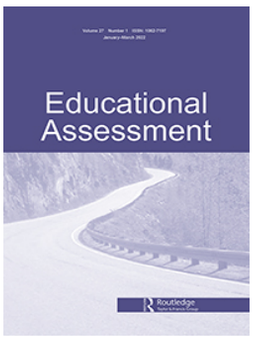

Routledge
Taylor & Francis Group

Educational Assessment

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/heda20

Health Education Teachers' Assessment Conceptions and Practices: Identifying Assessment Profiles

Olli Paakkari, Leena Paakkari, Henna Haapala & Mirja Hirvensalo

To cite this article: Olli Paakkari, Leena Paakkari, Henna Haapala & Mirja Hirvensalo (2022): Health Education Teachers' Assessment Conceptions and Practices: Identifying Assessment Profiles, Educational Assessment, DOI: 10.1080/10627197.2022.2063832

To link to this article: https://doi.org/10.1080/10627197.2022.2063832

Published online: 21 Apr 2022.

EDUCATIONAL ASSESSMENT
https://doi.org/10.1080/10627197.2022.2063832

Routledge
Taylor & Francis Group

Health Education Teachers' Assessment Conceptions and Practices: Identifying Assessment Profiles

Olli Paakkari [a], Leena Paakkari [a], Henna Haapala [b], and Mirja Hirvensalo [c]

[a]Research Centre for Health Promotion, University of Jyväskylä, Jyvaskyla, Finland; [b]Open University of the University of Jyväskylä, Jyvaskyla, Finland; [c]University of Jyväskylä, Jyvaskyla, Finland

ABSTRACT

The study explored the latent construct underlying the assessment conceptions and practices of Finnish Health Education teachers (n = 165) in the context of curricula, seeking thereby to identify the teachers' assessment profiles. Six underlying factors were found to encompass their assessment conceptions and practices, namely *Assessment supporting learning, Assessment of working, Self and peer assessment as part of grading, Common assessment criteria, Questionable assessment practices*, and *Norm-referenced assessment*. Via cluster analysis, three distinct assessment profiles were identified, labeled as *Problematic assessors, Learning supportive assessors*, and *Norm-based assessors*. These findings can be used to develop Health Education teacher training and facilitate teachers' assessment literacy.

Introduction

A decade ago, Popham (2009) regretted that when most teachers completed their teacher-education programs, there was "no requirement that they learn anything about educational assessment." He suggested that this could explain why many teachers then had a fairly limited understanding of educational assessment, and looked forward to an improvement in the situation over the next ten or twenty years. Indeed, assessment literacy – viewed as teachers' understanding and knowledge of assessment and the assessment practices they carry out (Popham, 2011) – can now be seen as an essential part of teachers' professionalism (Popham, 2009; Xu & Brown, 2016).

In parallel with the adoption of constructivist and sociocultural approaches to learning and teaching, there has been a move toward learning-focused models of assessment, including requirements for a balance between formative and summative assessment strategies (Hildén & Fröjdendahl, 2018). In fact, assessment can be said to have become an inseparable component of the teacher's work (Hildén & Fröjdendahl, 2018) and professional identity (Xu & Brown, 2016). As noted by Xu and Brown (2016):

> Under formative assessment policies, teachers assess students to make pedagogical decisions, and these decisions are sometimes used as part of formal certification processes. Under such circumstances, the teacher's role and identity now include "assessor." Understanding this role and successfully integrating it into the teacher's pedagogical function requires a new way of understanding what it means to be a teacher. (p. 158)

An important notion is that assessment should primarily be conducted to support learning (Black & Wiliam, 2012). In part, this constituent role of assessment, i.e. to support student learning in the classroom (Andrade, 2013; Black & Wiliam, 1998a, 1998b; Stiggins, 2010), explains the growing interest in exploring assessment literacy, as also does the general accountability policy of public education adopted in many countries (DeLuca, LaPointe-McEwan, & Luhanga, 2016).

CONTACT Olli Paakkari olli.paakkari@jyu.fi Faculty of Sport and Health Sciences, Research Centre for Health Promotion, University of Jyväskylä, P.O. Box 35, FI-40014, Finland

In Finland, it would have been unrealistic to expect such a major shift in perspective to occur over a single decade given that so many other interrelated changes had already taken place, all with implications for teachers' work. In fact, already in the 1980s it was possible to identify educational changes linked to new philosophical conceptions, involving shifts in ways of understanding teaching, learning, and knowing, which strongly influenced school curricula and daily practices in schools (Sahlberg, 2015). Elsewhere, several other European countries were facing challenges related to externally identified learning standards, school inspections, and competition. Furthermore, in the absence of externally monitored and annually conducted national assessments, a significant part of teachers' work has always involved assessment of students' learning and subject-related competencies (Sahlberg, 2015). Within Finland, a paradoxical educational situation has prevailed, describable – in the phrase used by Sahlberg (2015), – as "test less, learn more." This actually means that despite the lack of national tests, or perhaps because of it, teachers are involved in designing and conducting fairly versatile ongoing assessments. The assessments take place as part of normal classroom practice, and after each semester.

Teachers' conceptions of various teaching- and learning-related phenomena are intertwined with their teaching-related behavior (Barnes, Fives, & Dacey, 2015; Fulmer, Lee, & Tan, 2015; Marton & Booth, 1997). In accordance with this, good assessment literacy has been associated with better decisions regarding classroom assessment (Popham, 2009). Indeed, according to Stiggins (1995), assessment-literate educators "come to any assessment knowing what they are assessing, why they are doing so, how best to assess the achievement of interest, how to generate sound samples of performance, what can go wrong, and how to prevent those problems before they occur." On the basis that assessment practices "influence students' learning and classroom learning cultures" (DeLuca, Coombs, & LaPointe-McEwan, 2019, p. 159), it is important to study and develop teachers' assessment literacy (including conceptions and practices of assessment), both in general and in relation to individual subjects.

Assessment literacy is a complex construct with multiple dimensions, and hence, theoretical operationalizations vary (e.g. DeLuca et al., 2016; Pastore & Andrade, 2019; Xu & Brown, 2016). Xu and Brown (2016), for instance, have developed an assessment literacy framework consisting of six components as follows: 1) the knowledge base (e.g. pedagogical content knowledge, knowledge of assessment purposes/content/methods/grading/etc., and knowledge of assessment ethics); 2) teachers' conceptions of assessment; 3) institutional and socio-cultural contexts (e.g. policies, norms, regulations setting boundaries to teachers' assessment practices); 4) teacher assessment literacy in practice; 5) teacher learning; and 6) teacher identity (reconstruction (as an assessor).

