JYU DISSERTATIONS 223

Olli Paakkari

Developing an Instrument for Measuring Health Literacy Among School-Aged Children

UNIVERSITY OF JYVÄSKYLÄ
FACULTY OF SPORT AND
HEALTH SCIENCES

JYU DISSERTATIONS 223

Olli Paakkari

Developing an Instrument for Measuring Health Literacy among School-aged Children

Esitetään Jyväskylän yliopiston liikuntatieteellisen tiedekunnan suostumuksella
julkisesti tarkastettavaksi toukokuun 27. päivänä 2020 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Sport and Health Sciences of the University of Jyväskylä,
on May 27, 2020 at 12 o'clock noon.

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2020

Editors
Anne Viljanen
Faculty of Sport and Health Sciences, University of Jyväskylä
Timo Hautala
Open Science Centre, University of Jyväskylä

Permanent link to this publication: http://urn.fi/URN:ISBN:978-951-39-8171-6

ISBN 978-951-39-8171-6 (PDF)
URN:ISBN:978-951-39-8171-6
ISSN 2489-9003

ABSTRACT

Paakkari, Olli
Developing an Instrument for Measuring Health Literacy among School-aged Children
Jyväskylä: University of Jyväskylä, 2020, 59 p.
JYU Dissertations,
ISSN 2489-9003; 223
ISBN 978-951-39-8171-6

Health literacy (HL) is an important determinant of health and health behaviours. The aim of this research was to develop a brief, comprehensive, and theory-based instrument for the measurement of subjective HL among school-aged children. A further aim was to examine the cross-national measurement invariance of this instrument in four European countries, i.e. Finland, Poland, Slovakia, and Belgium.

The developmental process of the instrument was systematic, iterative, and validity- and reliability-driven. It involved 1) definition and elaboration of the construct, 2) choice of measurement method, 3) item generation, 4) a pilot study (N=401) including test-retest (N=117), and 5) field-testing, including construction of a brief instrument (N= 3853), and examination of the instrument's international applicability (N=1468). All the samples were taken from pupils aged 13 and 15.

Initially, 65 items were generated, of which 32 items were selected for the pilot study. After item reduction, the instrument contained 16 items. The test-retest showed high stability. For the field testing phase, a brief, 10-item instrument was constructed, referred to as *Health Literacy for School-aged Children* (HLSAC). The instrument, which exhibited a high Cronbach alpha (.93), and adequate fit with the data, included two items from each of five predetermined theoretical components (theoretical knowledge, practical knowledge, individual critical thinking, self-awareness, citizenship). In a cross-national measurement invariance examination, configural and metric invariance was established, but scalar invariance did not hold. Nevertheless, the HLSAC instrument's internal consistency was high (α=.80–.90) and the fit with the data was adequate in each country. A comparison of health literacy mean values showed significant mean value differences between countries and age groups.

Measurement of HL via reliable and valid instruments creates a basis for trustworthy assessment of HL levels and appropriate interventions. The HLSAC instrument is suitable for large-scale studies, and for use with children and adolescents. It constitutes a promising tool for subjective HL comparisons in the international context.

Keywords: health literacy, measurement, instrument, school-aged children, adolescent, health education

TIIVISTELMÄ (ABSTRACT IN FINNISH)

Paakkari, Olli
Mittarin kehittäminen lasten ja nuorten terveydenlukutaidon mittaamiseksi
Jyväskylä: Jyväskylän yliopisto, 2020, 59 s.
JYU Dissertations,
ISSN 2489-9003; 223
ISBN 978-951-39-8171-6

Terveyden lukutaito on tärkeä determinantti terveydelle ja terveyskäyttäytymiselle. Tämän tutkimuksen tarkoituksena oli kehittää lyhyt, moniulotteinen ja teoriaan perustuva mittari kouluikäisten lasten terveyden lukutaidon mittaamiseen. Lisäksi tarkoituksena oli tutkia mittausinvarianssitestin avulla mittarin soveltuvuutta maiden väliseen vertailuun. Testi tehtiin neljässä Euroopan maassa (Suomi, Puola, Slovakia, Belgia).

Mittarin kehittämisessä edettiin systemaattisesti ja huomioiden pätevyyteen ja luotettavuuteen liittyvät kysymykset. Kehitysprosessi sisälsi 1) terveyden lukutaidon käsitteellistämisen, 2) mittausmenetelmän valinnan, 3) väittämien luomisen, 4) pilottitutkimuksen (N=401) sisältäen toistomittauksen (N=117) sekä 5) mittarin viimeistelyn (N=3853) ja sen kansainvälisen toimivuuden tutkimisen (N=1468). Jokaisessa vaiheessa otos koostui 13- ja 15-vuotiaista oppilaista.

Aluksi kehitettiin 65 väittämää, joista 32 valittiin pilottitutkimukseen. Analyysien perusteella väittämien määrä supistettiin 16:een. Pilottivaiheen toistomittaus osoitti korkeaa pysyvyyttä. Tämän jälkeen rakennettiin lyhyt 10 väittämää sisältävä mittari, jonka nimeksi tuli *Health Literacy for School-aged Children* (HLSAC). Mittarin sisäinen konsistenssi oli korkea (α=.93), malli sopi dataan asianmukaisesti ja se sisälsi kaksi väittämää seuraavista terveyden lukutaidon osa-alueista: teoreettiset tiedot, käytännön taidot, kriittinen ajattelu, itsetuntemus ja eettinen vastuullisuus. Neljän maan välisessä mittausinvarianssitestissä konfiguraalinen ja metrinen mittausinvarianssi todennettiin, mutta skalaarinen mittausinvarianssi ei toteutunut. Mittarin sisäinen konsistenssi oli korkea (α=.80–.90) ja malli sopi dataan asianmukaisesti jokaisessa maassa. Terveydenlukutaitoa osoittavien keskiarvojen vertailu maiden välillä osoitti merkittäviä keskiarvoeroja maiden ja ikäryhmien välillä.

Pätevät ja luotettavat terveyden lukutaidon mittausmenetelmät luovat pohjan terveyden lukutaidon tasojen arvioinnille ja asianmukaisille interventioille. HLSAC mittari soveltuu laajamittaisiin tutkimuksiin lapsilla ja nuorilla, ja se on lupaava väline koetun terveydenlukutaidon vertailuun eri maiden välillä.

Asiasanat: terveyden lukutaito, mittaaminen, mittari, lapsi, nuori, terveyskasvatus

Author Olli Paakkari, LicSc
Faculty of Sport and Health Sciences
Research Centre for Health Promotion
University of Jyväskylä
Finland
olli.paakkari@jyu.fi

Supervisors Associate Professor Minna Torppa, PhD
Faculty of Education and Psychology
Department of Teacher Education
University of Jyväskylä
Finland

Professor Lasse Kannas, PhD
Faculty of Sport and Health Sciences
Research Centre for Health Promotion
University of Jyväskylä
Finland

Reviewers Professor Torbjørn Torsheim, PhD
Faculty of Psychology
Department of Psychosocial Science
University of Bergen
Norway

Associate Professor Unni Karin Moksnes, PhD
Faculty of Medicine and Health Sciences
Department of Public Health and Nursing
Norwegian University of Science and Technology
Norway

Opponent Professor Anna-Maija Pietilä, PhD
Faculty of Health Sciences
Department of Nursing Science
University of Eastern Finland
Finland

ACKNOWLEDGEMENTS

The starting point for this thesis – which went unrecognized at the time – arrived during 2011, when I and a colleague published an article conceptualizing health literacy as a learning outcome in schools. The conceptualization was later adopted as a theoretical basis for health education in Finland. At that time, I did not know that this theoretical academic paper and the discussions surrounding it would lead to my doctoral dissertation a good many years later. After publication of the theoretical framework for health literacy, our research group became more interested in the measurement of health literacy. At that time, there were indeed a number of tools to measure the health literacy of children and adolescents, but there was no comprehensive and generic instrument for this purpose. Development of a measurement instrument thus seemed like a worthwhile idea. Luckily, at that point, I didn't know how much work, energy, and time it would take to get this done. Yet now I can say with conviction that it has been a privilege to be a part of this national and international development work. I would like to thank the many people who have shared this journey with me, and who have helped and supported my research work in different ways.

First, I would like to express my sincere appreciation to my supervisors, Minna Torppa and Lasse Kannas. Minna, without your statistical guidance and contribution this work would be something utterly different. You were always well prepared for meetings, and spent your precious time discussing the issues I had in mind. I have learned from you that there is room for subjectivity and interpretation in statistics. Lasse, you have guided me towards the secrets of doing research with the assurance of an elder statesman, and have pushed me forward with precisely judged timetables and demands – because you have an instinct for such things. At every turn you have trusted me, and given me freedom to work the way I want. All in all, I do not know how to thank both of you as you deserve.

I would like to thank the two reviewers of my dissertation, Unni Karin Moksnes and Torbjørn Torsheim. You used your time to give carefully prepared and constructive feedback, which challenged my thinking and helped me to develop this thesis to the next level.

The collection of international data and the writing of a cross-country validation article would not have been possible without help. Many thanks to Zuzana Boberova, Gunter Maier, Joanna Mazur, and Laura Goeint for all of your input.

I would like to express my gratitude to Donald Adamson. You have proofread all the research articles and this thesis very diligently. With your valuable comments the language of the papers has improved significantly, and your questions tested my argumentation and thinking as a researcher.

I wish to thank Leena Paakkari and Jari Villberg. At the start of the thesis process, I had a totally different research theme in my mind – but Leena, you turned my ideas round. Your passion for doing research is impressive, and your constructive ideas and valuable contributions as a co-writer of articles have clarified many thoughts. Jari, you were always ready to discuss statistical subtleties and to give rapid responses to my questions. Every researcher should have such a room-mate.

Special thanks to my closest workmates, Raili Välimaa, Jorma Tynjälä, Kristiina Ojala, and Maija Kaisla, and to other colleagues in the Research Centre for Health Promotion. With you, I have spent many hours of academic, and most importantly, not-so-academic life during the working days. I appreciate the work atmosphere that allows everyone to be what they are. In addition, you have created good grant applications that obliged me to move out of my normal duties and to focus on writing the research articles.

I would like also to offer my thanks to the faculty clerical workers, to the editors, and to other personnel for assisting with completion of this thesis. Furthermore, I would like to express my gratitude to the National Institute for Health and Welfare, the Juho Vainio Foundation, and the University of Jyväskylä for their financial support.

Finally, I would like to dedicate this dissertation to my ‘past’ and my ‘future’. To my parents, Ossi (RIP) and Tuula – you have always let me to choose the direction I want to follow, and you have supported me in these choices. To my dear daughters, Oona and Eevi – I hope that I’ll be able to encourage and guide you, so that you always pursue your dreams fearlessly.

FIGURES

TABLES

LIST OF ORIGINAL PUBLICATIONS

1. Paakkari, O., Torppa, M., Kannas, L. & Paakkari, L. 2016. Subjective health literacy: Development of a brief instrument for school-aged children. Scandinavian Journal of Public Health, 44 (8), 751–757.

2. Paakkari, O., Torppa, M., Villberg, J., Kannas, L. & Paakkari, L. 2018. Subjective health literacy among school-aged children. Health Education, 118 (2), 182–195.

3. Paakkari, O., Torppa, M., Boberova, Z., Välimaa, R., Maier, G., Mazur, J., Kannas, L. & Paakkari, L. 2019. The cross-national measurement invariance of the health literacy for school-aged children (HLSAC) instrument. European Journal of Public Health, 29 (3), 432–436.

In all original publications, Olli Paakkari as the first author had the main responsibility of all phases. The statistical analyses were accomplished partly together with Minna Torppa and Jari Villberg, and partly independently. The author conducted the original draft preparation for the publications and submissions of the articles. All authors performed the review and editing of publications.

ABBREVIATIONS

ANOVA	Analysis of variance
CFA	Confirmatory factor analysis
CFI	Comparative Fit Index
HBSC	Health Behaviour in School-aged Children study
HL	Health literacy
HLQ	Health Literacy Questionnaire
HLSAC	Health Literacy for School-aged Children
HLS-EU-Q	European Health Literacy Survey Questionnaire
RMSEA	Root Mean Square Error of Approximation
SEM	Structural equation modelling
SRMR	Standardized Root Mean square Residual
TLI	Tucker-Lewis Index

CONTENTS

1 INTRODUCTION

During childhood and adolescence, many cognitive, emotional, and physical development processes take place, in parallel with changes in social relationships and interactions. These years are important for healthy development, and it is during them that the foundation for health literacy (HL) is laid. HL – which is the competence to obtain and construct health information, make sound health decisions, and change the factors that constitute one's own health and that of others – provides a basis for people's overall health behaviours and health. From a wider perspective, children's and adolescents' health are important not just for the individual in the present time, but also for the future wealth and wellbeing of society.

During the last few decades, health policies have tended to focus on the responsibility of the individual in health issues rather than that of society more broadly (Selkälä 2013, 145-146; Traina, Martinussen & Feiring 2019). If young citizens are to cope now and in the future in this kind of health policy climate, they will need to gain more competence in health-related matters. However, it appears that health policies and the need for individual responsibility are to some degree in conflict with the level of health expertise among the general population. The World Health Organization even speaks of an HL crisis in Europe and beyond, referring to the inconsistency between the complex requirements posed in taking care of one´s health and the competence capital of individuals (Kickbusch et al. 2013).

With such a background, the improvement of HL among the population has become one of the key goals of public health and health promotion worldwide (Nutbeam 2000; Tassi 2004; Wharf-Higgins 2012). In addition, the Organisation for Economic Co-operation and Development (OECD 2018) has highlighted HL as an essential element in a solid foundation that will advance the agency of children and adolescents, enabling them to navigate through a complex and uncertain world.

Improving the HL of individuals and of populations is critical to achieving health equity, and to tackling global health challenges, on the basis that HL may help to reduce health inequalities both between and within population groups

(Whitehead 1991; Kickbusch, Wait & Maag 2006; Batterham et al. 2016; van der Heide et al. 2016). The health behaviour adopted in the course of adolescence and early life circumstances can partly explain existing health inequalities in adulthood (Inchley et al. 2016, 5). Moreover, from the perspective of equality, the development of HL can be seen as a moral act, and therefore as something valuable in itself (Paakkari & George 2018).

One of the central arenas for promoting children's and adolescents' HL is the *education system* (Nutbeam 2000; St Leger & Nutbeam 2000; Begoray, Wharf-Higgins & MacDonald 2009; Paakkari & Paakkari 2012). School is a very important context for many reasons. School reaches most of the population within a certain phase of life, and in their juvenile years pupils spend more waking hours at school than in any other venue; they are thus significantly affected by school as an institution. Some school-aged children have cognitive, social, or emotional challenges that impact on their health, indicating that school is a suitable setting for reaching pupils with special requirements, and for developing their health-related ability. Hence, school has both the purpose and the potential to equalize differences between school-aged children from diverse backgrounds (UNESCO 2014). Studies have shown that if it is carefully designed and implemented, school-based health education can substantially improve pupils´ health related academic achievement, as well as their health skills (Flay, Allred & Ordway 2001; Li et al. 2011).

Low HL has consequences at both individual and societal level. In general, low HL has been recognized as an independent risk factor (Volandes & Paasch-Orlow 2007) and constitutive determinant for health (Berkman et al. 2011; van der Heide et al. 2016). Low HL has been shown to be associated with a lack of health knowledge, and further, with problems in interpreting health-related information, medication treatment errors, an increased use of medical and hospital services, and a decreased use of preventive health services. All of these factors lead to higher healthcare costs for society (Howard, Gazmararian & Parker 2005; Berkman et al. 2011). Studies have also reported associations between low HL and poor health status, mortality, risky health behaviours, and poor perceived health (Berkman et al. 2011). Among school-aged children, HL has been found to be an independent factor explaining health disparities, with a higher level of HL being related to more positive health outcomes (Paakkari et al. 2019b).

Low HL has become a global challenge, and it has attracted considerable attention in international research, practice, and policy-making. Generally speaking, two different approaches to HL can be identified. In the medical or healthcare setting, the perspective on HL is more risk-oriented, meaning that it focuses on low HL, viewed as a risk factor for poor health, and for poor compliance with advice on healthcare. On the other hand, in the field of public health and health promotion, HL is viewed in more positive way; thus it is seen as a personal asset that can offer greater control and autonomy over decision-making regarding health issues, with possibilities to increase the individual's empowerment (Nutbeam 2008; Pleasant & Kuruvilla 2008; Van den Broucke

2014). The present thesis can be seen as set within the public health and health promotion research tradition. Overall, health promotion can be defined as a comprehensive social and political process that enables people to increase control over, and to improve their own health (Ottawa Charter for Health Promotion 1986). It covers interventions designed to strengthen the health-related skills and capabilities of individuals, and further, actions to change social, environmental, and economic conditions in ways to support both individual and public health (Nutbeam 1998). Public health is the science and art of improving and protecting the health of entire populations and of prolonging life, throughout health promotion, disease prevention, and other organized endeavours of society (Nutbeam 1998).

The aim of the research reported in this thesis was to develop a brief, comprehensive, and theory-based instrument to measure school-aged children's HL. In fact, there are a number of HL measurement instruments for children and adolescents already available (Guo et al. 2018; Okan et al. 2018). However, there has been a lack of brief, comprehensive, generic (not focused merely on specific health topics), internationally comparable, and self-administered instruments, tested on a proper target group (school-aged children). From the perspective of large-scale surveys, where the purpose is to measure other phenomena in conjunction with HL, there is a need for a suitable instrument for children and adolescents. Such a measure will make it possible to explore the levels of HL of entire age groups – constituting an essential starting point for monitoring HL trends, designing effective interventions, and informing policy makers. Furthermore, cross-national research will allow examination of the instrument's applicability in different countries, permitting comparisons of HL levels.

The HL instrument developed in the research for this thesis has formed part of the international *Health Behaviour in School-aged Children* (HBSC) survey, which is a highly-valued and well-known health policy tool of the World Health Organization (Currie et al. 2009). In addition, the fact that in Finland HL has been adopted into the school curriculum (Finnish National Board of Education 2014; 2015), has stimulated further development work, deriving from the instrument. By means of this instrument it will be possible to assess school-aged children's perceived HL, and to examine the association of these results with health education learning outcomes.

2 HEALTH LITERACY, AND MEANS OF MEASURING IT

2.1 Health literacy as a concept

Health Literacy (HL) as a term was introduced in the mid 1970s, when Simonds (1974) suggested policy goals for health education. In the following 20 years the term was rarely used, but since the 1990s there has been expanding interest in HL. Indeed, during the last two decades there has been ongoing discussion on HL, with two main fields emerging with reference to the definition and use of HL. The development of the HL concept has evolved within medical and healthcare settings, and also in the field of public health and health promotion (Sørensen et al. 2012; Okan 2019). Regarding the medical or healthcare context, the HL assessment focus has been on the basic skills of reading, writing, and numeracy.This aspect of HL, which is relatively narrow in scope, is referred to as *functional HL* (Parker et al. 1995). It encompasses basic skills that are essential for individuals to operate within a healthcare system (Williams et al. 1995; Ratzan & Parker 2000). The significance of functional HL for the individual's health remains widely recognized, and research within this area continues.

Nevertheless, there is increased interest in searching for a broader construct of HL, i.e. one that does not limit its application purely to healthcare. Modern society demands a broad range of competences, if one accepts the aim that citizens should be ready to take care of and sustain health − their own and that of the community (Nutbeam 1998; Sørensen et al. 2012). In the field of public health and health promotion, HL was outlined in a more general sense at an early stage, and there is still a wide understanding of HL as involving a readiness to participate in social debate, with a view to promoting the health of the community (Chinn 2011; de Leeuw 2012; Sykes et al. 2013). The field of health promotion encompasses wide perspectives, including individuals' possibilities to live in a healthy way, and to have satisfactory living conditions (Abel 2008). It has been emphasized that individuals should be equipped with the knowledge and skills that help them to modify conditions affecting their health chances (Abel 2007). One can easily see that such a perspective must involve consideration of the need to construct an HL assessment tool.

As the premises behind HL vary according to the fields mentioned above, it is understandable that the definitions vary from fairly narrow to more broadly-focused definitions. Within reviews, numerous definitions of HL have been used (Sørensen et al. 2012; Malloy-Weir et al. 2016; Bröder et al. 2017; Sørensen & Pleasant 2017). The most commonly used definition of HL is that of Ratzan and Parker (2000), who refer to "the degree to which individuals have the capacity to obtain, process, and understand basic health information and the services needed to make appropriate health decisions" (Malloy-Weir et al. 2016). Another widely-used HL definition is that of the World Health Organization (1998), which refers to "the cognitive and social skills which determine the motivation and ability of individuals to gain access to, understand, and use information in ways which promote and maintain good health". Several definitions of children's and adolescents' HL also exist (Bröder et al. 2017). However, the concept of HL has been criticized on the grounds of inconsistent conceptualization and variable definition (Guzys et al. 2015). The lack of consensus in relation to conceptual dimensions, HL definitions, and HL measures also makes it harder to compare studies (Sørensen et al. 2012).

The ability requirements of society, both now and in the future, bring their own demands for a suitable definition of HL. From this perspective, it has been argued that individuals should be equipped with citizenship skills, which include critical thinking, problem solving, accessing and analysing information, and collaboration and initiative (Wagner 2008). Moreover, it is important that individuals should gain the competence to reflect on health matters from their own perspective, while being able also to understand the perspectives of others (Abel 2007; Nutbeam 2008). Individuals shoud be able to observe and understand the conditions that determine health, and have knowledge of how to change them (Abel 2007). This kind of awareness can equip young citizens to take responsible actions to sustain and promote their own health and that of others.

If we aim to develop individuals' HL to a more advanced level, it is important to be clear about the kind of competence we intend to develop, and to describe explicitly the constituent parts of HL. This will enable us to plan purposeful learning experiences for improving school-aged childrens' HL. Moreover, if we consider HL to be a learning outcome, the appropriate description of HL and its components will make it easier to assess how learning goals have been achieved.

With these considerations in mind, in the research reported here, the HL concept developed by Paakkari and Paakari (2012) was applied, as follows:

> Health literacy comprises a broad range of knowledge and competencies that people seek to encompass, evaluate, construct, and use. Through health literacy competencies people become able to understand themselves, others

> and the world in a way that will enable them to make sound health decisions, and to work on and change the factors that constitute their own and others' health chances.

This definition expresses the notion that individuals should become literate in health issues concerning themselves – but also the broader context we are part of, in other words, the environment in which we and others are immersed. The definition in question involves five core components, namely *theoretical knowledge, practical knowledge, individual critical thinking, self-awareness* and *citizenship* (Paakkari & Paakkari 2012; Paakkari et al. 2016, see Table 1). By nature, HL is a multidimensional, complex, and holistic construct, with core components that are partly overlapping, meaning that the separation of components is somewhat artificial. In fact, the components can be seen as broader competence fields. These form an expanding entity, ranging from mere literacy on health topics towards literacy concerning oneself, others, and the world beyond.

Theoretical knowledge can be understood as basic knowledge on health-related principles, theories, and conceptual models. It creates a necessary basis for other core components of HL, and it deepens one's understanding of health issues; nevertheless, taken on its own it is seldom sufficient for the adoption of healthy habits or of health-promoting actions.

Practical knowledge, also referred as procedural knowledge or skills (Bereiter & Scardamalia 1993, 45), is the competence to put theoretical knowledge into practice. It includes basic health-related skills that the individual needs in order to be able to behave in a health-promoting way in daily situations.

Individual critical thinking, i.e. the competence to think clearly and rationally, enables people to deal with large amounts of information and to have power over that knowledge. It allows individual to understand health issues widely and deeply, and to recognize the complex and multidimensional nature of health. People need theoretical and practical knowledge to think critically. This knowledge helps, for example, in seeing all the significant aspects of certain phenomena, and it enables one to search for information from reliable sources.

Self-awareness – considered as a competence to reflect on oneself (e.g. on thoughts, needs, behaviours, attitudes, values) – allows one to attribute personal meaning to health issues. This consciousness helps one to examine and evaluate why individuals behave or think in a particular way, and how these ways affects one's health-related choices. In addition, self-awareness involves the individual ability to reflect on the self as a learner; this supports the creation and initiation of purposeful learning strategies (e.g. goal setting, monitoring progress, finding suitable learning habits, evaluating the achievement of goals).

Citizenship means the ability to think and act in an ethically responsible way. People should be aware of the rights and responsibilities they have. Citizenship highlights the point that people should consider health-related issues beyond their own perspective. It can encompass e.g. what might be done to improve other people's health, and the possible effects of one's thoughts and actions on other people, the environment, and society (Paakkari & Paakkari 2012).

TABLE 1 The five core components of health literacy

The core components of health literacy (Paakkari et al. 2016)
Theoretical knowledge of health issues encompasses a range of principles, theories, and conceptual models. Knowledge is viewed as something explicit, factual, universal, formal, and declarative. It includes lower levels of thinking skills, such as remembering.
Practical knowledge (i.e. procedural knowledge, skills) can be seen as a competency that allows one to put theoretical knowledge into practice. Whereas theoretical knowledge is something applicable to many different situations, practical knowledge can be seen as usable in specific contexts. It is partly rooted in the individual's experiences, and thus it includes tacit, intuitive, or implicit knowledge. Practical knowledge includes basic health skills such as the ability to find health information, the ability to seek health services, and the ability to give first aid.
Individual critical thinking can be understood as the ability to think clearly and rationally. It is based on having a curious and investigative attitude towards the world, and a desire to understand health issues in a deeper way. In practice, critical thinking includes higher-level thinking skills, such as an ability to analyse, evaluate, and create something new; this could include e.g. the ability to search for the logical connections between health ideas, to solve problems, to argue, to draw conclusions, or to assess the validity of health information.
Self-awareness is the ability to reflect on oneself and it make possible the personal contextualization of health issues. Through self-reflection, the individual becomes conscious of his/her own thoughts, feelings, needs, motives, values, attitudes, and experiences, and is able to consider how these relate to ways of behaving in an individually health-enhancing way. An important part of self-awareness is the ability to reflect on oneself as a learner.
Citizenship involves the ability to take social responsibility, and to think of the probable consequences of one's own actions on others. The ability to act in an ethically responsible way means that individuals are able to consider health issues beyond their own perspective: they may become aware of their own rights and responsibilities, and the effects people's actions or thoughts may have on other people or on the environment. The component further includes the ability to identify and work on factors that influence one's own and other people's possibilities to achieve or maintain good health.

2.2 Measurement of school-aged children´s health literacy

Just as the definitions of HL vary, so also there are differences in the measurement of HL. In fact, there is no consensus how HL should be measured (Kiechle et al. 2015). The measurement differences are related to the conceptualization and purpose of measurement, i.e. whether the measurement is based on a narrower or broader concept of HL, and whether the measurement is aimed at general HL or domain-specific HL. Examples of specific HL aspects or sub-domains include the various specific illness groups, age groups, mental health literacy, oral health literacy, nutrition health literacy, and – at a time of rapid technological change – digital health literacy, referred to also as eHealth literacy and media health literacy. Functional health literacy is measured by frequently-used tools that are narrow in scope, notably the *Test of Functional Health Literacy in Adults* (TOFHLA, Parker et al. 1995; the adolescent version TOFHLAd, Chisolm & Buchanan 2007), the *Rapid Estimate of Adult Literacy in Medicine* (REALM, Davis et al. 1991; the adolescent version REALM-Teen, Davis et al. 2006), and *Newest Vital Sign* (NVS, Weiss et al. 2005); these relate to participants' reading comprehension and numeracy in the context of health. The instruments in question have been developed to provide brief and rapid screening for HL in medical or healthcare settings. Alternatively, there are more comprehensive measurement tools, based on a broader concept of HL and thus able to take into account more dimensions of HL. These include e.g. the *Health literacy Assessment Scale for Adolescents* (HAS-A, Manganello et al. 2015) and *Health Literacy Measure for Adolescents* (HELMA, Ghanbari et al. 2016).

From recent reviews on HL assessment one can see that an expanding interest in monitoring HL levels has led to an increasing number of instruments whose objective is to measure school-aged children´s HL (Guo et al. 2018, see Appendix 1; Okan et al. 2018). Nevertheless, a closer look indicates that there is still a lack of instruments that are simultaneously comprehensive, generic, internationally comparable, self-administered, and (most important of all) validated with the target group in question (i.e. children and adolescents; see Perry 2014).

