

**This is an electronic reprint of the original article.
This reprint *may differ* from the original in pagination and typographic detail.**

Author(s): Tang, Xin; Pakarinen, Eija; Lerkkanen, Marja-Kristiina; Kikas, Eve; Muotka, Joonas;
Nurmi, Jari-Erik

Title: Validating the early childhood classroom observation measure in first and third grade classrooms

Year: 2017

Version:

Please cite the original version:

Tang, X., Pakarinen, E., Lerkkanen, M.-K., Kikas, E., Muotka, J., & Nurmi, J.-E. (2017).
Validating the early childhood classroom observation measure in first and third grade
classrooms. *Scandinavian Journal of Educational Research*, 61(3), 275-294.
<https://doi.org/10.1080/00313831.2015.1120237>

All material supplied via JYX is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

Running Head: VALIDATING ECCOM IN PRIMARY SCHOOL

Validating the Early Childhood Classroom Observation Measure

in first and third grade classrooms

Xin Tang^a, Eija Pakarinen^a, Marja-Kristiina Lerkkanen^a, Eve Kikas^b, Joonna Muotka^a, & Jari-

Erik Nurmi^a

^a University of Jyväskylä and ^b Tallinn University

Author's Note

Xin Tang, Department of Psychology, University of Jyväskylä, Finland;

tangxin09@gmail.com

Eija Pakarinen, Department of Teacher Education, University of Jyväskylä, Finland;

eija.k.pakarinen@jyu.fi

Marja-Kristiina Lerkkanen, Department of Teacher Education, University of Jyväskylä,

Finland; marja-kristiina.lerkkanen@jyu.fi

Eve Kikas, Department of Psychology, Tallinn University, Estonia; eve.kikas@tlu.ee

Joonna Muotka, Department of Psychology, University of Jyväskylä, Finland;

joona.muotka@jyu.fi

Jari-Erik Nurmi, Department of Psychology, University of Jyväskylä, Finland;

jari-erik.j.nurmi@jyu.fi

Correspondence concerning this article should be addressed to Xin Tang, Department of Psychology, University of Jyväskylä, Finland. Contact: tangxin09@gmail.com

Abstract

The present study reports on the psychometric properties of the Early Childhood Classroom Observation Measure (ECCOM) in Finnish and Estonian first and third grade classrooms. The observation data were collected from 91 first grade teachers (32 from Finland, 59 from Estonia); and 70 third grade teachers (33 from Finland, 37 from Estonia). Teachers also self-rated variables, such as curriculum goals, their teaching experience and the classroom size. The results of confirmatory factor analysis (CFA) provided evidence which supported use of the three-factor model (Management, Climate, and Instruction) for each dimension that is, child-centred, teacher-directed and child-dominated, in both grades. The reliability of the dimensions and sub-scales was good, and some evidence was also found for criterion validity. The findings imply that ECCOM is thus a reliable and valid instrument for measuring teaching practices in primary school.

Keywords: teaching practice, child-centred classroom, teacher-directed classroom, child-dominated classroom, primary education

Validating the Early Childhood Classroom Observation Measure
in First and Third Grade Classrooms

An increasing amount of empirical research has indicated that teaching practices in preschool, kindergarten and the early years of school can play a significant role in student achievement later on (e.g., Hamre & Pianta, 2001; Pressley et al., 2003). However, much of what we know about classroom practices in schools comes from surveys and one-on-one standardised assessments (e.g., Flowers & Hancock, 2003; Ireson, Blatchford, & Joscelyne, 1995; Kiuru et al., 2012). The problem with those measures is that they lean heavily on self-reports from teachers and/or students, which means that they have the tendency to be biased. Consequently, it has been suggested that there is a need to find more concrete indicators of classroom practices through more effective classroom observation (e.g., Hamre & Pianta, 2005; La Paro, Pianta, & Stuhlman, 2004). What needed are observational methods that take into account the instructional and organisational choices that teachers make to create the classroom climate (e.g., Pianta, La Paro, & Hamre, 2008; Stipek & Byler, 2004). One such observational method is the Early Childhood Classroom Observation Measure (ECCOM; Stipek & Byler, 2005). However, previous studies on ECCOM have some limitations. First, it has so far been almost exclusively used in preschool and kindergarten (e.g., Lerkkanen et al., 2012b; Rao, Sun, Zhou, & Zhang, 2012; Yen & Ispa, 2000). Although there are a few studies (Kikas, Peets, & Hodegs, 2014; Lerkkanen, Kikas, Pakarinen, Poikonen, & Nurmi, 2013; Perry, Donohue, & Weinstein, 2007) which have used it in the first grade classrooms, there are no studies which have investigated the reliability and validity of ECCOM later on in primary school grades. Second, the age that children start kindergarten and primary school vary from one country to another. For example, first-graders in Finland and Estonia are somewhat older than first-graders in U.S., where the instrument was developed. Third, to develop observational instruments information is always needed on the applicability of the

instrument in other cultural and educational contexts. Consequently, the present study looks at the applicability of ECCOM measure in 1st and 3rd grade classrooms by using samples from Finland and Estonia.

Observational Studies in Classrooms

Classroom observation has been used as a measurement tool in educational research for over three decades now (Gage & Needels, 1989). In the 1970s and 1980s, many studies made an effort to examine the various typical behaviours of teachers and students in classrooms, and how they relate to students' cognitive development (Brophy & Good, 1986; Gage & Needels, 1989). In their landmark review of earlier studies that were done in this field, Brophy and Good (1986) placed teaching behaviours into five different categories: (i) organisational or management-related; (ii) giving information; (iii) questioning; (iv) reacting to how students respond; and (v) handling classwork and homework. Interestingly, none of the studies seemed to address one important set of classroom behaviours, namely the kind concerned with social-emotional support (Midgley, Feldlaufer, & Eccles, 1989; Voelkl, 1995).

More recently, standardised classroom observational measurements have been developed (e.g., Pianta & Hamre, 2009). For example, the Early Childhood Environment Rating Scale (ECERS and ECERS-R; Harms, Clifford, & Cryer, 1998), which focus primarily on the structural quality of early childhood programs and assess the quality and quantity, e.g. the space used for learning activities, number of books, and mathematical materials, in the classroom. However, the ECERS and ECERS-R are often regarded as measuring many distant factors for teacher-child interactions, e.g. furnishings, room arrangement, gross motor equipment, and do not focus too much on teacher-child interactions or the quality of teaching practices. The Classroom Assessment Scoring System (CLASS K-3; Pianta, La Paro, & Hamre, 2008) is one observational tool used widely in early childhood and

early primary education. It examines social-emotional, organizational, and instructional teacher-child interactions in a variety of classroom settings. The CLASS, which assesses teachers' behaviours through twenty-minute observational cycles, provides information about teacher's discrete scores on several dimensions of emotional, organizational and instructional supports. However, teachers' scores according to cycles may vary quite a lot and the more global view of teachers' behaviors and practices is missing.