In contrast with the above, Pastore and Andrade (2019) have suggested that assessment literacy comprises three interrelated dimensions, these being *conceptual* (knowledge of assessment), *praxeological* (assessment in practice), and *socio-emotional* (management of the social and emotional aspects of assessment). For their part, DeLuca et al. (2016) proposed key themes of assessment literacy, viewing it as comprising an understanding of the purposes, processes, fairness, and measurement theory pertaining to assessment. While there are differences between the conceptualizations, similarities can be found in the overall emphasis placed on teachers' conceptions related to assessment.

Teachers' conceptions of assessment are related to their epistemological beliefs concerning learning and teaching (Brown, 2004). The conceptions vary abundantly (Coombs, DeLuca, LaPointe-McEwan, & Chalas, 2018; Shepard, 2000). They influence how "an individual understands, responds to, and interacts with a phenomenon" (Brown, 2004, p. 303), and mediate the relationship between theoretical knowledge of the assessment and related practices (Xu & Brown, 2016). As noted by Xu and Brown (2016), "the conceptions of assessment denote the belief systems that teachers have about the nature and purposes of assessment" (p. 156). The conceptions cover an understanding of what is true or not in assessment (the cognitive aspect), and how assessment is emotionally perceived (i.e. the affective aspect). Overall, they further form an organizing, interpretive, and guiding framework that influences how conceptions change, and what changes are seen as relevant (Xu & Brown, 2016; see also Pastore & Andrade, 2019). According to Xu and Brown (2016), teachers' assessment literacy is in practice

"constantly negotiated between teachers' conceptions of assessment and the macro socio-cultural, micro institutional contexts and expected knowledge base." This being so, "it reflects a temporary equilibrium reached among tensions" (p. 157).

The assessment literacy of teachers, and particularly their conceptions of assessment, have been studied in the context of different school subjects, including science (Abell & Siegel, 2011) and language (Hildén & Fröjdendahl, 2018). However, no such research has been conducted in the context of Health Education or allied subjects. This may be due to the fairly small number of countries offering Health Education as an independent school subject. Indeed, Finland is one of the few countries where health issues are taught through an independent and statutory school subject, in basic education (students aged 7–15), and in upper secondary education (students aged 15–19).

General assessment policies and practices in Finnish compulsory education

Finnish compulsory education covers comprehensive education (basic education, grades 1–9) and post-comprehensive education (general upper secondary education or vocational education, lasting 3–4 years). Both post-comprehensive options provide a gateway to tertiary education (universities and polytechnics). Unlike the situation in many other countries, Finnish students are not frequently tested via nationwide standardized tests (Sahlberg, 2015). The matriculation examination conducted at the end of general upper secondary education is the only high-stakes test during compulsory education. For the test, the students have the freedom to select the subjects (e.g. Health Education) to be included in the test. A minimum of five subjects can be taken for the matriculation examination, and these can be divided into several semesters. Otherwise, assessments are organized and numerical grades given by an individual teacher for the classes or groups she or he teaches. The grades given reflect teachers' understanding of how well students have reached the learning goals and assessment criteria described in the national core curricula (Vainikainen et al., 2017).

To provide school-related information in support of national and local policy making and decision making, the Finnish Education Evaluation Center (FINEEC) organizes sample-based curricular summative assessments at the end of basic education (grade 9), at least once during the national curriculum period. The summative assessments monitor how well students have reached the learning goals mentioned in the national school curriculum for basic education. In Health Education the first – and so far the only assessment – took place in Spring 2013 (see Summanen, 2014). Unlike high-stakes standardized testing, the tests in these national summative assessments are neutral, in the sense that their purpose is not to categorize the students into less or more talented or the schools into poor or better schools (Vainikainen et al., 2017). Instead, they produce information about what is going on in the schools. On the basis of this information, the municipalities, schools, and teachers can tune schooling and arrange more effective teaching and learning (Vainikainen et al., 2017).

The fact that Finland does not have national level standardized tests creates a significant responsibility for teachers to conduct accurate and comparable assessments, particularly in the final assessment phase of basic education; the aim here is to secure equity in the grades, which serve as selection criteria for post-comprehensive education. Although the Finnish assessment system functions relatively well, there have nevertheless been concerns regarding teachers' assessment practices. Teachers most commonly use summative and individual-based assessment methods, with little use of formative assessment methods (Atjonen et al., 2019). Furthermore, there exist some norm-referenced assessment practices in which students' performances are compared to the rest of the group (Atjonen et al., 2019), and summative assessment practices where students with same level competencies are graded differently (Hilden, Ouakrim-Soivio, & Rautopuro, 2016), constituting a clear equity issue. To address these challenges the Finnish National Board of Education introduced slightly clarified assessment principles (FNBE, 2020a), and revised criteria for the final assessment in basic education (FNBE, 2020b). The aim was to emphasize the importance of assessment for learning, and to promote a somewhat more systematic approach in order to ensure assessment consistency. Despite this, the

basic idea remains unchanged; the teachers retain control and responsibility over curricula implementation and assessment methods, without having to submit to strong top-down external control (see Sahlberg, 2015; Vainikainen et al., 2017).

Assessment guidelines in the Finnish national school curricula: general and health education-specific guidelines

The curricula (FNBE, 2014, 2016) cover the guidelines for conducting learning assessment at the general level, and in individual subjects, including Health Education. The guidelines are based on the Basic Education Act (628/1998), on the General Upper Secondary Education Act (629/1998, revised 2018), and on current educational research. The acts state that assessment should support learning and develop students' abilities for self-assessment, and also give feedback to students on how they have met the objectives of the subject. The student's work, learning, and development should be assessed in various ways. All teachers, and schools more broadly, are obliged to follow the national curricula, including how assessment should be organized in schools.

A description of the subject named Health Education is written into the national core curriculum for basic education (FNBE, 2014) and for general upper secondary schools (FNBE, 2016). The curricula for Health Education (as for other subjects) includes the general aim of the subject, the objectives for instruction and learning, and a description of the form of assessment. The overall aim of Health Education is to improve students' health literacy. It covers the abilities to make reasonable health-related choices and decisions, in such a way that students can recognize and adjust factors that affect their health. Health literacy enables students to promote and maintain their own health, and also the health of other people and their environment (FNBE, 2014, 2016; Paakkari & Paakkari, 2012). Health Education learning objectives are formulated for the key dimensions of health literacy, which are taken to encompass theoretical knowledge, practical knowledge (skills), critical thinking, self-awareness, and citizenship (Paakkari & Paakkari, 2012).

A crucial principle is that assessment in all subjects should be ethically sound, and should be fair. The other main principles of assessment are that it should take place in an encouraging atmosphere, occur within interaction between teacher and students, be based on criteria derived from learning objectives, be conducted using a range of assessment practices, and be organized during the different stages of instruction. In addition, the curricula explicitly emphasizes that in Health Education one must ensure that the assessment does *not* focus on the student's ways of behaving, values, or attitudes; nor should it be influenced by her/his sociability, temperament, or other personal characteristics.

