

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Virtanen, Tuomo; Pakarinen, Eija; Lerkkanen, Marja-Kristiina; Poikkeus, Anna-Maija; Siekkinen, M.; Nurmi, Jari-Erik

Title: A Validation Study of Classroom Assessment Scoring System-Secondary in the Finnish School Context

Year: 2018

Version: Accepted version (Final draft)

Copyright: © The Authors 2017

Rights: In Copyright

Rights url: <http://rightsstatements.org/page/InC/1.0/?language=en>

Please cite the original version:

Virtanen, T., Pakarinen, E., Lerkkanen, M.-K., Poikkeus, A.-M., Siekkinen, M., & Nurmi, J.-E. (2018). A Validation Study of Classroom Assessment Scoring System-Secondary in the Finnish School Context. *Journal of Early Adolescence*, 38(6), 849-880.
<https://doi.org/10.1177/0272431617699944>

A Validation Study of Classroom Assessment Scoring System Secondary in the Finnish School
Context

Virtanen, T. E.^{1a}, Pakarinen, E.¹, Lerkkanen, M.-K.¹, Poikkeus, A.-M.¹, Siekkinen, M.² & Nurmi, J.-E.³

Author Notes

¹ University of Jyväskylä, Department of Teacher Education, P.O. Box 35, FI-40014 University of Jyväskylä, Finland; tuomo.e.virtanen@jyu.fi; eija.k.pakarinen@jyu.fi; marja-kristiina.lerkkanen@jyu.fi; anna-maija.poikkeus@jyu.fi.

² University of Eastern Finland, School of Applied Educational Science and Teacher Education, P.O. Box 111, FI-80101 Joensuu, Finland; martti.siekkinen@uef.fi.

³ University of Jyväskylä, Department of Psychology P.O. Box 35, FI-40014 University of Jyväskylä, Finland; jari-erik.nurmi@jyu.fi.

This study was supported by a grant from the Academy of Finland (No. 268586 for 2014-2017).

^a Corresponding author. Address: University of Jyväskylä, Department of Teacher Education, P.O. Box 35, FI-40014 University of Jyväskylä, Finland. Tel.+358-40-8053636.

E-mail address: tuomo.e.virtanen@jyu.fi

Abstract

This study examined the reliability and validity of the Classroom Assessment Scoring System-Secondary (CLASS-S; Pianta, Hamre, & Mintz, 2012) in Finnish classrooms. Trained observers coded classroom interactions based on video recordings of 46 Grade 6 classrooms (450 cycles). Concurrent associations were investigated with respect to teacher self-ratings (e.g., efficacy beliefs and teaching-related stress). Confirmatory factor analysis showed that the hypothesized three-factor structure of the original CLASS-S (Emotional Support, Organizational Support, and Instructional Support), with some modifications, provided a better fit for the data compared to one- and two-factor structures. Structural validity was demonstrated by mostly high factor loadings. Except for two interrater intraclass correlations, all item, scale, and interrater reliabilities were either acceptable or good. The study found some evidence for concurrent associations between the three CLASS-S factors and teacher self-ratings. The results provide evidence of the applicability of the CLASS-S instrument in educational contexts (Finland) outside the United States.

Keywords: Classroom Assessment Scoring System Secondary (CLASS-S), Teaching through Interactions, classroom observation, structure analysis

A Validation Study of the Classroom Assessment Scoring System-Secondary
in the Finnish School Context

Increasing evidence shows that the quality of classroom interactions is more important to explain student achievement and engagement in learning situations than, for example, teachers' qualifications and experience, or the curriculum (Schacter & Thum, 2004). This evidence highlights the relevance for a theory-based and empirically corroborated understanding of what constitutes high-quality classroom interactions (Pianta, Hamre, & Allen, 2012) in different cultures, as well as valid instruments to measure them (see Bell et al., 2012; Douglas, 2009). Measures are needed that are grounded in theory, valid for capturing the critical aspects of quality in classroom interactions between teachers and early adolescents (see Hafen et al., 2015), and which are anchored in observable indicators (see Pianta & Hamre, 2009). Given that the majority of studies examining teachers' instructional practices and classroom interactions have been conducted in the United States (US), there is an evident need to validate reliable structured assessment tools in other cultural and educational contexts, and thus obtain information about the instruments' cross-cultural applicability. Consequently, the present study set out to investigate the validity and reliability of an instrument measuring teacher–student interactions – the widely used Classroom Assessment Scoring System-Secondary (CLASS-S; Pianta et al., 2012) – in Finnish classrooms. The success of Finnish youth in cross-cultural comparative studies of student achievement makes Finland an interesting context to examine the structure of the CLASS-S and the quality of teacher–student interactions outside the US.