The Early Childhood Classroom Observation Measure (ECCOM), developed by Stipek and Byler (2004, 2005), is a unique observational tool. Its major strength is that it does not only provide us information about the global nature of teaching practices in terms of three different teaching approaches, namely, teacher-directed, child-centred, and child-dominated teaching practices, but it also enhances our understanding of teaching practices by providing three sub-scales - management, climate, and instruction for all the three dimensions. Another advantage of ECCOM is that no other observational measures had addressed the child-dominated practices, which had its roots in the philosophy of naturalism (Kikas et al., 2014; Valeski & Stipek, 2001). The ECCOM was created to assess the extent to which the child-centred, teacher-directed, and child-dominated dimensions of instruction, social climate and management are implemented in early year education.

A teacher who employs child-centred practices perceives children as active learners who construct knowledge based on their prior experiences and understanding. They will generally follow the principles of constructivist theories of teaching (Bransford, Brown, & Rodney, 2000) and value the children's interests highly when organising activities. Teacher-directed practices, meanwhile, have their basis in a more traditional view of education, with an emphasis on concrete didactics or rote learning (e.g., environmentalist/behaviourist; see Ormrod, 1999). High scores in the teacher-directed dimension therefore suggest that the teacher makes most of the decisions, controls instructional activities, does not encourage

communication between students, and emphasises the importance of facts and basic skills. Finally, child-dominated practices are ones in which the children are given almost complete autonomy regarding their learning in a classroom. This approach has its root in maturationist theory (Gesell et al., 1959) and naturalism (Kikas et al., 2014; Valeski & Stipek, 2001). The classroom rules are unclear and there are no systematically designed tasks. However, teachers may interrupt and control activities when children's behaviour is out of control.

Previous research has shown that child-centred practices have a positive impact on the teacher-child relationship (Stipek & Byler, 2004), on children's maths skills and self-performance competence (Perry et al., 2007), and on their interest in reading and mathematics (Lerkkanen et al., 2012b). Meanwhile, teacher-directed practices have been shown to have a beneficial effect on children's development (cognitive, in particular). For example, Stipek and her colleagues (1995) showed that children in a didactic program may in fact score much higher in reading achievement tests than those enrolled in a child-centred program. Pressley and his colleagues (2003), however, have argued that instruction which involves both direct transmission and constructivist behaviours may produce the best results for overall child development. There are also some studies that have revealed the negative effects of child-dominated practices on children's academic (Walker, 2008) and social development (Valeski & Stipek, 2001); while others (e.g., Kikas et al., 2014) found that child-dominated practices were actually positively related with spelling skills and task-persistent learning behaviour in classrooms with higher initial maths skills.

In addition to three dimensions of teaching practices, ECCOM also provides three subscales (i.e., management, climate, instruction; for detailed description see Table 1) for each dimension. Management includes items concerning 'child responsibility', the 'choice of activities' and 'discipline strategies'. Climate consists of items dealing with 'support for communication skills', 'support for interpersonal skills', 'student engagement', and the

‘individualisation of learning activities’. Finally, instruction relates to ‘learning standards’, the ‘coherence of instructional activities’, ‘teaching concepts’, ‘instructional conversation’, and ‘literacy and maths instruction’, in particular. Thus, ECCOM provides in-depth understanding of the teaching process by examining not only child-centred, teacher-directed, and child-dominated teaching practices but also the three components of management, climate, and instruction within each teaching practice.

INSERT TABLE 1 ABOUT HERE

The Relationship Between Teaching Practices and Other Teacher-Related Factors

Teachers’ self-reported beliefs as to which aspects they emphasize when instructing the child, and which curriculum goals they focus on have been shown to be related to their actual practices (e.g., Pakarinen et al., 2015; Stipek & Byler, 2004). For example, Stipek and Byler (1997) found that preschool and kindergarten teachers who held child-centred beliefs tended to emphasize basic skills less in the classroom and focus more on a positive climate, which in turn negatively correlated with their curriculum goals in basic skills, knowledge and facts; whereas teachers who used a more didactic approach emphasized basic skills, knowledge, and facts. Similar correlational pattern was found in a later study by Stipek and Byler (2004), which showed that teachers who used chiefly constructivist practices emphasised higher order thinking skills over basic skills, while those using more didactic practices focused the curriculum more on basic skills and less on higher order thinking skills and social development.

In addition, classroom and teacher characteristics have been found to relate to teaching practices, although the findings are somewhat inconsistent. Classroom size was found to be negatively associated with classroom quality in some studies (e.g., Allhusen et al., 2004), whereas others have not found this to be the case (e.g., NICHD ECCRN, 2002). In the same way, some studies have found that a teachers’ level of education and experience were

related to the quality of their early childhood classroom (Gerber, Whitebook, & Weinstein, 2007) and the level of emotional support given to children by them in primary school (NICHD ECCRN, 2002, 2005), while other studies have failed to confirm this finding (Early et al., 2007; Pianta et al., 2005).

Following with previous validation studies (e.g., Lerkkanen et al., 2012a), the present study used teachers' self-reported curriculum goals and teaching experience and the characteristics of their classroom (class size) as criterion variables to validate the ECCOM results.

Primary Education in Finland and Estonia

The present study was conducted in two countries, Finland and Estonia. Recently, both countries are regarded as the leading countries in education due to their top performance in the Programme for International Student Assessment (PISA; OECD, 2012). We chose these neighbouring countries because the structure of the educational system is quite similar; in both countries compulsory formal education consists of nine years of comprehensive school beginning the year children turn seven years of age, with the same classroom teacher for the main subjects during the first three school years. Moreover, the syllabic complexity and orthographic depth of these two highly transparent languages, Finnish and Estonian, have a strong effect on the word reading skill development in early school years: decoding is much easier to learn in shallow orthographies compared to more complex orthographies, such as English, which might take more than twice as long to learn to decode accurately (Dasinger, 1997; Seymour, Aro, & Erskine, 2003). Finally, the academic demands for the first three years are quite similar in both national core curriculum. Each curriculum outlines the objectives and addresses the concept of learning, different teaching approaches, and what makes a good learning environment.

Aims of the Present Study

Although some previous studies (Lerikkanen et al., 2012a) have examined the reliability and validity of ECCOM in kindergarten, no validation research has as yet been undertaken among primary school classrooms. Thus, the present article examines the results of using ECCOM in first and third grade classrooms in Finland and Estonia. Firstly, we look at the inter-rater reliabilities, internal consistency, and item reliability of the ECCOM dimensions and sub-scales. Secondly, the validity of ECCOM's constructs was analysed to see if the three factors suggested (i.e., management, climate, and instruction) actually fit the data. Finally, the concurrent validity of ECCOM was assessed by examining the associations between the ECCOM dimensions and teachers' self-reported curriculum goals, their work experience, and the size of their class.