Finnish teachers have a responsibility to conduct classroom assessment, and should be capable of designing and applying a range of assessment methods. Classroom assessment should be – as appropriate – diagnostic (at the pre-assessment stage, conducted before or at the start of a study module), formative (involving assessment for learning, conducted within a single study module or school year), or summative (involving an assessment of learning, conducted at the end of a study module and at the end of the school year) (FNBE, 2014, 2016, 2020a). For students, the main purpose of assessment during the studies as described in the curricula is to guide and encourage learning, to develop a capacity for self-assessment and support students as lifelong learners. For teachers, the assessment is a means to reflect on their own work. Its key purpose is to facilitate adaptation of the instruction to meet students' needs, and to develop teaching practices (FNBE, 2014, 2016, 2020a). The task of the final assessment is to define how well the students have achieved the objectives of the basic or upper secondary education syllabus in a given subject. Within basic education, the curriculum includes final assessment criteria for each core subject, aiming to support teachers' assessment work and to improve the national comparability of grades given by teachers. The curriculum provides the criteria for good performance (numerical grade 8) in the assessment at the end of grade six, and defines the required level of performance to receive a numerical grade 5, 7, 8, or 9 on the final assessment at grade nine. The grading scale is from 4 (fail) to 10 (excellent). Diverse evidence is to be used to assess the student's performance according to the stated criteria (FNBE, 2016, 2020b).

The aim of the study

The present study is the first to report Health Education teachers' conceptions and practices (related to assessment, particularly following the introduction of the most recent national core curriculum for Basic Education (FNBE, 2014) and for Upper Secondary Education (FNBE, 2016) in Finland. The more specific research questions were:

(1)) What kind of latent construct can be found regarding Finnish teachers' assessment conceptions and practices in Health Education, in the context of the curricula?
(2)) What kinds of teachers' assessment profiles can be identified on the basis of their assessment conceptions and practices?

The findings on teachers' assessment profiles can shed light on whether the teachers are following the assessment principles set out on the curricula (specific and general) and in the education acts, and further, identify the kinds of differences that may exist between teachers. The findings can be used to develop teacher training in Health Education, and teachers' assessment competencies in Health Education.

Methods

Participants and data collection

The data for the study were collected in Finland in the spring of 2019 via an online survey. The participants were recruited in collaboration with The Association of Physical and Health Educators in Finland, using their member registry. All the participants responded voluntarily and anonymously. They were aware of the confidentiality of the data, and the fact that only group-level results would be published. In total, the sample consisted of 165 Health Education teachers (female n = 136, male n = 29), out of an estimated 800 in total. The participants were from each province of Finland.

The background variables encompassed teachers' gender, age, teaching experience (years), school size, and average number of weekly Health Education lessons. In addition, the amount of Health Education studies and the school level were investigated. Around 67% of the teachers worked in basic education, and 33% in upper secondary education. The teachers represented schools of different sizes; approximately half of the teachers worked in schools with over 400 students, 44% worked in schools with 200–399 students, and the remainder (9%) taught in smaller schools. The sample consisted of teachers at different stages of their careers. Their teaching experience varied from under 10 years of instruction (45% of participants), through teachers who had taught for 11–20 years (36%), to highly experienced teachers (over 20 years of teaching, 19% of participants). On average, participants taught 6.3 hours of Health Education per week. Regarding the age profile of the respondents, 19% were aged 25–34, 69% were aged 35–54, while 12% were aged 55–64. The majority of the respondents (87%) had completed intermediate-level Health Education studies (60 ECTS), which is the minimum requirement for working as a subject teacher in secondary or upper secondary education. A further 10% had completed advanced studies in Health Education (120 ECTS).

Measures

An expert group generated an online questionnaire to collect data concerning Health Education teachers' assessment conceptions. The expert group consisted of eight university teacher educators, who were also researchers from the field of Health Promotion, and from Health and Physical Education. The members had experience in developing measurement instruments, and in teaching experience at different levels of the educational system, from basic education to higher education. In addition, the group members were experienced in developing national curricula for Health Education as a school subject.

The first step to in developing the questionnaire consisted of a literature review on the topic. The group of four professionals identified the main themes from earlier assessment literacy research, and from the aims and assessment criteria described in the Finnish national core curriculum for basic education (FNBE, 2014) and for upper secondary education (FNBE, 2016). Thereafter, the first draft of the questionnaire was piloted (30 university lecturers, 18 Health Education student teachers), which led to minor corrections to the final questionnaire.

The final questionnaire included in total 35 items, aimed at eliciting teachers' conceptions and practices of assessment in connection with the assessment principles set out in the national curricula and acts. The first set of items focused on issues such as learning, learning support, feedback, planning of teaching, and the definition of the students' initial level or competence. The items were formed by the statement "I think that assessment … " followed by a completion relating to the issue in question. The response options consisted of "never," "quite rarely," "sometimes," "quite often," "often." The second set of items covered teachers' conceptions of varying or uniform assessment criteria, norm-referenced assessment, self and peer assessment, and general assessment principles. The question "What do you think of the following statements related to assessment?" was followed by statements, plus the response options "strongly disagree," "disagree," "neither agree nor disagree," "agree," "strongly agree." The last set of items included teacher practices related to factors they include in assessment, for example success in exams or other assignments, active participation, independent studying, homework, health behavior, temperament and attitudes toward the school subject, and a healthy lifestyle. The question "To what extent do the following factors influence the grade you give in Health Education" was followed by the relevant factors, plus the response scale "not at all," "to a small extent," "to a moderate extent," "to a great extent," "to a very great extent."

Statistical analysis

Data analysis was performed in two phases. First of all, the underlying structure of the questionnaire was investigated via exploratory factor analysis (EFA). There was no definite theoretical assumption regarding the factorial structure, and the questionnaire was constructed in a fairly open way with a view to measuring the teachers' assessment-related conceptions. The factor extraction in EFA was conducted via the Alpha method, in order to maximize the alpha reliability of the factors. The oblique Promax method was selected for the rotation procedure, since it allows factors to be correlated. The item reduction was conducted one item at a time, with judgments based on the magnitudes of factor loadings, the internal consistency reliability for the items in each factor (Cronbach's alpha), and the interpretability of the factors. The Kaiser-Meyer-Olkin (KMO) test and Bartlett's test for sphericity were conducted to investigate how well the data suited the EFA. The KMO test measures sampling adequacy for each variable in the model, and for the complete model. A KMO value higher than 0.6 indicates that the sampling is adequate for EFA (Cerny & Kaiser, 1977; Field, 2009). Bartlett's test compares the correlation matrix with the identity matrix; if the correlation matrix is an identity matrix, this can indicate that the factor model is inappropriate. A low p-value in the Bartlett test (i.e. less than 0.05) indicates that the EFA is useful with respect to the data (Field, 2009). In addition, to evaluate the generated EFA model fit to the data, the Chi-square test, the Root Mean Square Error of Approximation (RMSEA), and the Root Mean Square of the Residuals (RMSR) were used. A good fit with the data is indicated when the RMSEA is lower than .06, and the RMSR lower than .08 (Hu & Bentler, 1999).