One key aspect in which HL measures differ is variation between subjective (i.e. self-reported and self-perceived HL) and performance-based measurements. Subjective measurements use self-reporting questionnaires, while performance-based measurements assess HL via performance in given tasks. Both methods contain pros and cons. It has been proposed that performance-based measurements should be prioritized in the development of HL measurements (McCormack et al. 2013). However, performance-based measurements involve a number of problems in measuring comprehensive and multidimensional HL – an aspect in which subjective measures have met with more success (Altin et al. 2014). It has been argued that performance-based measurements may also involve ethical concerns, on the grounds that the participants can experience embarrassment or shame if they have a low level of HL (Paasche-Orlov & Wolf 2007), while self-reported measures are more likely to preserve the respondent's dignity (Pleasant 2014). Instruments that measure self-reported HL can be more easily applied in large sample studies, and thus could provide a more effective means of examining HL at the population level (Kiechle et al. 2015). Self-reported measures have been seen as more time efficient, and also less resource intensive and expensive to administer than performance-based measures (Bowling 2005; Pleasant 2014).

Although the self-reporting approach has been used in many studies (Haun et al. 2014; Guo et al. 2018), a number of concerns remain. It has been argued that the self-reported answers of children and adolescents will tend to incur more measurement error than performance-based measures (Vaz et al. 2013). Nevertheless, this depends greatly on the comprehensibility of the instrument; if participants can understand the items, the response options, and what the HL instrument measures, they will be better able to self-assess their own HL accurately (Velardo & Drummond 2017). It is true that self-reported measurements may elicit overestimation, since participants may be drawn to give socially desirable responses (Altin et al. 2014; Paakkari et al. 2018). However, subjective measurements focused on individuals' perceived competence, i.e. their self-efficacy, and on the basis of over three decades of research, it has been shown that there is a clear link between self-efficacy and health behaviour (Conner & Norman 2005). As Bandura (2004) has argued, perceived self-efficacy (perceived competence) influences individuals' goal setting and aspirations, and makes them commit to the goals in question. These various considerations form a powerful argument for a self-reporting approach to HL measurement in large-scale surveys.

To sum up, it is essential to select a measure that suits the purpose and context, with consideration given to whether that purpose is in line with a narrower or a broader notion of HL, or favours a subjective or a performance-based measurement. The evaluation of the measure will be made accordingly.

3 AIMS OF THE RESEARCH

The overall aim of the research reported in this thesis was to develop a theory-based, generic HL instrument for school-aged children (aged 13–15 years), which would allow comparisons of subjective HL, within an international context, and as a component of large-scale studies. This aim was addressed by conducting three research with the following specific aims:

1. The first aim was (i) to develop a brief, comprehensive, and theory-based instrument for the measurement of subjective HL among school-aged children, and (ii) investigate with this new instrument the level of HL among Finnish school-aged children (Articles 1 and 2).
2. The second aim was to examine the cross-national measurement invariance of this new instrument in four European countries, i.e. Finland, Poland, Slovakia, and Belgium (Article 3).

4 METHODS

The following sections describe the developmental principles and detailed phases in creating the HL instrument. In general, the development process was guided by the conceptual framework of HL presented in Paakkari & Paakkari (2012). The applicability of the instrument was examined in a cross-national context (Figure 1).

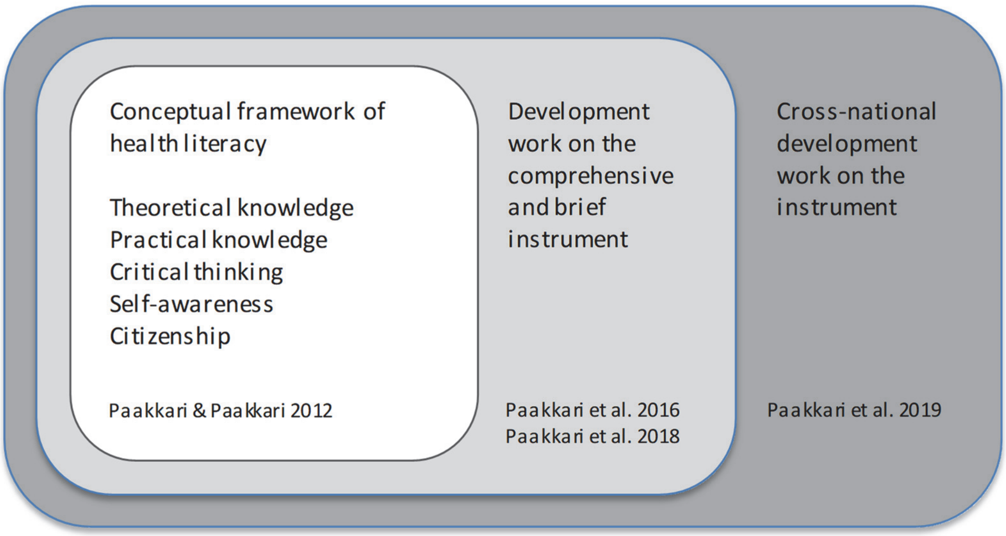

FIGURE 1 Empirical outline of the study

4.1 Principles in developing the health literacy instrument

HL measurement instruments vary greatly, depending on the purpose and context of the measure. Different disciplines have specific procedures for the development of measurement instruments. Nevertheless, from the methodological point of view, the basic steps follow the same developmental process for most instruments, relating to the adequacy of the instrument for its purpose. Typically, the steps for developing a new instrument are 1) definition and elaboration of the construct to be measured, 2) choice of a measurement method, 3) development of items and response options, 4) pilot testing including evaluation and adaptation, and finally 5) careful field-testing (de Vet et al. 2015, 30–32). These steps are intertwined, and the development process goes back and forth between these phases.

These generally accepted steps provide a good methodological basis for the development of an instrument. However, in the present case, the development had to take into account the specific characteristics of HL measurement. Pleasant, McKinney & Rikard (2011) and Jordan, Osborne & Buchbinder (2011) have proposed several attributes that a comprehensive HL measure should reflect, as follows:

- The instrument should be built on a testable HL theory or conceptual framework.
- It should be multidimensional in content (covering multiple conceptual domains) and in the methodology applied.
- It should measure HL on a continuous variable, and should treat HL as a latent construct (i.e. the instrument will contain multiple items drawn from the conceptual domains outlined by the underlying theory or conceptual framework).
- It should allow commensurate comparison across a variety of contexts (e.g. language, culture, the educational system, the population group).
- It should prioritize social research and public health applications as opposed to clinical screening (Jordan, Osborne & Buchbinder 2011; Pleasant, McKinney & Rikard 2011).

The general and specific principles set out above guided the development process of the HL instrument described in this thesis. Thus, the development followed a systematic and structured approach, containing the phases of definition and elaboration of the construct, choice of measurement method, item generation, a pilot study (including test-retest), and finally, field testing, including the construction of a brief instrument (*Health Literacy for School-aged Children*, HLSAC), plus examination of the instrument's international applicability (Table 2).

TABLE 2 Phases of the health literacy instrument development process

Phase	Item reduction
Definition and elaboration of the construct Contextualization, the five core components of health literacy (theoretical knowledge, practical knowledge, critical thinking, self-awareness, citizenship)	
Choice of measurement method Self-reported, multi-item questionnaire	
Item generation Reading and synthesis of relevant literature and existing instruments Formulation of the items and response scale Item selection for the pilot study (iterative process of evaluation and discussion by an expert group (6 persons)	65 items → 32 items
Pilot study N=401, 13- and 15-year-olds (7th and 9th graders), Finland	
Data analysis: inspection of distributions, reliability analysis/internal consistency, confirmatory factor analysis, factor loadings and content of the items, examination of model goodness and sufficiency	32 items → 16 items
Qualitative item analysis: four class-level discussions on the comprehensibility of the items, reformulation of three items	
Test-retest (N=117, 13- and 15-year-olds (7th and 9th graders), Finland): a two-week interval, the same pupils, test-retest reliability examined via structural equation modelling with one latent HL factor and with the five factors derived from the theoretical core components	16 items
Field-testing	
Construction of a brief Health Literacy for School-aged Children (HLSAC) instrument Nationally representative sample, N=3853, 13- and 15-year-olds (7th and 9th graders), Finland	16 items → 15 items → 10 items
Data analysis: inspection of distributions, reliability analysis/internal consistency, confirmatory factor analysis, factor loadings and content of the items, examination of model goodness and sufficiency, regression analysis to predict the relationship between 10- and 15-item instruments, 10-item test finalized	
Examining cross-national measurement invariance of the instrument (HLSAC) Total N=1468, 13- and 15-year-olds: Finland (N=351), Poland (N=642), Slovakia (N=291), Belgium (N=184)	10 items
Translation and back-translation of the questionnaire	
Data analysis: inspection of distributions, confirmatory factor analysis, factor loadings and content of the items, multigroup confirmatory factor analysis, examination of model goodness and sufficiency, reliability analysis/internal consistency, regression analysis to predict the relationship between 10- and 15-item instruments	

4.2 Ethical considerations

Ethical issues were taken into account at each phase of the development process. The study complied with accepted research principles, encompassing integrity, diligence, and accuracy in conducting research work. The methods applied conformed to scientific criteria; moreover, the data collection and analysis, and the presentation and evaluation of the results, were ethically sustainable. In the data collection (pilot study, test-retest, construction of a brief instrument, examination of cross-national measurement invariance) account was taken of the guidelines for the responsible conduct of research (Finnish Advisory Board on Research Integrity 2012), and the research protocol of the international HBSC study (Currie et al. 2014). Participation in the study was voluntary, and pupils responded anonymously to a paper-assisted questionnaire over the course of one lesson. The participants were aware of the confidentiality of the study, and the fact that only group-level results would be reported. The international survey followed each country's ethical and legal requirements.

4.3 Definition and elaboration of the construct

The development of the comprehensive HL instrument was based on Paakkari & Paakkari's (2012) conceptualization of HL as a learning outcome (for further details see Section 2.1 and Table 1). This conceptual framework captures the essential dimensions of HL, and highlights the multidimensional nature of the concept, defining HL as follows:

> [HL] comprises a broad range of knowledge and competencies that people seek to encompass, evaluate, construct, and use. Through health literacy competencies people become able to understand themselves, others, and the world in a way that will enable them to make sound health decisions, and to work on and change the factors that constitute their own and others' health chances (Paakkari & Paakkari 2012).

This definition incorporates the following core components or larger competence fields (also labelled as conceptual domains; see Pleasant, McKinney & Rikard 2011): *theoretical knowledge, practical knowledge, individual critical thinking, self-awareness*, and *citizenship* (Table 1).

4.4 Choice of measurement method

The construct to be measured affects the choice of measurement instrument, such that the instrument should correspond closely to the construct. A goal was measured individuals competence beliefs, i.e. perceived capability (Bandura

1997), in relation to HL. A self-reported questionnaire allows respondents to indicate how they see their competence with regard to HL, and it can be used in large-scale population studies. It has been argued that complex constructs (such as HL) should, in most cases, be measured with a multi-item instrument (de Vet et al. 2015, 36). One would hope that such an instrument would capture the multi-dimensional nature of HL (including all relevant dimensions of the construct), making it possible to test HL as a latent construct, while having content sufficiently specific for respondents to understand the items (Pleasant, McKinney & Rikard 2011; de Vet et al. 2015, 35–36).

4.5 Item generation

The next step was to operationalize the concept, from the abstract conceptual definition to concrete and observable measurement. The definition and core components of HL guided the item generation work of an expert group (six persons). This group included researchers from the field of health promotion, education, and psychology. The group members had teaching experience at different levels of the educational system (from comprehensive school to higher education), and members were experienced in developing national curricula for health education as a subject in schools.

The basis of item generation was the need to ensure the content validity, meaning that the content of any given item should match a certain core component of HL, and that the whole instrument should adequately reflect the construct to be measured (Mokkink et al. 2010). Efforts to achieve content validity proceeded on the basis of the expert group's judgement, logic, and reasoning, and no rigorous method or statistical test was available to asess it. Bearing in mind the concept and essential dimensions of HL, and previous studies on item design (covering e.g. *European Health Literacy Survey Questionnaire* (HLS-EU-Q), Sørensen et al. 2013; *Health Literacy Questionnaire* (HLQ), Osborne et al. 2013), the expert group started to generate items to measure each of the five core components. The pool of potential items for the final instrument had to be notably larger than the final tool, and the content of each item had to reflect the construct of the latent variable, HL (DeVellis 2003, 63–66).

The formulation of the items and the construction of the response scale were based on guidelines for measuring self-efficacy (Schwarzer & Fuchs 1995; Bandura 2006). Perceived self-efficacy can be defined as an individual's belief in his/her competence to perform in a given situation (Bandura 1997). In measuring these competence beliefs (perceived capability), the items should reflect the construct, and be phrased in terms of "can do", because this is a judgement of capability. Moreover, the basic rules of item formulation were taken into account, such as that the items should be understandable by the target population, be specific enough, contain only one question instead of more, and avoid terms with multiple meanings and negative wording – all of these being facets that increase the difficulty of answering (Bradburn, Sudman & Wansink 2004).

Perceived self-efficacy scales have to be formulated to fit the particular domain of functioning that is of interest. One frequently-used format is the Likert scale. This scale does not have a right, wrong, or favourable answer; instead, the participant chooses a response option to indicate the degree of agreement with a given statement. The range of response options is bipolar, typically from *strongly disagree* to *strongly agree*, and the compiler is recommended to include response options that are worded fairly strongly (DeVellis 2003, 63–66; de Vet et al. 2015, 46–48). Hence, all the items took the form "I am confident that…", and the response options were *not at all true, barely true, somewhat true*, and *absolutely true*. The respondents were forced to choose positive or negative options, due to the fact that the scale, having four response options, did not contain a middle option such as "no opinion" or "I can't say".

4.6 Pilot study

Pilot testing aims at examining the instrument's overall comprehensibility, relevance, and feasibility among the target population (de Vet et al. 2015, 57–59). An HL instrument with 32 items was piloted within the Finnish HBSC study in the autumn of 2013, in two ordinary upper secondary schools. These schools were chosen via a discretionary sampling method. In total, the sample contained 401 pupils, of whom 202 were aged 13 (7^{th} graders) and 199 were aged 15 (9^{th} graders). The respondents completed a paper-assisted survey during one school lesson.

After the survey session, all the items were discussed with participants, in order to secure maximum comprehensibility, clarity, and relevance regarding the items. These discussions were conducted in two 7^{th} grade classes (pupils aged 13) and two 9^{th} grade classes (pupils aged 15). Each class had 20–24 pupils. In the discussions (pupils still had questionnaire in front of them during the discussion), the participants were asked questions such as "Were the items comprehensible?", "Did you have any problems with the questions? – if yes, where?", "Were there any terms that were difficult to understand", "Did you understand the response options?", "On what basis did you choose a particular response option", and "Were any relevant issues missing from the questionnaire?". This all helped to improve the content validity.

With regard to feasibility, the pilot study and discussions made it possible to discover how long it took the pupils to complete the questionnaire, and what happened to the respondents' concentration and motivation while they were completing the questionnaire.

The analysis of responses started with inspection of the distributions of all the items. Four items were removed because of very low discrimination power. Thereafter, a confirmatory factor analysis (CFA) for the remaining 28 items was conducted, in combination with theoretical consideration of the item contents, to identify the best items. The a priori CFA model was specified on the basis of the theoretical conceptualization as having five fixed factors, in line with the HL core components. The items were removed one-by-one on the basis of the factor loading (poorest items removed), and with careful consideration of the item content.

The fit of the model to the data was tested with the Chi-square test, Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). A good fit with the data is indicated when the CFI and TLI are higher than .95, lower than .06 for RMSEA, and lower than .08 for SRMR, whereas values higher than .90 for CFI and TLI and less than .08 for RMSEA are considered to reflect acceptable levels of fit (Hu & Bentler 1999; Marsh, Balla & McDonald 1988; Marsh, Hau & Wen 2004). All the analyses were conducted with Mplus 7.3. The data had a hierarchical structure. This means that the pupils are nested in classrooms and in schools, for example pupils in the same class or school can be more similar for various reasons, e.g. variability in teachers' support, overall skill level in the class, or classroom atmosphere (resulting in so-called intra-class correlation). An unmeasured intra-class correlation can overestimate the statistical significance of parameter estimates (Goldstein 1995, 25–26; Snijders & Bosker 1999, 16–17). Because of this, models were constructed applying a COMPLEX option provided in Mplus, which corrects for standard errors in the models deriving from the nested data structure, and therefore corrects also for over-estimation of the significance of the parameter estimates.

The reliability of the instrument was examined. Reliability means the degree to which the measurement is free from measurement error. The most common type of reliability assessment is internal consistency reliability, i.e. the degree of the interrelatedness among the items (Mokkink et al. 2010). This was examined by Cronbach's alpha. Usually, it should be above 0.70 for acceptable internal consistency reliability, and values between 0.80 and 0.90 can be considered to be at a very good level (DeVellis 2003, 95–96; de Vet 2015, 81–83).

One form of reliability, i.e. test-retest reliability, relates to response consistency over time (Mokkink et al. 2010; de Vet et al. 2015, 125). The test-retest was the last component of the pilot testing. Measuring the stability of responses needs a sufficiently short time interval between measurements, such that one can assume that the underlying phenomenon is unlikely to have changed. The guidelines for the optimum test-retest interval differ, depending on the task, what the scale measures, and whether the scores could be expected to change rapidly/slowly over time. A frequently-used time interval is two weeks, but there is no standard rule for choosing the right time interval, with the choice depending on the considerations mentioned above. Broadly speaking, if the period is too long things can change, and if it is too short, the participants may remember their first responses (de Vet et al. 2015, 125; Streiner, Norman & Cairney 2015, 171–172).

After item reduction, the test-retest was conducted with the 16-item instrument. The test-retest was organized under the conditions of an ordinary Finnish upper secondary school, which was selected by discretionary sampling. The time interval between measurements points was two weeks. At the first measurement time, there were 131 respondents, and 14 pupils dropped out for the second measurement point. Because of this, the total sample contained 117 pupils, including both 13- and 15-year-old respondents (7th and 9th graders). Test-retest reliability was analysed via structural equation modelling (SEM) using Mplus 7.3. The latent HL factor for 16 items was estimated for both time points. Within the model, the retest latent factor was regressed on the latent HL factor at the first assessment time-point. In addition, the stability of the five theoretical components was examined by constructing the model with the same five latent factors at test and retest.

4.7 Field testing

After evaluation of instrument's performance on comprehensibility, relevance, feasibility (pilot study), and stability over time (test-retest), the study proceeded to field testing. The field testing started with construction of a brief instrument (HLSAC). Thereafter, the cross-national measurement invariance of the instrument (HLSAC) was examined.

4.7.1 Construction of a brief instrument

For multi-item instruments that are considered to measure unobservable constructs (latent factors), it is essential to apply a large-sample field test on the target population (de Vet et al. 2015, 65). The data were collected in Finland in 2014, with the 16-item HL instrument forming part of the broader HBSC study (see Currie et al. 2009). The nationally representative sample consisted in total of 3853 pupils from 359 schools, including 1918 pupils aged 13 (7th graders) and 1935 pupils aged 15 (9th graders). The schools were chosen from the Finnish school register using a cluster sampling method that took into consideration provinces, the type of municipality (urban, semi-urban, rural), and the size of the schools. From each school the participating class was randomly selected. The pupils responded to a questionnaire during a 45-minute lesson.

The goal for the field testing was explore the structure of the data and to make the final selection of the items per dimension (involving five factors, in line with the HL core components). As a first step, the instrument was reduced from 16 to 15 items, in order to have an equal number of items for each factor. One extra item that had remained from the previous round (involving the self-awareness factor) was removed after inspection of the distributions and the item loadings. The construction of a 10-item HL instrument was conducted via a process similar to that in the pilot study phase. The item reduction was based on inspection of the item distributions (items with homogeneous distribution/low discrimination power or with more missing values than other items were screened out), internal consistency reliability estimates (Cronbach's alpha), factor loadings, and CFA model fit, in addition to examination of the contents of the items (content validity).

In addition to content validity, the other commonly applied forms of validity (criterion validity, construct validity) were also taken into account (Mokkink et al. 2010; de Vet 2015, 150). Because we had no "gold standard" HL instrument for school-aged children, the validity assessment of the instrument could not be based on criterion validity. In this kind of situation, construct validation can provide evidence of the instrument's validity (de Vet et al. 2015, 169). Structural validity, meaning the degree to which the scores of an instrument are an adequate reflection of the dimensionality of the construct to be measured, is a common type of construct validity (Mokkink et al. 2010). Structural validity can be assessed by confirmatory factor analysis, with the model fit parameters indicating whether the data fit the hypothesized factor structure (de Vet et al. 2015, 169–170). If the structure of the instrument can be specified *a priori*, and if the contextualization is based on evidence from previous research, structural validation is a strong tool to inspect instrument's validity (de Vet et al. 2015, 197). To evaluate the CFA model fit to the data, the Chi-square test, RMSEA, CFI, TLI, and SRMR were used. In order to examine the sufficiency of the 10-item version of the instrument, a regression analysis was conducted in which the 10-item version was set to predict the longer version of the instrument. The analyses were conducted with Mplus 7.3.

The descriptive statistical analyses were conducted for the total sample, separately for boys and girls, and for both age groups (13-year-olds, 15-year-olds). The descriptive statistics for HL included means, standard errors, standard deviations, distributions of skewness and kurtosis, and percentage distributions of the HL levels. The differences between the group means (for gender and grade), and the gender and grade interaction effects on the HL were tested via a two-way analysis of variance (ANOVA). The analyses were conducted using SPSS (version 22).

Next, the brief 10-item instrument (HLSAC) was used to measure Finnish school-aged children´s subjective (perceived, self-reported) HL. The levels of HL were classified into three groups (low, moderate, high). The expert group (consisting of researchers and teachers in the field of health promotion, education, and psychology) set the thresholds, i.e. determined the HL scores required to reach a given level. The instrument contains 10 items and 4 response options in each item. To get a higher level of HL (including *moderate* or *high*), pupils had to choose a higher response option (scored at 3 or 4) in more than 50% of the items. Following consideration of the contents of the items mentioned above, and inspection of the response distribution, the resulting HL levels were set at *low* (score 10-25), *moderate* (score 26–35), and *high* (score 36–40). These cut-off based categories are useful for descriptive purposes. However, it should be noted that a pupil can move to the next threshold simply by gaining one point more. In reality, this pupil's competence might be at the same level as the pupil with a one-point lower score.

4.7.2 Examining the cross-national measurement invariance of the Health Literacy for School-aged Children (HLSAC) instrument

The final step of the field testing was inspection of the applicability of the 10-item HLSAC in the international context. Comparative data were collected in the spring of 2016 in Finland, Poland, Slovakia, and Belgium. The data collection was organized in collaboration with the HBSC Study Network. In each country, the participants responded to a standardized paper-assisted inquiry in the course of one lesson, with the HL instrument forming part of this broader questionnaire. The final sample consisted 1468 participants, including 690 pupils aged 13 (Finland N = 176, Poland N = 341, Slovakia N = 173), and 778 respondents aged 15 (Finland N = 175, Poland N = 301, Slovakia N = 118, Belgium N = 184). The number of 13-year-old Belgian pupils remained low, and therefore this group was excluded from the analyses.

For the inspection, the questionnaire was translated and back-translated by professional translators, aiming to ensure parity between languages. In this process it emerged that in the Slovak language there was almost no semantic difference between the response options *barely true* and *somewhat true.* Because of this, the second response option *barely true* was changed to *not quite true.* Finally a Likert-type scale was devised, which included the response options *not at all true, not quite true, somewhat true,* and *absolutely true.*

The adequate functioning of the HLSAC's original 10 items was examined in the different countries. It was necessary to check whether there could be items that could function better in some countries than the original items, and for this purpose, 14 new items were added to the questionnaire. First of all, the distribution of items was inspected. Because of low discriminatory power, six items were removed. Next, a confirmatory factor analysis was applied to the remaining 18 items. After inspection of the item loadings and consideration of the item contents, no evidence emerged suggesting a need to change the original ten HL items. In addition, to examine whether the predictive power of the short 10-item HLSAC instrument was similar to that of the longer 18-item instrument, a regression analysis was conducted, predicting the 18-item instrument with the 10-item instrument, using SPSS (version 24).

Multi-item self-reporting questionnaires are a common method to assess an underlying construct (latent factor). Thus, it is critical to inspect the measurement invariance (i.e. the factorial invariance, the measurement equivalence) of a questionnaire across the groups one is seeking to compare. Here it should be noted that equivalence of measurement is the foundation for meaningful and valid comparisons between different groups with reference to the means of a latent factor, as calculated from an instrument (Van De Schoot et al. 2015). Measurement invariance indicates that the instrument performs similarly in different groups or populations (de Vet et al. 2015, 182–185). Examination of the cross-national measurement invariance of the HLSAC instrument began with distribution screening, the aim being to ensure that each of the 10 items had adequate variance in each country, and to avoid the biases that can be caused by ceiling/floor effects.

The next step in the study was to apply a multigroup confirmatory factor analysis (CFA), which is one of the most common methods to inspect measurement invariance (Jöreskog 1971; Byrne 2004; Brown 2006). The goal was to test whether the HLSAC instrument functioned in a similar way in four

European countries (Finland, Poland, Slovakia, Belgium). Multigroup CFA allows simultaneous estimations and direct statistical comparisons of the measurement parameters across groups; it is therefore a useful method to test the structure of a scale (Little & Slegers 2005). The fits of the CFA models were evaluated using the Chi-square goodness-of-fit statistic (χ^2), the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). A good fit with the data is indicated when the CFI is higher than .95, lower than .06 for RMSEA, and lower than .08 for the SRMR (Hu & Bentler 1999). The fit of the model with the data can be adequate, even if the Chi-square test is significant, so long as the other indices suggest a good fit, particularly with large samples (on the grounds that the Chi-square test is sensitive to large sample sizes) (Hu & Bentler 1999). The levels of measurement invariance were tested using the χ^2 difference test statistic.

In the first phase of the modelling, the magnitude of the factor loadings for each country and each age were inspected, and a model with configural invariance with a freely estimated parameter was estimated (Horn & McArdle 1992; Vandenberg & Lance 2000). The model was a one-factor model with ten items. Configural invariance requires that the factor structure is similar across groups: this means that the construct (the latent factor) is similarly measured by the same items in each group, but the factor loading magnitudes can differ somewhat across groups. A precondition for clear comparisons is that there should be the same valid CFA construct in each group.

In the second phase of the modelling the equivalence of the factor loadings was inspected, to fit a metric invariance model. In order to examine the equality of the factor loadings across countries, the latent factor mean was set at zero, and the variance at one. Metric invariance (with the factor loadings equal across the groups) confirms that the measured construct has exactly the same meaning for the various participant groups (Vandenberg & Lance 2000). If the metric invariance model holds (models with and without equality constraints are compared using a χ^2 difference test), the third phase is to fit a scalar invariance model. This is achieved when both the factor loadings and the intercepts are equal across groups. If scalar invariance is not reached, one arrives at partial invariance (Byrne, Shavelson & Muthen 1989; Steenkamp & Baumgartner 1998).

In the fourth phase of the modelling, the equivalence of the HL mean values between countries was examined. To compare the mean values, the factor loadings were fixed at one, the corresponding intercept was set at zero, while the factor variance and the HL mean values were allowed to differ across countries. In setting the equality constraints, all the participant countries were first set to the model in each age-group. After implementing group comparisons, all possible paired country comparisons were conducted. All the analyses were conducted with Mplus (version 7.3).

Finally, the internal consistency reliability of the HLSAC instrument in each country was estimated using Cronbach's alpha.

5 RESULTS

The following sections present the results of the item generation, the pilot study (including test-retest), and the field testing. They cover construction of the brief instrument (Articles 1 and 2) and examination of the cross-national measurement invariance of the instrument (Article 3).

5.1 Item generation (Article 1)

The expert group generated in total 65 items, which were distributed evenly across the different core components of HL. The items were formulated in such a way that a four-point Likert-type scale could be applied. The preliminary reduction of the 65 items occurred after an iterative process, involving discussion with expert group, reading and re-reading, refining, and ranking and prioritizing the items per core component. At the end of this process, 32 items were included in the pilot study.

5.2 The pilot study (Article 1)

At the start of the pilot study (N=405) there were 32 items. After item reduction the instrument contained 16 items. The plan was to reach a valid and reliable 15-item solution with three items representing each theoretical HL component. However, one extra item was added to the self-awareness component, for content-based reasons, and with a view to finding the best possible combination of items.