Method

Participants

The sample consisted of 91 first grade teachers (32 in Finland and 59 in Estonia) and 70 third grade teachers (33 in Finland and 37 in Estonia). Both sets of data were collected as part of other ongoing longitudinal studies; these being the 'First Steps Study' in Finland (Lerikkanen et al., 2006), and the 'Kindergarten-School Study' and one other longitudinal study in Estonia (Kikas, 2007). Teachers in both countries were asked beforehand for their written consent to participate.

The Finnish sample. Thirty-two first grade teachers (26 female, 4 male, 2 unknown), and 33 third grade teachers (23 female, 8 male, 2 unknown) were observed in Spring semester, in 2008 and 2010, respectively. Seven teachers and most of the children were the same in both years. All of the first and third grade teachers had at least a bachelor's degree, while 86% of the first grade teachers and 97% of the third grade teachers had a master's degree or above. The teachers rated their work experience from 1 to 5 (1 = *less than a year*, 2

= 1-5 years, 3 = 6-10 years, 4 = 11-15 years, 5 = more than 15 years). On average teachers work experience was between 11 to 15 years in both grades (Median = 4). The schools were located in two medium-size towns and one less urban municipality in Finland. The average size of the class in first grade was 16.28 ($SD = 4.77$), and 19.94 in the third ($SD = 5.88$).

The Estonian sample. The grade 1 sample in Estonia comprised of 38 teachers (all female) from the Kindergarten-School Study and 21 teachers (all female) from another two-year longitudinal project. The first part of this sample was observed in 2008 and the second in 2012. Out of all 59 first grade teachers, 51 of them were classroom teachers and 8 were non-classroom (e.g., special education and language teachers). The Estonian teachers rated their work experience in similar manner as the Finnish teachers. On average, the work experience of the first grade teachers was more than 15 years (the median was 5). All of them had a master's level education. Meanwhile, 34 of the 37 third grade teachers observed were the same teachers as observed in the Grade 1 sample, with only three being different. The median work-experience of the third grade teachers was 5 (more than 15 years). Thirty-one of them were classroom teachers, and 6 were non-classroom. The average class size in first grade was 19.72 ($SD = 4.90$), and 16.67 in the third ($SD = 4.59$).

Measures

Classroom observations. The Early Childhood Classroom Observation Measure (ECCOM; Stipek & Byler, 2004; 2005) was used in this study. It measures classroom practice in terms of three dimensions, in other words, the degree to which they are child-centred, teacher-directed, and child-dominated practices. Each dimension is rated on 14 items and over the same three sub-scales: *management* (four items), *climate* (four items), and *instruction* (six items), as seen in table 1. The rating scale is based on the percentage of time that each particular kind of practice is demonstrated during the observation (1 = *practice is rarely seen* (0%–20% of the time) to 5 = *practice predominates* (80%–100% of the time)).

The observers independently rated the teaching practices using the same three sub-scales for each of the three teaching practice dimensions. For example, for one specific item (e.g., choice of activities), an observed practice might be rated as 4 on child-centred, 3 on teacher-directed, and 2 on child-dominated dimensions at the end of the observation day (Stipek & Byler, 2005).

Training. In the present study, the 12 observers at Grade 1 and 17 observers at Grade 3 in Finland and 6 observers in Estonia were carefully trained before each data collection. They were all research assistants with a master's degree in education or psychology, and they each received two additional training sessions and practice in the field beforehand. Before the first training session the observers were asked to read the scoring manual carefully; and at the second three-hour training session two 30-minute video tapes of classroom lessons were watched and coded independently by each observer. The ratings were then compared and discrepancies discussed among observers. Next, the observers went to code teaching practice in pairs for three-hour sessions in classrooms that were not part of the study. After the coding, the observers' interrater reliability was analysed (using intra-class correlations). Ratings that were within one point of each other were considered to reflect an acceptable degree of accuracy, and on only two occasions in Finland and on one occasion in Estonia, were the discrepancies between codings, greater than one point. For these observers, extra coding practice in classrooms was advised, and a meeting arranged to monitor their subsequent interrater agreement. Once the intra-class correlation between the two observers reached .81 or above, were the trained observers then allowed to conduct observations as part of the study.

Procedure. Observations were conducted in February and March in both countries, i.e., during the spring term of the first and third grades. Each observation session lasted half a day (about three lessons) and began at the start of the school day at either 8 or 9 am. After a session that lasted two or three hours, two observers independently rated the classroom, and

these two scores were averaged for further factor analyses. Questionnaires were then sent to the Finnish teachers after the observations in March, and to the Estonian teachers at the beginning of the next school year (that October).

Curriculum goals. The Finnish teachers were asked to rate 22 questions on a 5-point scale (1= *not important at all*, 5 = *very important*) about their first grade curriculum goals for reading and writing, and maths. The questions were based on those used in the study by Stipek and Byler (2004), and the descriptions were formulated according to descriptions in the National Core Curriculum for Basic Education in Finland. The final scores for the first grade were therefore organised and distributed accordingly within the following categories: ‘basic reading and writing skills’ (four items, e.g., reading fluency, and motor fluency of writing, $\alpha = .81$); ‘advanced reading and writing skills’ (three items, e.g., creative writing, $\alpha = .76$); ‘basic maths skills’ (five items, e.g., understanding numerical concepts and numbers, $\alpha = .62$); ‘problem solving skills’ (three items, e.g., solving verbal assignments, $\alpha = .85$); ‘motivation and self-esteem’ (four items, e.g., positive learner self-image as a reader and writer/ related to mathematics, $\alpha = .86$); and ‘working skills’ (two items, e.g., concentration, and endurance, $\alpha = .74$). For the third grade, the distribution was ‘basic reading and writing skills’ (five items, $\alpha = .81$); ‘advanced reading and writing skills’ (seven items, $\alpha = .87$); ‘basic maths skills’ (four items, $\alpha = .56$); ‘problem solving skills’ (four items, $\alpha = .88$); ‘motivation and self-esteem’ (four items, $\alpha = .73$); and ‘working skills’ (two items, $\alpha = .69$).

Background variables. In both countries, teachers were asked to choose one of five options to quantify their work experience (1 = *less than a year*, 2 = *1-5 years*, 3 = *6-10 years*, 4 = *11-15 years*, 5 = *more than 15 years*). The number of students in the classroom during each observation session was also recorded.

Analysis Strategy

This study examined the reliability and validity of ECCOM in the specific contexts of these Finnish and Estonian schools. To do this, descriptive statistics for the ECCOM dimensions, their sub-scales, and the items therein, were calculated separately for each country's samples. Then, to establish inter-rater agreement, the intra-class correlation between each pair of observers was then calculated by combining the Finnish and Estonian data to obtain cross-coder reliabilities (McGraw & Wong, 1996). Finally, to analyse the construct validity we carried out a confirmatory factor analysis for the combined Finnish and Estonian data of child-centred, teacher-directed, and child-dominated dimensions according to the suggested three factor model. In other words, each model assumed three positively correlated latent constructs (management, climate, and instruction). This theoretical model is presented in figure 1.