Thereafter, the teachers' assessment profiles (i.e. different classes of teachers on the EFA- generated factors) were analyzed via two-step cluster analysis (CA). With this exploratory analysis method it is possible to identify natural homogenous structures within the data, and thus to classify teachers into different assessment profiles. The CA algorithm aims to find the solution that maximizes the between-profile variance while minimizing the within-profile variance (Ketchen Jr. & Shook, 1996). The predictor importance value (range 0 to 1) indicates how well the factors are able to estimate the cluster model. As

stated in IBM (2016), the greater the value, the higher the importance measure will be, and the less likely it is that the variation for a variable (i.e. a factor) between clusters is due to chance and more likely due to some underlying difference.

The overall goodness-of-fit of the cluster structure was analyzed via the silhouette measure of cohesion and separation. In this measure the index ranges from −1 to 1, with a fair fit represented by 0.2 or higher, and a good fit by 0.5 or higher (IBM, 2011). In addition, Schwarz's (1978) Bayesian information criterion (BIC) and the size of the clusters were used to identify an adequate cluster solution. In order to compare the clusters (i.e. the teachers' assessment profiles) with background variables, a one-way analysis of variance was performed using Tukey's test for honestly significant difference (applying pairwise comparisons between groups).

All the analyses were conducted using SPSS (version 24), except for the goodness-of-fit analyses of the EFA model, which were made with R project (psych-package).

Results

Analysis of the latent structure of the data

The latent structure of the questionnaire was analyzed via EFA. The KMO value (.661) and Bartlett's test of sphericity ($\chi^2(630) = 1726.30$, $p = .000$) showed that the data were adequate for the EFA. A 6-factor solution (Table 1) was chosen after investigating the eigenvalues, the scree plot, and the interpretability of solutions, and the factor model was found to have a proper fit to the data ($\chi^2((130) = 184.45$, $p = .001$; RMSEA = .05; RMSR = .04). Four factors had an eigenvalue above 2.3, while the remaining two factors had an eigenvalue slightly under 2.0. The scree plot showed a clear drop

Table 1. Item loadings from the exploratory factor analysis, and Cronbach alphas per factor.

Factors and items	Factor loading	α
Assessment supporting learning		
Helps to identify student's needs for learning support	.74	.74
Supports planning of teaching	.63	
Helps me to define student's competence level	.62	
Enables provision of the continuous feedback to a student about his/her learning	.60	
Helps a student to learn	.46	
Helps me to map out the starting level of the student's competence	.38	
Assessment of working		
Independent studying and taking the responsibility for one's own studying	.60	.70
Doing homework	.54	
Active participation in class activities	.53	
Taking others into account in collective learning situations	.43	
Success in other written assignments (e.g. portfolio, essay) than in exam	.42	
Self and peer assessment as part of grading		
Grading should also include pupil's self-assessment	.67	.71
Peer assessment can be included as part of the numerical assessment	.59	
Student's own evaluation of his/her competence in HE influence to grading	.57	
Common assessment criteria		
All teachers who teach HE in our school have uniform criteria for assessing HE	.79	.76
All teachers who teach HE in our school have common principles for assessing pupils in need of support	.54	
Questionable assessment practices		
Student's health-related behavior	.74	.73
Student's attitude toward healthy lifestyle	.67	
Student's attitude toward the subject	.55	
Student's shyness or other factor related to the temperament	.54	
The assessment criteria may vary depending on the pupil, even if the pupil is studying according to the general syllabus	.48	
Norm-referenced assessment		
Factors included grading: student's competence in relation to other students in the class	.67	.69
During the studies, the assessment of individual pupils should always be commensurate with the competence of other students	.64	

from factor 6 to factor 7. Items with low loadings (0.3 or less) were eliminated, as were items that strongly reduced the internal consistency of the current factor. After item reduction, all the factors showed adequate internal consistency reliability (with Cronbach's alphas varying between .69 and .76), and a 6-factor solution was reasonably interpretable.

The factors were named according to the items that loaded onto a given factor. The following factors were generated: 1) Assessment supporting learning; 2) Assessment of working; 3) Self and peer assessment as part of grading; 4) Common assessment criteria; 5) Questionable assessment practices; 6) Norm-referenced assessment. The items in the factor *Assessment supporting learning* related to e.g. students' needs for learning support, feedback, the defining of students' competence or starting level, and the planning of teaching. The factor *Assessment of working* covered items which described students' working skills, such as competence in independent and responsible studying, active participation, and taking other students into account in collective learning. The factor *Self and peer assessment as part of grading* comprised items which related to the notion of using self and peer assessment in decisions on grading. The items in the factor *Common assessment criteria* linked to having uniform criteria and general principles of assessment in the school. The *Questionable assessment practices* factor included items highlighting dubious assessment variables, such as health-related behavior, students' attitudes, temperament, and inconsistent criteria. The factor *Norm-referenced assessment* was based on items which indicated an emphasis on comparing students to each other during formative or summative assessment.

Identification of the assessment profiles

The teachers' assessment profiles were identified via CA on the basis of the latent classes, i.e. the six factors found via EFA. Solutions with two to five clusters were constructed and critically investigated. The model with the lowest BIC is considered best: thus, a 3-cluster solution (Figure 1) gave the best fit, since the BIC value started to show a clear increase after cluster 3. The silhouette measure of cohesion and separation showed an average value of 0.3 for the 3-cluster solution, indicating a fair fit with the data.

The importance of the factors for predicting the cluster model varied. The strongest predictor (on a possible value range of 0 to 1) was for *Questionable assessment practices* (1.00), followed by *Assessment of working* (0.90), *Norm-referenced assessment* (0.85), and *Assessment supporting learning* (0.41). The lowest predictor values occurred with *Self and peer assessment as part of grading* (0.14), and *Common assessment criteria* (0.10).

The largest cluster (n = 78, 47.3% of the sample) was named Problematic assessors (Figure 1). Teachers on this profile had above-average scores on each of the six assessment factors. Particularly high scores were found on the factors *Questionable assessment practices* and *Assessment of working*. The other two clusters were referred to as Learning supportive assessors (n = 45, 27.3%), and Norm-based assessors (n = 42, 25.5%). Learning supportive assessors had especially high scores on the *Assessment supporting learning* factor, and particularly low scores on the factors *Questionable assessment practices* and *Norm-referenced assessment*. Teachers with the profile Norm-based assessors had below average scores on every factor except *Norm-referenced assessment*, the score which emerged as highest for all three profiles. Norm-based assessors received particularly low scores on the factors *Assessment supporting learning* and *Assessment of working*.