Teaching through Interactions (TTI) Framework and the CLASS

The Teaching through Interactions (TTI; Hamre et al., 2013) framework and the Classroom Assessment Scoring System (CLASS) that operationalizes this framework focus on the patterns of interactions between teachers and students as central drivers for student learning (Bell et al., 2012; Hafen et al., 2015; Hamre et al., 2013), and the dynamic transaction between an individual's skills

and capabilities and the contextual resources (Bronfenbrenner & Morris, 1998). Various studies in the US have shown the framework to have strong empirical and theoretical support (Allen et al., 2013; Sandilos & DiPerna, 2014), which is set in a broader framework of effective teaching (Pressley et al., 2013). The CLASS operationalizes classroom interactions into three major domains capturing emotional, organizational, and instructional features of classroom teacher–student interactions (Pianta & Hamre, 2009).

The first CLASS-S domain, Emotional Support, draws from attachment (Bowlby, 1979) and self-determination theories (Connell & Wellborn, 1991), highlighting the importance of fulfilling an individual’s need for relatedness (Deci & Ryan, 2000). The CLASS-S (Pianta et al., 2012) characterizes the Emotional Support domain as warm and caring relationships in the classroom, and the teacher’s sensitivity to the students’ academic, behavioral, and affective needs, as well as their perspectives and ideas (Hamre et al., 2013). Recent evidence suggests that a high level of organizational and instructional support is contingent upon a high level of emotional support (AUTHORS, 2015; Hagelskamp, Brackett, Rivers, & Salovey, 2013).

The second CLASS-S domain, Organizational Support, refers to predictable, efficient, and goal-oriented activities and disciplinary practices in the classroom with the goal of engaging students in learning activities (Hamre, Pianta, Mashburn, & Downer, 2007). The domain of Organizational Support encompasses effective behavior management through clear expectations, routines, and rules that are positively reinforced (Gettinger & Walter, 2012), and teachers’ strategies to encourage desirable behavior, and prevent and redirect misbehavior, thereby maximizing learning time (Pianta et al., 2012). Moreover, classes with high levels of organizational support are characterized by students transitioning smoothly and effectively from one activity to another without any disrespectful words or actions (Pianta et al., 2012). A clear classroom structure has, for instance, been found to support students’ self-regulatory skills, which, in turn, positively contributes to their learning (see Paris & Paris, 2001).

The third CLASS-S domain, Instructional Support, draws from cognitive learning theories (Yilmaz, 2011), which view students as active participants whose development is supported by effective teaching that focuses on understanding and learning ideas. Effective instructional support entails the teacher being aware of aspects students typically misunderstand, transforming content to make it easily accessible for the students, presenting both key ideas and a broad framework, and using aids, such as metaphors, analogies, problems, pictures, and diagrams, to foster understanding (Gibbs & Poskitt, 2010). Classrooms with high levels of instructional support are characterized by teachers who actively facilitate students' higher-level thinking by keeping lessons interesting and engaging, and who provide feedback that expands the students' learning and understanding (Pianta et al., 2012). A cognitively stimulating classroom environment has been shown to advance students' higher-order thinking skills and bolster their cognitive development (Hamre et al., 2007).

Empirical Studies on the Factor Structure of the CLASS-S

Different versions of the CLASS instrument have been widely employed to assess classroom interactions among younger children and have been validated in several cultural contexts (e.g., Leyva et al., 2015; Pakarinen et al., 2010; Reyes, Brackett, Rivers, White, & Salovey, 2012). However, studies conducted with the CLASS-S (Pianta et al., 2012) are still sparse, and it has not yet been validated outside of the US. The few studies that have tested the structure of the CLASS-S have provided support for the a priori assumption of three interrelated factors of Emotional Support, Organizational Support, and Instructional Support (see Figure 1).