Next, we examined the reliabilities of the ECCOM dimensions and subscales. The internal consistency was investigated by estimating factor-score scale reliabilities and Cronbach's alphas for the unweighted summary scores. Meanwhile, item reliability was measured by estimating the squared correlation between the item and factor (Bollen, 1989).

Finally, concurrent validity was investigated by examining whether teachers' self-reported curriculum goals, their teaching experience and the size of the class were associated with the ECCOM dimensions. This investigation (in both first and third grades) was done separately for the Finnish and Estonian teachers, as the questionnaires were slightly different in the two countries.

Statistical analyses were carried out using Mplus (Version 7; Muthén & Muthén, 1998-2012), and the standard MAR (missing at random) approach was applied (Muthén & Muthén, 1998-2012), because it uses all of the data available to assess theoretical models without imputing the data. Meanwhile, the model parameters were evaluated for their

robustness using the Maximum Likelihood Robust (MLR) estimator. The goodness of fit of the estimated models was then evaluated using the following three absolute goodness-of-fit indices: (a) the chi-square test, (b) the root mean square error of approximation (RMSEA), and (c) the standardised root-mean-square residual (SRMR). If the chi-square value divided by the number of degrees of freedom is less than 2, this typically indicates a sufficiently good fit (Kline, 1998). In addition, we used these other relative goodness-of-fit indices to evaluate model fit: (a) the comparative fit index (CFI), and (b) the Tucker-Lewis index (TLI). The cutoff criteria for the fit indices were a CFI $>.95$, a TLI $>.95$ and an RMSEA $<.06$ (Hu & Bentler, 1999). As the chi-square test is sensitive to sample size and deviations from underlying assumptions (e.g., multivariate normality), other goodness-of-fit indices have also been considered to better assess the model fit (e.g., Hu & Bentler, 1998).

INSERT FIGURE 1 ABOUT HERE

Results

Descriptive Statistics of ECCOM

Table 2 shows the means and standard deviations for the ECCOM dimensions, subscales and items, as they apply to the Finnish and Estonian samples in first and third grades. Independent t-tests of the samples showed that Finnish teachers scored lower on the child-dominated dimension than Estonian teachers both in first ($t = -3.95, p < .001$) and third ($t = -2.37, p < .05$) grades. No statistically significant differences between the two countries were found for the child-centred and teacher-directed dimensions in either grade though.

INSERT TABLE 2 ABOUT HERE

Interrater Reliability

The results showed that the interrater reliabilities (table 3) varied between .60 and .86, which can be regarded as ranging from good to excellent (Hallgren, 2012).

INSERT TABLE 3 ABOUT HERE

Structure of ECCOM

First grade. A three-factor theoretical model (see figure 1) was estimated separately for each of the teaching practice dimensions in each grade, taking into account the three positively correlated factors of *management*, *climate*, and *instruction*. These are detailed below.

Child-centred dimension. Results for a three-factor model for the child-centred dimension showed it did not fit the data well ($\chi^2(74, N = 91) = 181.18, p < .001; CFI = 0.85; TLI = 0.82; RMSEA = 0.13; SRMR = 0.08$). However, the modification indices (*MI*) indicated that adding three estimates would increase the fit of the model. So the residual of the item ‘support for communication skills’ was correlated with (a) the residual of ‘instructional conversation’ ($MI = 27.39$); and with (b) the residual of the ‘student engagement’ item ($MI = 18.67$); and (c) the residual of ‘individualisation of learning activities’ was correlated with the residual of the ‘learning standards’ item ($MI = 18.46$). After this the results were better ($\chi^2(71, N = 91) = 107.25, p < .001; CFI = 0.95; TLI = 0.94; RMSEA = 0.08; SRMR = 0.06$), which suggests an acceptable model fit overall (see figure 2).

Because the three factors now correlated highly with each other, we decided to go one step further and test a one-factor model. However, the fit indices indicated that not only would such a model not fit the data ($\chi^2(77, N = 91) = 248.76, p < .001; CFI = 0.76; TLI = 0.72; RMSEA = 0.16; SRMR = 0.08$); but that even after estimating three more parameters using the modification indices (*MI*), the model fit was still not good ($\chi^2(74, N = 91) = 165.82; CFI = 0.87; TLI = 0.84; RMSEA = 0.12; SRMR = 0.07$). Furthermore, when we conducted a Satorra-Bentler scaled chi-square test¹ for difference (Satorra & Bentler, 2010), the results

¹ For an accurate comparison, we used the same residual correlations pattern for both the three-factor and one-factor model.

showed that the goodness of fit of the three-factor model was superior to that of the one-factor model ($\chi^2(3) = 78.12, p < .001$).

INSERT FIGURE 2 ABOUT HERE

Teacher-directed dimension. The results of a three-factor model for the teacher directed dimension showed it did not fit the data well either ($\chi^2(74, N = 91) = 194.58, p < .001$; $CFI = 0.83$; $TLI = 0.79$; $RMSEA = 0.13$; $SRMR = 0.08$); but the *MI* also suggested that adding estimates would improve this. With this in mind, correlations were therefore calculated between the residuals of the following items: (a) ‘individualisation of learning activities’ and ‘learning standards’ ($MI = 13.58$); (b) ‘teaching concepts’ and ‘maths instruction’ ($MI = 13.90$); and (c) ‘support for communication skills’ and ‘student engagement’ ($MI = 11.94$). After calculating these, the fit of the model was $\chi^2(71, N = 91) = 149.42, p < .001$; $CFI = 0.89$; $TLI = 0.86$; $RMSEA = 0.11$; $SRMR = 0.07$.

Although the model fit was still not good enough, the results were better than for the one-factor model after *MI* modifications were done ($\chi^2(74, N=91) = 177.59$; $CFI = 0.85$; $TLI = 0.82$; $RMSEA = 0.12$; $SRMR = 0.08$). In addition, the Satorra-Bentler scaled chi-square test for difference ($\chi^2(3) = 26.39, p < .001$) suggested the three-factor model had a better fit than the one-factor model (see figure 3).

INSERT FIGURE 3 ABOUT HERE

Child-dominated dimension. The results of a three-factor model for the child-dominated dimension however, did fit the data acceptably ($\chi^2(74, N=91) = 120.95, p < .001$; $CFI = 0.94$; $TLI = 0.92$; $RMSEA = 0.08$; $SRMR = 0.04$); as did the results of the one-factor model ($\chi^2(77, N=91) = 131.63, p < .001$; $CFI = 0.93$; $TLI = 0.92$; $RMSEA = 0.09$; $SRMR = 0.04$). Nevertheless, as for the other two first grade dimensions, the results of the Satorra-Bentler scaled chi-square test for difference ($\chi^2(3) = 11.95, p < .01$) showed that the three-factor model had a better fit than the one-factor model (see figure 4).

INSERT FIGURE 4 ABOUT HERE

Third grade. Analogous analyses were conducted for the third grade data.