The Cronbach alpha for the entire 16-item HL measure was very high (.94). In the final one-factor CFA model (with 16 items and no error covariances), the fit to the data was reasonably good ($\chi^2(104)=261.69$, $p=.000$; RMSEA=.06, CFI=.93, TLI=.92, SRMR=.04). The standardized factor loadings were between .57 and .74, with the majority above .70. The Cronbach alphas for each of the five factors (based on the theoretical HL components) were between .75 and .84. Hence, the internal consistency of each of the HL components was at an adequate level.

The pilot phase also included class-level discussions with the pupils. In general the pupils considered the items to be understandable. However, three items required some reformulation to avoid ambiguity.

At the end of the pilot testing phase, a test-retest procedure was conducted. The reformulated 16-item instrument was utilized with 117 pupils. The test-retest was conducted with both a one-factor and a five-factor theory-based model. In the Structural Equation Model (SEM) with one latent HL factor at both assessments, the standardized stability estimate was .83. The SEM with five factors also exhibited high stability estimates: theoretical knowledge .88, practical knowledge .81, critical thinking .81, self-awareness .84, and citizenship .90. These estimates suggested adequate test-retest reliability for HL, and also for the predetermined factors.

5.3 Construction of the 10-item Health Literacy for School-aged Children (HLSAC) instrument (Articles 1 & 2)

As an initial step in the construction of a brief HL tool (N=3853), the instrument was reduced to 15 items, with each factor represented by three items. Thereafter, the instrument was reduced to the 10-item *Health Literacy for School-aged Children* (HLSAC) instrument, with each factor represented by two items. The distributions of all the items were reasonable, and none of the answer options accounted for more than 58% of the answers (Table 4). For the 10-item instrument, Cronbach's alpha was high (.93), suggesting high internal consistency. The Cronbach alphas for the five core components (each with two items) were also reasonable (.69–77) (Figure 2).

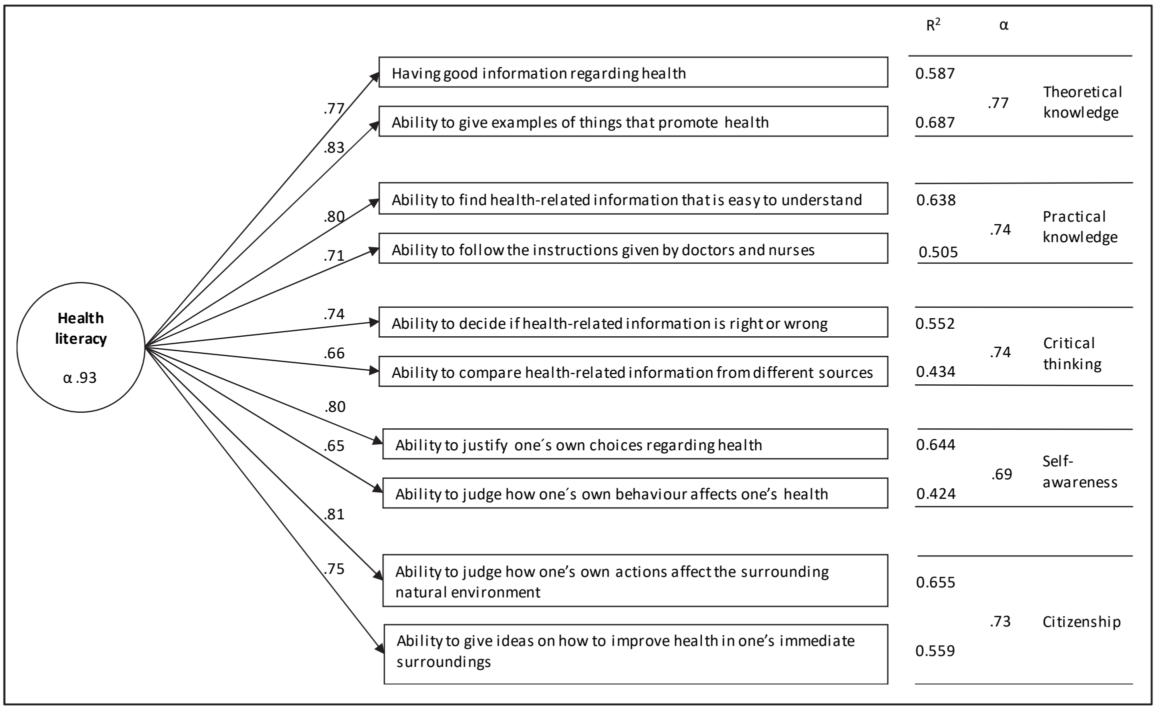

FIGURE 2 The Health Literacy for School-aged Children (HLSAC) instrument: Cronbach alphas, item loadings, and R^2 values

A 5-factor model with 10 items ($\chi^2(25)=681.41$, $p<.001$; RMSEA=.08, CFI=.96, TLI=.92, SRMR=.03) showed strong correlations (.95–1) between the factors. For this reason, the final model (Figure 2) was constructed as a one-factor model. The model had good item loadings. Moreover, considering the large sample size, and the fact that no error covariances were allowed between any of the items, it also had a reasonably good fit to the data ($\chi^2(35)=948.64$, $p<.001$; RMSEA=.08; CFI=.94; TLI=.92; SRMR=.04).

A regression analysis was conducted in order to examine the relationship between the 10- and the 15-item instrument. The 10-item HLSAC instrument predicted approximately 97% of the variance of the 15-item instrument (R2=.97, $p<.001$).

The final HLSAC instrument consists two items from each of the five predetermined theoretical components (Paakkari & Paakkari 2012): *theoretical knowledge* (items 1 and 5), *practical knowledge* (items 4 and 7), *critical thinking* (items 3 and 9), *self-awareness* (items 8 and 10), and *citizenship* (items 2 and 6) (Figure 3).

From the following options, choose the one that best describes your opinion

I am confident that…	Not at all true	Not quite true	Somewhat true	Absolutely true
1. I have good information about health	❑	❑	❑	❑
2. When necessary I am able to give ideas on how to improve health in my immediate surroundings (e.g. a nearby place or area, family, friends)	❑	❑	❑	❑
3. I can compare health-related information from different sources	❑	❑	❑	❑
4. I can follow the instructions given to me by healthcare personnel (e.g. nurse, doctor)	❑	❑	❑	❑
5. I can easily give examples of things that promote health	❑	❑	❑	❑
6. I can judge how my own actions affect the surrounding natural environment	❑	❑	❑	❑
7. When necessary I find health-related information that is easy for me to understand	❑	❑	❑	❑
8. I can judge how my behaviour affects my health	❑	❑	❑	❑
9. I can usually figure out if some health-related information is right or wrong	❑	❑	❑	❑
10. I can give reasons for choices I make regarding my health	❑	❑	❑	❑

FIGURE 3 The instruction, final items, and response options of the Health Literacy for School-aged Children (HLSAC) instrument

Table 3 provides the descriptive HL statistics for girls and boys in the 7th and 9th grades, separately. The skewness and kurtosis indicated that the data were approximately normally distributed. In every group the minimum score was 10 and the maximum 40. The overall mean HL score was 32.55. The lowest subjective HL was found among 7th grade boys, while the 9th grade girls reported the highest subjective HL. The girls' HL mean score was higher than that of the boys, and according to a two-way analysis of variance (ANOVA) the difference was statistically significant (F(df1)=8.214, p=.004). The mean HL score was higher among the 9th graders than among 7th graders, and this difference was also statistically significant (F(df1)=10.555, p=.001).

TABLE 3 Descriptive statistics and the Cronbach alphas for health literacy (HLSAC)

	N	Min.	Max.	Mean	SE	SD	Skewness	SE	Kurtosis	SE	α
Boys 7th grade	880	10	40	31.90	0.20	5.91	-0.65	0.08	0.52	0.17	.94
Boys 9th grade	882	10	40	32.39	0.20	6.06	-0.96	0.08	1.42	0.16	.95
Girls 7th grade	894	10	40	32.51	0.17	5.13	-0.54	0.08	0.25	0.16	.91
Girls 9th grade	963	10	40	33.32	0.16	4.88	-0.61	0.08	0.38	0.16	.91
Total	3619	10	40	32.55	0.09	5.53	-0.76	0.04	0.93	0.08	.93

Table 4 shows the response distributions of the items. The Finnish respondents indicated that their theoretical and practical knowledge was, generally speaking, at a good level. Around 90% reported that they had a good knowledge of health: they felt that they could easily find understandable health information and could follow the instructions of doctors or nurses (response options: *somewhat true* or *absolutely true*). More difficulties were indicated regarding critical thinking and citizenship. About 15% reported difficulties in the ability to compare the information from different sources, or in the ability to decide if information is right or wrong. Approximately one in five indicated problems in terms of being able to give ideas on how to improve health in their environment. Gender comparison showed that the boys reported more difficulties (response options *not at all true* or *not quite true*) than the girls on almost every HL item. Similarly, more girls than boys reported having good competence in the cases that were asked; in other words, they chose the response option *absolutely true* more often than the boys in most of the HL items.

TABLE 4 Percentage distributions of the items in the Health Literacy for School-aged Children (HLSAC) instrument, divided by gender

	Boys (N=1820)				**Girls** (N=1912)			
	Not at all true	Not quite true	Somewhat true	Absolutely true	Not at all true	Not quite true	Somewhat true	Absolutely true
Theoretical knowledge								
Having good information regarding health	2.6	9.3	44.6	43.5	0.5	7.5	45.4	46.6
Ability to give examples of things that promote health	2.3	13.6	49.9	34.1	0.9	12.2	49.9	37.0
Practical knowledge								
Ability to find health-related information that is easy to understand	2.1	9.3	44.3	44.3	0.6	6.3	42.5	50.7
Ability to follow the instructions given by doctors and nurses	2.3	9.6	43.9	44.3	0.7	5.7	35.3	58.2
Critical thinking								
Ability to decide if health-related information is right or wrong	2.3	12.7	49.7	35.4	1.5	12.9	51.9	33.7
Ability to compare health-related information from different sources	2.8	13.8	48.5	34.9	1.9	13.8	51.7	32.6
Self-awareness								
Ability to justify one´s own choices regarding health	2.5	12.2	48.7	36.6	0.9	9.3	49.7	40.2
Ability to judge how one´s own behaviour affects one's health	3.2	12.0	45.3	39.5	1.1	9.0	48.8	41.1
Citizenship								
Ability to judge how one's own actions affect the surrounding natural environmen	2.5	10.6	49.4	37.5	0.9	8.8	48.2	42.1
Ability to give ideas on how to improve health in one's immediate surroundings	3.5	16.5	51.8	28.2	2.4	16.3	52.4	28.9

The HL scores were subsequently categorized into three levels (*low* = score 10–25, *moderate* = score 26–35, *high* = score 36–40). Around one tenth of the participants had low HL, 57% had moderate HL, and approximately one third achieved a high level of HL (Figure 4). In both age groups there were more boys than girls with low HL. In both genders, the proportion of pupils who had a high level of HL increased towards the 9th grade.

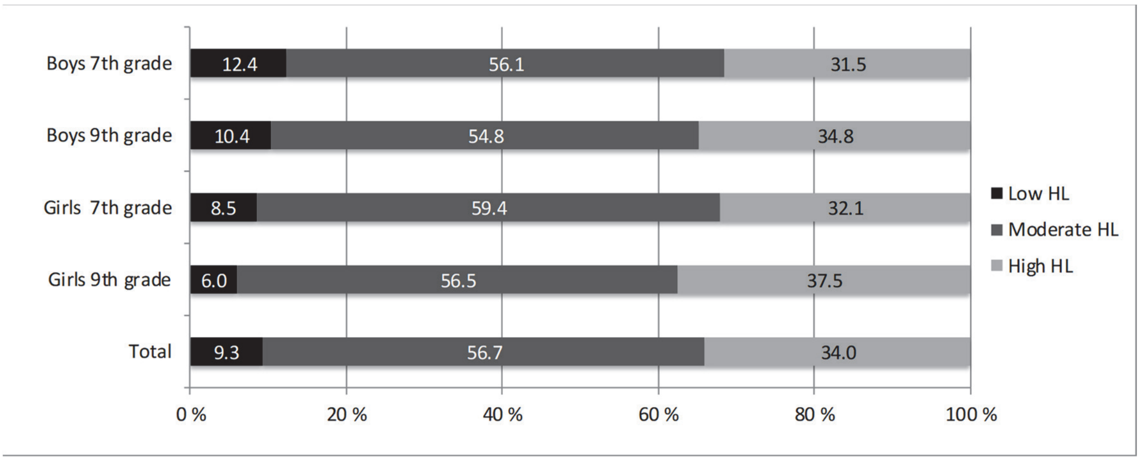

FIGURE 4 Levels of subjective health literacy (HL) by gender and grade, and for the total sample (percentage distribution)

5.4 Cross-national measurement invariance of the Health Literacy for School-aged Children (HLSAC) instrument (Article 3)

The applicability of the HLSAC instrument in international contexts was examined on the basis of the data collected from Finland, Poland, Slovakia, and Belgium (N=1468). Table 5 presents the descriptive HL statistics for the different age groups and countries. The total HL mean score was 31.02. The highest mean score was found in Finland, and the lowest in Belgium. The HL mean score was higher among pupils aged 15 than among those aged 13.

TABLE 5 Descriptive cross-national statistics and the Cronbach's alphas for health literacy (HLSAC)

	N	Mean	SE	SD	α
Finland, aged 13	176	32.45	0.40	5.28	.90
Finland, aged 15	175	33.11	0.41	5.37	.90
Poland, aged 13	341	30.30	0.27	5.08	.85
Poland, aged 15	301	30.85	0.28	4.94	.85
Slovakia, aged 13	173	31.12	0.37	4.90	.81
Slovakia, aged 15	118	31.33	0.42	4.59	.79
Belgium, aged 15	184	29.33	0.38	5.15	.81
Total	1468	31.02	0.13	5.18	.85

A regression analysis was conducted in order to examine what proportion of the variance the short 10-item HLSAC instrument was explained about the 18-item instrument. The proportion of explained variance was 97% in Finland and Poland, 96% in Slovakia, and 95% in Belgium.

The 10-item HLSAC instrument exhibited an adequate Cronbach alpha (.85) for the data as a whole, and the internal consistency of the instrument was also at a high level in each participant country (Finland α=.90, Poland α=.85, Slovakia α=.80, Belgium α=.81). The CFA models for the 10-item instrument showed adequate fit with the overall data ($\chi^2(35)=200.65$, $p=.000$; RMSEA = .06, CFI=.96, SRMR=.03), and also with the data for each country.

5.4.1 Comparisons of the factor loadings across the countries

Table 6 indicates that the item loadings were in general at the same level in the different age groups (pupils aged 13 and 15) and countries, although there were a few items for which the loadings seemed to vary slightly between countries.

TABLE 6 The Health Literacy for School-aged Children (HLSAC) instrument: the standardized item loadings from the confirmatory factor analysis, by age group and country

Items per core component	13-year-olds			15-year-olds			
	Finland	Poland	Slovakia	Finland	Poland	Slovakia	Belgium
Theoretical knowledge							
Having good information regarding health	.36	.45	.31	.32	.35	.16	.39
Ability to give examples of things that promote health	.58	.42	.52	.52	.49	.43	.49
Practical knowledge							
Ability to find health-related information that is easy to understand	.53	.48	.54	.47	.45	.58	.40
Ability to follow the instructions given by doctors and nurses	.37	.36	.33	.42	.33	.21	.21
Critical thinking							
Ability to decide if health-related information is right or wrong	.52	.47	.44	.44	.50	.45	.44
Ability to compare health-related information from different sources	.56	.52	.53	.51	.50	.60	.43
Self-awareness							
Ability to justify one´s own choices regarding health	.47	.55	.44	.43	.46	.42	.34
Ability to judge how one´s own behaviour affects one's health	.50	.48	.26	.47	.48	.34	.28
Citizenship							
Ability to judge how one's own actions affect the surrounding natural environment	.50	.47	.50	.51	.44	.41	.42
Ability to give ideas on how to improve health in one's immediate surroundings	.52	.47	.51	.54	.42	.42	.39

The analysis of configural and metric invariance provided evidence that the corresponding common factors had the same meaning across groups. The baseline model (against which the restricted models were compared) had free loadings across all countries ($\chi^2(105)=204.60$, RMSEA=.06, CFI=.95, SRMR=.04). The country comparison between Finland, Slovakia, and Poland in the data for pupils aged 13 showed that while all ten factor loadings were equivalent in Finland and Poland (χ^2diff(10)=15.62, *p*=.11), some loadings were different in Slovakia. We continued the analysis by examining which of the factor loadings were different in the Slovakian data, setting each factor loading free one by one. The analysis showed that there was one loading that was lower in the Slovakian data ("ability to judge how one´s own behaviour affects one's health"). A model in which all ten factor loadings were set as equivalent in Finland and Poland, and in which all factor loadings except for one item were set as equivalent in Slovakia, Finland, and Poland, proved to be as well fitted to the data as a model with all the loadings free in each country (χ^2diff(19)=21.01, *p*=.33). However, the effect of setting all ten items as equivalent in all three countries was very close to being acceptable (χ^2diff(20)=45.51, *p*=.01).

The baseline model for pupils aged 15 had free loadings across all countries ($\chi^2(140)=261.54$; RMSEA=.07, CFI=.95, SRMR=.05). The comparison of the factor loadings for pupils aged 15 indicated that the loadings were equivalent in Finland, Poland, and Belgium (χ^2diff(20)=27.56, *p*=.12). The model fit fell only slightly below the accepted limit when all the Slovakian factor loadings were set as equal to the loadings for the other countries (χ^2diff(30)=53.87, *p*=.01). After releasing one item ("having good information regarding health") in the Slovakian data, the Chi-square difference test already approached non-significance (χ^2diff(29)=44.64, *p*=.03). After releasing three items in the Slovakian data ("having good information regarding health", "ability to follow the instructions given by doctors and nurses", "ability to judge how one´s own behaviour affects one's health") the difference test showed non-significance (χ^2diff(27)=38.73, p=.07). Note, however, that the loading for the item "ability to follow the instructions given by doctors and nurses" was equivalent in Slovakia and Belgium. After releasing the item, "having good information regarding health" in the Slovakian data, and setting the item "ability to follow the instructions given by doctors and nurses" as equal in Belgium and Slovakia, the Chi-square difference test showed non-significance (χ^2diff(28)=37.80, p=.10). These analyses suggest that the 10-item model with equal loadings across countries fitted well with the data, and particularly well in Finland, Poland, and Belgium.

Examination of intercept equivalence across the three countries showed that the intercepts were not equal across countries. This was the case both for the 13-year-olds (χ^2diff(39)=258.79, $p<.001$), and for the 15-year-olds (χ^2diff(57)=268.10, $p<.001$).

The analyses suggested that configural and metric invariance holds for comparisons between Poland and Finland (for 13-year-olds), and between Poland, Finland, and Belgium (for 15-year-olds). However, the model fit was good, and the chi-square difference tests very close to the significance level also for the Slovakian data. Based on these analyses, we considered the factorial invariance to be sufficient to conduct mean value comparisons between the countries in question. However, all the comparisons involving the Slovakian data should be interpreted with caution, given the slight differences that emerged.

5.4.2 Comparisons of the health literacy mean values across the countries

Comparison of the mean values for 13-year-olds (Table 5) across the countries revealed differences (χ^2diff(123)=238.08, p=.000). All the paired country differences were tested, and all the paired country comparisons were significant (Finland>Poland, Finland>Slovakia, Poland<Slovakia). Comparison of the mean values for 15-year-olds (see Table 5) also indicated differences between countries (χ^2diff(167)=310.92, p=.000). Subsequent paired testing showed significant differences between the following pairs: Finland>Slovakia, Finland>Poland, Finland>Belgium, Poland<Slovakia. Poland and Slovakia showed no difference from Belgium. These findings were confirmed in the overall multigroup models. Comparison of the baseline model (with all means free in all countries) to the model in which the means of Poland and Belgium were set as equal showed a non-significant Chi-square difference (χ^2diff(1)=1.77, p=.18).

6 DISCUSSION

The first aim of the study was to develop a brief, comprehensive, theory-based, and feasible instrument to measure subjective HL among school-aged children, and to use this instrument to investigate the level of HL among Finnish participants. The second aim was to examine the cross-national measurement invariance of the HLSAC instrument in four European countries. For this purpose, a 10-item multi-dimensional (based on five predetermined core components) instrument (HLSAC) was constructed.

The results showed that the tool is suitable for school-aged children and for large-scale studies, enabling comparison of subjective HL in international contexts. The main development principles and research findings are discussed in the following sections, as are also the limitations of the study and future implications.

6.1 General elements of the health literacy instrument development process

The development process of the HL instrument was iterative, systematic, and validity- and reliability-driven. It took into consideration the constituent principles of an HL instrument (Jordan, Osborne & Buchbinder 2011; Pleasant, McKinney & Rikard 2011). The development work, which was based on a definition and a conceptual framework of HL, was multi-dimensional in content (encompassing relevant aspects of HL). It treated HL as a latent construct, and took into account the context in which the instrument would be used. Grounded in the fields of health promotion and public health, the conceptual framework (Paakkari & Paakkari 2012) guided the elaboration of the instrument. The intention was to have items with a multi-dimensional content, in order to measure a wide range of competencies, namely *theoretical knowledge, practical knowledge, individual critical thinking, self-awareness,* and *citizenship*. The instrument was constructed also with the aim of permitting comparisons across a variety of settings (involving e.g. different languages and cultures).

The development process contained generally-accepted phases, with the validity of the instrument being taken into account at every step of the study. A multidisciplinary expert group conducted item generation. The group generated a large pool of items, seeking to ensure content validity via detailed conversation, plus judgement, logic, and reasoning. The aim was that every single item should match a certain HL core component. Discussions with participants in the pilot phase also influenced the formulation of the items, with efforts to ensure that the participants understood the items in the same way as the researchers.

The complex nature of HL imposes requirements for the development and design of a brief instrument. It involves continuous discussion on the essential components of HL, their interconnections, what theoretical component a certain item may represent, and how items should be formulated to measure the entity in question. The 10-item instrument with five factors showed correlations between the factors. Thus, the final model was constructed as a one-factor model, such that this model describes a single phenomenon, labelled as HL. The discovery of cross-correlations among the factors in the HLSAC instrument came as no surprise; after all, an assumption of the background theory was that the core components of HL are partly overlapping, and have some hierarchical elements (Paakkari & Paakkari 2012). Note, however, that the one-factor solution does not mean that the model cannot be based on a theoretical framework containing multiple dimensions. Importantly, a one-factor model offers the possibility to calculate a sum score and to define a single overall HL index. This is a clear benefit as compared to multifactorial models which, since they have several HL subscales, can violate the requirement of additivity, and in so doing question the plausibility of the sum score of composite HL scales (Altin et al. 2014; Finbråten et al. 2017).

6.2 The psychometric quality of the Health Literacy for School-aged Children (HLSAC) instrument

The quality of any measurement instrument plays an important role in research. Detailed descriptions on how the measurement properties have been assessed give the kind of information that researchers need in choosing the right instrument for a particular purpose. If studies are to have any value, it is crucial that the instrument should be able to reproduce consistent results over time, and measure what it proposes to measure. The instrument's ability to provide scientifically robust results is also the starting point for sound practical implications (e.g. political decisions, interventions).

In the present study, all the phases of the item reduction were based on inspection of item distributions, factor loadings, internal consistency reliability estimates, and CFA model fit, in conjunction with consideration of the item contents. This procedure was followed on the grounds that if item selection is based merely on factor loadings, there is a danger of capitalizing on chance fluctuations, with the possibility that small differences in loadings may become overly influential.

From the perspective of structural validity, the results indicated that the final instrument (10 items, two items per theoretical core component) had a reasonably good fit to the data, and good item loadings. The instrument's internal consistency reliability was found to be adequate: the high value of Cronbach's alpha indicated the inter-relatedness of the items, hence the items measured the same construct. It should be noted that the number of the items in the instrument has a strong influence on the value of Cronbach's alpha coefficient (Sijtsma 2009). A low number of items reduces the alpha; hence, given the relatively small number of items of in HLSAC instrument, the instrument demonstrated an adequate level of internal consistency.

The test-retest reliability of the instrument showed high stability estimates; hence, consistency over time was at an adequate level. The difficult element in test-retest is the choice of an appropriate time interval between measurement points, but in fact there are no clear guidelines for an optimum test-retest interval (de Vet et al. 2015, 125; Streiner, Norman & Cairney 2015, 171–172). As a phenomenon, one cannot expect HL to show very rapid change over time, so it was essential to ensure that respondents could not remember their first answers. Based on these factors, a two-week interval between measurement points was chosen.

There are a wide range of fit indices for measuring a model's fit to the data. There is no single criterion or golden rule for model fit evaluation, and different indices reflect a different aspect of model fit. Thus, it is necessary to present a variety of indices. In this study, commonly recommended indices are reported (χ^2, RMSEA, CFI, TLI, SRMR) (Hu & Bentler 1999; Marsh, Hau & Wen 2004; Kline 2005). The overall goodness and sufficiency of the instrument were clearly at an acceptable level: even if not all the index values reached the precise recommended cut-off values for a good fit (Hu & Bentler 1999), they nevertheless came close, and were better than the values recommended as providing an acceptable fit for the data (Marsh, Balla & McDonald 1988; Hu & Bentler 1999; Marsh, Hau & Wen 2004). Here it should be noted that there can be different proposals for cut-off values (Niemand & Mai 2018). Fit indices are a useful way to analyse a model's fit, but strict adherence to cut-off values can lead to incorrect rejection of an acceptable model; hence, cut-off values should be used more as rough guidelines than golden rules (Marsh, Hau & Wen 2004; Marsh et al. 2018). There are many factors with the potential to distort fit values (e.g. sample size, model size, normality of the data distribution). In this study the large sample size, and the fact that no error covariances were allowed between the items, could have affected achievement of the cut-off values. The Chi-square test is sensitive to sample size, with a large sample size bringing about a greater risk of rejection of the model (Bentler & Bonnet 1980; Hu & Bentler 1999). A large sample size also influences impairing to SRMR and TLI. By contrast, RMSEA and CFI have been determined to be more independent of sample size (Cangur & Ercan 2015). Viewed against this background, the model showed an adequate fit with the data.

Although several tools already exist for measuring the HL of children and adolescents (Guo et al. 2018; Okan et al. 2018), there was no instrument which would concurrently satisfy the requirements for a brief, comprehensive, generic, self-administered, and target group-validated HL instrument. The results showed that the brief 10-item HLSAC instrument predicted the variance of the longer 15-item instrument very well. It is important that the short instrument can capture the multi-dimensional nature of HL, since otherwise, the instrument will not measure what it aims to measure. The availability of a brief and comprehensive HL instrument will allow it to be used in large-scale surveys aimed at measuring levels and trends in HL, including the association between HL and other phenomena, such as health behaviour. A more ambitious, detailed, or specific instrument would run the risk of survey becoming too long and burdensome for young respondents.

6.3 The level of health literacy among Finnish pupils

An important task in the development process was to define the HL thresholds. Setting the threshold scores for a given level (low, moderate, high) has a number of benefits. It makes it possible to visualize the current HL situation and to monitor changes in HL over time. In this form, it provides condensed information to different stakeholders (e.g. policy makers, researchers, educators, parents, pupils), so that they can make comparisons across different population groups (e.g. by age, gender, country). It further facilitates studies (meta-analyses), and interventions targeted at specific populations. In general, threshold-based categories work well on a population level, but can be problematic at an individual level. A single point on the scale can cause the respondent to move from one category to another, even if in reality there has been no significant change in the person's competence.

Finnish school-aged children manifested a fairly high subjective HL level. About one-tenth had low HL, 60% had moderate HL, and around one third reported a high level of HL. Girls showed a higher HL level than boys, and 15-year-old pupils had a higher HL level than 13-year-old pupils. One possible explanation for these results can be found from the Finnish school system. Health issues are taught within a statutory independent subject labelled health education, and are also integrated with other subjects. Every school follows the national curriculum, which contains the objectives for primary school (grades 1–6, ages 7–12) and for secondary school (grades 7–9, ages 13-15). Older pupils have received more teaching than younger ones, and this can partly explain the higher HL levels shown by older pupils.