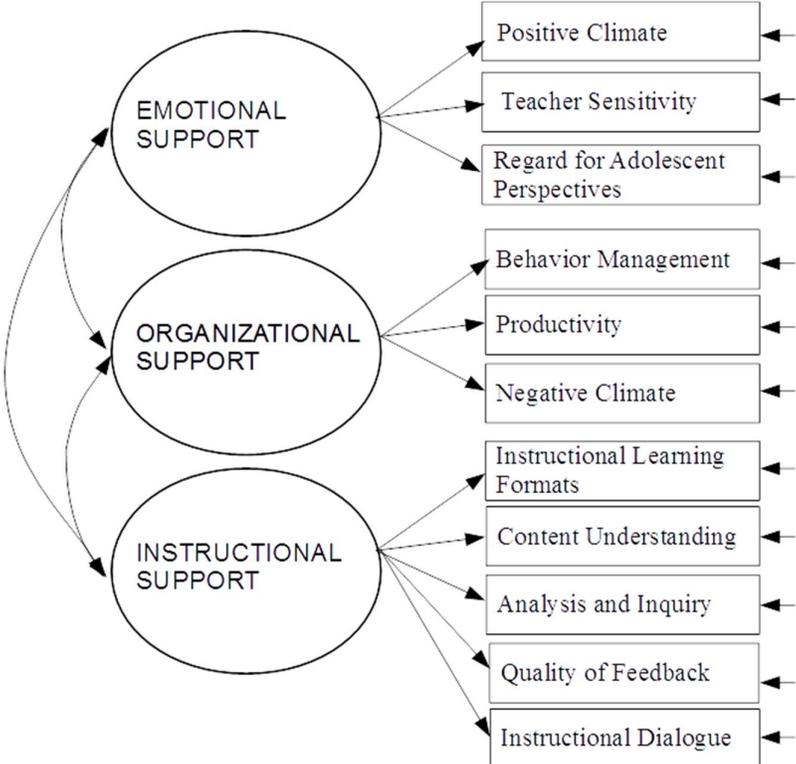


Figure 1. The hypothesized CLASS-S three-factor model.

These three correlated latent domains of the CLASS are composed of dimensions of classroom interactions operationalized and rated at the behavioral level (see Table 1).

Table 1

Description of Classroom Assessment Scoring System Dimensions (Pianta et al., 2012)

Domain	Dimension	Description
Emotional Support	Positive Climate	Enjoyment and emotional connection between teachers and students and in peer interactions
	Teacher Sensitivity	Teacher responsiveness to the academic and social/ emotional needs of individual students
	Regard for Adolescent Perspectives	The degree to which teachers meet the social and developmental needs and goals of adolescents for decision-making and autonomy, relevance, having their opinions valued, and meaningful interactions with peers
Organizational Support	Behavior Management	Encouragement of positive behaviors and monitoring, preventing, and redirecting misbehavior
	Productivity	How well the classroom runs with respect to routines, and the degree to which teachers provide activities that allow maximum time to be spent in learning activities
	Negative Climate	Level of expressed negativity such as anger, hostility, aggression, or disrespect by teachers and/or students in the classroom
Instructional Support	Instructional Learning Formats	How teachers engage students in and facilitate activities so that learning opportunities are maximized
	Content Understanding	Emphases and approaches used to help students understand the broad framework and key ideas in an academic discipline

Analysis and Inquiry	Promotion of higher-order thinking skills (e.g., analysis and integration of information, hypothesis testing, metacognition) and opportunities for application in novel contexts
Quality of Feedback	Feedback given to extend and expand students' learning through their responses and participation in activities
Instructional Dialogue	Use of structured, cumulative questioning and discussion to guide and prompt students' understanding of content

Previous CLASS-S studies have reported contradictory findings concerning the factor structure of the instrument. In their seminal study of CLASS-S with a sample of 37 secondary school classrooms, Allen et al. (2013) found a good fit for the expected three-factor structure formed by Emotional Support (positive climate, negative climate, teacher sensitivity, and regard for adolescent perspectives), Classroom Organization (classroom management, productivity, and instructional learning formats), and Instructional Support (content understanding, analysis and problem solving, and quality of feedback). Hafen and his colleagues (2015) examined several factor structures of the CLASS-S with a large-scale sample consisting of 1,482 middle and high school classrooms across the US. The results of their exploratory and confirmatory factor analyses indicated superiority of the three-factor solution over: a) the one-factor structure; b) the two-factor structure, which combines Emotional Support and Instructional Support into one factor while leaving Organizational Support as a separate factor; c) the three-factor structure, which includes Negative Climate in the Organizational Support factor rather than the Emotional Support factor); and d) the bi-factor approach, specifying both general and domain-specific factors that are uncorrelated, which hypothesizes a global factor of responsive teaching with sub-factors representing the three-factor approach.