Child-centred dimension. The results of a three-factor model for the child-centred dimension showed, as in the first grade, that it did not fit the data well ($\chi^2(74, N = 70) = 228.45, p < .001; CFI = 0.82; TLI = 0.78; RMSEA = 0.17; SRMR = 0.08$). Similarly, the *MI* suggested that adding three estimates would improve the fit. Therefore, correlations were calculated between the residuals of the following items: (a) ‘management’ and ‘disciplinary strategies’ (*MI* = 26.15); (b) ‘student engagement’ and ‘coherence of instructional activity’ (*MI* = 13.33); and (c) ‘support for communication skills’ and ‘instructional conversation’ (*MI* = 11.81). However, after this procedure, the fit of the model was still not good enough ($\chi^2(71, N=70) = 170.87, p < .001; CFI = 0.89; TLI = 0.86; RMSEA = 0.14; SRMR = 0.07$); and yet it was better than for the one-factor model, when the same procedure following the *MI* was applied ($\chi^2(74, N = 70) = 200.58; CFI = 0.85; TLI = 0.82; RMSEA = 0.16; SRMR = 0.07$). Meanwhile, the Satorra-Bentler scaled chi-square test for difference between the one and three-factor models ($\chi^2(3) = 32.74, p < .001$) suggested the three-factor model was better (see figure 2).

Teacher-directed dimension. The results of a three-factor model for the teacher-directed dimension at third grade showed that the model did not fit the data well either ($\chi^2(74, N = 70) = 174.38, p < .001; CFI = 0.87; TLI = 0.84; RMSEA = 0.14; SRMR = 0.07$). The *MI* indicated again, however, that adding three estimates would improve the fit, so correlations were calculated between the residuals of the following items: (a) ‘individualisation of learning activities’ and ‘learning standards’ (*MI* = 14.08); (b) ‘choice of activities’ and ‘support for communication skills’ (*MI* = 13.14); and (c) ‘coherence of instructional activity’ and ‘math instruction’ (*MI* = 10.57). After that, the fit of the model was acceptable ($\chi^2(71, N = 70) = 131.75, p < .001; CFI = 0.92; TLI = 0.90; RMSEA = 0.11; SRMR = 0.06$). We then

tested the one-factor model with the same three *MI* estimated parameters, and got $\chi^2 (74, N = 70) = 153.51$; $CFI = 0.90$; $TLI = 0.88$; $RMSEA = 0.12$; $SRMR = 0.06$. Again, the Satorra-Bentler scaled chi-square test for difference ($\chi^2 (3) = 24.79, p < .001$) suggested a three-factor model was better than one (Figure 3).

Child-dominated dimension. Finally, the result of a three-factor model for the child-dominated dimension showed it did not fit the data well ($\chi^2 (74, N = 70) = 232.68, p < .001$; $CFI = 0.77$; $TLI = 0.71$; $RMSEA = 0.18$; $SRMR = 0.07$), so we followed the *MI*, as for other dimensions, and added three estimates to increase the fit. Correlations were thus calculated between the residuals of the following items: (a) ‘choice of activities’ and ‘coherence of instructional activity’ ($MI = 19.43$); (b) ‘support for communication skills’ and ‘support for interpersonal skills’ ($MI = 14.29$); and (c) ‘individualisation of learning activities’ and ‘teaching concepts’ ($MI = 12.58$). After this procedure, the three-factor model still did not adequately fit to the data ($\chi^2 (71, N = 70) = 172.99, p < .001$; $CFI = 0.85$; $TLI = 0.81$; $RMSEA = 0.14$; $SRMR = 0.06$), but then nor did the one-factor model even after it was modified with the same procedure ($\chi^2 (74, N = 70) = 192.52$; $CFI = 0.83$; $TLI = 0.78$; $RMSEA = 0.15$; $SRMR = 0.06$). Finally, the Satorra-Bentler scaled chi-square test for difference ($\chi^2 (3) = 81.55, p < .001$) also suggested the three-factor model was the better of the two (see figure 4).

Item Reliability

The item reliability coefficients and standardised factor loadings of the final three-factor models also feature in figures 2 -4. For the first grade, the latent factors of child-centred practice explained between 20 and 90% of the variance in items, suggesting that all the rated items had good reliability ($> 30\%$; Bollen, 1989), except for ‘choice of activities’ and ‘student engagement’ (see figure 2). Furthermore, with these two items again proving the only exceptions, the high standardised factor loadings suggest that all ECCOM items had a high construct validity. Meanwhile, figure 3 shows that the latent factors of teacher-directed

practice in first grade explained between 30% and 76% of the variance in items, suggesting that all of the ECCOM items here had good reliability; while the high standardised factor loadings indicated that they also had high construct validity. Finally (with regard to the first grade dimensions), figure 4 shows that the latent factors of child-dominated practice explained between 41 and 88% of the variance in items, which again suggests that all of the rated items had good reliability, while the high standardised factor loadings suggest that they also had high construct validity.

For the third grade, figure 2 shows that the latent factors of child-centred practice explained between 50 and 91% of the variance in items, suggesting that all the rated items had very good reliability. Similarly, the high standardised factor loadings suggest they had very high construct validity. Meanwhile, figure 3 shows that the latent factors of teacher-directed practice explained between 24 and 92% of the variance in items, which also suggests that all the ECCOM items had good reliability, just as the high standardised factor loadings suggest they had high construct validity. Lastly, figure 4 shows that the latent factors of child-dominated practice explained between 13 and 91% of the variance in items, suggesting that all of the items, except for ‘support for communication skills’ and ‘literacy instruction’, had quite good reliability. Furthermore, with these two items again proving the only exceptions, the high standardised factor loadings suggest that the ECCOM items had high construct validity.

Internal Consistency

Next, we calculated the Cronbach’s alpha reliabilities (i.e., internal consistency) for the three subscales of each ECCOM dimension, as well as for each of the three dimensions overall. The results in table 3 show that the coefficients for the three sub-scales of each ECCOM dimension ranged from .72 to .93. Moreover, the overall coefficients for each

dimension ranged from .94 to .97. These results suggest that all of the ECCOM sub-scales had good internal consistency.

Concurrent Validity

Finally, we examined whether the teachers' self-ratings would provide evidence for concurrent validity of the ECCOM sub-scales. For this purpose, we calculated correlations between the ECCOM sub-scales and the first and third grade teachers' self-reports of their curriculum goals for reading, writing, and maths. The teacher's work experience and the size of the class were also used as criterion variables.

Child-centred dimension. For the Finnish first grade sample, the results show that child-centred teaching was, overall, positively associated with teacher curriculum goals in 'advanced reading and writing skills' ($r = .55, p < .01$), and 'working skills' ($r = .39, p < .05$). Meanwhile, on the level of the dimension's sub-scales, *management, climate and instruction* were all positively associated with curriculum goals in 'advanced reading and writing skills' ($r = .37, p < .05$; $r = .47, p < .01$; $r = .51, p < .01$; respectively); while *instruction* was positively associated with curriculum goals in 'problem solving skills' ($r = .39, p < .05$), and 'working skills' ($r = .45, p < .01$).