According to the analysis of variance, the mean values for the three assessment profiles showed statistically significant differences (p = .000) regarding the factors *Assessment supporting learning, Assessment of working*, and *Norm-referenced assessment*. (Table 2). Statistically significant differences were also found between the Problematic assessors and the Norm-based assessors, regarding the factors *Self and peer assessment as part of grading* (p = .002) and *Common* assessment criteria (p = .010). The teachers denoted as Learning supportive assessors and Norm-based assessors differed significantly from the Problematic assessors on the factor *Questionable assessment practices* (p = .000).

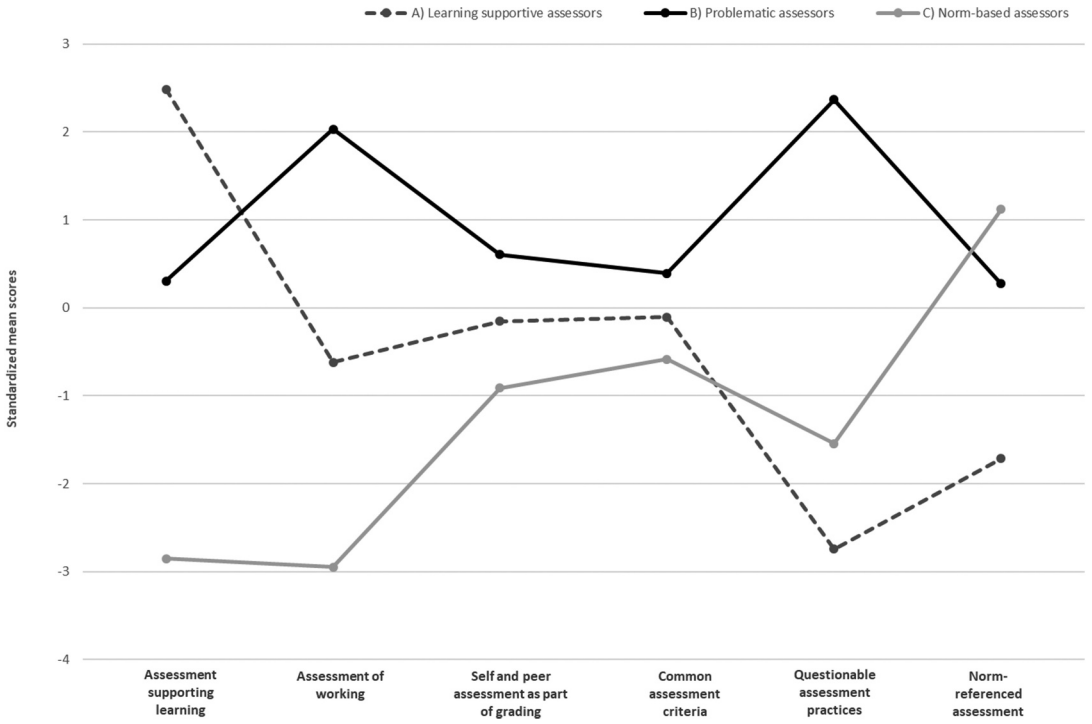

Figure 1. Teacher assessment profiles with standardized mean scores on factors.

Comparison of background variables between assessment profiles

An analysis of variance was conducted to investigate how different background variables were related to the assessment profiles. Significant differences between assessment profiles were found in relation to age, teaching experience, and school level (Table 2). Problematic assessors (M = 5.77, SD = 1.87, CI: 5.35–6.19) were older than Learning supportive assessors (M = 4.81, SD = 1.60, CI: 4.31–5.31, p = .013) or Norm-based assessors (M = 4.96, SD = 1.69, CI: 4.45–5.46, p = .038). Learning supportive assessors (M = 2.55, SD = 1.09, CI: 2.21–2.89) and Problematic assessors (M = 2.68, SD = 0.99, CI: 2.46–2.90) had more teaching experience than Norm-based assessors (M = 2.20, SD = 1.10, CI: 1.87–2.53), but a statistically significant difference was observed only between Problematic assessors and Norm-based assessors (p = .040). From the perspective of the school level, there were more Norm-based assessors (M = 1.49, SD = 0.50, CI: 1.34–1.64) in upper secondary education as compared to Learning supportive assessors (M = 1.24, SD = 0.43, CI: 1.10–1.37, p = .033) or Problematic assessors (M = 1.28, SD = 0.45, CI: 1.18–1.38, p = .047). Learning supportive assessors had on average a higher number of weekly HE lessons, and were slightly more educated than teachers in the other assessment profiles, but these differences were not statistically significant. There were no statistically significant differences between the assessment profiles in terms of gender or school size.

Discussion

The first research question concerned the type of latent construct pertaining to Finnish teachers' assessment conceptions and practices in Health Education, in the context of curricula. A 6-factor solution was found, and the factors were named as follows: *Assessment supporting learning, Assessment of working, Self and peer assessment as part of grading, Common assessment criteria, Questionable assessment practices*, and *Norm-referenced assessment*.

Table 2. Mean values for the teacher assessment profiles and related background variables, and significant profile differences.

	Assessment profiles					
	A	B	C			
	Learning	Problematic assessors supportive assessors	Norm-based assessors	Total mean	*F*(df2) (Tukey HSD)	Pairwise comparisons significant differences
Assessment supporting learning (z)	2.49	0.31	−2.85	0	20.600***	A > B > C
Assessment of working (z)	−0.62	2.03	−2.95	0	52.001***	A > B > C
Self and peer assessment as part of grading (z)	−0.15	0.61	−0.91	0	6.245**	B > C
Common assessment criteria (z)	−0.11	0.39	−0.59	0	4.467**	B > C
Questionable assessment practices (z)	−2.74	2.37	−1.55	0	59.144***	A > B, B > C
Norm-referenced assessment (z)	−1.71	0.28	1.12	0	48.371***	A > B > C
Gender[a]	1.26	1.10	1.22	1.18	2.902	n.s.
Age[b]	4.81	5.77	4.96	5.30	5.302**	A > B, B > C
Teaching experience[c]	2.55	2.68	2.20	2.52	3.037*	B > C
Size of school[d]	4.02	4.03	4.20	4.07	0.427	n.s.
HE studies[e]	2.14	2.04	2.07	2.07	1.166	n.s.
Average amount of weekly HE lessons[f]	2.00	1.83	1.78	1.86	1.754	n.s.
School level[g]	1.24	1.28	1.49	1.33	3.900*	A > C, B > C