However, it should be noted that not all studies have confirmed the theoretically presumed three-factor structure. After analyzing a total of 376 segments from 144 secondary school lessons (17 teachers) in the United Kingdom (UK), Malmberg et al. (2010) found that the three-factor structure provided a poor fit for the data. Furthermore, Kane et al. (2012) found that a single-factor solution was the most parsimonious fit with data consisting of math and English language arts teachers in Grades 4 through 8.

Associations between Teacher–Student Interactions and Teacher and Classroom Characteristics

Previous studies have shown that teachers' beliefs affect their instructional approach (Stipek & Byler, 1997). For example, teachers' high self-efficacy beliefs have been shown to be associated with effective classroom management (Woolfolk, Rogoff, & Hoy, 1990) and teaching practices (Parker & Neuharth-Pritchett, 2006; Rimm-Kaufman & Sawyer, 2004), high-quality literacy instruction (Justice, Mashburn, Hamre, & Pianta, 2008), a close relationship with children, and a positive impression of children (Mashburn, Hamre, Downer, & Pianta, 2006). Moreover, Guo et al. (2012) reported that teachers with higher levels of self-efficacy were more likely to interact with their students in sensitive and responsive ways to support their learning (e.g., provide constructive feedback).

Previous studies have also shown that teacher reports of high levels of stress or burnout may be linked to challenges in building supportive relationships in their classroom, sensitively supporting students' learning, and effectively redirecting students' misbehaviors (Hamre, Pianta, Downer, & Mashburn, 2008; Hoglund, Klingle, & Hosan, 2015; McLean, & Connor, 2015). In their review, Jennings and Greenberg (2009) suggested that teacher well-being is important for the development and maintenance of supportive teacher–student relationships and effective classroom management. Findings by Li Grining et al. (2010) indicated that teacher reports of high levels of stress were moderately predictive of a lower use of effective strategies of behavior management in the

classroom. Psychologically controlling teaching (i.e., including guilt induction, shaming, and expressing disappointment) is another teaching dimension that has been suggested to correlate negatively with teacher–student interactions, in particular, teachers’ abilities to support student autonomy and involvement (Soenens et al., 2012). In a previous study, no significant associations were found between kindergarten teachers’ self-reported psychological control and the CLASS domain scores (Pakarinen et al., 2010); however, it is possible that aspects of psychological control will be more relevant to the quality of teacher–student relationships in the secondary school period.

In addition, some contextual factors have been shown to be related to observed classroom interactions. Smaller class sizes enable teachers to organize and implement high-quality classroom interactions in terms of emotional support, classroom organization, and instructional support (Graue, Rauscher, & Sherfinski, 2009). In small classes, teachers have more time for individualized teaching (Blactford et al., 2003), and children have more opportunities to interact with the teacher and with other students (Graue et al., 2009). Results concerning the links between teachers’ work experience and the quality of classroom interactions are somewhat mixed. While some studies have found that more years of teaching are related to higher quality classroom interactions (e.g., NICHD, 2002), other studies have shown that less experience is associated with a higher quality of classroom interactions (e.g., Guo et al., 2012).

The previous literature on teacher–student interactions has at least two major limitations. Information on the classroom processes contributing to effective teaching in the later school years is scant. Second, the psychometric properties of CLASS-S have seldom been analyzed outside the US (for an exception, see Malmberg et al., 2010). Since culture may partly affect the patterns of classroom interactions (see Schoorman, Mayer, & Davis, 2007), empirical studies from other cultural and educational contexts are needed to corroborate the suggested structure of the measure, and to investigate the universality of the critical domains proposed in the TTI framework and operationalized in the CLASS-S.

Education in Finland

The present study was conducted in Finland, which has consistently high student performance outcomes in the international comparative education studies of achievement, such as the Program for International Student Assessment (PISA; OECD, 2013). The primary (Grades 1–6, ages 7–12) and lower secondary (Grades 7–9, ages 13–15) schools form an integrated, compulsory basic education from Grades 1 to 9 and ages 7 to 15. Instead of selective admissions, student enrollment in basic education is based on catchment areas where the child is allocated a place in a school nearest to where he/she lives. Class size in Finland is typically relatively small (on average, 20 students per classroom; OECD, 2014), and the vast majority of teachers have at least a master's degree. Finnish schools are almost exclusively public schools and the differences between the schools are minor. There are no high-stakes national standardized tests or test-based school accountability. Instead, formative student assessment is emphasized, with the focus on supporting students' skill development and individual growth. Access to intensified or special support is flexible. Grade 6, the grade level of the classroom observations of the present study, is still part of primary school, and class teachers are mainly responsible for giving instruction, while subject teachers teach only some specific subjects, such as physical education and foreign languages. Teachers follow the national core curriculum for basic education (Finnish National Board of Education, 2014), but they are accorded a great deal of autonomy regarding teaching methods and choosing teaching material. Student-centered approaches that are stressed with younger children (especially Grades 1–3) give way to relatively more traditional teacher-led instruction with older students (see Andrews, Ryve, Hemmi, & Sayers, 2014; Moate, 2017). Collaboration rather than competition among students is highlighted (Sahlberg, 2015).