Meanwhile, the results for the Finnish third grade sample show that child-centred practice was, overall, negatively associated with the work experience of teachers ($r = -.40, p < .05$). Meanwhile, on the dimension's sub-scale level, *management* and *climate* were negatively associated with teachers' work experience ($r = -.47, p < .01$; $r = -.42, p < .05$; respectively). For the Estonian samples, no statistically significant associations were found however.

Teacher-directed dimension. The results for the Finnish first grade sample show that teacher-directed practice was, overall, negatively associated with curriculum goals in 'advanced reading and writing skills' ($r = -.64, p < .01$), 'problem solving skills' ($r = .40, p$

< .05), and ‘working skills’ ($r = .40, p < .05$). On the sub-scale level, *management*, *climate* and *instruction* in this dimension were all negatively associated with goals in ‘advanced reading and writing skills’ ($r = -.37, p < .05$; $r = -.62, p < .01$; $r = -.64, p < .01$; respectively); but only *climate* and *instruction* were negatively associated with them in ‘problem solving skills’ ($r = -.38, p < .05$; $r = -.46, p < .05$; respectively). Lastly, *instruction* alone was negatively associated with curriculum goals in ‘working skills’ ($r = -.40, p < .05$).

The results for the Finnish third grade sample in this dimension show that, whereas the *management* sub-scale was negatively associated with curriculum goals in ‘advanced reading and writing skills’ ($r = -.36, p < .05$), it was positively associated with teacher’s experience ($r = .38, p < .05$). For the Estonian samples, no statistically significant associations were found however.

Child-dominated dimension. The results for the Finnish first grade sample show that *instruction* in child-dominated practice was positively associated with teachers’ stress levels ($r = .37, p < .05$), but negatively associated with curriculum goals in ‘basic reading and writing skills’ ($r = -.42, p < .05$) and ‘problem solving skills’ ($r = -.36, p < .05$).

Results for the Finnish third grade sample show that, not only was child-dominated practice overall positively associated with teacher goals in ‘problem solving skills’ ($r = .43, p < .05$); but that each of the sub-scales in this dimension were too (*management*, $r = .41, p < .05$; *climate*, $r = .37, p < .05$; and *instruction*, $r = .39, p < .05$).

In this dimension, statistically significant associations were found for the Estonian samples. Child-dominated practice for the first grade teachers was negatively associated with their work experience, both generally ($r = -.42, p < .01$), and on the sub-scale level (*management*, $r = -.33, p < .01$; *climate*, $r = -.40, p < .01$; and *instruction*, $r = -.31, p < .05$). Meanwhile, the results for the Estonian third grade show that the only sub-scale negatively associated with teachers’ work experience was *management* ($r = -.42, p < .05$).

Discussion

This study aimed to contribute to the field by extending this context to first and third grade classrooms in Finland and Estonia. The results showed evidence which supports the suggested three-factor model of management, climate, and instruction for each of the ECCOM dimensions (child-centred, teacher-directed, and child-dominated). The results showed good reliabilities, not only for each of the three sub-scales in the three dimensions, but also for the total scores of each ECCOM dimension. Moreover, some further criteria validity was found with respect to other self-rated variables from teachers.

The Reliability of ECCOM

The present study not only examined the inter-rater and item reliability in ECCOM, but also the internal consistency of the sub-scales, and total scores for each of the three dimensions. Firstly, the results showed that inter-rater reliabilities in both countries and in both grades were good for all the sub-scales (management, climate, and instruction) in each ECCOM dimension (child-centred, teacher-directed, and child-dominated). Our results also showed high internal consistencies for each of the ECCOM sub-scales, and for the total scores of each ECCOM dimension in each grade, and in each country. These reliabilities were very similar to those obtained by Stipek and Byler (2004) and Lerkkanen et al. (2012a) in kindergarten settings. Overall, such good or excellent reliabilities indicate that ECCOM is an instrument that is applicable not only in kindergarten, but also in primary school classrooms.

Moreover, almost all the items showed good reliability for ECCOM across the board, irrespective of whether this applied to first grade, third grade, Finland, or Estonia. For example, each latent factor in each of the three dimensions (i.e., management, climate, and instruction) explained at least 30% of the variance in items, which suggests that the ECCOM items are reliable. The only exceptions were the items ‘choice of activities’ and ‘student

engagement' in child-centred practices at first grade, and 'support for communication skills' and 'literacy' child-dominated practices at third grade. These results indicate that these items should therefore be omitted from the final three-factor model for ECCOM, especially when it is to be used in Finland or Estonia.

Factor Structure

We also examined the factor structure of ECCOM in first and third grade classrooms in Estonia and Finland. In other words, we tested whether the three-factor model would actually fit the child-centred, teacher-directed, and child-dominated dimensions of the measure. The results of our confirmatory factor analysis showed that such a model (assuming the three positively correlated factors of management, climate, and instruction) does indeed fit the data well - not only for each dimension, but for each grade too, when compared to the one-factor model. This result suggests that ECCOM is indeed a multi-componential instrument that can be used in primary school classrooms, as was originally suggested by those who created the instrument (Stipek & Byler, 2005).

Previous ECCOM studies in kindergarten have shown that it is actually quite difficult to differentiate between child-centred, teacher-directed, and child-dominated practice (Lerkkanen et al., 2012a;). Our study confirmed that this was also the case in primary school, as the three factors correlated very highly, suggesting that they provide very similar information irrespective of the dimension. This result provides evidence for the interrelated nature of management, climate, and instruction in primary school context. In other words, it is difficult to provide high-quality management and instruction without there simultaneously being a climate of high quality in the classroom. Another possible explanation our and other scholars' findings is that although they represent different domains, it is difficult for observers to tease these dimensions apart while in the process of actually observing classroom practice.

Overall, our results suggested that ECCOM can be used in two ways, either as three-factor model or as one general factor (second-order factor). The way to use ECCOM should be depended on the research aim. For example, if researchers are trying to examine the complex relationship between teaching and child outcomes, they could use ECCOM as multi-factor way to examine which aspects of teaching are the most influential. If they use ECCOM as background/controlling variable, then a total score might be enough.

Concurrent Validity

The final aim of our study was to examine whether ECCOM would also show good criteria validity with respect to teachers' self-reports concerning their teaching style, support strategies, and curriculum goals; as well as what they felt about time restrictions and exhaustion experienced at work. The results provided some evidence of concurrent validity.