Background variables scoring and distributions in total sample

a) 1 = female (82,4%), 2 = male (17,6%)

b) 1 = 25–29 yr (4.2%), 2 = 30–34 yr (15.2%), 3 = 35–39 yr (13.9%), 4 = 40–44 yr (23.0%), 5 = 45–49 yr (17,0%), 6 = 50–54 yr (14.5%), 7 = 55–59 yr (7.3%), 8 = 60–64 yr (4.8%)

c) 1 = under 5 yr (23.6%), 2 = 6–10 yr (20.6%), 3 = 11–20 yr (36.4%), 4 = over 20 yr (19.4%)

d) 1 = under 99s (1.8%), 2 = 100–199 s (7.3%), 3 = 200–299 s (21.8%), 4 = 300–399 s (20.0%), 5 = 400 or over (49,1%)

e) 1 = basic HE studies/25 ECTS (3.0%), 2 = intermediate HE studies/60 ECTS (86.7%), 3 = advanced HE studies/120 ECTS (10.3%)

f) 1 = 1–3 hrs (24.8%), 2 = 4–10 hrs (64.2%), 3 = 11 hrs or over (10.9%)

g) 1 = basic education (67.3%), 2 = upper secondary education (32.7%)

The second research question concerned the type of teachers' assessment profiles exhibited on the basis of their assessment conceptions and practices (see the six latent factors above). Three distinct assessment profiles were recognized, labeled as *Problematic assessors, Learning supportive assessors*, and *Norm-based assessors*. All in all, the teachers' assessment conceptions varied considerably in this study, in line with previous studies (Coombs et al., 2018; Shepard, 2000). The variation was manifested in the form of different assessment profiles for the Health Education teachers under study.

The Problematic assessors had above-average scores in all the measures, but showed particularly high scores on the factors *Questionable assessment practices* and *Assessment of working*. The teachers with this profiled exhibited many questionable issues as an element in their grading, regarding notably students' health behavior and temperament, together with attitudes toward a healthy lifestyle and Health Education as a subject. They also allowed the criteria to vary depending on the student. These practices are not aligned with the national curricula (FNBE, 2014, 2016, 2020a), which clearly emphasize that such factors should not be included within the Health Education assessment. Despite this, the finding was expected. It confirms earlier findings indicating that half of Finnish Health Education teachers at the basic education level do (according to self-reports) include health behavior within their assessment (Summanen, 2014). This questionable type of assessment is clearly in conflict with the curricula, but more importantly, it is problematic from the perspective of equity in student assessment (DeLuca et al., 2016; Pastore & Andrade, 2019; Xu & Brown, 2016).

The Problematic assessors were older, and their conceptions of assessment and related practices may reflect the assessment practices they themselves experienced as school students. If teachers have not received enough theoretical and practical training on assessment during their teacher education, their assessment practices may be influenced by their experiences of assessment during their own student days, and the models used in these times (Guskey & Bailey, 2001; Mertler & Campbell, 2005; Siegel & Wissehr, 2011). Overall, the clear implication for Health Education teacher training is that assessment education should be sufficiently long-term and continuous, if the aim is to persuade teachers to challenge and change their own assessment conceptions and practices (Koh, 2011; Xu & Brown, 2016). The continuous professional development that the teachers had undergone showed a clear decrease in line with the age of the teacher (OECD, 2009). Moreover, previous findings have indicated that early-career teachers emphasize assessment fairness more than is the case among established in-service teachers (Coombs et al., 2018). Interestingly, the Problematic assessors clearly gave more weight to the *Assessment of working* factor than did the teachers within the other assessment profiles. This can be considered a positive sign, since working skills are part of formative assessment and grade formulation, and the curricula emphasize diversity in assessment (FNBE, 2014, 2016).

Learning supportive assessors had particularly high scores on the factors *Assessment supporting learning* and low scores in *Questionable assessment practices* and *Norm-referenced assessment*. Assessment can be a powerful catalyst for learning. It has been shown that the deliberate use of assessment to contribute to learning improves student achievement (Andrade, 2013; Black & Wiliam, 1998a, 1998b; Wiliam, Lee, Harrison, & Black, 2004). Learning supportive assessors had a versatile view on the purposes of assessment. More often than the assessors in other profiles they viewed assessment as helping students to learn, and as helping the teacher to identify students' needs for learning support; furthermore they saw assessment as facilitating feedback, and as supporting the planning of teaching (formative assessment/assessment for learning). In addition, other purposes of assessment were recognized by those teachers: they saw assessment as helping to define the students' starting level (diagnostic assessment) and competence (summative assessment/assessment of learning). Teachers of this profile could be said to have achieved many of the ideas raised by researchers; they had embraced learning-focused models of assessment and were able to use different forms of assessment (formative, summative) in a balanced way. Furthermore, they were willing to make pedagogical decisions based on the information they received from the assessment. Overall, the assessment had become a natural and integral part of their way of working (Hildén & Fröjdendahl, 2018; Xu & Brown, 2016).

Norm-based assessors scored below average in almost every measure, with particularly low scores on *Assessment supporting learning* and *Assessment of working*. They also had the highest scores on *Norm-referenced assessment*. Norm-referenced assessment means that the teacher compares the students' performances with the achievement of other students, or changes the scoring to get what she/he regards as an appropriate grade distribution (or appropriate differences) between students (Hailikari, Postareff, Tuononen, Räisänen, & Lindblom-Ylänne, 2014). In fact, the curricula emphasize that assessment should be based on criteria derived from learning objectives (FNBE, 2014, 2016, 2020a), but as earlier studies have shown, the phenomenon of norm-referenced assessment does exist to some degree in Finnish basic and upper secondary education (Atjonen et al., 2019). In criterion-referenced assessment the learning of the students will be assessed against pre-specified criteria (Brown, 1988), with a consequent increase in the transparency and openness of the assessment. Criterion-based assessment also indicates to students the nature of the competence (quality, extent) expected of them, and the aim of the assessment. The students' understanding of assessment criteria guides their approaches to studying, and influences their learning (Entwistle & Entwistle, 1991; Struyven, Dochy, & Janssens, 2005). If students perceive that they can qualify in school assessments via memorization and recall of information, they are more likely to adopt a superficial approach to learning. A deep approach to learning has been found to be associated with the learner's awareness of the learning objectives and of the assessment criteria used (Prosser & Trigwell, 1999). Overall, Norm-

based assessors do not have the possibility to take advantage of all the opportunities to advance students' learning. This, combined with the fact the teachers in the study were less interested in the assessment of learning and working, is a disquieting finding.

In all three profiles the scores were closest in the factors *Self and peer assessment as part of grading* and *Common assessment criteria*. Nevertheless, there were statistically significant differences between the Problematic assessors (scores above average) and Norm-based assessors (scores below average). According to the core curricula, students should be given opportunities for self- and peer-assessment (FNBE, 2014, 2016). With criteria-referenced self-assessment students can obtain information about their own competence level. This information can lead students to reflect on the reasons for their own learning, with better possibilities for a deep learning approach, and enhancement of their ability to self-regulate the learning process (Ozogul & Sullivan, 2007; Prosser & Trigwell, 1999).

