal preparation is associated with food preference and self-efficacy among Canadian children. Public Health Nutr. 2013;16:108–12. https://doi.org/10.1017/S1368980012001218.

POSTADRESSE:
OsloMet – storbyuniversitetet
Pilestredet 46
Postboks 4, St. Olavs Plass
0130 Oslo

OsloMet Avhandling 2021 nr 43

ISSN 2535-471X
ISBN 978-82-8364-347-3