The results of the present study - concerning the curriculum goals of teachers, and observations of their teaching practice - would seem to agree with the previous findings (Stipek & Byler, 2004; Stipek & Byler, 1997; Stipek et al., 2001). At Finnish grade 1, the goals in 'advanced reading and writing', 'problem solving' and 'working skills' showed positive association with almost every component of child-centred practice, and a negative correlation with most sub-scales within teacher-directed practice. One possible explanation for this result is that teachers who practise child-centred activities are mostly relying on a constructivist method of teaching which focuses more on high-order thinking skills in their students; while teachers who typically practise teacher-directed activities are more likely to emphasise knowledge of basic facts, and use drills or rote learning to get this across - requiring less advanced skills in their students (Stipek & Byler, 2004; Stipek et al., 1995).

One interesting result was that domain of instruction within child-dominated practices was negatively correlated with curriculum goals at first grade in both 'basic reading and writing' and 'problem solving', and yet positively associated with 'problem solving skills' at

third grade. One possible explanation for this is that teachers who practise child-dominated activities are more likely to subscribe to maturationist theory (Gesell et al., 1959) and will therefore give children complete autonomy in the classroom. Consequently, in the first year of primary school they are more specifically concerned about how children adapt to school than their literacy and math skills. However, in later school years, child-dominated teachers may well feel it's time to focus more on students' academic skills (NICHD ECCRN, 2002, 2005). Additionally, Lerkkanen et al. (2012a) found that kindergarten teachers who encouraged more child-dominated practices used less instruction, showed less affection, and had a lower belief in efficacy than other teachers.

Another interesting result, was we found that teaching experience had a negative correlation with child-centred practice, but a positive one with teacher-directed practice among the Finnish third grade teachers. This result suggests that, in Finland, child-centred teaching is used more often by less-experienced teachers, while teacher-directed practice is seen more often among teachers with greater experience. These results echo previous studies which have suggested that the longer their teaching experience is, the less time teachers spend on non-academic activity (NICHD ECCRN, 2005). One reason for this result might simply be that teachers with more experience are generally older and thus received a more traditional pedagogical teacher training from teacher colleges emphasizing more teacher-directed practice compared to their younger colleagues with recent university level studies on education and child development.

In our study, the Estonian results also showed a significant association between teaching experience and the kind of teaching practised in first and third grades - those whose practice was observed as child-dominated were the less experienced teachers. One possible explanation for this result is that less experienced teachers have a more passive role since they have not yet been able to organise their classroom practice as efficiently as more

experienced teachers. This was a pattern that was also shown by Lerkkanen et al. (2012a), at kindergarten level, who found that child-dominated practice was negatively associated with teaching experience among teachers in Estonia, but not in Finland. Moreover, in our study, no associations were found between the classroom size and teacher practice in first or third grade classrooms. This was also shown in the study by NICHD ECCRN (2002), implied that general small size of class had no effect on teacher's practices.

Limitations

This study is not without its limitations. First, self-reported variables were used here as indicators for the concurrent validity of ECCOM, but perhaps it would have been better to use another standardised observation instrument for this purpose. For example, the study by Salminen et al. (2012) found that ECCOM shows convergent validity with other observational measures of classroom practices, such as the CLASS (Pianta et al., 2008). Second, the sample size in our study was relatively small which decreased the power of the statistical analyses made. Future studies should therefore be undertaken of larger samples of teachers and classrooms. Third, the present data only consisted of one-day observations, which meant it was hard to verify the stability of observational practices. It would therefore be important in future studies to gather cross-day data to provide a more stable ECCOM score.

Conclusion

Overall, the present study suggests that ECCOM is not only valid instrument just at preschool or in kindergarten, but is also a reliable and valid method to measure classroom practice in primary school classrooms. Moreover, the previously suggested three-factor structure of ECCOM provides a helpful tool to examine and promote teachers' further professional development and could also be useful in teacher-training programs.

Acknowledgement

This study was supported by grants from the Academy of Finland (No. 268 586 for 2013–2017 and No. 277 299 for 2015–2017) and institutional research funding IUT (3-3) of the Estonian Ministry of Education and Research.

References

- Allhusen, V., Belsky, J., Booth-LaForce, C. L., Bradley, R., Brownwell, C. A., Burchinal, M., ... Weinraub, M. (2004). Does class size in first grade relate to children's academic and social performance or observed classroom processes? *Developmental Psychology, 40*(5), 651–64. doi:10.1037/0012-1649.40.5.651
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.
- Bransford, J. D., Brown, A. L., & Rodney, C. R. (2000). *How people learn: Brain, mind, experience, and school* (2nd ed., p. 374). National Academies Press.
- Brophy, J., & Good, T. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.), *Handbook of research on teaching* (pp. 328–375). New York: Macmillan.
- Dasinger, L. (1997). Issues in the acquisition of Estonian, Finnish, and Hungarian: A cross-linguistic comparison. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition, Volume 4* (pp. 1–86). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Early, D. M., Maxwell, K. L., Burchinal, M., Alva, S., Bender, R. H., Bryant, D., ... Zill, N. (2007). Teachers' education, classroom quality, and young children's academic skills: results from seven studies of preschool programs. *Child Development, 78*(2), 558–80. doi:10.1111/j.1467-8624.2007.01014.x
- Flowers, C. P., & Hancock, D. R. (2003). An interview protocol and scoring rubric for evaluating teacher performance. *Assessment in Education: Principles, Policy & Practice, 10*(2), 161–168. doi:10.1080/0969594032000121261
- Gage, N. L., & Needels, M. C. (1989). Process-product research on teaching: A review of criticisms. *The Primary School Journal, 89*(3), 253–300.
- Gerber, E. B., Whitebook, M., & Weinstein, R. S. (2007). At the heart of child care: Predictors of teacher sensitivity in center-based child care. *Early Childhood Research Quarterly, 22*(3), 327–346. doi:10.1016/j.ecresq.2006.12.003

- Gesell, A., Amatruda, C. S., Ames, L. B., Castner, B. M., Halverson, H. M., Ilg, F. L., & Thompson, H. (1959). *The first five years of life: A guide to the study of the preschool child*. London: Methuen.
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23–34.
- Hamre, B. K., & Pianta, R. C. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development*, 72(2), 625–638. doi:10.1111/1467-8624.00301
- Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development*, 76(5), 949–67. doi:10.1111/j.1467-8624.2005.00889.x
- Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early childhood environment rating scale, revised edition*. New York, NY: Teacher College Press.
- Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4), 424–453. doi:10.1037/1082-989X.3.4.424
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. doi:10.1080/10705519909540118
- Ireson, J., Blatchford, P., & Joscelyne, T. (1995). What do teachers do? Classroom activities in the initial teaching of reading. *Educational Psychology*, 15(3), 245–256. doi:10.1080/0144341950150302
- Kikas, E. (2007). *Kindergarten-School Study*. Unpublished data, Department of Psychology, Tallinn University, Tallinn, Estonia.