Peer assessment, for its part, has a number of favorable characteristics. When peer assessment is used, the student becomes immersed in the output of others. In that case, the student is led to take another look at the subject to be learned, adopting a broader or new perspective, with a more complete grasp of both the entity and the relationships between concepts (Vu & Dall'Alba, 2007). Peer assessment also encourages the active participation of students, which itself can support learning (Gouli, Gogoulou, & Grigoriadou, 2008). Overall, the use of self and peer assessment is highly desirable from the perspective of learning, though not as a component in grade formulation (FNBE, 2014, 2016), due to the possibility of unreliability. The Problematic assessors had the highest score in the factor *Self and peer assessment as part of grading*.

The Problematic assessors had the highest score and the Norm-based assessors the lowest score on the factor *Common assessment criteria*. To ensure equality in assessment, national assessment criteria for numerical grades 5, 7, 8 and 9 have been defined within the basic education curriculum (FNBE, 2014, 2020a, 2020b). These criteria for summative assessment can also be used to build up formative assessments. It is clear that each teacher should use these national criteria. Nevertheless, as shown by this study and also by previous research on all teacher groups (Atjonen et al., 2019) and on Health Education teachers (Summanen, 2014), teachers differ in the use of the criteria available. With regard to Health Education, the problem is that if teachers do not use the relevant subject-specific assessment criteria, students with the same level of competence will receive different grades, involving a real threat to assessment fairness and students' equality in Health Education. This is a particular concern at the final assessment stage at the end of basic education, since these are the grades that students use in applying for secondary education. The issue will have even wider significance in parallel with any increase in the number of problematic assessors. Here it should be noted that the equality problem exists in a range of subjects, and not only in Health Education (Hilden et al., 2016); hence, one can expect increasing debate on the need for national standardized tests. The hope is that if teacher training includes high-quality assessment education, and if slightly revised assessment principles (FNBE, 2014, 2020a) and criteria (FNBE, 2014, 2020b) are included in the core curriculum, it will be possible to overcome this problem, and to maintain the Finnish assessment system on the basis of teachers' responsibility and assessment literacy, without the need for strong external control (Sahlberg, 2015; Vainikainen et al., 2017).

It is interesting to note that although there have been some difficulties with the use of the criteria in basic education, upper secondary education has, in fact, manifested greater problems. As the results showed, there were more Norm-based assessors in upper secondary education than in basic education. This was mainly due to the fact that no assessment criteria had been defined in the upper secondary curriculum (FNBE, 2016), meaning that the teachers themselves had been obliged to resolve the situation in their own way.

The present study has some limitations which should be taken into account when interpreting the findings. The sample could have contained more male teachers, in which case credible differences might have emerged between male and female teachers. Voluntary online data collection can cause some bias, in that teachers who are interested in assessment may participate in the study more readily

than other teachers. On the other hand, the design and purpose of questionnaire were both neutral, and any possible positive bias was reduced by allowing respondents to answer anonymously. In the future, it would be important to collect data in a variety of ways (involving e.g. self-reported questionnaires, observations, interviews, student data), and to construct teachers' assessment profiles by comparing findings from multidimensional data. In addition, it would be interesting to examine the associations between teachers' assessment profiles and students' learning outcomes; also to investigate the factors that might influence teachers toward appropriate assessment practices, and encourage them to switch from one profile to another.

Conclusion

Almost two decades ago Health Education was introduced as an independent school subject in Finland. However, there has been no previous research on Health Education teachers' understandings of how best to assess the subject. This study is the first to investigate Health Education teachers' assessment profiles on the basis of their conceptions; thus, it gives important information on the differences between teachers in their assessment orientations. A theoretical understanding of teachers' assessment conceptions and profile differences is important if one is to develop teachers' assessment competencies within pre- and in-service teacher training.

The present study, outlining three different assessment profiles, has described in some detail the differences between individual Health Education teachers as curriculum implementers. It underlines the need to better understand teachers as individuals, bearing in mind how much variation exists in teachers' previous conceptions and experiences of assessment, and how (by modifying previous conceptions) it may be possible to make teacher assessment training more effective in practice (Coombs et al., 2018; Xu & Brown, 2016). As Brown (2004) has noted, for effective teacher training it is essential to explore teachers' conceptions along with assessment-related knowledge and skills. The complex structure of teachers' assessment conceptions has to be taken into account in efforts to implement the assessment policies built into the national curricula. If teachers are unaware of their own assessment-related conceptions, they cannot challenge them, or make any necessary changes to themselves and to their assessment profile.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Ministry of Education and Culture, Finland [#OKM/63/592/2018]. The funding source had no involvement in the study design, the collection, analysis, or interpretation of the data, the writing of the report, or the decision to submit the article for publication.

ORCID

Olli Paakkari http://orcid.org/0000-0001-9798-5042
Leena Paakkari http://orcid.org/0000-0002-4130-9202
Henna Haapala http://orcid.org/0000-0001-7989-3910
Mirja Hirvensalo http://orcid.org/0000-0003-4841-2250

References

Abell, S. K., & Siegel, M. A. (2011). Assessment literacy: What science teachers need to know and be able to do. In D. Corrigan, J. Dillon, & R. Gunstone (Eds.), *The professional knowledge base of science teaching* (pp. 205–221). Dordrecht, Netherlands: Springer.

Andrade, H. L. (2013). Classroom assessment in the context of learning theory and research. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 17–34). Los Angeles, California: SAGE.

Atjonen, P., Laivamaa, H., Levonen, A., Orell, S., Saari, M., Sulonen, K., , and Immonen, J. (2019). *"Että tietää missä on menossa" Oppimisen ja osaamisen arviointi perusopetuksessa ja lukiokoulutuksessa. ["So that we know where we stand" Assessment of learning and competence in basic education and general upper secondary education]*. Kansallinen koulutuksen arviointikeskus. Tampere, Suomi: PunaMusta.

Barnes, N., Fives, H., & Dacey, C. M. (2015). Teachers' beliefs about assessment. In H. Fives & M. G. Gill (Eds.), *International handbook of research on teacher beliefs* (pp. 284–300). New York, New York: Routledge.

Basic education Act 628/1998. Accessed 9.4.2022. Retrieved from https://www.finlex.fi/fi/laki/kaannokset/1998/en19980628.pdf

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education*, *5*(1), 7–74.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *80*(2), 1–13.

Black, P., & Wiliam, D. (2012). Assessment for learning in the classrooms. In J. Gardner (Ed.), *Assessment and learning* (pp. 11–32). London, United Kingdom: SAGE Publications.

Brown, S. (1988) Criterion referenced assessment: what role for research? In Black, H. D., & Dockerell, W. D. (Eds.). *New developments in educational assessment*. British Journal of Educational Psychology, Monograph series no. 3 (pp. 1-14). Edinburgh, Scotland: Scottish Academic Press.

Brown, G. T. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, *11*(3), 301–318.