The Present Study

By examining teachers' instructional practices and classroom interactions in Finland, this study provides evidence for the cross-cultural applicability of CLASS in a cultural and educational

context other than the US and validation of its correlates with teacher and context characteristics. The aim of the present study was to examine the psychometric properties of the CLASS-S in the Finnish educational context.

1. To determine the structural validity of the CLASS-S, we fit several factor models to our data. As suggested by previous studies, we hypothesized that the best fit for the CLASS-S is acquired by three correlated factors: Emotional Support, Organizational Support, and Instructional Support (Hypothesis 1; Hafen et al., 2015; Pianta et al., 2012).
2. To examine the reliability of the CLASS-S, we calculated item, scale, and interrater reliabilities. In line with previous studies (e.g., Allen et al., 2013; Gregory, Allen, Mikami, Hafen, & Pianta, 2014), we expected to find good interrater reliabilities (IRR) as well as item and scale reliabilities (Hypothesis 2).
3. To investigate concurrent associations to teacher and classroom factors, we calculated correlations between the CLASS-S domains (Emotional Support, Organizational Support, and Instructional Support) and teacher self-ratings of their efficacy beliefs, psychological control, teaching-related stress and exhaustion at work, as well as teacher work experience and class size. The following hypotheses were formulated: High self-efficacy beliefs would be linked to high-quality Classroom Organization (Hypothesis 3a; Justice et al., 2008). High psychological control would be associated with low-quality Emotional Support (Hypothesis 3b; Pakarinen et al., 2010; Soenens et al., 2012). High teaching-related stress and exhaustion at work would be associated with low observed Emotional Support and Classroom Organization (Hypothesis 3c; Hamre et al., 2008; Hoglund et al., 2015; McLean, & Connor, 2015). Longer work experience would be linked to high-quality Classroom Organization and Instructional Support (Hypothesis 3d; NICHD, 2002). Finally, small class size would be associated with high-quality Emotional and Instructional Support (Hypothesis 3e; Graue et al., 2009).

Method

Participants

The present study is part of a longitudinal study (AUTHORS, 2006–2016) that has followed approximately 2,000 children from kindergarten to Grade 9. The original sample was recruited from four municipalities in Finland: two in central, one in western, and one in eastern Finland. Forty-six Grade 6 class teachers (24 female and 22 male) from 32 schools participated in classroom video recordings. The participating teachers were selected on a voluntary basis from a total of 153 teachers for the longitudinal study. The teachers gave written consent for their own participation in the study, and parents gave written consent to allow their children to participate in the video recordings. The percentage of students with parental consent to take part in the video recordings was 73.5%. Students without written consent to participate in the study attended their lessons normally and were not removed from the classroom; however, their faces were later blurred on the video recordings. Most of the classrooms ($n = 42$) were mainstream classes, and four were special education classes. The class sizes ranged from 7 to 30 students ($M = 20.64$, $SD = 5.93$; information was missing for four classes). All the classrooms were Finnish speaking.

Measures

Classroom interactions. The students' ages were used as the criterion for choosing the appropriate version of the CLASS instrument for this study. Because of differences between the Finnish and US educational systems, Finnish 12-year-old students are in Grade 6 and attending their final year of primary school, while in the US, they are typically 7th graders in middle school or junior high school. The CLASS-S (Pianta et al., 2012) assesses classroom interactions along three domains – Emotional Support, Organizational Support, and Instructional Support – which consist of eleven dimensions. The dimensions are anchored in behavioral indicators provided in the CLASS-S manual (Pianta et al., 2012), which are scored along a continuous rating scale between 1 and 7 (1–2 low range, 3–5 middle range, and 6–7 high range). The scores for each dimension were averaged

across cycles, lessons, and days. Following previous CLASS-S validation studies (Allen et al., 2013; Hafen et al., 2015), the dimension of Student Engagement was not included in the present study.