There could be several reasons underlying the higher HL among girls. They could be related to aspects of society, culture, the school environment, and pedagogy that may favour girls (Stoet & Geary 2013). According the PISA study, girls seem to invest more effort in school and homework, and there are differences between girls and boys in their attitudes to learning and to school (OECD 2015). Within Finland, boys are less interested than girls in the health issues discussed in health education lessons (Aira et al. 2014). A high level of self-regulation can explain better performance, and girls tend to be more disciplined and self-regulated than boys, i.e. they appear to be better able to set goals, plan ahead, and deal with setbacks and frustrations (Duckworth & Seligman 2006; Kenney-Benson et al. 2006). Although these studies shed some light on the overall phenomenon, one must remember that group-level results are far from the whole truth. In both groups (girls and boys) there are always some who manage education well, and others who cannot cope well with school.

6.4 Cross-national measurement invariance

Examination of the cross-national measurement invariance of the HLSAC instrument showed that the instrument's psychometric properties were at an adequate level. The internal consistency of the instrument was at a sound level in each of the four countries in which it was administered. Inspection of the distributions, factor structure, and reliability estimates showed no compelling evidence that the original items of the HLSAC instrument should be changed. The brief 10-item instrument predicted 95–97% of the variance of the longer 18-item instrument, depending on the country.

The fit with the data on the CFA models was at an adequate level in every country, being highest in Slovakia. Configural and metric invariance was established. The factor loadings were equivalent in Finland, Poland, and Belgium at age 15, and in Finland and Poland at age 13. In both age-groups, the models were slightly poorer when the Slovakian factor loadings were set to be equal with those in the other countries. It must be borne in mind that overall, it is very difficult to attain full measurement invariance; indeed this applies to most empirical studies (Van De Schoot et al. 2015; Davidov, Muthen & Schmidt 2018; Marsh et al. 2018). As Marsh et al. (2018) have recently noted, "scalar invariance is an unachievable ideal that in practice can only be approximated". They nevertheless presented some tentative solutions to the issue, while making it clear that these should be examined further via applied research and simulations.

There can be many reasons why full measurement invariance is not achieved. The respondents' understanding of the questions can vary, and a lack of equivalence can affect, for example, how well the adaptation (i.e. the translation) of the instrument has succeeded; aspects of social desirability, or of familiarity with the item response formats can also play a part (Oberski Weber & Revilla 2012; Davidov et al. 2014). As indicated above, the original English version of questionnaire was translated into the native language of each country. In association with the translation process or with cultural differences, the meanings of the questions or of the response options might have differed to some extent in the target languages.

In connection with the above, it is important to bear in mind how challenging it is to develop a meaningful conceptual framework, and an instrument that is internationally comparable. Harmonization and translation across countries needs considerable effort and sufficient financial recourses. On the other hand, cross-national research does have strong potential: it can contribute to conceptual and methodological development and mutual learning through collaboration, the exchange of ideas among researchers, and collective problem solving. In the best case, this can raise awareness of the embedded nature of concepts that are associated with e.g. social or historical factors, and bring to the fore assumptions that may have been taken for granted.

The comparisons of the HLSAC instrument mean values highlighted significant differences between the countries in both age groups, the only exceptions being Slovakia-Belgium and Poland-Belgium at age 15. In both age groups subjective HL was highest in Finland. Among the 15-year-olds, Finland was followed by Slovakia, Belgium, and Poland, while among the 13-year-olds, Finland was followed by Slovakia and Poland. The systematic statistically significant differences between Finland and the other countries could have several reasons. One major factor could be the school system. In the Finnish comprehensive school, health issues are taught within an independent and obligatory school subject labelled health education. The national core curriculum details learning objectives that guide the teacher's work, with the result that in principle, every pupil should receive the same amount of health education during his/her school path. Health issues are also addressed in a number of other school subjects; in addition, general health promotion (e.g. via the school nurse, school welfare officer, catering personnel, the involvement of parents) contributes to knowledge on health-related matters.

6.5 Limitations of the study

Certain limitations should be noted in this research. It can be argued that self-reported measurements may produce over-estimated results, in that the participants will tend to give socially desirable responses (Altin et al. 2014). Here it is worth considering the nature and construction of subjective HL instruments, and the question of whether items intended to test HL as a latent construct may actually be "too easy ", insofar as participants may prefer to choose alternatives (response options) that are slightly more favourable. This present study does not contain any comparison of subjective and performance-based HL measurements. Thus, it is not possible to compare results for the same respondents, or to gain insights on whether this self-reported instrument would give different scores from those obtained via a performance-based measurement. However, it should be noted that subjective instruments that are based on self-efficacy (perceived competence) have been found to work well (Bandura 2004; Conner & Norman 2005).

It should be noticed the nature of general and comprehensive instrument; it gives good overview about HL, whereas domain specific instruments can give more focused picture of certain HL pattern.

The HLSAC instrument was based on a single conception about constituent HL components. Because variations in conceptualization remain, it is possible that these could generate a range of different instruments. Indeed, this situation is reflected in the increasing number of HL instruments. These have the potential to enrich HL research, while at the same time causing difficulties in comparing research findings.

The target group requires careful consideration. The present instrument was developed and validated for persons aged 13–15. Hence, the applicability of the HLSAC instrument for younger or older respondents will require further research.

The cross-national examination of the HLSAC instrument covered four languages, and one language per country. Although the findings indicated that the instrument worked adequately in different settings (i.e. had external validity), it would have been better to include more countries in the study, with possibilities to understand how well the HLSAC instrument performs in different cultural or educational contexts. Due to limited resources this was not possible. Moreover, as noted above, although the translation process was handled carefully, it is challenging to translate the questionnaire in such a way that participants understand the items and response options in a similar way, regardless of country, culture, or language.

The results indicated that the instrument is applicable in the participant countries (having good factor loadings, with models adequately fitting the data). Nevertheless, slight differences in countries did have effects. Configural and metric invariance was established, but as often happens in empirical invariance studies, scalar invariance did not hold. As a result, mean value comparisons should be conducted with a degree of caution.

6.6 Future perspectives

The implications of the study relate to both research and practical issues. Overall, research on the measurement of school-aged children´s HL is very important. HL has consequences at both the individual and the societal level, underlining the need for measurement, and increasing demands for HL measuring instruments. In future, if will be important to investigate HLSAC measurement invariance in more countries and age groups. Translation process could constitute one factor preventing the achievement of scalar invariance, so future research will have to give greater attention to this issue. Standardization of the translation process can be aided by qualitative methods, in conjunction with quantitative approaches that can shed light on how respondents in different countries and age groups understand items and response options. In addition, there is a need to research the differences between self-reported and performance-based HL instruments. Exploration of this issue will need specifically designed studies, with examination of the same target group via several HL instruments.

Trustworthy measurements create a basis for the assessment of HL levels across a range of settings, leading to appropriate national and international interventions, and to meaningful concrete practices. Measurement is essential for the progress of HL research, and importantly, for the health of the children and adolescents themselves. Individuals who have adequate HL will have a wider range of opportunities and options for health and health equality. They will have the skills to improve their own health, to change health behaviours, to avoid health risks, and to influence others towards healthy decisions. In the field of public health and health promotion, population group comparisons are frequently a point of interest. According to this study, the HLSAC instrument has good applicability for these purposes. Overall, it functions as promising tool for comparing the subjective HL of school-aged children in international contexts.

REFERENCES

Abel, T. 2007. Cultural capital in health promotion. In McQueen, D.V. & Kickbusch, I. (eds) Health and modernity: the role of theory in health promotion. New York: Springer, 43–73.

Abel, T. 2008. Measuring health literacy: moving towards a health-promotion perspective. International Journal of Public Health, 53 (4), 169-170.

Aira, T., Välimaa, R., Paakkari, L., Villberg, J. & Kannas, L. 2014. Finnish pupils' perceptions of health education as a school subject. Global Health Promotion, 21 (3), 6–18.

Altin, S.V., Finke, I., Kautz-Freimut, S. & Stock, S. 2014. The evolution of health literacy assessment tools: a systematic review. BMC Public Health, 14 (1), 1207.

Bandura, A. 1997. Self-efficacy: The exercise of control. New York: Freeman & Co.

Bandura, A. 2004. Health promotion by social cognitive means. Health Education & Behavior, 31 (2), 143–164.

Bandura, A. 2006. Guide for constructing self-efficacy scales. In Pajares, F. & Urdan, T. (eds) Self-efficacy beliefs of adolescents. Greenwich: Information Age Publishing, 307–337.

Batterham, R., Hawkins, M., Collins, P.A., Buchbinder, R. & Osborne, R.H. 2016. Health literacy: applying current concepts to improve health services and reduce health inequalities. Public Health, 132, 3–12.

Begoray, D.L., Wharf-Higgins, J. & MacDonald, M. 2009. High school health curriculum and health literacy: Canadian student voices. Global Health Promotion, 16 (4), 35-42.

Bentler, P.M. & Bonnet, D.C. 1980. Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88 (3), 588–606.

Bereiter, C. & Scardamalia, M. 1993. Surpassing ourselves – an inquiry into the nature and implications of expertise. Chicago (IL): Open Court.

Berkman, N.D., Sheridan, S.L., Donahue, K.E., Halpern, D.J. & Crotty, K. 2011. Low health literacy and health outcomes: an updated systematic review. Annals of Internal Medicine, 155 (2), 97–107.

Bowling, A. 2005. Mode of questionnaire administration can have serious effects on data quality. Journal of Public Health, 27 (3), 281–91.

Bradburn, N., Sudman, S. & Wansink, B. 2004. Askin questions: the definitive guide to questionnaire design – for market research, political polls, and social and health questionaires, revised edition. San Francisco: Jossey-Bass, A Wiley Imprint.

Brown, T. A. 2006. Confirmatory factor analysis for applied research. New York: Guilford.

Bröder, J., Okan, O., Bauer, U., Bruland, D., Schlupp, S., Bollweg, T.M., Saboga-Nunes, L., Bond, E., Sørensen, K., Bitzer, E.M., Jordan, S., Domanska, O., Firnges, C., Carvalho, G.S., Bittlingmayer, U.H., Levin-Zamir, D., Pelikan, J., Sahrai, D., Lenz, A., Wahl, P., Thomas, M., Kessl, F. & Pinheiro, P. 2017. Health literacy in childhood and youth: a systematic review of definitions and models. BMC Public Health, 17 (1), 361.

Byrne, B.M. 2004. Testing for multigroup invariance using AMOS graphics: a road less traveled. Structural Equation Modeling, 11 (2), 272–300.

Byrne, B.M., Shavelson, R.J. & Muthen, B. 1989. Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. Psychological Bulletin, 105 (3), 456–466.

Cangur, S. & Ercan, I. 2015. Comparison of model fit indices used in structural equation modeling under multivariate normality. Journal of Modern Applied Statistical Methods, 14(1), 152–167.

Chinn, D. 2011. Critical health literacy: a review and critical analysis. Social Science & Medicine, 73 (1), 60–67.

Chisolm, D.J. & Buchanan, L. 2007. Measuring adolescent functional health literacy: a pilot validation of the test of functional health literacy in adults. Journal of Adolescent Health, 41 (3), 312–314.

Conner, M. & Norman, P. 2005. Predicting health behavior: a social cognition approach. In Conner, M. & Norman, P. (eds) Predicting health behaviour: research and practice with social cognition models. 2nd edition. Maidenhead: Open University Press, 1–27.

Currie, C., Inchley, J., Molcho, M., Lenzi, M., Veselska, Z. & Wild, F. (eds) 2014. Health Behaviour in School-Aged Children (HBSC) study protocol: background, methodology and mandatory items for the 2013/14 survey. St Andrews: Child and Adolescent Health Research Unit (CAHRU).

Currie, C., Nic Gabhainn, S., Godeau, E. & the International HBSC Network Coordinating Committee. 2009. The Health Behaviour in School-aged Children: WHO Collaborative Cross-National (HBSC) Study: origins, concept, history and development 1982–2008. International Journal of Public Health, 54 (2), 131-139.

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P. & Billiet, J. 2014. Measurement equivalence in cross-national research. Annual Review of Sociology, 40 (1), 55–75.

Davidov, E., Muthen, B. & Schmidt, P. 2018. Measurement invariance in cross-national studies: challenging traditional approaches and evaluating new ones. Sociological Methods & Research, 47 (4), 631–636.

Davis, T.C., Crouch, M.A., Long, S.W., Jackson, R.H., Bates, P., George, R.B. & Bairnsfather, L.E. 1991. Rapid assessment of literacy levels of adult primary care patients. Family Medicine, 23 (6), 433-435.

Davis, T.C., Wolf, M.S., Arnold, C.L., Byrd, R.S., Long, S.W., Springer, T., Kennen, E. & Bocchini, J.A. 2006. Development and validation of the rapid estimate of adolescent literacy in medicine (REALM-Teen): a tool to screen adolescents for below-grade reading in health care settings. Pediatrics, 118 (6), 1707–1714.

DeVellis, R.F. 2003. Scale development: theory and applications. Thousand Oaks: SAGE Publications.

de Leeuw, E. 2012. The political ecosystem of health literacies. Health Promotion International, 27 (1), 1-4.

de Vet, H.C.W., Terwee, C.B., Mokkink, L.B. & Knol, D.L. 2015. Measurement in medicine. Practical guides to biostatistics and epidemiology. Cambridge: Cambridge University Press.

Duckworth, A. L. & Seligman, M. E. P. 2006. Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. Journal of Educational Psychology, 98 (1), 198—208.

Finbråten, H.S., Pettersen, K.S., Wilde-Larsson, B., Nordström, G., Trollvik, A. & Guttersrud, Ø. 2017. Validating the European Health Literacy Survey Questionnaire in people with type 2 diabetes: Latent trait analyses applying multidimensional Rasch modelling and confirmatory factor analysis. Journal of Advanced Nursing, 73 (11), 2730–2744.

Finnish Advisory Board on Research Integrity. 2012. Responsible conduct of research and procedures for handling allegations of misconduct in Finland. Guidelines of the Finnish Advisory Board on Research Integrity. Available at: https://www.tenk.fi/sites/tenk.fi/files/HTK_ohje_2012.pdf (accessed 7 April 2019).

Finnish National Board of Education. 2014. National core curriculum for basic education 2014. Helsinki: Next Print.

Finnish National Board of Education. 2015. National core curriculum for general upper secondary schools 2015. Helsinki: Next Print.

Flay, B.R., Allred, C.G. & Ordway, N. 2001. Effects of the positive action program on achievement and discipline: two matched-control comparisons. Prevention Science, 2 (2), 71-89.

Ghanbari, S., Ramezankhani, A., Montazeri, A. & Mehrabi, Y. 2016. Health Literacy Measure for Adolescents (HELMA): Development and Psychometric Properties. PLoS One, 11 (2), e0149202.

Goldstein, H. 1995. Multilevel statistical models. London: Arnold.

Guo, S., Armstrong, R., Waters, E., Sathish, T., Alif, S. M., Browne, G. R. & Yu, X. 2018. Quality of health literacy instruments used in children and adolescents: a systematic review. BMJ open, 8 (6), e020080.

Guzys, D., Kenny, A., Dickson-Swift, V. & Threlkeld, G. 2015. A critical review of population health literacy assessment. BMC Public Health, 15 (1), 215.

Haun, J.N., Valerio, M.A., McCormack, L.A., Sørensen, K. & Paasche-Orlow, M.K. 2014. Health literacy measurement: an inventory and descriptive summary of 51 instruments. Journal of Health Communication, 19 (2), 302–333.

Horn, J.L. & McArdle, J.J. 1992. A practical and theoretical guide to measurement invariance in aging research. Experimental Aging Research, 18 (3-4), 117–144.

Howard, D.H., Gazmararian, J. & Parker, R.M. 2005. The impact of low health literacy on the medical costs of medicare managed care enrollees. The American Journal of Medicine, 118 (4), 371–377.

Hu, L.T. & Bentler, P.M. 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Modelling, 6 (1), 1–55.

Inchley, J., Currie, D., Young, T., Samdal, O., Torsheim, T., Augustson, L., Mathison, F., Aleman-Diaz, A., Molcho, M., Weber, M. and Barnekow, V. (eds) 2016. Growing up unequal: gender and socioeconomic differences in young people's health and well-being. Health Policy For Children and Adolescents, No 7. Copenhagen: World Health Organization.

Jordan, J.E., Osborne, R.H. & Buchbinder, R. 2011. Critical appraisal of health literacy indices revealed variable underlying constructs, narrow content and psychometric weaknesses. Journal of Clinical Epidemiology, 64 (4), 366-379.

Jöreskog, K. G. 1971. Simultaneous factor analysis in several populations. Psychometrika, 36 (4), 409–426.

Kenney-Benson, G. A., Pomerantz, E. M., Ryan, A. M. & Patric H. 2006. Sex differences in math performance: The role of children’s approach to schoolwork. Developmental Psychology, 42 (1), 11—26.

Kickbusch I., Pelikan J.M., Apfel, F. & Tsouros, A.D. 2013. Health literacy: The solid facts. Copenhagen: World Health Organization.

Kickbusch, I., Wait, S. & Maag, D. 2006. Navigating health: The role of health literacy. London, Alliance for Health and the Future, International Longevity Centre-UK.

Kiechle, E. S., Bailey, S. C., Hedlund, L. A., Viera, A. J. & Sheridan, S. L. 2015. Different measures, different outcomes? A systematic review of performance-based versus self-reported measures of health literacy and Numeracy. Journal of General Internal Medicine, 30 (10), 1538-1546.

Kline, R.B. 2005. Principles and Practice of Structural Equation Modeling. 2nd edition. New York: The Guilford Press.

Little, T. D. & Slegers, D. W. 2005. Factor analysis: multiple groups. In Everitt, B.S., Howell, D.C. (eds) Encyclopedia of statistics in behavioral science. Chichester: Wiley, 617–623.

Malloy-Weir, L., Charles, C., Gafni, A. & Entwistle, V. 2016. A review of health literacy: Definitions, interpretations, and implications for policy initiatives. Journal of Public Health Policy, 37 (3), 334–352.

Manganello, J., DeVellis, R., Davis, T. & Schottler-Thal, C. 2015. Development of the Health literacy Assessment Scale for Adolescents (HAS-A). Journal of Communication in Healthcare, 8 (3), 172-184.

Marsh, H.W., Balla, J.R. & McDonald, R.P. 1988. Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. Psychological Bulletin, 103 (3), 391–410.

Marsh, H.W., Guo, J., Parker, P.D., Nagengast, B., Asparouhov, T., Muthén, B. & Dicke, T. 2018. What to do when scalar invariance fails: The extended alignment method for multi-group factor analysis comparison of latent means across many groups. Psycholgical Methods, 23 (3), 524–545.

Marsh, H.W., Hau, K.T. & Wen, Z. 2004. In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's findings. Structural Equation Modeling, 11 (3), 320–41.

McCormack, L., Haun, J., Sørensen, K. & Valerio, M. 2013. Recommendations for advancing health literacy measurement. Journal of Health Communnication, 18 (1), 9–14.

Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., Stratford, P.W., Knol, D.L., Bouter, L.M. & de Vet, H.C. 2010. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement

properties for health-related patient-reported outcomes. Journal of Clinical Epidemiology, 63 (7), 737–745.
Niemand, T. & Mai, R. 2018. Flexible cutoff values for fit indices in the evaluation of structural equation models. Journal of the Academy of Marketing Science, 46 (6), 1148–1172.
Nutbeam, D. 1998. Health promotion glossary. Health Promotion International, 13 (4), 349-364.
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.
Nutbeam, D. 2008. The evolving concept of health literacy. Social Science and Medicine, 67 (12), 2072-2078.
Oberski, D.L., Weber, W. & Revilla, M. 2012. The effect of individual characteristics on reports of socially desirable attitudes towards immigration. In Salzborn, S., Davidov, E. & Reinecke, J. (eds) Methods, theories, and empirical applications in the social sciences: festschrift for Peter Schmidt. Wiesbaden: Springer, 151–158.
OECD. 2015. The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence. PISA. OECD Publishing. Available at: https://www.oecd.org/pisa/keyfindings/pisa-2012-results-gender-eng.pdf (accessed 29 April 2019)
OECD. 2018. The future of education and skills. Education 2030. OECD Publishing. Available at: https://www.oecd.org/education/2030/E2030PositionPaper(05.04.2018).pdf (accessed 2 May 2019)
Okan, O. 2019. From Saranac Lake to Shanghai: A brief history of health literacy. In O. Okan, U. Bauer, D. Levin-Zamir, P. Pinheiro and K. Sørensen (eds) International handbook of health literacy. Research, practice and policy across the lifespan. Bristol: Policy Press, 21–37.
Okan, O., Lopes, E., Bollweg, T.M., Bröder, J., Messer, M., Bruland, D., Bond, E., Carvalho, G.S., Sørensen, K., Saboga-Nunes, L., Levin-Zamir, D., Sahrai, D., Bittlingmayer, U.H., Pelikan, J.M., Thomas, M., Bauer, U. & Pinheiro, P. 2018. Generic health literacy measurement instruments for children and adolescents: a systematic review of the literature. BMC Public Health, 18 (1), 166.
Osborne, R.H., Batterham, R.W., Elsworth, G.R., Hawkins, M. & Buchbinder, R. 2013. The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). BMC Public Health, 13 (1), 658.
Ottawa Charter for Health Promotion. 1986. World Health Organization. Available at: http://www.euro.who.int/__data/assets/pdf_file/0004/129532/Ottawa_Charter.pdf?ua=1 (accessed 5 March 2020)
Paakkari, L. & George, S. 2018. Ethical underpinnings for the development of health literacy in schools: ethical premises (‘why’), orientations (‘what’) and tone (‘how’). BMC Public Health, 18 (1), 326.
Paakkari, L. & Paakkari, O. 2012. Health literacy as a learning outcome in schools. Health Education, 112 (2), 133-152.

Paakkari, L., Torppa, M., Paakkari, O., Välimaa, R., Ojala, K. & Tynjälä, J. 2019b. Does health literacy explain the link between structural stratifiers and adolescent health? European Journal of Public Health, 1-6. [Epub ahead of print]

Paakkari O, Torppa M, Boberova Z, Välimaa R, Maier G, Mazur J, Kannas L, Paakkari L. 2019. The cross-national measurement invariance of the health literacy for school-aged children (HLSAC) instrument. European Journal of Public Health, 29 (3), 432–436.

Paakkari, O., Torppa, M., Kannas, L. & Paakkari, L. 2016. Subjective health literacy: Development of a brief instrument for school-aged children. Scandinavian Journal of Public Health, 44 (8), 751-757.

Paakkari, O., Torppa, M., Villberg, J., Kannas, L. & Paakkari, L. 2018. Subjective health literacy among school-aged children. Health Education, 118 (2), 182–195.

Parker, R.M., Baker, D.W., Williams, M.V. & Nurss, J.R. 1995. The test of functional health literacy in adults: a new instrument for measuring patients' literacy skills. Journal of General Internal Medicine, 10 (10), 537–541.

Perry, E.L. 2014. Health literacy in adolescents: an integrative review. Journal for Specialists in Pediatric Nursing, 19 (3), 210-218.

Pleasant, A. 2014. Advancing health literacy measurement: a pathway to better health and health system performance. Journal of Health Communication, 19 (12), 1481–1496.

Pleasant, A. & Kuruvilla, S. 2008. A tale of two health literacies: public health and clinical approaches to health literacy. Health Promotion International, 23 (2), 152–159.

Pleasant, A., McKinney, J. & Rikard, R.V. 2011. Health literacy measurement: a proposed research agenda. Journal of Health Communication, 16 (3), 11–21.

Ratzan, S.C. & Parker, R.M. 2000. Introduction. In C.R. Selden, M. Zorn and R.M. Ratzan (eds) National library of medicine current bibliographies in medicine: health literacy. NLM Pub. No. CMB 2000-1. Bethesda, MD: National Institutes of Health, U.S. Department of Health and Human Services, v–vi.

Selkälä, A. 2013. Kansalaisuuden hallinta suomalaisessa sosiaali- ja terveyspolitiikassa. Acta Universitatis Lapponiensis 248. Rovaniemi: Lapin yliopistokustannus.

Simonds, S.K. 1974. Health education as social policy. Health Education Monograph, 2, 1-25.

Snijders, T. & Bosker, R. 1999. Multilevel analysis: An introduction to basic and advanced multilevel modeling. London: SAGE Publications.

Steenkamp, J.B. & Baumgartner, H. 1998. Assessing measurement invariance in cross-national consumer research. Journal of Consumer Research, 25 (1), 78–90.

St Leger, L.S. & Nutbeam, D. 2000. A model for mapping linkages between health and education agencies to improve school health. Journal of School Health, 70 (2), 45-50.

Stoet, G. & Geary, D. C. 2013. Sex differences in mathematics and reading achievement are inversely related: within- and across-nation assessment of 10 years of PISA data. PloS One, 8 (3), e57988.

Streiner, D.L., Norman, G.R. & Cairney, J. 2015. Health measurement scales: a practical quide to their development and use. Fifth edition. Oxford: Oxford University Press.

Sykes, S., Wills, J., Rowlands, G. & Popple, K. 2013. Understanding critical health literacy: a concept analysis. BMC Public Health, 13 (1), 150.

Schwarzer, R. & Fuchs, R. 1995. Self-efficacy and health behaviour. In Conner M. & Norman P. (eds) Predicting health behavior. Buckingham: Open University Press, 163–196.

Sijtsma, K. 2009. On the use, the misuse, and the very limited usefulness of Cronbach's alpha. Psychometrika, 74 (1), 107–120.

Sørensen, K., Pelikan, J. M., Röthlin, F., Ganahl, K., Slonska, Z., Doyle, G., Fullam, J., Kondilis, B., Agrafiotis, D., Uiters, E., Falcon, M., Mensing, M., Tchamov, K., van den Broucke, S. & Brand, H. 2015. Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU). European Journal of Public Health, 25 (6), 1053–1058.

Sørensen, K. & Pleasant, A. 2017. Understanding the conceptual importance of the differences among health literacy definitions. Studies in Health Technology and Informatics, 240 (1), 3–14.

Sørensen, K., Van den Broucke, S., Fullam, J., Doyle, G., Pelikan, J., Slonska, S. & Brand, H. 2012. Health literacy and public health: A systematic review and integration of definitions and models. BMC Public Health 12 (1), 80.

Sørensen, K., Van den Broucke, S., Pelikan, J., Fullam, J., Doyle, G., Slonska, Z., Kondilis, B., Stoffels, V., Osborne, R.H., Brand, H. & HLS-EU Consortium. 2013. Measuring health literacy in populations: illuminating the design and development process of the European Health Literacy Survey Questionnaire (HLS-EU-Q). BMC Public Health, 13 (1), 948.

Tassi, A. 2004. The emergence of health literacy as a public policy priority: from research to consensus to action. Literacy Harvest, 11 (1), 5–10.

Traina, G., Martinussen, P.E. & Feiring, E. 2019. Being healthy, being sick, being responsible: attitudes towards responsibility for health in a public healthcare system. Public Health Ethics, 12 (2), 145–157.

UNESCO. 2014. UNESCO education strategy 2014-2021. Paris: United Nations Educational, Scientific and Cultural Organization.

Vandenberg, R.J. & Lance, C.E. 2000. A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. Organizational Research Methods, 3 (1), 4–70.

Van De Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K. & Zondervan-Zwijnenburg, M. 2015. Editorial: Measurement Invariance. Frontiers in Psychology, 6 (883), 1064.

Van den Broucke, S. 2014. Health literacy: a critical concept for public health. Archives of Public Health, 72 (1), 10.

van der Heide, I., Uiters, E., Sørensen, K., Röthlin, F., Pelikan, J., Rademakers, J., Boshuizen, H. & EPHORT consortium. 2016. Health literacy in europe: the development and validation of health literacy prediction models. European Journal of Public Health, 26 (6), 906–911.

Vaz, S., Parsons, R., Passmore, A.E., Andreou, P. & Falkmer, T. 2013. Internal consistency, test-retest reliability and measurement error of the self-report version of the social skills rating system in a sample of Australian adolescents. PLoS One, 8 (9), e73924.

Velardo, S. & Drummond, M. 2017. Emphasizing the child in child health literacy research. Journal of Child Health Care, 21 (1), 5–13.

Volandes, A.E. & Paasche-Orlow, M.K. 2007. Health literacy, health inequality and a just healthcare system. The American Journal of Bioethics, 7 (11), 5–10.

Wagner, T. 2008. Rigor redefined. Educational Leadership, 66 (2), 20-24.

Wharf-Higgins, J. 2012. Preface: A health literacy syllabus. In R. Marks (ed) Health literacy and school-based health education. Bingley: Emerald Group Publishing Limited, xi—xvi.