- Kikas, E., Peets, K., & Hodges, E. V. E. (2014). Collective student characteristics alter the effects of teaching practices on academic outcomes. *Journal of Applied Developmental Psychology, 35*(4), 273–283. doi:10.1016/j.appdev.2014.04.004
- Kiuru, N., Poikkeus, A.-M., Lerkkanen, M.-K., Pakarinen, E., Siekkinen, M., Ahonen, T., & Nurmi, J.-E. (2012). Teacher-perceived supportive classroom climate protects against detrimental impact of reading disability risk on peer rejection. *Learning and Instruction, 22*(5), 331–339. doi:10.1016/j.learninstruc.2011.12.003
- Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York, NY: The Guilford Press.
- La Paro, K. M., Pianta, R. C., & Stuhlman, M. (2004). The classroom assessment scoring system: Findings from the prekindergarten year. *The Primary School Journal, 104*(5), 409–426. doi:10.2307/3202821
- Lerkkanen, M., Kikas, E., Pakarinen, E., Poikonen, P., & Nurmi, J. (2013). Mothers' trust toward teachers in relation to teaching practices. *Early Childhood Research Quarterly, 28*(1), 153–165. doi:10.1016/j.ecresq.2012.04.005
- Lerkkanen, M.-K., Kikas, E., Pakarinen, E., Trossmann, K., Poikkeus, A.-M., Rasku-
Puttonen, H., ... Nurmi, J.-E. (2012a). A validation of the early childhood classroom observation measure in Finnish and Estonian kindergartens. *Early Education & Development, 23*(3), 323–350. doi:10.1080/10409289.2010.527222
- Lerkkanen, M.-K., Kiuru, N., Pakarinen, E., Viljaranta, J., Poikkeus, A., Rasku-
Puttonen, H., ... Nurmi, J.-E. (2012b). The role of teaching practices in the development of children's interest in reading and mathematics in kindergarten. *Contemporary Educational Psychology, 37*(4), 266–279. doi:10.1016/j.cedpsych.2011.03.004

- Lerikkanen, M.-K., Niemi, P., Poikkeus, A.-M., Poskiparta, M., Siekkinen, M., & Nurmi, J.-E. (2006). *The First Steps Study*. Unpublished data, Department of Psychology and Department of Teacher Education, University of Jyväskylä, Jyväskylä, Finland.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30.
- Midgley, C., Feldlaufer, H., & Eccles, J. S. (1989). Student/teacher relations and attitudes toward mathematics before and after the transition to junior high school. *Child Development, 981–992*.
- Muthén, L. K., & Muthén, B. O. (n.d.). Mplus users guide and Mplus version 7.0. Retrieved January 10, 2014, from <http://www.statmodel.com/index.shtml>
- National Institute of Child Health and Human Development and Early Child Care Research Network. (2002). The relation of global first-grade classroom environment to structural classroom features and teacher and student behaviors. *The Primary School Journal, 102*(5), 367–387. doi:10.2307/1002181
- National Institute of Child Health and Human Development and Early Child Care Research Network. (2005). A day in third grade: A large-scale study of classroom quality and teacher and student behavior. *The Primary School Journal, 105*(3), 305–323. doi:10.1086/428746
- OECD. (2012). *PISA 2012 results*. Retrieved from <http://www.oecd.org/pisa/keyfindings/pisa-2012-results.htm>
- Ormrod, J. E. (1999). *Human learning*. Upper Saddle River, N.J: Merrill.
- Pakarinen, E., Lerikkanen, M.-K., Poikkeus, A.-M., Rasku-Puttonen, H., Eskelä-Haapanen, S., Siekkinen, M., & Nurmi, J.-E. (2015). High Quality Classroom Interactions Promote the Development of Readings Skills in the 1st Grade. Submitted for publication, University of Jyväskylä.

- Perry, K. E., Donohue, K. M., & Weinstein, R. S. (2007). Teaching practices and the promotion of achievement and adjustment in first grade. *Journal of School Psychology, 45*(3), 269–292. doi:10.1016/j.jsp.2007.02.005
- Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*(2), 109–119. doi:10.3102/0013189X09332374
- Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science, 9*(3), 144–159. doi:10.1207/s1532480xads0903_2
- Pianta, R., La Paro, K., & Hamre, B. (2008). *CLASS: Classroom assessment scoring system. assessment*. Baltimore: Paul H. Brookes.
- Pressley, M., Roehrig, A. D., Raphael, L., Dolezal, S., Bohn, C., Mohan, L., ... Hogan, K. (2003). Teaching processes in primary and secondary education. In *Handbook of Psychology*. John Wiley & Sons, Inc. doi:10.1002/0471264385.wei0708
- Rao, N., Sun, J., Zhou, J., & Zhang, L. (2012). Early achievement in rural China: The role of preschool experience. *Early Childhood Research Quarterly, 27*(1), 66–76. doi:10.1016/j.ecresq.2011.07.001
- Salminen, J., Lerkkanen, M.-K., Poikkeus, A.-M., Pakarinen, E., Siekkinen, M., Hännikäinen, M., Poikonen, P.-L., & Rasku-Puttonen, H. (2012). Observed classroom quality profiles in kindergarten classrooms in Finland. *Early Education and Development, 23*, 654-677.
- Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference Chi-square test statistic. *Psychometrika, 75*(2), 243–248. doi:10.1007/s11336-009-9135-y

- Seymour, P. H. K., Aro, M., Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143–174.
doi:10.1348/000712603321661859
- Stipek, D., & Byler, P. (1997). Early childhood education teachers: Do they practice what they preach? *Early Childhood Research Quarterly*, *12*(3), 305–325.
doi:10.1016/S0885-2006(97)90005-3
- Stipek, D., & Byler, P. (2004). The early childhood classroom observation measure. *Early Childhood Research Quarterly*, *19*(3), 375–397. doi:10.1016/j.ecresq.2004.07.007
- Stipek, D., & Byler, P. (2005). The early childhood classroom observation measure. Unpublished coding manual, School of Education, Stanford University.
- Stipek, D., Feiler, R., Daniels, D., & Milburn, S. (1995). Effects of different instructional approaches on young children's achievement and motivation. *Child Development*, *66*(1), 209–223. doi:10.2307/1131201
- Stipek, D., Givvin, K., Salmon, J., & MacGyvers, V. (2001). Teachers' beliefs and practices related to mathematics instruction. *Teaching and Teacher Education*, *17*(2), 213–226.
doi:10.1016/S0742-051X(00)00052-4
- Valeski, T. N., & Stipek, D. J. (2001). Young children's feelings about school. *Child Development*, *72*(4), 1198–213. Retrieved from
<http://www.ncbi.nlm.nih.gov/pubmed/11480942>
- Walker, J. M. T. (2008). Looking at teacher practices through the lens of parenting style. *The Journal of Experimental Education*, *76*(2), 218–240. doi:10.3200/JEXE.76.2.218-240
- Voelkl, K. E. (1995). School warmth, student participation, and achievement. *The Journal of Experimental Education*, *63*(2), 127–138.

Yen, S.-C., & Ispa, J. M. (2000). Children's temperament and behavior in Montessori and constructivist early childhood programs. *Early Education & Development, 11*(2), 171–186. doi:10.1207/s15566935eed1102_3