Procedure. Classroom observation data were collected between mid-March and the end of April 2013 from 46 classrooms. In each classroom, two video recordings were conducted on two separate days in 31 classrooms and during one day in 15 classrooms. Video recordings on separate days were typically conducted within five days of each other ($N = 23$), and the maximum time period between the two video recordings was 17 days. This resulted in video recordings of 150 lessons, which were typically 45 minutes long, with 2–4 lessons recorded for each classroom (average 3.3 lessons). Calculations of within classroom stability across lessons in the 31 classrooms that were observed on two separate days indicated a moderate test-retest reliability for Emotional Support ($r = .33, p = .117$); and somewhat higher reliability for Organizational Support ($r = .50, p = .020$) and Instructional Support ($r = .44, p = .066$). Lessons on which video recordings were conducted were mainly Finnish language and literacy (45% of the lessons) and mathematics (43%). The subjects of the remaining video-recorded lessons were English, biology, geography, religion, physics, chemistry, and arts-and-crafts (12% in total). The lessons were video recorded by research assistants. The camera was positioned so that it captured the teacher as well as the majority or all of the students. The research assistants were instructed to record lessons that involved teachers' actual instruction (as opposed to students taking a test, watching a video, etc.). Trained coders performed the coding based on the video recordings. When coding, the recorded lessons were divided into three cycles that varied between 8 and 15 minutes ($M = 13$ min. 18 sec., $SD = 1$ min. 21 sec.), depending on the total length of the lesson (varied between 24–45 min). The coders first observed the interactions in a cycle while making notes on indicators. Coding was recorded on a separate scoring sheet before beginning the next observation cycle. Of the lessons, 15% were double-coded by two independent coders.

Training. Coders were doctoral and postdoctoral students with previous experience in CLASS coding. They participated in a comprehensive CLASS-S training, which consisted of several

sessions led by more experienced researchers with CLASS-S training. The coders watched several videos on classroom interactions, rated them, received feedback about their coding, discussed discrepancies, and clarified coding criteria, carefully following the guidelines presented in the CLASS-S manual (Pianta et al., 2012). The training was conducted mainly in Finnish using both the original English version of the CLASS-S manual and guidelines translated into Finnish (not published) to support the training. When necessary, the research team discussed the coding procedure with the TeachStone team (for example, how to score Instructional Dialogue when students work silently and independently with their working sheets/books, etc.). At the end of the training, observer agreement exceeded 80% as required in the CLASS-S manual (Pianta et al., 2012).

Teacher questionnaires. Teachers were asked to complete questionnaires on their efficacy beliefs for classroom management, psychological control, teaching-related stress and exhaustion at work, as well as their work experience and class size. The proportion of missing data for teacher questionnaires was small (four out of 46 teachers did not fill in the questionnaires).

Efficacy beliefs for classroom management. Teacher efficacy beliefs were assessed using a 12-item short form of the Teacher Efficacy Scale by Tschannen-Moran and Woolfolk Hoy (2001) adapted into Finnish. The items were rated on a 5-point scale from 1 (*not at all*) to 5 (*to a great extent*). The present analyses focused on the items belonging to the classroom management subscale. A sum score was calculated as a mean of the following four items: “How much can you do to calm a student who is disruptive or noisy?”; “How well can you keep a few problem students from ruining an entire lesson?”; “How well can you respond to defiant students?”; “How much can you do to get children to follow classroom rules?” (Cronbach’s alpha = .77).

Psychological control. Psychological control was measured using four questions from the Teacher Interactional Styles Scale (Aunola, Lerkkanen, Poikkeus, & Nurmi, 2005; see also Pakarinen, Lerkkanen et al., 2010). Teachers were asked to rate the items on a 5-point scale (1 = *does not fit me at all*, 5 = *fits me very well*). The items reflected the teachers’ efforts to control students

through guilt and expressing disappointment in teaching situations: “Children in my class should know how much I sacrifice for them”; “I believe that the children in my class should be aware of how much I do for them”; “I let the children in my class see how disappointed and ashamed I am if they misbehave”; “The children in my class need to learn to respect how good their situation is.” The sum score was calculated as the mean of these items (Cronbach’s alpha = .76).