Weiss, B.D., Mays, M.Z., Martz, W., Castro, K.M., DeWalt, D,A., Pignone, M.P., Mockbee, J. & Hale, F.A. 2005. Quick assessment of literacy in primary care: the Newest Vital Sign. Annals of Family Medicine, 3 (6), 514–522.

Whitehead, M. 1991. The concepts and principles of equity and health. Health Promotion International, 6 (3), 217–228.

World Health Organization. 1998. Health promotion glossary. Geneva: World Health Organization.

Appendix 1 Classification of health literacy instruments used in children and adolescents (Guo et al. 2018)

Table 2 General and important characteristics of included instruments used in children and adolescents

No	HL instrument	HL domain and component (item number)	Participant characteristics consideration	Health topic and content (readability level)	Response category	Scoring system	Burden	Administration form
1	NVS[39 50 52]	Functional HL 1. Reading comprehension (2). 2. Numeracy (4).	Demographic patterns.	Nutrition-related information about the label of an ice cream container (na).	Open-ended.	Score range: 0–6; ordinal category: 0–1: high likelihood of limited literacy; 2–3: possibility of limited literacy; 4–6: adequate literacy.	No longer than 3 min.	Interviewer-administered and performance-based.
2	TOFHLA[48]	Functional HL 1. Reading comprehension (50). 2. Numeracy (17).	Developmental change. Demographic patterns.	Instruction for preparation for an upper gastrointestinal series (4.3 grade), a standard informed consent form (10.4 grade), patients' rights and responsibilities section of a Medicaid application form (19.5 grade), actual hospital forms and labelled prescription vials (9.4 grade).	4 response options.	Score range: 0–100; ordinal category: 0–59: inadequate health literacy; 60–74: marginal health literacy; 75–100: adequate health literacy.	12.9 min (8.9–17.3 min).	Interviewer-administered and performance-based.
3	s-TOFHLA[50]	Functional HL 1. Reading comprehension (36).	Demographic patterns.	Instruction for preparation for an upper gastrointestinal series (fourth grade), patients' rights and responsibilities section of a Medicaid application form (10th grade).	4 response options	Score range: 0–36; ordinal category: 0–16: inadequate literacy; 17–22: marginal literacy; 23–36: adequate literacy.	na.	Interviewer-administered and performance-based.
4	c-sTOFHLAd[51]	Functional HL 1. Reading comprehension (36).	Developmental change. Demographic patterns.	Instruction for preparation for an upper gastrointestinal series (fourth grade), patients' rights and responsibilities section of a Medicaid application form (10th grade).	4 response options.	Score range: 0–36; ordinal category: 0–16: inadequate literacy; 17–22: marginal literacy; 23–36: adequate literacy.	20 min class period.	Self-administered and performance-based.
5	REALM-Teen[41 50]	Functional HL 1. Reading recognition (66).	Developmental change. Demographic patterns.	66 health-related words such as weight, prescription and tetanus (sixth grade).	Open-ended.	Score range: 0–66; ordinal category: 0–37: ≤3rd; 38–47: 4th–5th; 48–58: 6th–7th; 59–62: 8th–9th; 63–66: ≥10th	2–3 min.	Interviewer-administered and performance-based.
6	HLAB[40]	Functional and critical HL 1. Understanding health information (30). 2. Evaluating health information (17).	Developmental change. Demographic patterns.	A range of topics such as nutrition and sexual health (pilot-tested).	Open-ended.	Score range: 0–107; continuous score.	Two regular classroom sessions.	Self-administered and performance-based.
7	MMAHL[44]	Functional, interactive and critical HL 1. Patient–provider encounter (4). 2. Interaction with the healthcare system (5). 3. Rights and responsibilities (7). 4. Confidence in using health information from personal source (3). 5. Confidence in using health information from media source (3). 6. Health information seeking competency using the internet (2).	Developmental change. Demographic patterns. Dependency.	Experiences of how to access, navigate and manage one's healthcare and preventive health needs (sixth grade).	5-point Likert scale.	Score range: na; continuous score.	na.	Self-administered and self-reported.
8	MHL[49]	Functional, interactive and critical HL 1. Content identification (6). 2. Perceived influence on behaviour (6). 3. Critical analysis and intended (6). 4. Action/reaction (6).	Dependency. Demographic patterns.	Nutrition/dieting, physical activity, body image, sexual activity, cigarette smoking, alcohol consumption, violent behaviour, safety habits and/or friendship and family connectedness (pilot-tested).	Open-ended and multiple choice.	Score range: 0–24; continuous score.	na.	Video-assisted interviewer-administered and performance-based.
9	DNT-39[53]	Functional health literacy 1. Health numeracy (39).	Differential epidemiology. Demographic patterns.	Nutrition, exercise, blood glucose monitoring and insulin administration (na).	Open-ended.	Score range: 0–100; continuous score.	na.	Interviewer-administered and performance-based.
10	DNT-14[53]	Functional health literacy 1. Health numeracy (14).	Differential epidemiology. Demographic patterns.	Nutrition, exercise, blood glucose monitoring and insulin administration (na).	Open-ended.	Score range: 0–100; continuous score.	na.	Interviewer-administered and performance-based.
11	eHEALS[43]	Functional, interactive and critical HL 1. Accessing health information (4). 2. Evaluating health information (2). 3. Applying health information (2).	Developmental change. Dependency. Demographic patterns.	General health topics about online health information (pilot-tested).	5-point Likert scale.	Score range: na; continuous score	na.	Self-administered and self-reported.
12	CHC Test[47]	Critical HL 1. Understanding medical concepts (15). 2. Searching literature skills (22). 3. Basic statistics (18). 4. Design of experiments and sampling (17).	Developmental change. Demographic patterns.	Echinacea and common cold, MRI in knee injuries, treatment of acne, breast cancer screening (pilot-tested).	Open-ended and multiple choice.	na.	Less than 90 min.	Interviewer-administered and performance-based.
13	HKACSS[46]	Functional and interactive HL 1. Health knowledge (3). 2. Health attitudes (4). 3. Health communication (3). 4. Self-efficacy (3).	Developmental change. Dependency. Demographic patterns.	Physical activities, nutrition, smoking, vaccination, tooth health and general health (na).	2 response options; 5-point Likert scale; 4-point Likert scale.	Score range: na; continuous score.	na.	Self-administered and performance-based and self-reported.
14	HLAT-51[42]	Functional, interactive and critical HL 1. Comprehension skill (20). 2. Health numeracy (11). 3. Media literacy (8). 4. Digital literacy (12).	Developmental change. Dependency. Demographic patterns.	Health topics such as gout and uric acid, high cholesterol and triglyceride levels, health information-seeking skills (na).	Yes/no; multiple choice.	na.	30–45 min.	Self-administered and performance-based and self-reported.
15	HLAT-8[45]	Functional, interactive and critical HL 1. Understanding health information (2). 2. Finding health information (2). 3. Interactive health literacy (2). 4. Critical health literacy (2).	Dependency. Demographic patterns.	General health topics in people's daily life (na).	5-point Likert scale; 4-point Likert scale.	Score range: 0–37; continuous score.	na.	Self-administered and self-reported.
16	CHLT[54]	Functional, interactive and critical HL 1. Health knowledge (11). 2. Health attitude (16). 3. Health skills (5).	Developmental change. Dependency. Demographic patterns.	Personal hygiene, growth and ageing, sexual education and mental health, healthy eating, safety and first aid, medicine safety, substance abuse prevention, health promotion and disease prevention, consumer health, health and environment (pilot-tested).	Multiple choice.	Score range: 0–32; continuous score.	na.	Self-administered and performance-based.
17	VOHL[55]	Functional HL 1. Health knowledge (2).	Developmental change	Oral health for tooth and gingiva (na).	Visual drawing.	Score range: 0–6; continuous score.	na.	Self-administered and performance-based.

Table 2 Continued

No	HL instrument	HL domain and component (item number)	Participant characteristics consideration	Health topic and content (readability level)	Response category	Scoring system	Burden	Administration form
18	HAS-A[56]	Functional, interactive and critical HL 1. Understanding health information (6). 2. Communication health information (5). 3. Confusion about health information (4).	Developmental change. Demographic patterns. Dependency.	General health topics in daily life (pilot-tested).	5-point Likert scale.	Score range: 0–24 (understanding), 0–20 (communication), 0–16 (confusion); continuous score.	na.	Self-administered and self-reported.
19	MaHeLi[57]	Functional, interactive and critical HL 1. Health seeking-behaviour (1). 2. Competence and coping skills (6). 3. Appraisal of health information (5).	Developmental change. Demographic patterns. Dependency. Differential epidemiology.	General health and maternal health topics (na).	6-point Likert scale.	na.	na.	Interviewer-administered and self-reported.
20	QuALiSMental[64]	Functional, interactive and critical HL 1. Recognition disorders (14). 2. Knowledge about the professionals and treatments available (16). 3. Knowledge of the effectiveness of self-help strategies (12). 4. Knowledge and skills needed to provide support and first aid to others (10). 5. Knowledge of how to prevent mental disorders (8).	Developmental change. Demographic patterns. Dependency.	Mental health vignettes (na).	Yes/no; multiple choice.	na.	40–50 min.	Self-administered and performance-based.
21	FCCHL-AYAC[65]	Functional, interactive and critical HL 1. Functional HL (6). 2. Communicative HL (3). 3. Critical HL (4).	Developmental change. Dependency. Differential epidemiology.	Health topics regarding cancer in daily life (second grade).	4-point Likert scale.	Score range: 13–52; continuous score.	na.	Self-administered and self-reported.
22	ICHL[58]	Interactive and critical HL 1. Interactive HL (2). 2. Critical HL (7).	Developmental change. Demographic patterns. Dependency. Differential epidemiology.	General health topics in daily life (pilot-tested).	5-point Likert scale.	na.	40–55 min (together with other measures).	Self-administered and self-reported.
23	HELMA[59]	Functional, interactive and critical HL 1. Access (5). 2. Reading (5). 3. Understanding (10). 4. Appraise (5). 5. Use (4). 6. Communication (8). 7. Self-efficacy (4). 8. Numeracy (3).	Developmental change. Dependency.	General health topics in daily life (pilot-tested).	5-point Likert scale.	Score range: 0–100; ordinal category: 0–50: inadequate; 50.1–66: problematic; 66.1–84: sufficient; 84.1–100: excellent.	15 min.	Self-administered and self-reported.
24	HLSAC[60]	Functional, interactive and critical HL 1. Theoretical knowledge (2). 2. Practical knowledge (2). 3. Individual critical thinking (2). 4. Self-awareness (2). 5. Citizenship (2).	Developmental change. Dependency.	General health topics in daily life (seventh grade).	4-point Likert scale.	na.	45 min (together with the HBSC survey).	Self-administered and self-reported.
25	REALM-TeenS[66]	Functional HL 1. Reading recognition (10).	Developmental change. Demographic patterns.	10 health-related words such as diabetes (sixth grade).	Open-ended.	Score range: 0–10; ordinal category: 0–2: ≤3rd; 3–4: 4th–5th; 5–6: 6th–7th; 7–8: 8th–9th; 9–10: ≥10th.	13.6 s (range: 7.8–23.0).	Interviewer-administered and performance-based.
26	funHLS-YA[61]	Functional HL 1. Word recognition and comprehension (19).	Developmental change.	Diseases and symptoms, nutrition and diet, biology of the human body (na).	Multiple choice.	Score range: na; continuous score.	5 min.	Self-administered and performance-based.
27	HLS-TCO[62]	Functional, interactive and critical HL 1. Health knowledge and understanding (10). 2. Accessing information and services (5). 3. Communicating skills (6). 4. Managing health conditions (5). 5. Media literacy (5). 6. Making decisions (4).	Developmental change. Demographic patterns. Dependency. Differential epidemiology.	Health information for obesity preventive behaviours (pilot-tested).	Multiple choice; Likert scale.	Score range: 0–135; ordinal category: low:<21 for FHL, <33 for IHL, <27 for CHL; fair: 21–27.99 for FHL, 33–43.99 for IHL, 27–35.99 for CHL; high: 28–35 for FHL, 44–54.9 for IHL, 36–45 for CHL.	na.	Self-administered and performance-based and self-reported.
28	HLRS-Y[63]	Functional, interactive and critical HL 1. Knowledge (10). 2. Self-advocacy/support (14). 3. Resiliency (13).	Developmental change. Demographic patterns. Differential epidemiology.	Health information about chronic health conditions (pilot-tested).	4-point Likert scale.	Score range: na; continuous score.	15–20 min.	Self-administered and self-reported.
29	p_HLAT-8[67]	Functional, interactive and critical HL 1. Understanding health information (2). 2. Searching health information (2). 3. Communicating health information (2). 4. Appraising health information (2).	Dependency.	General health topics in daily life (pilot-tested).	5-point Likert scale; 4-point Likert scale.	Score range: 0–37; continuous score.	na.	Self-administered and self-reported.

Table 2 Continued

c-sTOFHLAd, Chinese version of short-form Test of Functional Health Literacy in Adolescents; CHC Test, Critical Health Competence Test; CHL, critical health literacy; CHLT, Child Health Literacy Test; DNT, Diabetes Numeracy Test; eHEALS, eHealth Literacy Scale; FCCHL-AYAC, Functional, Communicative, and Critical Health Literacy-Adolescents and Young Adults Cancer; FHL, functional health literacy; funHLS-YA, Functional Health Literacy Scale for Young Adults; HAS-A, Health Literacy Assessment Scale for Adolescents; HBSC, Health Behaviour in School-aged Children; HELMA, Health Literacy Measure for Adolescents; HKACSS, Health Knowledge, Attitudes, Communication and Self-efficacy Scale; HL, health literacy; HLAB, Health Literacy Assessment Booklet; HLAT-8, 8-item Health Literacy Assessment Tool; HLAT-51, 51-item Health Literacy Assessment Tool; HLRS-Y, Health Literacy and Resiliency Scale: Youth Version; HLS-TCO, Health Literacy Scale for Thai Childhood Overweight; HLSAC, Health Literacy for School-aged Children; ICHL, Interactive and Critical Health Literacy; IHL, interactive health literacy; MaHeLi, Maternal Health Literacy; MHL, Media Health Literacy; MMAHL, Multidimensional Measure of Adolescent Health Literacy; na, no information available; NVS, Newest Vital Sign; p_HLAT-8, Portuguese version of the 8-item Health Literacy Assessment Tool; QuALiSMental, Questionnaire for Assessment of Mental Health Literacy; REALM-Teen, Rapid Estimate of Adolescent Literacy in Medicine; REALM-TeenS, Rapid Estimate of Adolescent Literacy in Medicine Short Form; s-TOFHLA, short-form Test of Functional Health Literacy in Adults; TOFHLA, Test of Functional Health Literacy in Adults; VOHL, Visual Oral Health Literacy.

ORIGINAL PAPERS

I

SUBJECTIVE HEALTH LITERACY: DEVELOPMENT OF A BRIEF INSTRUMENT FOR SCHOOL-AGED CHILDREN

by

Paakkari, O., Torppa, M., Kannas, L. & Paakkari, L. 2016.

Scandinavian Journal of Public Health, 44 (8), 751–757.

https://doi.org/10.1177/1403494816669639

.

Scandinavian Journal of Public Health, 2016; 44: 751–757

ORIGINAL ARTICLE

Subjective health literacy: Development of a brief instrument for school-aged children

OLLI PAAKKARI, MINNA TORPPA, LASSE KANNAS & LEENA PAAKKARI

University of Jyväskylä, Jyväskylä, Finland

Abstract

Aims: The present paper focuses on the measurement of health literacy (HL), which is an important determinant of health and health behaviours. HL starts to develop in childhood and adolescence; hence, there is a need for instruments to monitor HL among younger age groups. These instruments are still rare. The aim of the project reported here was, therefore, to develop a brief, multidimensional, theory-based instrument to measure subjective HL among school-aged children. *Methods*: The development of the instrument covered four phases: item generation based on a conceptual framework; a pilot study (n = 405); test–retest (n = 117); and construction of the instrument (n = 3853). All the samples were taken from Finnish 7th and 9th graders. *Results*: Initially, 65 items were generated, of which 32 items were selected for the pilot study. After item reduction, the instrument contained 16 items. The test–retest phase produced estimates of stability. In the final phase a 10-item instrument was constructed, referred to as *Health Literacy for School-Aged Children* (HLSAC). The instrument exhibited a high Cronbach alpha (0.93), and included two items from each of the five predetermined theoretical components (theoretical knowledge, practical knowledge, critical thinking, self-awareness, citizenship). ***Conclusions*: The iterative and validity-driven development process made it possible to construct a brief multidimensional HLSAC instrument. Such instruments are suitable for large-scale studies, and for use with children and adolescents. Validation will require further testing for use in other countries.**

Key Words: *Health literacy, measurement, school-aged children, adolescent, subjective*

Introduction

According to a World Health Organization (WHO) report [1] there is a health literacy (HL) crisis in Europe and beyond, due to a mismatch between individuals' competence capital and the increasingly complex requirements posed in taking care of one's health. According to a recent study, over 10% of Europeans have insufficient HL, while almost 50% have limited (insufficient or problematic) HL [2]. Inadequate HL has consequences at both the individual and societal level. It has been reported as an independent risk factor for health [3], being associated, for example, with problems in interpreting health-related information, more frequent hospitalization, incorrect use of medicines, poorer health status, and increased mortality [4] – all this in addition to higher healthcare costs for society as a whole [5].

During childhood, a foundation is laid for many health behaviours, and for overall health and well-being. Moreover, HL itself starts to develop in childhood, and there is, therefore, a need to focus on HL among younger age groups. At the present time, a particularly important issue for children and adolescents is the complexity of the challenges faced in taking care of one's health. The modern environment exposes children and adolescents to many unhealthy behaviours, such as physical inactivity [6] or excessive gain of energy [7]. Increasingly, the Internet and other media expose children to information of varying quality [8], with media

Correspondence: Olli Paakkari, University of Jyväskylä, PO Box 35 (L), Jyväskylä, 40014, Finland. E-mail: olli.paakkari@jyu.fi

(Accepted 23 August 2016)

Reprints and permissions: sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/1403494816669639
sjp.sagepub.com
SAGE

platforms for both positive and negative learning about health issues [9].

There are numerous contested conceptualizations of HL. The definitions vary from a fairly narrow focus on the reading, writing, and numeracy needed in a healthcare or medical context (i.e. functional health literacy) [10], toward a broader focus involving cognitive skills, social skills, and the motivation to promote and maintain health in general [11,12]. It seems unlikely that any definition of HL will work in every situation; thus, any useful or meaningful definition will be highly dependent on the purpose and context of its use. This also holds true for measures of HL. In fact, in parallel with the lack of an agreed HL definition, there is no generally accepted view on how HL should be measured [13].

Just as there have been differences in HL conceptualizations, there have been varying measures of HL. These have ranged from a narrow focus on functional skills, as with the Test of Functional Health Literacy in Adults (TOFHLA) [14], toward broader competence operationalizations, as with the Health Literacy Questionnaire (HLQ) [15]. Moreover, variation exists according to whether one is seeking to measure subjective (i.e. self-reported, perceived) HL, or objective (performance-based) HL. There has been a tendency to attach more value to the objective measures [16]. However, current subjective measures have succeeded in measuring comprehensive (i.e. multidimensional) HL, which is still not the case with many objective tests [17]. Kiechle et al. [13] go so far as to argue that subjective instruments 'could allow more efficient research for health literacy', on the grounds that they can be more easily applied to large samples [18]. For the most part, subjective measures focus on individuals' perceived competence, that is, their self-efficacy. Many of these measures are based on more than three decades of extensive research, showing a clear link between self-efficacy (i.e. perceived competence) and health behaviour [19]. According to Bandura [20] 'efficacy beliefs influence goals and aspirations' and 'the stronger the perceived self-efficacy, the higher the goals people set for themselves and the firmer their commitment to them'. If this is so, it would indeed indicate a need to study how people perceive their competence regarding health-related issues. All in all, one can say that it is crucial to select or develop an instrument that fits with one's purposes, whether that purpose involves focusing on a narrow or broader notion of HL, or an objective or subjective means of measurement.

Currently, we are in need of reliable and valid HL instruments for different age groups and contexts. Despite an increasing interest in measuring HL, tools for examining children's and adolescents' subjective HL at the population level are still rare [21,22]. Here it should be noted that in many studies aimed at measuring adolescents' HL, the instruments applied have not been validated for the target group [23]. Moreover, the majority of studies on adolescents' HL have focused purely on functional HL, highlighting the need for instruments based on a broader notion of HL. Brevity is also a consideration, since the instruments used should be brief enough to be reliably used with children. The length of the instrument becomes particularly important when the aim is to use it in more extensive surveys, involving an examination of the links between HL and other phenomena, such as health behaviour or perceived health among children. The aim of the present study was, thus, to develop a brief, comprehensive, and theory-based instrument for measuring subjective HL among school-aged children.

Methods

The development of the HL instrument comprised four main phases [24]: item generation, piloting, test–retest, and construction of the instrument (Table I).

Item generation (1st phase)

Item generation involved two steps, namely the conceptualization and synthesizing of relevant literature, and the formulation of the items. In the first place, as suggested by Pleasant et al. [25], the instrument was explicitly developed on a testable conceptualization of HL; that is, one which highlights the multidimensional nature of the concept. Due to an interest in assessing (i) multidimensional HL, in (ii) an educational setting, among (iii) school-aged children, Paakkari and Paakkari's [26] conceptualization of HL as a learning outcome was chosen to guide the construction of the items. Their conceptualization of HL and its components is based on an understanding that HL can be explicitly defined, operationalized, and translated into pedagogical practices. This is particularly important when the intention is to move beyond HL assessment towards *how* HL can be developed further among the target group. According to such a conceptualization, HL

> comprises a broad range of knowledge and competencies that people seek to encompass, evaluate, construct and use. Through HL competencies people become able to understand themselves, others and the world in a way that will enable them to make sound health decisions, and to work on and change the factors that constitute their own and others' health chances. [26,136]

Table I. Overview of the HL instrument development process.

Phase	Item reduction
1. Item generation Contextualization, the five core components of HL (theoretical knowledge, practical knowledge, critical thinking, self-awareness, citizenship) Reading and synthesis of relevant literature and existing instruments Formulation of the items Item selection for the pilot study (iterative process of evaluation and discussion by an expert group)	65 items → 32 items
2. Pilot study (*n* = 405, 7th and 9th graders) Data analysis: inspection of distributions, reliability analysis/internal consistency, confirmatory factor analysis, factor loadings and content of the items, examination of model goodness and sufficiency Qualitative item analysis: four class-level discussions on the comprehensibility of the items, reformulation of three items	32 items → 16 items
3. Test–retest (*n* = 117, 7th and 9th graders) A two-week interval, the same pupils, structural equation modelling with one latent HL factor and with the five factors derived from the theoretical core components	16 items
4. Construction of a brief instrument (*n* = 3853, 7th and 9th graders) Data analysis: inspection of distributions, reliability analysis/internal consistency, confirmatory factor analysis, factor loadings and content of the items, examination of model goodness and sufficiency, regression analysis to predict the relationship between 10- and 15-item instruments, 10-item test finalized	16 items → 15 items → 10 items

HL: health literacy.

Table II. The contextualization of HL: the core components [26].

Five core components of HL
The *theoretical knowledge* of health issues contains a range of principles, theories, and conceptual models. Knowledge is viewed as something explicit, factual, universal, formal, and declarative. It includes lower levels of thinking skills, such as remembering.
Practical knowledge (i.e. procedural knowledge, skills) can be seen as a competency that allows one to put theoretical knowledge into practice. Whereas theoretical knowledge is something applicable to many different situations, practical knowledge can be seen as usable only in specific contexts. It is partly rooted in the individual's experiences, and thus it includes tacit, intuitive, or implicit knowledge. Practical knowledge contains basic health skills such as the ability to find health information, the ability to seek health services, and the ability to give first aid.
Individual critical thinking can be understood as the ability to think clearly and rationally. It is based on having a curious and investigative attitude towards the world, and a desire to understand health issues in a deeper way. In practice, critical thinking contains higher-level thinking skills, such as an ability to analyse, evaluate, and create something new; this could include, e.g., the ability to search for the logical connections between health ideas, to solve problems, to argue, to draw conclusions, or to assess the validity of health information.
Self-awareness is the ability to reflect on oneself and to make possible the personal contextualization of health issues. Through self-reflection, the individual becomes conscious of his/her own thoughts, feelings, needs, motives, values, attitudes, and experiences, and is able to consider how these relate to ways of behaving in an individually health-enhancing way. An important part of self-awareness is the ability to reflect on oneself as a learner.
Citizenship involves the ability to take social responsibility, and to think of the probable consequences of one's own actions on others. The ability to act in an ethically responsible way means that individuals are able to consider health issues beyond their own perspective: they may become aware of their own rights and responsibilities, and the effects our actions or thoughts may have on other people or on the environment. The component further includes the ability to identify and work on factors that influence one's own and other people's possibilities to achieve or maintain good health.

HL: health literacy.

This definition of HL incorporates five core components: theoretical knowledge, practical knowledge, individual critical thinking, self-awareness, and citizenship (Table II). These components (in other words, broader competence areas) are partly overlapping, and they are not necessarily fully hierarchical, even if some elements of a hierarchy can be found. The components encompass how HL expands from mere literacy on certain health topics towards a deeper literacy, in terms of realizing the potentials in oneself, others, and the world beyond.

Thereafter, existing theoretical and empirical HL literature (and also relevant health promotion and educational literature) was critically examined by an expert group. The group consisted of researchers in the field of health promotion, education, and psychology. The group members were experienced in teaching at different levels of the educational system (comprehensive school, upper secondary school, and higher education), and in developing national curricula for health education in schools. There was also consultation with external experts in developing HL instruments.

Previous studies on HL instrument development (such as HLQ [15] and the European Health Literacy Survey Questionnaire (HLS-EU-Q) [27]) afforded ideas, patterns, and examples for item design, as did Bloom's taxonomy [28]. The expert group aimed to ensure that the content of the items matched the core components of HL (content validity [29]).

The expert group then generated 65 items covering all five core components of HL. After an iterative process of reading and re-reading, discussion with the expert group, and consultation with external researchers, 32 items were included in the pilot study. The item formulation and the response scale were based on the guidelines for measuring general self-efficacy [30]. Thus, all the items took the form 'I am confident that…', with the response options ranging from 'not at all true' to 'absolutely true' (Likert scale with 4 options). There were no non-committal options (such as 'I can't say' or 'No opinion'). In the instrument itself, the items corresponding to a given component (out of the five) were mixed among items from the other components.

The pilot study (2nd phase)

In the autumn of 2013, the HL instrument with 32 items was piloted in two upper secondary schools in Finland. This formed part of a pilot study for the collaborative cross-national project of the WHO entitled *Health Behaviour in School-aged Children* (HBSC) [31]. The schools were chosen via a discretionary sampling method. The sample included a total of 401 pupils (7th graders, n = 202; 9th graders, n = 199). The pupils filled the survey anonymously in the course of one lesson (lasting 45 minutes).

In analysing the responses, the distributions of all the items were examined, and four items with very low discrimination power were removed. Next, a confirmatory factor analysis (CFA) was applied to the remaining 28 items (in conjunction with theoretical consideration of the item contents) to identify the best items. The model to which the CFA was applied was specified a priori, on the basis of a theoretical conceptualization involving five fixed factors. Items were removed one by one, on the basis of the factor loadings, plus careful consideration of the item content.

The fit of the model to the data was tested via the following: Chi-square test, Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), and Standardized Root Mean Square Residual (SRMR). All the analyses were conducted with Mplus 7.3. Because the data had a hierarchical structure (pupils nested in classrooms), models were built with a COMPLEX option provided in Mplus; this was done to correct for the standard errors in the models according to the nested data structure.

With a view to content validity, all the items were discussed with adolescents, in order to secure maximum comprehensibility and clarity of the items. The discussions were conducted in four classes (two 7th-grade classes and two 9th-grade classes, with 20–24 pupils in each class).

Test–retest (3rd phase)

The version used in the test–retest phase consisted of 16 items. The sample consisted of 117 pupils (7th and 9th graders) from one school. The school was a normal Finnish upper secondary school, and it was selected via discretionary sampling. Fourteen pupils dropped out between the measurements. The test–retest was organized under similar conditions, but with a two-week interval. There are differences in the guidelines for the optimum test–retest interval, but it has been argued that a shorter interval may help students to remember their answers, and that a longer interval between the measurements can indicate the stability of the phenomenon rather than the reliability of the instrument [24].