Cerny, C. A., & Kaiser, H. F. (1977). A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate Behavioral Research*, *12*(1), 43–47. doi:10.1207/s15327906mbr1201_3

Coombs, A., DeLuca, C., LaPointe-McEwan, D., & Chalas, A. (2018). Changing approaches to classroom assessment: An empirical study across teacher career stages. *Teaching and Teacher Education*, *71*(1), 134–144. doi:10.1016/j.tate.2017.12.010

DeLuca, C., Coombs, A., & LaPointe-McEwan, D. (2019). Assessment mindset: Exploring the relationship between teacher mindset and approaches to classroom assessment. *Studies in Educational Evaluation*, *61*(1), 159–169. doi:10.1016/j.stueduc.2019.03.012

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability*, *28*(3), 251–272. doi:10.1007/s11092-015-9233-6

Entwistle, N. J., & Entwistle, A. (1991). Contrasting forms of understanding for degree examinations: The student experience and its implications. *Higher Education*, *22*(3), 205–227. doi:10.1007/BF00132288

Field, A. (2009). *Discovering statistics using SPSS* (Third ed.). London, United Kingdom: SAGE Publication.

FNBE. (2014). *National core curriculum for basic education 2014*. Helsinki, Finland: Finnish National Board of Education.

FNBE. (2016). *National core curriculum for general upper secondary schools 2015*. Helsinki, Finland: Finnish National Board of Education.

FNBE. (2020a). *Oppilaan oppimisen ja osaamisen arviointi perusopetuksessa. Perusopetuksen opetussuunnitelman perusteiden 2014 muutokset. [Assessment of student learning and competence in basic education. Changes in the National Core Curriculum for Basic Education 2014]*. Helsinki, Suomi: Opetushallitus.

FNBE. (2020b). *Perusopetuksen päättöarvioinnin kriteerit. [Criteria for the final assessment of basic education]*. Helsinki, Suomi: Opetushallitus.

Fulmer, G. W., Lee, I. C. H., & Tan, K. H. K. (2015). Multi-level model of contextual factors and teachers' assessment practices: An integrative review of research. *Assessment in Education: Principles, Policy, and Practice*, *22*(4), 475–494.

General Upper Secondary Education Act 629/1998. Accessed 9.4.2022. Retrieved from https://www.finlex.fi/fi/laki/ajantasa/2018/20180714

Gouli, E., Gogoulou, A., & Grigoriadou, M. (2008). Supporting self-, peer-, and collaborative-assessment in e-Learning: The case of the PEer and Collaborative ASSessment Environment (PECASSE). *Journal of Interactive Learning Research*, *19*(4), 615–647.

Guskey, T. R., & Bailey, J. M. (2001). *Developing grading and reporting systems for student learning*. Thousand Oaks, California: Corwin Press.

Hailikari, T., Postareff, L., Tuononen, T., Räisänen, M., & Lindblom-Ylänne, S. (2014) Students' and teacher's perceptions of fairness in assessment. In Hounsell, C., Kreber, C., Anderson, C., Entwistle, N., McArthur, J. (Eds.), *Advances and innovations in university assessment and feedback : A Festschrift in honour of Professor Dai Hounsell* (pp. 99-113). Edinburgh, Scotland: Edinburgh University Press.

Hildén, R., & Fröjdendahl, B. (2018). The Dawn of assessment literacy – Exploring the conceptions of Finnish student teachers in foreign languages. *Apples – Journal of Applied Language Studies*, *12*(1), 1–24. doi:10.17011/apples/urn.201802201542

Hilden, R., Ouakrim-Soivio, N., & Rautopuro, J. (2016). Kaikille ansionsa mukaan? Perusopetuksen päättöarvioinnin yhdenvertaisuus Suomessa. [Fair marks for all? Equal and equitable grading in the end of basic education in Finland]. *Kasvatus*, *47*(4), 342–357.

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. doi:10.1080/10705519909540118

IBM. (2011). Cluster analysis. In *IBM SPSS statistics 19 guide* (pp. 375–404). New York, New York: Author.

IBM. (2016). Cluster evaluation algorithms. In *IBM SPSS modeler 18.0 algorithms guide* (pp. 87–91). New York, New York: Author.

Ketchen, D. J., Jr., & Shook, C. L. (1996). The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal*, *17*(6), 441–458. doi:10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G

Koh, K. H. (2011). Improving teachers' assessment literacy through professional development. *Teaching Education*, *22*(3), 255–276. doi:10.1080/10476210.2011.593164

Marton, F., & Booth, S. (1997). *Learning and awareness*. Mahwah, NJ: Erlbaum.

Mertler, C. A., & Campbell, C. (2005). *Measuring teachers' knowledge & application of classroom assessment concepts: Development of the assessment literacy inventory*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada. Accessed 9.4.2022. Retrieved from https://files.eric.ed.gov/fulltext/ED490355.pdf

OECD. (2009). *Creating effective teaching and learning environments. First results from TALIS*. Paris, France: Author.

Ozogul, G., & Sullivan, H. (2007). Student performance and attitudes under formative evaluation by teacher, self- and peer-evaluators. *Education Technology Research and Development*, *57*(3), 393–410. doi:10.1007/s11423-007-9052-7

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

Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education*, *84*(1), 128–138. doi:10.1016/j.tate.2019.05.003

Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, *48*(1), 4–11. doi:10.1080/00405840802577536

Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, *46*(4), 265–273. doi:10.1080/08878730.2011.605048

Prosser, M., & Trigwell, K. (1999). *Understanding learning and teaching: The experience in higher education*. Buckingham: SRHealth Education, Open University Press.

Sahlberg, P. (2015). *Finnish lessons 2.0: What can the world learn from educational change in Finland*. New York, New York: Teachers College Press.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*(2), 461–464. doi:10.1214/aos/1176344136

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, *29*(7), 4–14. doi:10.3102/0013189X029007004

Siegel, M. A., & Wissehr, C. (2011). Preparing for the plunge: Preservice teachers' assessment literacy. *Journal of Science Teacher Education*, *22*(4), 371–391. doi:10.1007/s10972-011-9231-6

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, *77*(3), 238.

Stiggins, R. J. (2010). Essential formative assessment competencies for teachers and school leaders. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 233–250). New York, New York: Taylor & Francis.

Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education*, *30*(4), 325–341. doi:10.1080/02602930500099102

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. Tampere, Suomi: Juvenes Print.

Vainikainen, M.-P., Thuneberg, H., Marjanen, J., Hautamäki, J., Kupiainen, S., & Hotulainen, R. (2017). How do Finns know? Educational monitoring without inspection and standard setting. In S. Blömeke & J.-E. Gustafsson (Eds.), *Standard setting in education. The Nordic countries in an international perspective* (pp. 243–259). Cham, Switzerland: Springer International Publishing.

Vu, T. T., & Dall'Alba, G. (2007). Students' experience of peer assessment in a professional course. *Assessment & Evaluation in Higher Education*, *32*(5), 541–556. doi:10.1080/02602930601116896

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles Policy and Practice*, *11*(1), 49–65.

Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*(1), 149–162. doi:10.1016/j.tate.2016.05.010