Teaching-related stress. Teaching-related stress was assessed using a modified inventory originally developed to assess parenting stress (Gerris et al., 1993). The modification involved changing from the home context to the school context. This inventory was adapted earlier to use with teachers in a Finnish sample, and reliability and validity data were collected (see Pakarinen, Kiuru et al., 2010; Pakarinen, Lerkkanen et al., 2010). In the present sample, the correlation with teacher exhaustion at work was .55 ($p < .001$). The items assessed teachers’ feelings of teaching-related stress due to experiencing inadequacy and guilt regarding their ability to cope with teaching demands. The ratings were given on a 5-point scale (1 = *does not fit me at all*, 5 = *fits me very well*). The three items were as follows: “I have a lot more problems guiding the children than I expected”; “I often feel guilty or inadequate when thinking about what kind of teacher I am”; “I sometimes feel that guiding children is an overwhelming task for me.” The sum score was calculated as the mean of these items (Cronbach’s alpha = .67).

Teacher exhaustion at work. Teachers’ exhaustion at work was measured using a shortened version of the Bergen Burnout Indicator 15 (BBI-15) (Näätänen, Aro, Matthiesen, & Salmela-Aro, 2003). Teachers rated the following five items from the exhaustion subscale: “I feel I am drowning in work”; “I often sleep poorly because of different work affairs”; “I continuously have a bad conscience because I have to neglect my family because of my work”; “I think about work matters during my leisure time”; “The pressure of work has caused problems in my close relationships.” The ratings were given on a 6-point scale (1 = *strongly disagree*, 6 = *strongly agree*). The sum score was calculated as the mean of these items. (Cronbach’s alpha = .85).

Teacher work experience. Length of work experience as a schoolteacher was assessed by an item with the following six choices: 0 = *none at all*, 1 = *less than a year*, 2 = *1–5 years*, 3 = *6–10 years*, 4 = *11–15 years*, 5 = *more than 15 years*.

Analysis strategy

The analyses were conducted with Mplus version 7.3, using maximum likelihood estimation with non-normality robust standard errors (Muthén & Muthén, 1998–2012). The model parameters were estimated using the full-information maximum likelihood estimation, allowing all present data to be used (Muthén & Muthén, 1998–2012). The goodness-of-fit of the estimated models was evaluated using five indicators: three absolute fit indices consisting of χ^2 , standardized root mean square residual (SRMR), and root mean square error of approximation (RMSEA), and two comparative fit indices consisting of the comparative fit index (CFI) and Tucker-Lewis index (TLI). The cutoff values for well-fitting models were as follows: $\chi^2 = ns$ ($p > .05$), SRMR $< .05$, RMSEA $< .05$, CFI $> .95$, and TLI $> .95$ (Byrne, 2012).

The analysis comprised the following steps. First, the analysis of structural validity of the CLASS-S in the Finnish data of Grade 6 classrooms was conducted using confirmatory factor analysis (CFA) to test multiple factor structures. Second, reliabilities were calculated at the level of items, scales, and raters. Item level reliabilities were estimated as squared standardized factor loadings (Bollen, 1989), while reliabilities of the scales were estimated as Cronbach alphas and factor score reliabilities. Interrater reliability between pairs of coders was determined on 15% of the lessons coded independently by two coders as a cycle-level percentage of codings that represented a perfect match or agreement within 1 point. Additionally, intraclass correlations (ICCs) were computed in order to estimate the agreement between raters (a high ICC indicates a high interrater agreement). The advantage of using ICCs is that they do not quantify interrater reliability based on an all-or-nothing agreement, but rather incorporate the magnitude of the disagreement to compute interrater agreement estimates (Hallgren, 2012). ICCs were computed as recommended by McGraw

and Wong (1996). Classroom observations were specified as random, but raters were fixed in that they were not randomly selected from a larger population of raters, warranting a two-way mixed-effect model. Raters' consistency (i.e., similarity in rank orders) in scoring the classroom interactions was estimated. Third, concurrent associations were estimated by calculating correlations between the three CLASS domains (Emotional Support, Organizational Support, and Instructional Support) and teacher self-ratings.

Results

Table 2 presents descriptive statistics and correlations among the CLASS-S items.