Response consistency over time [29] was examined via structural equation modelling (SEM), using Mplus 7.3. The latent HL factor for the 16 items was estimated for both time points. In the model, the retest latent factor was regressed on the latent HL factor at the first assessment time point. No error covariances between the items were allowed in the models. In addition, the stability of the five theoretical components was examined by constructing the model with the same five latent factors at test and retest.

Construction of a brief instrument (4th phase)

The data were collected in Finland in 2014, as part of the HBSC study. The HBSC study provides information on adolescents' demographic factors, health behaviours (health-promoting behaviours, risk behaviours), perceived health, lifestyles, and life circumstances; thus, the HL instrument (having reduced it to 16 items) was part of this broader survey. The nationally representative sample consisted, in total, of 3853 pupils (for 7th graders, n = 1918; for 9th graders, n = 1935) from 359 schools. The schools were chosen from the Finnish school register using a cluster sampling method that took into consideration provinces, the type of municipality (urban, semi-urban, rural), and the size of the schools. The participating class in each school was randomly selected. Thirteen- and 15-year-old participants responded voluntarily and anonymously to a standardized questionnaire administered in the course of one 45-minute lesson. Pupils were informed about the confidentiality of the study, and the fact that only group-level results would be reported.

As a first step, the instrument was reduced from 16 to 15 items, with each factor having an equal number of items. One extra item that had remained from the previous round (involving the self-awareness factor) was removed after inspection of the distributions and the item loadings. The 10-item version of the instrument was constructed using the same item reduction process as in the pilot study phase; that is, inspection of the item distributions, internal consistency estimates (Cronbach's alpha), factor loadings, and CFA model fit, in addition to examination of the contents of the items. In order to examine the sufficiency of the 10-item version of the instrument, a regression analysis was conducted in which the 10-item version was set to predict the longer version of the instrument.

Results

Pilot study

At the start of the pilot study there were 32 items, then after item reduction the instrument contained 16 items. The plan had been to reach a valid and reliable 15-item solution with three items representing each theoretical HL component. However, one extra item was added to the self-awareness component, for content-based reasons, and with a view to finding the best possible combination of items.

The Cronbach alpha for the entire 16-item HL measure was very high (0.94). In the final one-factor CFA model with 16 items, no error covariances between the items were allowed (χ^2 (104) = 261.69, p =.000; RMSEA = 0.06, CFI = 0.93, TLI = 0.92, SRMR = 0.04). The standardized factor loadings were between 0.57 and 0.74, with the majority being above 0.70. The Cronbach alphas for each of the five factors (based on the theoretical HL components) were between 0.75 and 0.84. Hence, the internal consistency of each of the HL components was at an adequate level.

The pilot phase also included class-level discussions with the pupils. In general, the pupils considered that the items were understandable. However, three items required some reformulation to avoid ambiguity.

Test–retest

The reformulated 16-item instrument was utilized with 117 pupils in the test–retest phase. In the SEM with one latent HL factor at both assessments, the standardized stability estimate was 0.83. The SEM with five factors also exhibited high stability estimates: theoretical knowledge 0.88, practical knowledge 0.81, critical thinking 0.81, self-awareness 0.84, and citizenship 0.90. These estimates suggested adequate test–retest reliability for HL, and also for the predetermined factors.

Construction of the 10-item Health Literacy for School-Aged Children (HLSAC) instrument

As an initial step, the instrument was reduced to 15 items, with each factor addressed by three items. Thereafter, the instrument was reduced to a 10-item Health Literacy for School-Aged Children (HLSAC) instrument (Figure 1). Five of the 10 items were informed by the HLQ [15]. The distributions of all the items were reasonable, and none of the answer options accounted for more than 50% of the answers. For the 10-item instrument, Cronbach's alpha was high (0.93), suggesting high internal consistency. The Cronbach alphas for the five core components (each with two items) were also reasonable (0.69–0.77).

The five-factor model with 10 items (χ^2 (25) = 681.41, $p < .001$; RMSEA = 0.08, CFI = 0.96, TLI = 0.92, SRMR = 0.03) showed strong correlations (0.95–1) between the factors. For this reason, the final model (Figure 1) was constructed as a one-factor model. The model had good item loadings and, considering the large sample size and the fact that no error covariances were allowed between any of the items, it also had a reasonably good fit to the data (χ^2 (35) = 948.64, $p < .001$, RMSEA = 0.08, CFI = 0.94, TLI = 0.92, SRMR = 0.04.).

Finally, a regression analysis was conducted in order to examine the relationship between the 10- and the 15-item instrument. The 10-item HLSAC instrument predicted approximately 97% of the variance of the 15-item instrument ($R^2 = 0.97, p < .001$).

Discussion

The objective of this study was to develop a brief and comprehensive theory-based instrument (HLSAC) to measure school-aged children's subjective HL. The instrument was based on a definition of HL [26], and was explicitly operationalized to measure five theoretical components. The goal of the iterative, systematic, and validity-driven development process [24] was that the instrument should encompass a wide range of competencies, covering the demands made on pupils by present and future society. In conjunction with this, bearing in mind that the theoretical background is based on seeing HL as a learning outcome, the items should be capable of being transformed into pedagogical practices, with a view to developing the competence that any given item is taken to represent.

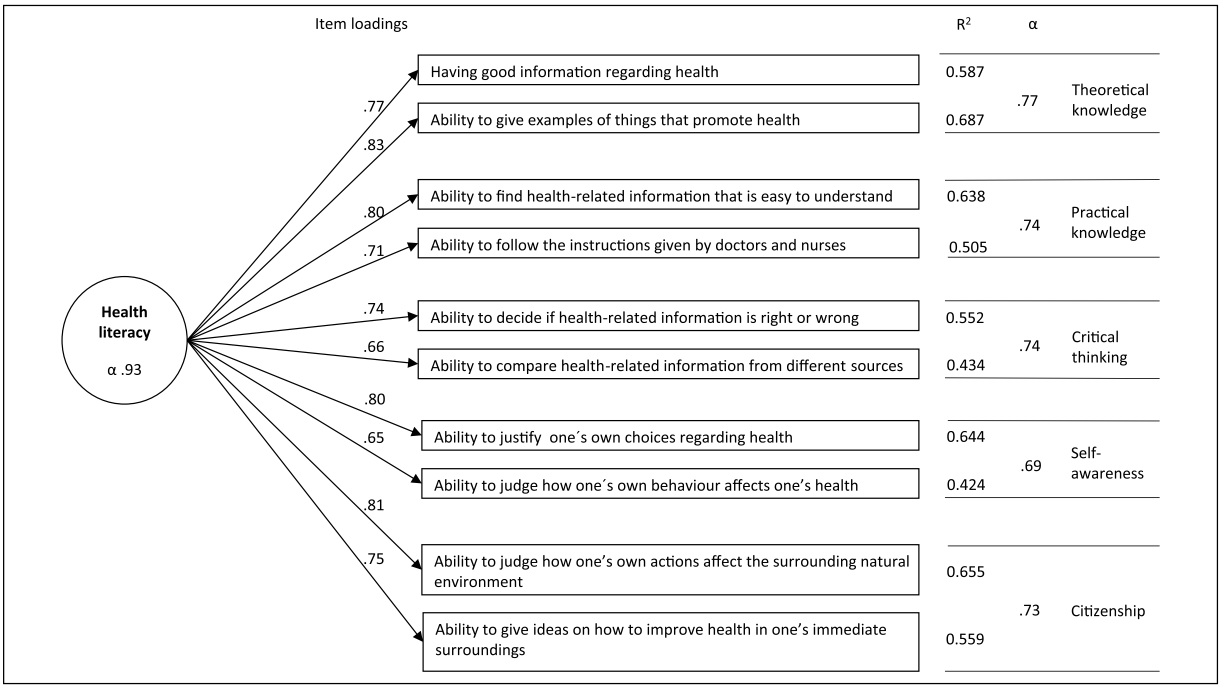

Figure 1. The Health Literacy for School-aged Children (HLSAC) instrument: Cronbach alphas, item loadings, and R^2 values.

There seems to be a clear need for this sort of instrument, given the lack of population-level instruments for inspecting school-aged children's subjective HL [21,22]. Given that a low HL level seems to be a risk factor for health [3] and associated with a poorer health status [4], and also that the foundation for health is built during childhood, it is important to gain information on the overall level of children's HL, and further, on how the level of HL varies between different subgroups. It should then be possible to design and direct effective interventions. The availability of a brief, valid, and reliable instrument is important for large-scale studies, and is particularly useful in studies with children and adolescents. If the survey is too long, or if the instrument contains too many items, pupils get tired of responding, and this is likely to reduce the reliability of the study. Since the instrument we finally arrived at contains only 10 items, it can be included in large-scale surveys to measure the association between HL and other phenomena, such as health behaviour or perceived health, without extending the overall length of the survey too far.

The approach to developing an instrument followed generally accepted steps and principles. The major challenge was to determine items and combinations that represented wide theoretical core components; hence, the validity of the instrument was taken into account at every phase of study. Item generation was based on the iterative work of a broad, experienced, and multidisciplinary expert group, also involving external specialists. Discussions with adolescents on the comprehensibility of the items influenced the further formulation of the items. Moreover, findings from the pilot test and from the test–retest phase directed and contributed to the development of the final brief instrument. The results indicated that the short HLSAC instrument (i.e. 10 items, with two items per theoretical core component) gives a good representation of the longer instrument. This is a significant issue when one has to consider how short an instrument can be without losing important aspects of HL (viewed as multidimensional in nature). The reliability of the instrument was found to be adequate. The high value for Cronbach's alpha indicated that the selected items all measured the same construct; furthermore, the test–retest of the instrument showed high consistency over time, in terms of encompassing one latent factor, and also the five preselected factors. According to the results, the overall goodness and adequacy of the instrument were clearly at an acceptable level.

Nevertheless, it is possible to identify limitations, and challenges for further development, even if the instrument itself worked well with the participants. There is a clear need now to test how far the instrument is applicable to younger pupils. Furthermore, given that HL is a complex phenomenon, there will inevitably be ongoing discussion of the essential HL components and their associations, of what a certain item may represent, and of how items should be formulated. Discussion may touch on the fact that in the model with five factors and 10 items there were correlations between the factors, with the final model thus being constructed as a one-factor model. The one-factor model describes a single phenomenon, here taken to denote 'health literacy'. It should be noted that on the basis of the background theory, the core components are partly overlapping and have some hierarchical elements; hence, it is not surprising that cross-correlations were found.

One significant issue involves comparing the results of subjective and objective HL measurements, and studying how the level of HL is connected to adolescents' self-reported health behaviour, or perceived health. Obtaining international comparative data could deepen our understanding of how well the HLSAC instrument works in different cultural and educational contexts. Overall, research on the measurement of adolescents' HL is very important, not just for the sake of making a methodological contribution to the field, but also for the health of the adolescents themselves.

Declaration of conflicting interests

None declared.

Funding

This work was supported by the National Institute for Health and Welfare, Finland.

References

[1] Kickbusch L, Pelikan JM, Apfel F, et al. *Health literacy: the solid facts*. Copenhagen: World Health Organization, 2013, p.1.
[2] Sørensen K, Pelikan JM, Rothlin F, et al. Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU). *Eur J Public Health* 2015;25:1053–8.
[3] Volandes AE and Paasche-Orlow MK. Health literacy, health inequality and a just healthcare system. *The Am J Bioeth* 2007;7:5–10.
[4] Berkman ND, Sheridan SL, Donahue KE, et al. Low health literacy and health outcomes: an updated systematic review. *Ann Int Med* 2011;155:97–107.
[5] Howard DH, Gazmararian J and Parker RM. The impact of low health literacy on the medical costs of medicare managed care enrollees. *Am J Med* 2005;118:371–7.
[6] Davison KK and Lawson CT. Do attributes in the physical environment influence children's physical activity? A review of the literature. *Int J Behav Nutr Phys Act* 2006;3:19.
[7] Patrick H and Nicklas TA. A review of family and social determinants of children's eating patterns and diet quality. *J Amer Coll Nutr* 2005;24:83–92.
[8] Mackey TK, Liang BA, Kohler JC, et al. Health domains for sale: the need for global health internet governance. *J Med Internet res* 2014;16:e62.
[9] Begoray DL, Banister EM, Wharf Higgins J, et al. Puppets on a string? How young adolescents explore gender and health in advertising. *JMLE* 2015;6:57–64.
[10] Parker RM, Baker DW, Williams MV, et al. The test of functional health literacy in adults. *J Gen Inter Med* 1995;10:537–41.
[11] Nutbeam D. Health promotion glossary. *Health Promot Int* 1998;13:349–64.
[12] Sørensen K, Van den Broucke S, Fullam J, et al. Health literacy and public health: a systematic review and integration of definitions and models. *BMC Public Health* 2012;12:80.
[13] Kiechle ES, Bailey SC, Hedlund LA, et al. Different measures, different outcomes? A systematic review of performance-based versus self-reported measures of health literacy and numeracy. *J Gen Intern Med* 2015;30:1538–46.
[14] Baker DW, Williams MV, Parker RM, et al. Development of a brief test to measure functional health literacy. *Patient Educ Couns* 1999;38:33–42.
[15] Osborne RH, Batterham RW, Elsworth GR, et al. The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). *BMC Public Health* 2013;13:658.
[16] McCormack L, Haun J, Sørensen K, et al. Recommendations for advancing health literacy measurement. *J Health Commun* 2013;18:9–14.
[17] Altin SV, Finke I, Kautz-Freimut S, et al. The evolution of health literacy assessment tools: a systematic review. *BMC Public Health* 2014;14:1207.
[18] Manganello JA, DeVellis RF, Davis TC, et al. Development of the Health Literacy Assessment Scale for Adolescents (HAS-A). *J Comm Healthcare* 2015;8:172–84.
[19] Conner M and Norman P. Predicting health behavior: a social cognition approach. In: Conner M and Norman P (eds) *Predicting health behaviour: research and practice with social cognition models*. 2nd ed. Maidenhead: Open University Press, 2005, pp.1–27.
[20] Bandura A. Health promotion by social cognitive means. *Health Educ Behav* 2004;31:143–64.
[21] Haun JN, Valerio MA, McCormack LA, et al. Health literacy measurement: an inventory and descriptive summary of 51 instruments. *J Health Commun* 2014;19:302–33.
[22] Ormshaw M, Paakkari L and Kannas L. Measuring child and adolescent health literacy: a systematic review. *Health Education* 2013;113:433–55.
[23] Perry EL. Health literacy in adolescents: an integrative review. *J Spec Pediatr Nurs* 2014;19:210–18.
[24] De Vet HCW, Terwee CB, Mokkink LB, et al. *Measurement in medicine. Practical guides to biostatistics and epidemiology*. Cambridge: Cambridge University Press, 2015, pp.30–2, 125.
[25] Pleasant A, McKinney J and Rikard RV. Health literacy measurement: a proposed research agenda. *J Health Commun* 2011;16:11–21.
[26] Paakkari L and Paakkari O. Health literacy as a learning outcome in schools. *Health Education* 2012;112:133–52.
[27] Sørensen K, Van den Broucke S, Pelikan J, et al. Measuring health literacy in populations: illuminating the design and development process of the European Health Literacy Survey Questionnaire (HLS-EU-Q). *BMC Public Health* 2013;13:948.
[28] Bloom BS. *Taxonomy of educational objectives: the classification of educational goals. Handbook I: Cognitive domain*. New York: McKay, 1956.
[29] Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–45.
[30] Schwarzer R and Fuchs R. Self-efficacy and health behaviour. In: Conner M and Norman P (eds) *Predicting health behavior*. Buckingham: Open University Press, 1995, pp.163–96.
[31] Currie C, Nic Gabhainn S, Godeau E, et al. The health behaviour in school-aged children: WHO collaborative cross-national (HBSC) study: origins, concept, history and development 1982–2008. *Int J Public Health* 2009;54:131–9.

II

SUBJECTIVE HEALTH LITERACY AMONG SCHOOL-AGED CHILDREN

by

Paakkari, O., Torppa, M., Villberg, J., Kannas, L. & Paakkari, L. 2018.

Health Education, 118 (2), 182–195.

https://doi.org/10.1108/HE-02-2017-0014

.

Received 21 February 2017
Revised 3 April 2017
2 May 2017
Accepted 8 May 2017

Subjective health literacy among school-aged children

Olli Paakkari
Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland
Minna Torppa
Faculty of Education, University of Jyväskylä, Jyväskylä, Finland, and
Jari Villberg, Lasse Kannas and Leena Paakkari
Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland

Abstract
Purpose – The purpose of this paper is to explore Finnish adolescents' subjective health literacy (HL) in association to school achievement, learning difficulties, educational aspirations, and family affluence.
Design/methodology/approach – Nationally representative data were collected in Finland as a part of the international Health Behaviour in School-aged Children study. The respondents consisted in total of 3,833 adolescents (7th and 9th graders) from 359 schools. The Health Literacy for School-aged Children instrument was applied to measure adolescents' subjective HL, while the Family Affluence Scale was used to measure adolescents' socioeconomic status. Information was gathered on school achievement, learning difficulties, and educational aspirations.
Findings – Approximately one-third of the adolescents manifested a high level of HL, around 60 per cent had a moderate level of HL, and about one-tenth had low HL. The HL level was lower for boys than for girls, and lower for 7th graders than for 9th graders. In the total sample, the strongest explanatory variables for HL were school achievement in the first language, and educational aspirations.
Originality/value – This study provides the first nationally representative examination of adolescents' subjective HL levels, and how these vary across age and gender groups. In drawing conclusions and presenting suggestions for HL interventions, it is important to verify the nature of the HL examined in any given study, and how it was researched.

Keywords Health education, Learning difficulties, School health promotion, Adolescence

Paper type Research paper

Introduction
In the field of public health and health promotion, there has been increasing interest in the health literacy (HL) of various age groups. HL as "the degree to which individuals have the capacity to obtain, process and understand basic health information and the services needed to make appropriate health decisions" (Ratzan and Parker, 2000, p. vi) has been reported to be a clear risk factor for poor health (Volandes and Paasche-Orlow, 2007). The development of HL among the population can be seen as an important means to decrease health disparities (Kickbusch *et al.*, 2006). The advancement of HL among the broader population requires a focus on HL, with age-appropriate measurements, across various age groups and settings (Kickbusch *et al.*, 2013). Children within schools comprise one such target.

School comprises a valuable setting for supporting HL, since the school reaches most of the population within a certain age demographic. The foundation for HL, health behaviour, and health and well-being in general is laid during childhood and the school years. Adolescence is generally understood to be a significant period of life in many respects, including that of independent decision making (Ghanbari *et al.*, 2016). Health inequalities among the adult population can be partly explained via health behaviours adopted in adolescence, and with reference to early life circumstances (Inchley *et al.*, 2016, p. 5).

This study is funded by National Institute for Health and Welfare, Finland and Academy of Finland (No. 276239).

Health Education
Vol. 118 No. 2, 2018
pp. 182-195

0965-4283
DOI 10.1108/HE-02-2017-0014

One of the main purposes of the school is to reduce inequalities, including those that are already present in adolescence (UNESCO, 2014a). Education contributes to the health and well-being of pupils in a general sense (UNESCO, 2014b); however, when health-related competences (i.e. HL) are given specific attention when formulating and putting into practise school health education learning objectives and standards, health disparities are likely to further decrease (see Parker *et al.*, 2003). As Perry (2014, p. 217) has aptly noted, "the task of improving HL amongst millions of adolescents is daunting; but, ultimately, improving HL in adolescents is imperative for achieving better health outcomes". There is a clear and increasing gap between the demands to take care of one's health and the actual skills that people possess (Kickbusch *et al.*, 2013; Gazmararian *et al.*, 2005; Parker *et al.*, 2003). From this perspective, the development of HL among schoolchildren is not merely desirable, but could be regarded as a moral act.

The level of school-aged children's HL – what have we learned so far?

Over the years, most studies on HL have been conducted on patients in a health care context, with a research focus on the basic skills of reading, writing, and numeracy. These are often referred to collectively as functional HL, as defined by Parker *et al.* (1995). Research on these domains continues, and the importance of functional HL skills on a person's health remains widely recognised. Nevertheless, there has been an increasing willingness to move beyond these skills and to monitor a broader construct of HL, one that would encompass the HL competences (e.g. critical thinking, problem solving and advocacy skills) that are nowadays needed if individuals are to take care of and sustain their own and community health in modern society (Nutbeam, 1998; Sørensen *et al.*, 2012).

There has only been limited monitoring and reporting of HL of any kind among school-aged children overall (Ghanbari *et al.*, 2016; Ormshaw *et al.*, 2013), with studies focusing primarily on the functional HL of adolescents. Moreover, there have not been many cross-national comparative studies on adolescents' HL levels, although some information has been obtained from countries or regions in various parts of the world, including Asia (Taiwan, China) and the USA. Studies focusing on functional HL have shown that up to 90 per cent of children in Taiwan and China have a moderate or high level of HL, while 10 per cent have low HL (Chang, 2010; Lam and Yang, 2014). In the USA, the proportion of young people with "below basic" HL has been the same as in Taiwan and China (Kutner *et al.*, 2006; see also Ghaddar *et al.*, 2012). The findings reflect a good general literacy level among young people in those areas (UNESCO, 2014c). Interestingly, when another study in the same area of Asia assessed HL – via an instrument which addressed the broader construct of HL – the proportion of children with low HL was higher, at over 25 per cent (Shih *et al.*, 2016). This is consistent with the view that functional HL tools such as Test of Functional Health Literacy in Adults (Parker *et al.*, 1995) and Rapid Estimate of Adult Literacy in Medicine (Davis *et al.*, 1991) focus on phenomena (e.g. reading ability; DeWalt and Pignone, 2005) that are narrower in scope than the HL domains addressed in the fields of public health and health promotion.

In Europe, HL has been measured mainly in adult populations via the European Health Literacy Survey. This has been conducted within eight European countries, and it has also included adolescents and young adults (15 years and older) (Sørensen *et al.*, 2015). However, since age-specific findings have not been reported, there is no information on HL among the young people represented in that sample.

In Finland, HL among school-aged children has been studied as part of the national assessment of learning applied to health education as a distinct school subject. Using an objective measure (a pen and paper exam) based on a broad construction of HL, a nationally representative sample of 9th graders gave responses in various tasks (Summanen, 2014). The pupils showed only a satisfactory level of competence. However, it should be noted that the assessment was first and foremost an evaluation of how well pupils had met the learning criteria identified in the national curriculum. Hence, the main starting point for the development of the exam was the curriculum, not the concept of HL, even if the findings do indeed reflect HL.

According to a review by Perry (2014, p. 215), the majority of studies on the current HL status of adolescents has used "health literacy instruments that have not been validated for use in adolescents", hence caution is seen as necessary in interpreting the findings. Moreover, most of the studies have focused on functional literacy, with a concomitant lack of studies addressing the broader construct of HL, plus related factors. It is only recently that efforts have been made to develop adolescent-specific instruments that go beyond the evaluation of basic literacy skills (Ghanbari *et al.*, 2016; Paakkari *et al.*, 2016; Shih *et al.*, 2016). One of such instruments is the Health Literacy for School-aged Children (HLSAC) instrument (Paakkari *et al.*, 2016). It was developed based on a broader construct of HL and for the purpose of measuring HL of the adolescents.

The aim of this study was to investigate the level of subjective HL among adolescents (boys and girls, 7th and 9th grade, ages 13 and 15) in Finland, on the basis of the HLSAC instrument. In addition, the study sought to determine the associations between HL, school achievement, learning difficulties, educational aspirations, and family affluence.

Methods

Participants and data collection

The empirical data for the study were collected in Finland in 2014, as part of the cross-national collaborative study entitled Health Behaviour in School-aged Children (HBSC). The general objective of the HBSC research is to gain a better understanding of lifestyles, health behaviours, and the surrounding context, insofar as they affect children and adolescents (Currie *et al.*, 2009). The research thus covers various aspects of adolescence, including demographic factors, health behaviours, perceived health, HL, learning, lifestyles, and life circumstances. All the measures are based on self-reports.

The data for the present study were nationally representative, being obtained from a total of 3,833 adolescents (boys 7th grade $n = 880$, boys 9th grade $n = 882$, girls 7th grade $n = 894$, girls 9th grade $n = 963$) in 359 schools. The schools were chosen from the Finnish school register using a cluster sampling method. Sampling was adjusted to take into account the province within Finland, the type of municipality (urban, semi-urban, rural), and the size of the school. Within each school the participating class was randomly selected.

The data collection followed the general guidelines of responsible conduct of research (Finnish Advisory Board on Research Integrity, 2012), and the research protocol of the international HBSC study (Currie *et al.*, 2014): 13- and 15-year-old participants responded voluntarily and anonymously to a standardized paper-and-pen questionnaire, administered in the course of one lesson. The pupils were informed of the confidentiality of the data, and of the fact that only group-level results would be reported.

The response rate of the pupils was 85 per cent, while the response rate for the schools was 68 per cent.

Measures

HL instrument. A brief HLSAC instrument (Paakkari *et al.*, 2016) was used to measure the adolescents' subjective (self-reported, perceived) HL. The validated ten-item instrument (Table I) contains two items from each of the five core components (theoretical knowledge, practical knowledge, critical thinking, self-awareness, citizenship). The HLSAC instrument has been found to have high internal consistency (overall Cronbach's α 0.93, Table II).

	Boys ($n = 1,820$)				Girls ($n = 1,912$)			
	Not at all true	Barely true	Somewhat true	Absolutely true	Not at all true	Barely true	Somewhat true	Absolutely true
Theoretical knowledge								
Having good information regarding health	2.6	9.3	44.6	43.5	0.5	7.5	45.4	46.6
Ability to give examples of things that promote health	2.3	13.6	49.9	34.1	0.9	12.2	49.9	37.0
Practical knowledge								
Ability to find health-related information that is easy to understand	2.1	9.3	44.3	44.3	0.6	6.3	42.5	50.7
Ability to follow the instructions given by doctors and nurses	2.3	9.6	43.9	44.3	0.7	5.7	35.3	58.2
Critical thinking								
Ability to decide if health-related information is right or wrong	2.3	12.7	49.7	35.4	1.5	12.9	51.9	33.7
Ability to compare health-related information from different sources	2.8	13.8	48.5	34.9	1.9	13.8	51.7	32.6
Self-awareness								
Ability to justify one´s own choices regarding health	2.5	12.2	48.7	36.6	0.9	9.3	49.7	40.2
Ability to judge how one´s own behaviour affects one's health	3.2	12.0	45.3	39.5	1.1	9.0	48.8	41.1
Citizenship								
Ability to judge how one's own actions affect the surrounding natural environment	2.5	10.6	49.4	37.5	0.9	8.8	48.2	42.1
Ability to give ideas on how to improve health in one's immediate surroundings	3.5	16.5	51.8	28.2	2.4	16.3	52.4	28.9

Table I. Percentage distributions of the items in the HLSAC instrument, divided by gender

	n	Min.	Max.	Mean	SE	SD	Skewness	SE	Kurtosis	SE	α
Boys 7th grade	880	10	40	31.90	0.20	5.91	−0.65	0.08	0.52	0.17	0.94
Boys 9th grade	882	10	40	32.39	0.20	6.06	−0.96	0.08	1.42	0.16	0.95
Girls 7th grade	894	10	40	32.51	0.17	5.13	−0.54	0.08	0.25	0.16	0.91
Girls 9th grade	963	10	40	33.32	0.16	4.88	−0.61	0.08	0.38	0.16	0.91
Total	3,619	10	40	32.55	0.09	5.53	−0.76	0.04	0.93	0.08	0.93

Table II. Descriptive statistics and the Cronbach's αs for health literacy (HLSAC)

All the items took the form "I am confident that […]", and the Likert-type response scale included four options: "not at all true", "barely true", "somewhat true", and "absolutely true". For the analysis of the HL levels the response options "not at all true" and "barely true" were combined to describe "low" HL.

The levels of HL were classified in such a way as to fall into three groups. The thresholds were set by an expert group (consisting of researchers and teachers in the field of health promotion, education, and psychology) who determined the HL scores required to reach a given level. Following consideration of the contents of the items, and inspection of the response distribution, the resulting HL levels consisted of "low" (score 10-25), "moderate" (score 26-35), and "high" (score 36-40).

School achievement. Participants were asked to indicate their school achievement in their first language and in mathematics via the following question: "In my latest school report the mark was […]". The response scale (marks) ranged from 4 (fail) to 10 (excellent). The marks thus obtained were regrouped to form three categories (marks 4-6, marks 7-8, and marks 9-10).

Learning difficulties. Respondents were asked to indicate whether they had learning difficulties in two areas, namely reading or spelling, and mathematics. The response options for both questions were no, some, and yes.

Educational aspirations. To assess educational aspirations, the adolescents were asked what they would do when they finished comprehensive school (at age 15). The response options were: "upper secondary school (age 16-19)", "vocational school or other vocational training", "an apprenticeship, double examination (upper secondary school and vocational school)", "get a job", "be unemployed", or "don't know". Only the options "upper secondary school" and "vocational school or other" were included in the statistical analysis (dummy variable), because of very low frequencies for the other responses. The upper secondary school in Finland mainly represents an academic orientation, while the vocational school or vocational training can be seen as having a practical orientation.

Family affluence. The Family Affluence Scale (Torsheim *et al.*, 2015) included six items that are associated with parental income and hence function as measures of adolescents' socioeconomic status. The questions encompassed the material conditions of the household, covering: occupancy of bedrooms; number of bathrooms; number of computers and dishwasher ownership; ownership of a car; and holidays abroad. The response options for the questions on the dishwasher and on having own bedroom were "no" and "yes". For the other questions the response scale was "none", "one", "two", and "more than two". The respondents were divided into three affluence groups according to the HBSC protocol (Currie *et al.*, 2016): low affluence (lowest 20 per cent), medium affluence (middle 60 per cent), and high affluence (highest 20 per cent).

Statistical analysis

All the statistical analyses were conducted for the total sample, separately for boys and girls, and for 7th and 9th graders. The descriptive statistics for HL included means, standard errors, standard deviations, distributions of skewness and kurtosis, and percentage distributions of the HL levels. The differences between the group means (for gender and grade), gender and grade interaction effect on the HL were tested via a two-way analysis of variance (ANOVA).

The analyses were conducted using SPSS (version 22). The relationships between subjective HL and school achievement, learning difficulties, educational aspirations, and family affluence were tested via a mixed-effects multilevel regression analysis, because the data had a hierarchical structure (pupils nested in classrooms). The analyses in this case were conducted using Stata (version 14).

Results

Level of subjective HL

The HLSAC instrument used contains ten items (Table I).

Respondents indicated that their theoretical and practical knowledge was, generally speaking, at the "good" level. Around 90 per cent reported having a good knowledge of health; they felt that they could easily find understandable health information and could follow the instructions of doctors or nurses (response options: "somewhat true" or "absolutely true").

More difficulties were indicated regarding critical thinking and citizenship. About 15 per cent reported difficulties in the ability to compare the information from different sources, or in the ability to decide if information is right or wrong. Approximately one in five indicated problems in terms of being able to give ideas on how to improve health in their environment.

Gender comparison showed that the boys reported more difficulties (response options "not at all true" or "barely true") than the girls on almost every HL item. Similarly, more girls than boys reported having good competence in the cases that were asked, that is, they chose the response option "absolutely true" more often than the boys in most of the HL items.

Table II reports the descriptive HL statistics for girls and boys in the 7th and 9th grades, separately. The HL distributions were approximately normal. In every group, the minimum score was 10 and the maximum 40. The overall mean HL score was 32.55. The lowest subjective HL was found among 7th grade boys (mean score 31.90), and the 9th grade girls reported highest subjective HL (mean score 33.32). The girls HL mean score was higher than boys, and according a two-way ANOVA the difference was statistically significant ($F(\text{df}1) = 8.214, p = 0.004$). The mean score of the HL was higher among the 9th graders than among 7th graders, and this difference was also statistically significant ($F(\text{df}1) = 10.555, p = 0.001$).

Thereafter, we categorised HL into three levels (low = score 10-25, moderate = score 26-35, high = score 36-40). We observed that around one-tenth of the participants had low HL, 57 per cent had moderate HL, and approximately one-third achieved a high level of HL (Figure 1). In both age groups, there were more boys than girls with low HL. In both genders, the proportion of pupils who had a high level of HL increased towards the 9th grade.

HL associations with school achievement, learning difficulties, educational aspirations, and family affluence

The descriptive statistics for the predictors (Table III) showed that poorer results were more common among boys, and in general more frequent for mathematics than for the first language. Adolescents reported more difficulties in mathematics than in reading or spelling. About four out of five respondents reported that they had no difficulties in

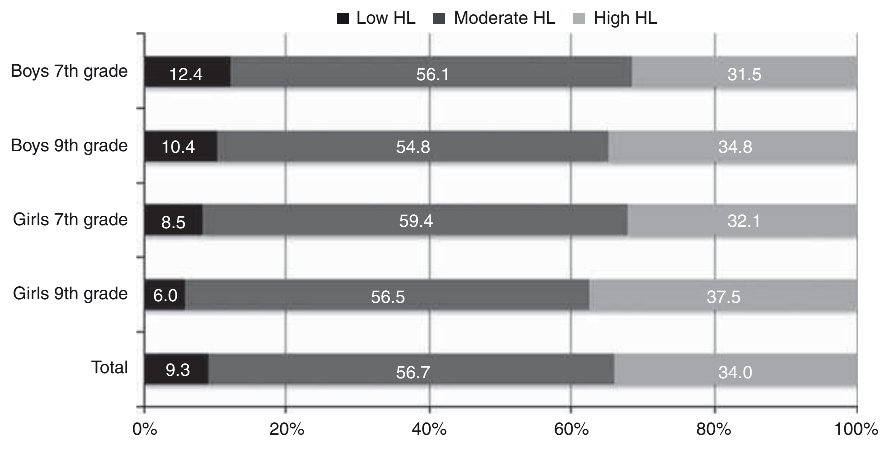

Figure 1.
Levels of subjective HL by gender and grade, and for the total sample (percentage distribution)

Table III. Percentage distributions for school achievement, learning difficulties, educational aspirations, and family affluence (predictors), categorised by gender

	Boys ($n = 1,870$)			Girls ($n = 1,928$)		
Predictor	Marks 4-6	Marks 7-8	Marks 9-10	Marks 4-6	Marks 7-8	Marks 9-10
School achievement in the first language	17.7	61.7	20.6	4.0	46.9	49.1
School achievement in the math	19.7	48.1	32.2	12.4	47.6	40.0
	No	Some	Yes	No	Some	Yes
Difficulties in reading or spelling	78.2	18.7	3.1	78.2	17.7	4.1
Difficulties in mathematics	56.6	35.7	7.6	45.6	40.5	14.0
	Upper secondary school	Vocational school	Other	Upper secondary school	Vocational school	Other
Educational aspirations	51.3	39.4	9.3	70.3	23.3	6.4
	Low	Medium	High	Low	Medium	High
Family affluence	20.0	60.0	20.0	20.0	60.0	20.0

reading or in spelling. In mathematics, on average half of the participants reported difficulties, with girls reporting more difficulties than the boys. The majority of the participants intended to apply to upper secondary school, or to vocational school, or other vocational training. Among boys the intention to go vocational school or vocational training was higher than for the girls, most of whom reported that they intended to go to upper secondary school.

In the total sample, the strongest correlations between HL and the other variables were for school achievement in the first language, and for educational aspirations ($r = 0.22$) (Table IV). The correlations for the other variables (difficulties in reading, spelling or mathematics, family affluence) varied between 0.12 and 0.14. Better performance in the first language or mathematics, family affluence, and higher educational aspirations predicted higher HL, whereas lower HL was associated with difficulties in reading, spelling, or mathematics. Overall, the correlations were somewhat higher among the 9th grade pupils than among the 7th grade pupils.

In the total sample, the predictive variables for HL were school achievement in the first language, educational aspirations, difficulties in reading or spelling, difficulties in mathematics, and family affluence. Among the respondents who were planning to go to upper secondary school after finishing the comprehensive school, the HL score was 1.3 points higher than among those who were planning to enter vocational school or other vocational training.

The HL predictors varied between class and gender groups. For the girls (7th and 9th graders) the strongest predictive variable for HL was school achievement in the first language, and for the boys difficulties in mathematics. The multilevel effect of the school was significant only for the total sample and for 9th grade girls.

Discussion

To recap, the first aim of the study was to ascertain the level of school-aged children's subjective HL, and the associations of HL with school achievement, learning difficulties, educational aspirations, and family affluence.

Subjective HL

The adolescents' subjective HL level proved to be fairly high according to the HLSAC-scale: about 60 per cent had moderate HL, around one-third reported a high level of HL, and no

	Coefficient	*t*	Sig.	95% confidence interval		Pearson correlation
Total sample						
School achievement in the first language	0.56	5.54	0.000	0.36	0.76	0.22
Educational aspirations (upper secondary school)	1.25	5.24	0.000	0.78	1.71	0.22
Difficulties in reading or spelling	−0.76	−3.83	0.000	−1.15	−0.37	−0.14
Family affluence	0.22	4.26	0.000	0.12	0.32	0.12
Difficulties in mathematics	−0.45	−3.02	0.003	−0.74	−0.16	−0.14
Boys 7th graders						
Educational aspirations (upper secondary school)	0.95	2.08	0.038	0.05	1.85	0.13
Difficulties in mathematics	−0.98	−2.55	0.011	−1.73	−0.22	−0.13
Family affluence	0.30	2.50	0.013	0.06	0.54	0.12
Boys 9th graders						
Educational aspirations (upper secondary school)	1.41	2.93	0.003	0.47	2.36	0.25
Difficulties in mathematics	−1.05	−3.23	0.001	−1.69	−0.41	−0.23
Difficulties in reading or spelling	−1.12	−2.62	0.009	−1.95	−0.28	−0.19
Family affluence	0.33	2.98	0.003	0.11	0.54	0.15
School achievement in the first language	0.53	2.53	0.012	0.12	0.95	0.24
Girls 7th graders						
Educational aspirations (upper secondary school)	1.22	2.57	0.010	0.29	2.15	0.17
School achievement in the first language	0.62	2.79	0.005	0.19	1.06	0.17
Family affluence	0.27	2.70	0.007	0.07	0.47	0.13
Girls 9th graders						
School achievement in the first language	0.89	4.97	0.000	0.54	1.24	0.30
Educational aspirations (upper secondary school)	1.80	4.43	0.000	1.00	2.60	0.29
Difficulties in reading or spelling	−0.76	−2.65	0.008	−1.33	−0.20	−0.17

Table IV. The mixed-effects multilevel regression model for HL (dependent variable)

more than about one-tenth had low HL. The girls' HL level was higher than that of the boys, and HL was higher among the 9th graders than the 7th graders. According to the regression analysis, the strongest predictive variables for HL in the total sample were school achievement in the first language and educational aspirations. For the boys, the most important predictor of HL was difficulties in mathematics, and for the girls, school achievement in the first language. A regression model across gender and grade level explained 8 per cent of the HL variance. The model predicted more of the HL variance for the 9th grade than for the 7th grade.

The fact that, in general, most of the pupils reported a fairly high level of HL may be because in the Finnish school system health issues are taught within health education as a school subject, which is a statutory independent subject both at primary school (grades 1-6, ages 7-13) and at secondary school (grades 7-9, ages 13-15). Schools have to follow the national curriculum and the objectives for the subject. Moreover, every school has to offer the same amount of HE teaching to every pupil. Nevertheless, pupils gain health knowledge in other contexts as well, such as within media, guardians, and peers or other school subjects. It suggests that general health promotion work in the school community (involving e.g. health-promoting schools/whole-school approach) can advance HL among adolescents. Since the 9th graders have received more teaching than the 7th graders, this may partly explain the finding that the older pupils had better HL than the younger ones. On the other hand, we do not as yet have cross-national research results on the levels of HL in countries where there is no systematic teaching of HE. Hence, we do not know how having HE as a school subject affects the level of HL.

Overall, girls showed a higher level of HL than boys. The result was consistent with the national HE examination, which showed a large gender difference (Summanen, 2014). Similar gender differences have been found among adults (Sørensen *et al.*, 2015), but the comparison of the results is problematic because of different age groups and instruments. There is ongoing debate on the size of the gender gap and the explanations for it (Hyde, 2014; Voyer and Voyer, 2014). From a wider perspective, potential reasons for gender differences could relate to aspects of society, the culture, the school environment, and pedagogy that may favour girls (Stoet and Geary, 2013). The results of the PISA study indicate consistent differences between boys and girls in reading, doing homework, and investing effort at school (OECD, 2015), and in attitudes to learning and school (OECD, 2015; Summanen, 2014). In Finland, girls tend to be more interested than boys in the health issues discussed in HE lessons (Aira *et al.*, 2014).

Pupils with a higher level of self-regulation (i.e. the ability to control, direct, and plan their thinking, emotions, and behaviours, Schunk and Zimmerman, 1997) generally perform better than students with lower levels of self-regulation (OECD, 2015). Girls tend to be more self-regulated and disciplined than boys, and have better ability to set goals, plan ahead, and deal with setbacks and frustrations (Duckworth and Seligman, 2006; Kenney-Benson *et al.*, 2006). All these reasons could partly explain the higher HL levels among the girls in our study. However, it is important to remember that boys and girls are not homogenous groups: both boys and girls include pupils who do not cope with school, and others who manage education well.

Associations between HL and other variables

The second aim of this study was to explore the association of HL with school achievement, learning difficulties, educational aspirations, and family affluence.

There were statistically significant associations between HL and the variables in question.

School achievement in the first language was the strongest predictive variable both in the total sample and for girls; however, for boys the best predictor was difficulties in mathematics. This finding was consistent with the study of the Finnish National Board of Education, in which school achievement in mathematics explained 21 per cent of the success in the HE national exam, while the first language explained as much as 31 per cent (Summanen, 2014). School achievement in general is linked to perceived competence in health issues; thus, the finding is unsurprising, given the nature of HLSAC, which is based on beliefs in one's own competences (i.e. self-efficacy).

In this study, difficulties in reading, spelling, or mathematics, and also low educational aspirations, predicted a lower level of HL. Learning difficulties have been found to be associated with adolescents' academic achievement, and this also predicts their educational aspirations (Rimkute *et al.*, 2013). Adolescents with higher educational aspirations are more willing than others to seek an academic professional career in the future (Jodl *et al.*, 2001). The study by Summanen (2014) indicated that pupils who were planning to continue to upper secondary school succeeded better in assessments of learning within HE (i.e. the school subject) than those who were planning otherwise. In particular, pupils who were thinking of taking a break for a year, or going directly to a job, succeeded poorly in the examination (Summanen, 2014).

Studies have shown that, compared to students who move on to upper secondary school after comprehensive school, unhealthy behaviours are more common among adolescents who go on to vocational school or training (Grotvedt *et al.*, 2008; Luopa *et al.*, 2014; Vereecken *et al.*, 2004). In Finland, among those adolescents who study at upper secondary school, 8 per cent smoke daily, while for young people studying in vocational school the proportion is 36 per cent. A similar picture emerges with regard to other unhealthy behaviours such as alcohol consumption, having less sleep, physical inactivity, and experimentation with drugs (Luopa *et al.*, 2014). The aspiration to study in a vocational school indicated a lower level of HL overall. Thus, a non-academic educational path, in conjunction with low HL (which is a risk factor *per se*), can reinforce health inequalities among children and adolescents.

In this study, family affluence predicted a higher level of HL. It has been suggested that low socio-economic status is a risk factor for academic performance (Frederickson and Petrides, 2008). According the PISA study (OECD, 2015), there are differences in academic achievement (reading, mathematics, science, problem-solving) which are associated with parents' education, occupation, and family wealth, or household possession. Boys, in particular, are in a disadvantaged position when their parents' socio-economic status is low (OECD, 2015). Parents' education was also linked to HE learning outcomes in a study conducted for the Finnish National Board of Education (Summanen, 2014). There are multiple reasons for the correlation between family affluence and adolescent HL. Highly affluent families may possess forms of capital that are conducive to HL. These include general knowledge of health issues, and a high value given to health in general, in addition to the higher school achievement and the higher income of the families (Reardon, 2011).

The amount of explained variance remained low in the regression model with school success, educational aspirations, family affluence, and learning difficulties explaining HL. This has been noticed also in previous studies. The comparative study on HL in EU (HLS-EU Consortium, 2012) showed that age, gender, education, financial deprivation, and social status explained about 10 per cent of variance in HL in countries like Austria, Germany, or the Netherlands. However, in other countries, e.g. Poland, Greece, and Bulgaria, the coefficient of determination was higher, over 20 per cent. It seems that the socio-economic indicators influence to HL more in certain countries than in others (HLS-EU Consortium, 2012).

Limitations of the study

This study had certain limitations. In the best case, the same instrument will contain both subjective and objective measurements of HL. It then becomes possible to compare results for the same respondents. Because the respondents may give so-called "socially expected answers", the self-reported questionnaires could give higher scores than objective measurements. The measurement of educational difficulties was here based on the adolescents' report of their experience, and the report could be different from what would be revealed by actual tests in reading, mathematics, or problem solving. Moreover, the marks in the first language and in mathematics were reported by the pupils themselves, and there could have been some problems with retention of the information.

The research design was cross-sectional, meaning that the results do not address change across time, instead describing differences between two (separate) samples of 7th and 9th graders. A longitudinal study could provide more detailed information on the development of each participant's HL during the last years of secondary school.

Comparison of the findings between different studies using different instruments (e.g. Chang, 2010; Ghaddar *et al.*, 2012; Lam and Yang, 2014; Shih *et al.*, 2016; Sørensen *et al.*, 2015) is difficult, since the studies in question may be measuring various constructs of HL. Previous research has shown that the tool used to measure HL has an influence on the level of HL identified (Barber *et al.*, 2009).

Conclusions

This paper provides the first results on the level of adolescents' subjective HL in Finland, measured by the instrument specifically developed for the target group. It is based on nationally representative data, and it offers important insights into how HL levels vary between different class and gender groups. The study of subjective HL is clearly important, since one's perceived competence – i.e. self-efficacy – has been found to be a clear and independent factor explaining various health-related behaviours (Bandura, 2004; Conner and Norman, 2005). The overall findings of this study are somewhat different from those of Finnish National Board of Education (Summanen, 2014); when measured objectively adolescents' HL was at the satisfactory level (Summanen, 2014), whereas this study using a subjective measure showed a good level of HL among the target group. This indicates that both objective and subjective measures are needed in order to construct strong policy recommendations.

It is likely that reducing school achievement gaps could contribute to a reduction in HL disparities among school-aged children. If this is so, it confirms the important role of education and of schools in tackling health disparities. The establishment of school-based learning standards for HL could assist in tackling health disparities overall (Parker *et al.*, 2003).

In drawing conclusions and making suggestions for HL interventions, there is a clear need for researchers and politicians to be clear about the different kinds of HL covered in various studies, and the methods and measures applied. Caution is needed, insofar as general discussion of HL – divorced from careful consideration of what is actually focused on – could lead to impractical conclusions, false generalisations, and unhelpful concrete practices. Future avenues could include monitoring children's HL across various settings, using the same instrument to check the influence of the context. It may be possible to utilise the findings of this study within cross-national research, for example as a part of the international HBSC study.

References

Aira, T., Välimaa, R., Paakkari, L., Villberg, J. and Kannas, L. (2014), "Finnish pupils' perceptions of health education as a school subject", *Global Health Promotion*, Vol. 21 No. 3, pp. 6-18.

Bandura, A. (2004), "Health promotion by social cognitive means", *Health Education & Behavior*, Vol. 31 No. 2, pp. 143-164.

Barber, M.N., Staples, M., Osborne, R.H., Clerehan, R., Elder, C. and Buchbinder, R. (2009), "Up to a quarter of the Australian population may have suboptimal health literacy depending upon the measurement tool: results from a population-based survey", *Health Promotion International*, Vol. 24 No. 3, pp. 252-261.

Chang, L.-C. (2010), "Health literacy, self-reported status and health promoting behaviours for adolescents in Taiwan", *Journal of Clinical Nursing*, Vol. 20 Nos 1-2, pp. 190-196.

Conner, M. and Norman, P. (2005), "Predicting health behavior: a social cognition approach", in Conner, M. and Norman, P. (Eds), *Predicting Health Behaviour: Research and Practice with Social Cognition Models*, 2nd ed., Open University Press, Maidenhead, pp. 1-27.

Currie, C., Nic Gabhainn, S. and Godeau, E., the International HBSC Network Coordinating Committee (2009), "The health behaviour in school-aged children: WHO collaborative cross-national (HBSC) study: origins, concept, history and development 1982-2008", *International Journal of Public Health*, Vol. 54 No. 2, pp. 131-139.

Currie, C., Inchley, J., Molcho, M., Lenzi, M., Veselska, Z. and Wild, F. (Eds) (2014), *Health Behaviour in School-Aged Children (HBSC) Study Protocol: Background, Methodology and Mandatory items for the 2013/14 Survey*, Child and Adolescent Health Research Unit (CAHRU), St Andrews.

Currie, D., Elgar, F., Augustson, L. and Torsheim, T. (2016), "Key data", in Inchley, J., Currie, D., Young, T., Samdal, O., Torsheim, T., Augustson, L., Mathison, F., Aleman-Diaz, A., Molcho, M., Weber, M. and Barnekow, V. (Eds), *Growing up Unequal: Gender and Socioeconomic Differences in Young People's Health and Well-Being. Health Health Behaviour in School-Aged Children (HBSC) Study: International Report from the 2013/2014 Survey*, WHO, Copenhagen, pp. 13-19.

Davis, T.C., Crouch, M.A., Long, S.W., Jackson, R.H., Bates, P., George, R.B. and Bairnsfather, L.E. (1991), "Rapid assessment of literacy levels of adult primary care patients", *Family Medicine*, Vol. 23 No. 6, pp. 433-435.

DeWalt, D.A. and Pignone, M.P. (2005), "Reading is fundamental: the relationship between literacy and health", *Archives of Internal Medicine*, Vol. 165 No. 17, pp. 1943-1944.

Duckworth, A.L. and Seligman, M.E.P. (2006), "Self-discipline gives girls the edge: gender in self-discipline, grades, and achievement test scores", *Journal of Educational Psychology*, Vol. 98 No. 1, pp. 198-208.

Finnish Advisory Board on Research Integrity (2012), "Responsible conduct of research and procedures for handling allegations of misconduct in Finland", Guidelines of the Finnish Advisory Board on Research Integrity, available at: www.tenk.fi/sites/tenk.fi/files/HTK_ohje_2012.pdf (accessed 30 March 2017).

Frederickson, N. and Petrides, K.V. (2008), "Ethnic, gender, and socio-economic group differences in academic performance and secondary school selection: a longitudinal analysis", *Learning and Individual Differences*, Vol. 18 No. 2, pp. 144-151.

Gazmararian, J.A., Curran, J.W., Parker, R.M., Bernhardt, J.M. and DeBuono, B.A. (2005), "Public health literacy in America: an ethical imperative", *American Journal of Preventive Medicine*, Vol. 28 No. 3, pp. 317-322.

Ghaddar, S.F., Valerio, M.A., Garcia, C.M. and Hansen, L. (2012), "Adolescent health literacy: the importance of credible sources for online health information", *Journal of School Health*, Vol. 82 No. 1, pp. 28-36.

Ghanbari, S., Ramezankhani, A., Montazeri, A. and Mehrabi, Y. (2016), "Health Literacy Measure for Adolescents (HELMA): development and psychometric properties", *PLoS ONE*, Vol. 11 No. 2, pp. 1-12.

Grotvedt, L., Stigum, H., Hovengen, R. and Graff-Iversen, S. (2008), "Social differences in smoking and snuff use among Norwegian adolescents: a population based survey", *BMC Public Health*, Vol. 8 No. 1, pp. 1-322.

HLS-EU Consortium (2012), "Comparative report of health literacy in eight EU member states: The European Health Literacy Survey (HLS-EU)", available at: http://ec.europa.eu/chafea/documents/news/Comparative_report_on_health_literacy_in_eight_EU_member_states.pdf (accessed 30 March 2017).

Hyde, J.S. (2014), "Gender similarities and differences", *Annual Review of Psychology*, Vol. 65 No. 1, pp. 373-398.

Inchley, J., Currie, D., Young, T., Samdal, O., Torsheim, T., Augustson, L., Mathison, F., Aleman-Diaz, A., Molcho, M., Weber, M. and Barnekow, V. (Eds) (2016), *Growing up Unequal: Gender and Socioeconomic Differences in Young People's Health and Well-Being. Health Policy For Children and Adolescents*, No. 7, World Health Organization, Copenhagen.

Jodl, K.M., Michael, A., Malanchuk, O., Eccles, J.S. and Sameroff, A. (2001), "Parents' roles in shaping early adolescents' occupational aspirations", *Child Development*, Vol. 72 No. 4, pp. 1247-1265.

Kenney-Benson, G.A., Pomerantz, E.M., Ryan, A.M. and Patric, H. (2006), "Sex differences in math performance: the role of children's approach to schoolwork", *Developmental Psychology*, Vol. 42 No. 1, pp. 11-26.

Kickbusch, I., Wait, S. and Maag, D. (2006), *Navigating Health: The Role of Health Literacy*, Alliance for Health and the Future, International Longevity Centre-UK, London.

Kickbusch, I., Pelikan, J.M., Apfel, F. and Tsouros, A.D. (2013), *Health Literacy: The Solid Facts*, World Health Organization, Copenhagen.

Kutner, M., Greenberg, E., Jin, Y. and Paulsen, C. (2006), *The Health Literacy of America's Adults: Results From the 2003 National Assessment of Adult Literacy (NCES 2006-483)*, US Department of Education, National Center for Education Statistics, Washington, DC.

Lam, L. and Yang, L. (2014), "Is low health literacy associated with overweight and obesity in adolescents: an epidemiology study in a 12-16 years old population, Nanning, China, 2012", *Archives of Public Health*, Vol. 72 No. 1, p. 11.

Luopa, P., Kivimäki, H., Matikka, A., Vilkki, S., Jokela, J., Laukkarinen, E. and Paananen, R. (2014), "Nuorten hyvinvointi Suomessa 2000-2013. Kouluterveyskyselyn tulokset. (Wellbeing of adolescents in Finland 2000-2013. The results of the school health promotion study). Terveyden ja hyvinvoinnin laitos. Raportit 25/2014", Juvenes Print, Tampere, pp. 35-39.

Nutbeam, D. (1998), "Health promotion glossary", *Health Promotion International*, Vol. 13 No. 4, pp. 349-364.

OECD (2015), *The ABC of Gender Equality in Education: Aptitude, Behaviour, Confidence*, OECD Publishing, Pisa, avialable at: www.oecd.org/pisa/keyfindings/pisa-2012-results-gender-eng.pdf

Ormshaw, M., Paakkari, L. and Kannas, L. (2013), "Measuring child and adolescent health literacy: a systematic review", *Health Education*, Vol. 113 No. 5, pp. 433-455.

Paakkari, O., Torppa, M., Kannas, L. and Paakkari, L. (2016), "Subjective health literacy: development of a brief instrument for school-aged children", *Scandinavian Journal of Public Health*, Vol. 44 No. 8, pp. 751-757.

Parker, R.M., Ratzan, S.C. and Lurie, N. (2003), "Health literacy: a policy challenge for advancing high-quality health care", *Health Affairs*, Vol. 22 No. 4, pp. 147-153.

Parker, R.M., Baker, D.W., Williams, M.V. and Nurss, J.R. (1995), "The test of functional health literacy in adults", *Journal of General Internal Medicine*, Vol. 10 No. 10, pp. 537-541.

Perry, E.L. (2014), "Health literacy in adolescents: an integrative review", *Journal for Specialists in Pediatric Nursing*, Vol. 19 No. 3, pp. 210-218.

Ratzan, S.C. and Parker, R.M. (2000), "Introduction", in Selden, C.R., Zorn, M., Ratzan, S.C. and Parker, R.M. (Eds), *Current Bibliographies in Medicine – Health Literacy*, National Library of Medicine, Bethesda, MD, pp. v-vi.

Reardon, S.F. (2011), "The widening academic achievement gap between the rich and the poor: new evidence and possible explanations", in Duncan, G.J. and Murnane, R.J. (Eds), *Whither Opportunity. Rising Inequality, Schools, and Children's Life Chances*, Russell Sage Foundation, New York, NY, pp. 91-116.

Rimkute, L., Torppa, M., Eklund, K., Nurmi, J.-E. and Lyytinen, H. (2013), "The impact of adolescents' dyslexia on parents' and their own educational expectations", *Reading and Writing*, Vol. 27 No. 7, pp. 1231-1253.

Schunk, D.H. and Zimmerman, B.J. (1997), "Social origins of self-regulatory competence", *Educational Psychologist*, Vol. 32 No. 4, pp. 195-208.

Shih, S.-F., Liu, C.-H., Liao, L.-L. and Osborne, R.H. (2016), "Health literacy and the determinants of obesity: a population-based survey of sixth grade school children in Taiwan", *BMC Public Health*, Vol. 16 No. 280, pp. 1-8.

Sørensen, K., Van den Broucke, S., Fullam, J., Doyle, G., Pelikan, J., Slonska, Z. and Brand, H. (2012), "Health literacy and public health: a systematic review and integration of definitions and models", *BMC Public Health*, Vol. 12 No. 1, p. 80.

Sørensen, K., Pelikan, J.M., Röthlin, F., Ganahl, K., Slonska, Z., Doyle, G., Fullam, J., Kondilis, B., Agrafiotis, D., Uiters, E., Falcon, M., Mensing, M., Tchamov, K., van den Broucke, S. and Brand, H. (2015), "Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU)", *European Journal of Public Health*, Vol. 25 No. 6, pp. 1053-1058.

Stoet, G. and Geary, D.C. (2013), "Sex differences in mathematics and reading achievement are inversely related: within- and across-nation assessment of 10 years of PISA data", *PLoS One*, Vol. 8 No. 3, pp. 1-10.

Summanen, A.-M. (2014), "Terveystiedon oppimistulokset perusopetuksen päättövaiheessa 2013. (The health education learning outcomes at the end of basic education in 2013)", Koulutuksen seurantaraportit 2014, No. 1, Opetushallitus, Juvenes Print, Tampere.

Torsheim, T., Cavallo, F., Levin, K.A., Schnohr, C., Mazur, J., Niclasen, B. and Currie, C. (2015), "Psychometric validation of the Revised Family Affluence Scale: a latent variable approach", *Child Indicators Research*, Vol. 9 No. 3, pp. 771-784.

UNESCO (2014a), "UNESCO education strategy 2014-2021", Paris.

UNESCO (2014b), "Teaching and learning: achieving quality for all", EFA Global Monitoring Report 2013/14, Paris.

UNESCO (2014c), “Adult and youth literacy”, fact sheet, No. 29, Unesco Institute for Statistics, September, available at: www.uis.unesco.org/literacy/Documents/fs-29-2014-literacy-en.pdf (accessed 29 August 2016).

Vereecken, C.A., Maes, L. and De Bacquer, D. (2004), “The influence of parental occupation and the pupils’ educational level on lifestyle behaviors among adolescents in Belgium”, *Journal of Adolescent Health*, Vol. 34 No. 4, pp. 330-338.

Volandes, A.E. and Paasche-Orlow, M.K. (2007), “Health literacy, health inequality and a just healthcare system”, *American Journal of Bioethics*, Vol. 7 No. 11, pp. 5-10.

Voyer, D. and Voyer, S.D. (2014), “Gender differences in scholastic achievement: a meta-analysis”, *Psychological Bulletin*, Vol. 140 No. 4, pp. 1174-1204.

Corresponding author
Olli Paakkari can be contacted at: olli.paakkari@jyu.fi

III

THE CROSS-NATIONAL MEASUREMENT INVARIANCE OF THE HEALTH LITERACY FOR SCHOOL-AGED CHILDREN (HLSAC) INSTRUMENT

by

Paakkari, O., Torppa, M., Boberova, Z., Välimaa, R., Maier, G., Mazur, J., Kannas, L. & Paakkari, L. 2019.

European Journal of Public Health, 29 (3), 432–436.

https://doi.org/10.1093/eurpub/cky229

Title page

Title of the article: The cross-national measurement invariance of the *Health Literacy for School-aged Childre*n (HLSAC) instrument

Olli Paakkari, Faculty of Sport and Health Sciences, Research Centre for Health Promotion, University of Jyväskylä, Finland.
Minna Torppa, Faculty of Education and Psychology, University of Jyväskylä, Finland.
Zuzana Boberova, Faculty of Science, P. J. Safarik University, Slovakia.
Raili Välimaa, Faculty of Sport and Health Sciences, Research Centre for Health Promotion, University of Jyväskylä, Finland.
Gunter Maier, Institute for Health Promotion and Disease Prevention, Vienna, Austria.
Joanna Mazur, Department of Child and Adolescent Health, Institute of Mother and Child, Warsaw, Poland.
Lasse Kannas, Faculty of Sport and Health Sciences, Research Centre for Health Promotion, University of Jyväskylä, Finland.
Leena Paakkari, Faculty of Sport and Health Sciences, Research Centre for Health Promotion, University of Jyväskylä, Finland.

Corresponding author: Olli Paakkari, P.O. Box 35, FI-40014 University of Jyväskylä, Finland, tel: +358408053582, e-mail: olli.paakkari@jyu.fi

The cross-national measurement invariance of the *Health Literacy for School-aged Children* (HLSAC) instrument

Abstract

Background: Health literacy (HL) is an important determinant of health and health behaviours, and there is a need to monitor HL levels among all population groups. It is therefore essential to develop instruments to assess HL during childhood and adolescence. The aim of this study was to examine the cross-national measurement invariance of the instrument *Health Literacy for School-aged Children* (HLSAC) in four European countries. **Methods:** The data were collected via standardized self-administered anonymous questionnaires within classrooms in Finland, Poland, Slovakia, and Belgium. There were in total 1468 respondents (aged 13, N=690; aged 15, N= 778). The HLSAC instrument was used to measure the subjective HL of adolescents in each country. A multigroup confirmatory factor analysis was applied to test measurement invariance. **Results:** Configural and metric invariance was established, but scalar invariance did not hold. However, the instrument exhibited high internal consistency (α=.85) and showed adequate fit with the data. Moreover, the partial invariance allowed comparison of mean values across the countries in question. There were significant mean value differences between countries and age-groups. **Conclusion:** HL mean values (as assessed via the HLSAC instrument) can be compared across countries. The instrument has utility for large-scale international HL studies on adolescents.

Keywords: health literacy, adolescent, measurement invariance

Introduction

Health literacy (HL) can be defined in various ways.[1-4] It can be understood as a capacity to obtain, process, and construct health information, enabling people to make sound health decisions, and to work on factors that constitute their own and others' health chances. HL is both a constitutive determinant[5,6] of and an independent risk factor[7] for health. Hence, HL may help in tackling global health challenges and in reducing health inequalities between and within population groups.[6,8] HL has consequences at both the societal and the individual level; hence, it is essential to promote and assess the HL of both populations and individuals, especially during childhood and adolescence, when the foundations of many health behaviours are laid. However, little is known about how HL is related to health disparities among adolescents.

Various suggestions have been made for developing an appropriate HL monitoring instrument. Pleasant, McKinney, and Rikard[9] emphasized that a comprehensive instrument should be based on key attributes, including a conceptual framework, multi-dimensionality, and the treatment of HL as a latent construct. One tool along these lines is the instrument entitled *Health Literacy for School-aged Children* (HLSAC).[10] The relevant conceptualization[11] emphasized the multidimensional nature of health literacy (involving several domains, a latent construct, and multiple items)[10]. The instrument was developed to meet the needs of adolescent HL. It was validated using a nationally representative target sample in Finland, where its psychometric properties were shown to be at an adequate level.[10,12] However, the applicability of the instrument (including performance and feasibility) beyond Finland remained to be tested.

It is common in behavioural and social sciences to assess relevant traits, habits, behaviours, attitudes, or conceptions with reference to self-reporting multi-item questionnaires. Usually, the items in such questionnaires are developed to assess an underlying phenomenon (latent factor). In making meaningful and valid comparisons between different groups with reference to the means of a latent factor (as calculated from an instrument), it becomes critical to examine the *measurement invariance* (i.e. the factorial invariance, the measurement equivalence).[13] The attribute of measurement invariance indicates that the instrument performs similarly in different groups or populations.[14] There are several methods to inspect measurement invariance.[15] In the present study we applied multigroup confirmatory factor analysis, which is one of the most commonly used methods to test invariance.[16–18] Our aim was to examine the cross-national measurement invariance of the HLSAC instrument in four European countries, i.e. Finland, Poland, Slovakia, and Belgium.

Methods

Participants and data collection

The research was conducted in collaboration with the *Health Behaviour in School-aged Children* (HBSC) Study Network.[19] The data were collected in Finland, Poland, Slovakia, and Belgium in the spring of 2016. All the participants responded voluntarily and anonymously to a standardized paper-and-pen questionnaire administered during one lesson. The pupils were aware of the confidentiality of the data, and they were informed that only group-level results would be published. Each country followed national ethical/legal requirements for the survey. The final data retained 1468 participants. Of these, 690 pupils were aged 13 (Finland N=176, Poland N=341, Slovakia N=173). In Belgium the number of 13-year-olds remained low; hence, no Belgian 13-year-olds were included in the analysis. There were in total 778 pupils aged 15 (Finland N=175, Poland N=301, Slovakia N=118, Belgium N=184).

Measures

The *Health Literacy for School-aged Children* (HLSAC) instrument, which has been validated among 13- and 15-year-old pupils, was used to measure self-reported health literacy.[10] The brief 10-item instrument includes five core components (theoretical knowledge, practical knowledge, critical thinking, self-awareness, and citizenship), each component being tapped by two items (see Table 2). All the items took the form ‘I am confident that…’. Originally the response options (on a Likert-type scale) were (1) ‘not at all true’, (2) ‘barely true’, (3) ‘somewhat true’, and (4) ‘absolutely true’. The HLSAC score range was thus 10–40. The scores were divided into three categories, covering levels ‘low’ (scores 10–25), ‘moderate’ (scores 26–35), and ‘high’ (scores 36–40).[12] However, it became clear that there is almost no semantic difference in the Slovak language between the options ‘barely true’ and ‘somewhat true’, and the second option ‘barely true’ was therefore changed to ‘not quite true’. All the countries ensured translation and back-translation of items by professional translators, to ensure parity between languages. In Belgium, the questionnaire was translated into Dutch.

To confirm adequate functioning of the original HLSAC items in different countries, 14 items were added to the survey, aiming thus to check whether there could be items that functioned better than the original ten items. The analysis started with inspection of the item distributions. Six items were removed because of low discriminatory power. Next, a confirmatory factor analysis was applied to the remaining 18 items. We inspected the item loadings, and considered the item contents, but found no evidence suggesting a need to change the original HL items. Using regression analysis, we also examined what proportion of the variance of the 18-item instrument was explained by the short 10-item HLSAC instrument. In Finland and Poland, the proportion of explained variance was 97%, in Slovakia 96%, and in Belgium 95%.

Statistical analysis

The distributions were first screened carefully to ensure that each item had adequate variance in each country, and to avoid the biases that can be caused by ceiling/floor effects.. A multigroup confirmatory factor analysis (CFA) was used to test measurement invariance, i.e. whether the HLSAC instrument functioned in a similar way in four European countries (Finland, Poland, Slovakia, Belgium). Multigroup CFA allows simultaneous estimations and direct statistical comparisons of the measurement parameters across groups; it is therefore a useful method to test the structure of a scale.[20] The fits of the CFA models were evaluated using the chi-square goodness-of-fit statistic (χ^2), the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). A good fit with the data is indicated when CFI is higher than .95, lower than .06 for RMSEA, and lower than .08 for SRMR.[21] We considered that the fit with the data would be adequate, even if the chi square test was not non-significant, so long as other indexes suggested a good fit (on the grounds that the chi-square test is sensitive to large sample sizes).[21] The levels of measurement invariance were tested using the χ^2 difference test statistic.

In the first phase of the modelling, we inspected the magnitude of the factor loadings for each country and each age, and fitted the configural invariance model with freely estimated parameters.[22,23] The model was a one-factor model with ten items. Configural invariance requires that the factor structure should be similar across groups, i.e. the construct (latent factor) can be measured by the same items, but the factor loading magnitudes can differ across groups. A precondition for clear comparisons is that there should be the same valid CFA construct in each group.

In the second phase of the modelling we inspected the equivalence of the factor loadings, fitting a metric invariance model. To examine the equality of the factor loadings across countries, the latent factor mean was set at zero, and the variance at one. Metric invariance (with the factor loadings equal across the groups) confirms that the measured construct has the same meaning for the various participant groups.[23] If the metric invariance model holds (via a non-significant χ^2 difference test), we continue by fitting a scalar invariance model (third phase). This is realized when both the factor loadings and the intercepts are equal across groups. If scalar invariance is not reached, one arrives at partial invariance.[24,25]

In the fourth phase of the modelling, the equivalence of the HL mean values between countries was examined. To compare the mean values, the factor loadings were fixed at one, and the corresponding intercept was set at zero, while the factor variance and the HL mean values were allowed to differ across countries. In setting the equality constraints, we first set all the participant countries to the model in each age group. After implementing group comparisons, we conducted all possible paired country comparisons. All the analyses were conducted using Mplus (version 7.3).

Finally, the internal consistency (reliability estimate) of HLSAC instrument in each country was estimated using Cronbach's alpha. In addition, to examine whether the predictive power of the short 10-item HLSAC instrument was similar to that of the longer 18-item instrument, we ran a regression analysis, predicting the 18-item instrument with the 10-item instrument (using SPSS version 24).

Results

Table 1 presents the descriptive HL statistics for the different age-groups. The total HL mean score was 31.02. The highest mean score was found in Finland, and the lowest in Belgium. The HL mean score was higher among pupils aged 15 than among those aged 13.

Table 1. The descriptive statistics and Cronbach alphas for health literacy (HLSAC).

	N	**Mean**	**SE**	**SD**	**α**
Finland, aged 13	176	32.45	0.40	5.28	.90
Finland, aged 15	175	33.11	0.41	5.37	.90
Poland, aged 13	341	30.30	0.27	5.08	.85
Poland, aged 15	301	30.85	0.28	4.94	.85
Slovakia, aged 13	173	31.12	0.37	4.90	.81
Slovakia, aged 15	118	31.33	0.42	4.59	.79
Belgium, aged 15	184	29.33	0.38	5.15	.81
Total	1468	31.02	0.13	5.18	.85

The 10-item HLSAC instrument exhibited an adequate Cronbach alpha (.85) for the data as a whole, and the internal consistency of the instrument was also at a high level in each participant country (Finland α=.90, Poland α=.85, Slovakia α=.80, Belgium α=.81). The CFA models for the 10-item instrument showed adequate fit with the overall data ($\chi^2(35)=200.65$, p=.000; RMSEA = .06, CFI = .96, SRMR = .03), and also with the data for each country (*Finland* $\chi^2(35)=86.53$, p=.000; RMSEA = .07, CFI = .97, SRMR = .03; *Poland* $\chi^2(35)=168.83$, p=.000; RMSEA = .08, CFI = .93, SRMR = .04; *Slovakia*

$\chi^2(35)=51.42$, $p=.036$; RMSEA = .04, CFI = .97, SRMR = .04; *Belgium* $\chi^2(35)=69.23$, $p=.000$; RMSEA = .07, CFI = .92, SRMR = .05).

Comparisons of the factor loadings

Table 2 indicates that the item loadings were in general at the same level in the different age-groups (pupils aged 13 and 15) and countries, although there were a few items for which the loadings seemed to vary slightly between countries.

Table 2. The HLSAC instrument: the standardized item loadings of the confirmatory factor analysis, by age-group and country.

Items per core component	13-year-olds			15-year-olds			
	Finland	Poland	Slovakia	Finland	Poland	Slovakia	Belgium
Theoretical knowledge							
Having good information regarding health	.36	.45	.31	.32	.35	.16	.39
Ability to give examples of things that promote health	.58	.42	.52	.52	.49	.43	.49
Practical knowledge							
Ability to find health-related information that is easy to understand	.53	.48	.54	.47	.45	.58	.40
Ability to follow the instructions given by doctors and nurses	.37	.36	.33	.42	.33	.21	.21
Critical thinking							
Ability to decide if health-related information is right or wrong	.52	.47	.44	.44	.50	.45	.44
Ability to compare health-related information from different sources	.56	.52	.53	.51	.50	.60	.43
Self-awareness							
Ability to justify one´s own choices regarding health	.47	.55	.44	.43	.46	.42	.34
Ability to judge how one´s own behaviour affects one's health	.50	.48	.26	.47	.48	.34	.28
Citizenship							
Ability to judge how one's own actions affect the surrounding natural environment	.50	.47	.50	.51	.44	.41	.42
Ability to give ideas on how to improve health in one's immediate surroundings	.52	.47	.51	.54	.42	.42	.39

The analysis of configural and metric invariance provided evidence that the corresponding common factors had the same meaning across groups. The baseline model (against which the restricted models were compared) had free loadings across all countries ($\chi^2(105)=204.60$, RMSEA=.06, CFI=.95, SRMR=.04). The country comparison between Finland, Slovakia, and Poland in the data for pupils aged 13 showed that while all ten factor loadings were equivalent in Finland and Poland (χ^2diff(10)=15.62, $p=.11$), some loadings were different in Slovakia. We continued the analysis by examining which of the factor loadings were different in the Slovakian data, setting each factor loading free one by one. The analysis showed that there was one loading that was lower in the Slovakian data ('ability to judge how one´s own behaviour affects one's health'). A model in which all ten factor loadings were set as equivalent in Finland and Poland, and in which all factor loadings except for one item were set as equivalent in Slovakia, Finland, and Poland, proved to be as well fitted to the data as a model with all the loadings free in each country (χ^2diff(19)=21.01, p=.33).

The baseline model for pupils aged 15 had free loadings across all countries (χ^2(140)=261.54; RMSEA=.07, CFI=.95, SRMR=.05). The comparison of the factor loadings for pupils aged 15 indicated that the loadings were equivalent in Finland, Poland, and Belgium (χ^2diff(20)=27.56, p=.12). The model fit fell below the accepted limit when all the Slovakian factor loadings were set as equal to the loadings for the other countries (χ^2diff(30)=53.87, p=.01). After releasing one item ('having good information regarding health') in the Slovakian data, the chi square difference test approached nonsignificance (χ^2diff(29)=44.64, p=.03). After releasing three items in the Slovakian data ('having good information regarding health', 'ability to follow the instructions given by doctors and nurses', 'ability to judge how one´s own behaviour affects one's health') the difference test showed nonsignificance (χ^2diff(27)=38.73, p=.07). Note, however, that the loading for the item 'ability to follow the instructions given by doctors and nurses' was equivalent in Slovakia and Belgium. After releasing the item, 'having good information regarding health' in the Slovakian data, and setting the item 'ability to follow the instructions given by doctors and nurses' as equal in Belgium and Slovakia, the chi square difference test showed nonsignificance (χ^2diff(28)=37.80, p=.10). These analyses suggest that the 10-item model with equal loadings across countries fitted well with the data, and particularly well in Finland, Poland, and Belgium.

Examination of intercept equivalence across the three countries showed that the intercepts were not equal across countries, and that scalar invariance did not hold. This was the case both for the 13-year-olds (χ^2diff(39)=258.79, p<.001), and for the 15-year-olds (χ^2diff(57)=268.10, p<.001).

The analyses suggested that configural and metric invariance holds for comparisons between Poland and Finland (for 13-year-olds), and between Poland, Finland, and Belgium (for 15-year-olds), and that models fitted the data well. Based on these analyses, partial invariance[24,25] holds. We considered this to be sufficient to conduct mean value comparisons between the countries in question. However, all the comparisons should be interpreted with caution, given the slight differences that emerged.

Comparisons of HL mean values

Comparison of the mean values for 13-year-olds (see Table 1) across the countries revealed differences (χ^2diff(123)=238.08, p=.000). All the paired country differences were tested, and all the paired country comparisons were significant (FIN>POL, FIN>SLO, POL<SLO). Comparison of the mean values for 15-year-olds (see Table 1) also indicated differences between countries (χ^2diff(167)=310.92, p=.000). Subsequent paired testing showed significant differences between the following pairs: FIN>SLO, FIN>POL, FIN>BEL, POL<SLO. Poland and Slovakia showed no difference from Belgium. These findings were confirmed in the overall multigroup models. Thus, comparison of the baseline model (with all means free in all countries) to the model in which the means of Poland and Belgium were set as equal showed a nonsignificant chi-square difference (χ^2diff(1)=1.77, p=.18).

Discussion

The aim of this study was to examine the cross-national measurement invariance of the instrument *Health Literacy for School-aged Children* (HLSAC) in four European countries. The psychometric properties of the instrument were at a sound level, with configural and metric invariance accomplished. The original HLSAC instrument has been found to have adequate internal consistency in nationally representative Finnish HBSC data (α= .93)[10] and in a Turkish study (α= .77)[26]. This was the case also in the present study (total α=.85, variation between countries α=.80 – .90).

Furthermore, in this study, the 10-item HLSAC instrument predicted 95–97 % of the variance of the longer 18-item instrument, and there was no compelling evidence (involving poor item distributions, factor structure, or reliability estimates) to change the items of the original HLSAC instrument in any of the countries.

In the CFA models for the 10-item instrument, the fit with the data was highest in Slovakia, and the fit with the other countries was at an adequate level. Metric invariance was established, and the tests showed that the factor loadings were equivalent in Finland, Poland, and Belgium at age 15, and in Finland and Poland at age 13. The models at both ages were slightly poorer when the Slovakian factor loadings were set as equal to those of the other countries. It should be noted that it is very difficult to reach full measurement invariance in most empirical studies.[13] From the perspective of measurement, it has been proposed that more liberal criteria in assessing invariance might be appropriate, on the grounds that the traditional criteria for testing exact measurement invariance may be too strict.[27] There are many issues that can cause variation in how respondents react to questions, and bias or lack of equivalence can cause issues, involving for example instrument adaptation (translation), social desirability, or familiarity with item response formats.[28,29] The questionnaire was translated from the original English version to the target language. Due the translation process or cultural differences, the meanings might have differed to some extent in the target languages.

Comparison of the HLSAC mean values showed significant differences between the countries in both age groups, except for Slovakia/Belgium and Poland/Belgium (15-year-olds). For both age groups, the self-reported health literacy was highest in Finland, followed by the 13-year-olds in Slovakia and Poland. The order was the same among the 15-year-olds (Finland, Slovakia, Belgium, Poland). A number of reasons could underlie the systematic statistically significant differences between Finland and the other countries. In Finland, health issues are taught at comprehensive school (grades 7–9, ages 13–15), within the subject of health education, which is an independent and obligatory school subject. The national core curriculum lays down the objectives of health education. Schools are obliged to follow these, and must also offer the same amount of health education to all pupils. In addition, pupils can gain health knowledge within other school subjects, and from general health promotion work in the schools (as part of the whole-school approach).

The present study has some limitations. It covered only four languages and one language per country. In future, it would be important to investigate measurement invariance in other countries and languages, as well as in different age groups. It is always challenging to obtain translations that are adequately similar, in such a way that participants can understand items in the same way, regardless of country, culture, or language. In future studies, it will be important to continue to pay careful attention to the translation process, as this could have been one key reason for not achieving scalar invariance in this sample.

To conclude, it is essential to develop HL measurement instruments, and to assess the levels across various settings, as this creates a basis for appropriate interventions and meaningful concrete practices. There is clearly a need for instruments such as HLSAC in the fields of public health and health promotion, in which group comparisons are frequently of interest. We suggest that the HLSAC Instrument has good applicability for these purposes, and strong potential for use in international contexts. In future, the aim will be to apply the instrument in more countries, and to examine the association between HLSAC and health behaviour. Surveys such as the HBSC study will offer good opportunities for such comparative research, with possibilities to develop the instrument further if necessary.

Conflict of interest

None declared.

Acknowledgements

We thank Professor Thomas Abel and M.Sc. Laura Goeint for their help with the research.

Funding

Funded by Juho Vainio Foundation [#21000037611], Academy of Finland [#276239], and the Scientific Grant Agency of the Ministry of Education Science, Research, and Sport of the Slovak Republic and the Slovak Academy of Sciences [reg. no. 1/0427/17].

Key points

- The applicability of the HLSAC instrument was tested across a number of European countries for the first time.
- The instrument is applicable in the participant countries, with a degree of caution.
- The HLSAC Instrument is a promising tool for comparing the subjective HL of adolescents in international contexts.

References

1 Berkman ND, Davis TC, McCormack L. Health Literacy: What Is It?. *J Health Commun* 2010;15:9–19.

2 Bröder J, Okan O, Bauer U, et al. Health literacy in childhood and youth: a systematic review of definitions and models. *BMC Public Health* 2017; 17:361.

3 Sørensen K, Van den Broucke S, Fullam J, et al. Health literacy and public health: a systematic review and integration of definitions and models. *BMC Public Health* 2012;12: 80.

4 Sørensen K, Pleasant A. Understanding the Conceptual Importance of the Differences Among Health Literacy Definitions. *Stud Health Technol Inform* 2017;240:3–14.

5 Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. *Ann Int Med* 2011;155:97–107.

6 Heide I, Uiters E, Boshuizen H, Rademakers J. Health Literacy in Europe: the development and validation of health literacy prediction models. *Eur J Public Health* 2016;26:906–11.

7 Volandes AE, Paasche-Orlow MK. Health literacy, health inequality and a just healthcare system. *The Am J Bioeth* 2007;*7*: 5–10.

8 Batterham RW, Hawkins M, Collins PA, Buchbinder R, Osborne RH. Health literacy: applying current concepts to improve health services and reduce health inequalities. *Public Health* 2016;132:3–12.

9 Pleasant A, McKinney J, Rikard RV. Health literacy measurement: a proposed research agenda. *J Health Commun* 2011;16:11–21.

10 Paakkari O, Torppa M, Kannas L. Paakkari L. Subjective health literacy: development of a brief instrument for school-aged children. *Scand J Public Health* 2016;44:751–757.

11 Paakkari L, Paakkari O. Health literacy as a learning outcome in schools. *Health Education* 2012;112:133–152.

12 Paakkari O, Torppa M, Villberg J, Kannas L, Paakkari L. Subjective health literacy among school-aged children. *Health Education* 2018;118.

13 Van De Schoot R, Schmidt P, De Beuckelaer A, Lek K, Zondervan-Zwijnenburg M. Editorial: Measurement Invariance. *Frontiers in Psychology* 2015;6:1064.

14 De Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in medicine. Practical guides to biostatistics and epidemiology.* Cambridge: Cambridge University Press, 2015.

15 Teresi JA. Overview of quantitative measurement methods. Equivalence, invariance, and differential item functioning in health applications. *Med Care* 2006;44:39–49.

16 Jöreskog KG. Simultaneous factor analysis in several populations. *Psychometrika* 1971;36:409–426.

17 Byrne BM. Testing for multigroup invariance using AMOS graphics: a road less traveled. *Struct Equ Model* 2004;11:272–300.

18 Brown, T. A. *Confirmatory Factor Analysis for Applied Research.* New York: Guilford Press, 2006.

19 Currie C, Nic Gabhainn S, Godeau E and the International HBSC Network Coordinating Committee. The Health Behaviour in School-aged Children: WHO Collaborative Cross-National (HBSC) Study: origins, concept, history and development 1982–2008. *Int J Public Health* 2009;54:131—139.

20 Little TD, Slegers DW. Factor analysis: Multiple groups. In: Everitt BS, Howell DC, editors. *Encyclopedia of statistics in behavioral science.* Chichester: Wiley, 2005:617–623.

21 Hu, LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct equ modeling* 1999; 6:1–55.

22 Horn JL, McArdle JJ. A practical and theoretical guide to measurement invariance in aging research. *Exp Aging Res* 1992;18:117–144.

23 Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organ Res Methods* 2000;3:4–70.

24 Byrne BM, Shavelson RJ, Muthén B. Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychol Bull* 1989;105:456–466.

25 Steenkamp JB, Baumgartner H. Assessing measurement invariance in cross-national consumer research. *J Consum Res* 1998;25:78–90.

26 Haney, MO. Psychometric testing of the Turkish version of the Health Literacy for School-Aged Children. *J Child Health Care* 2017;1–11.

27 Muthén BO, Asparouhov T. *BSEM Measurement Invariance Analysis.* Mplus Web Notes: No. 17, 2013. https://www.statmodel.com/examples/webnotes/webnote17.pdf Accessed 9.10.2017.

28 Davidov E, Meuleman B, Cieciuch J, Schmidt P, Billiet J. Measurement equivalence in cross-national research. *Annu Rev Sociol* 2014;40:55–75.

29 Oberski DL, Weber W, Revilla M. The effect of individual characteristics on reports of socially desirable attitudes towards immigration. In: Salzborn S, Davidov E, Reinecke J editors. *Methods, Theories, and Empirical Applications in the Social Sciences: Festschrift for Peter Schmidt*. Wiesbaden: Springer, 2012:151–158.