Table 2

Correlation Matrix of the Raw scores of Items of the CLASS-S, and Descriptive Statistics

CLASS-S items	1	2	3	4	5	6	7	8	9	10	11
1. PC											
2. TS	.70 ^a										
3. RAP	.62 ^a	.40 ^b									
4. BM	.42 ^b	.46 ^b	.19								
5. P	.39 ^b	.35 ^c	.06	.81 ^a							
6. NC ¹	.64 ^a	.48 ^b	.30 ^c	.66 ^a	.60 ^a						
7. ILF	.69 ^a	.73 ^a	.45 ^b	.64 ^a	.71 ^a	.57 ^a					
8. CU	.54 ^a	.68 ^a	.53 ^a	.36 ^b	.43 ^b	.36 ^c	.76 ^a				
9. AI	.38 ^b	.50 ^a	.63 ^a	.36 ^b	.27 ^c	.31 ^c	.60 ^a	.74 ^a			
10. QF	.41 ^b	.56 ^a	.41 ^b	.38 ^b	.36 ^c	.36 ^c	.56 ^a	.76 ^a	.73 ^a		
11. ID	.47 ^b	.54 ^a	.63 ^a	.23	.21	.20	.50 ^a	.70 ^a	.61 ^a	.67 ^a	
Mean	4.93	5.14	3.07	5.79	5.89	6.78	4.80	3.84	2.46	3.02	2.73
Variance	0.45	0.42	0.76	0.61	0.38	0.09	0.31	0.62	0.40	0.51	0.64
Min	3.78	3.83	1.00	3.83	4.42	5.83	3.75	2.36	1.38	1.67	1.17
Max	6.17	6.45	5.60	6.83	6.83	7.00	6.17	5.42	4.00	5.08	4.75

*Note.*¹ Reversed scale. ^a $p < .001$; ^b $p < .01$; ^c $p < .05$. PC = Positive Climate; TS = Teacher Sensitivity; RAP = Regard for Adolescent Perspectives; BM = Behavior Management; P = Productivity; NC = Negative Climate; ILF = Instructional Learning Formats; CU = Content Understanding; AI = Analysis and Inquiry; QF = Quality of Feedback; ID = Instructional Dialogue.

Structural Validity

To investigate the factorial structure of the CLASS-S, we investigated several factor models. As can be seen in Table 3, the original three-factor model provided a relatively poor fit for

the data: $\chi^2(41) = 151.18$, $p < .001$, SRMR = .10, RMSEA = .24, CFI = .75, and TLI = .66.

Consequently, we tested a single-factor model, which also showed a poor fit for the data: $\chi^2(44) = 218.60$, $p < .001$, SRMR = .12, RMSEA = .29, CFI = .60, and TLI = .50. An alternative factor structure with two factors, whereby Emotional and Instructional Support are combined into one factor while Organizational Support forms a separate factor (Hafen et al., 2015), also showed a poor fit for the data: $\chi^2(43) = 158.76$, $p < .001$, SRMR = .10, RMSEA = .24, CFI = .73, and TLI = .66.

After careful inspection of the modification indices and residual correlations, we omitted two items because of their poor discriminant validity. The correlated residuals exceeded .10 in absolute value (c.f. criterion in Kline, 2011) between Regard for Adolescent Perspectives and Productivity, Analysis and Inquiry, and Instructional Dialogue in the three-factor model, and between Regard for Adolescent Perspectives and Positive Climate, Productivity, and Quality of Feedback in the two-factor model. The correlated residuals were greater than .10 between Instructional Learning Formats and Productivity and Behavior Management in the three-factor model, and between Instructional Learning Formats and Productivity in the two-factor model. However, the correlated residuals between Instructional Learning Formats and Behavior Management reached .09 in the two-factor model. High residual correlations indicated that the models predicted the corresponding correlations poorly. For other dimensions, no residual correlations exceeded .10.

After excluding Regard for Adolescent Perspectives and Instructional Learning Formats from the analysis, we reran the three-factor model. The fit of the revised three-factor model was acceptable: $\chi^2(24) = 41.43$, $p = .015$, SRMR = .06, RMSEA = .13, CFI = .94, and TLI = .91. The modification indices suggested that allowing a residual correlation between Positive Climate and Negative Climate would improve the model fit. The fit of the revised three-factor model was excellent: $\chi^2(23) = 28.58$, $p = .195$, CFI = .98, TLI = .97, RMSEA = .07, SRMR = .10, except for RMSEA and SRMR which were somewhat below acceptable (Byrne, 2012). Due to the relatively

Table 3

Estimated Item Reliabilities, Standardized Factor Loadings (in Parentheses), and Fit-indices for the CLASS-S Models

	Three-factor model			Single factor	Two-factor model		Revised three-factor model ^b		
	ES	OS	IS	ET	ES-IS	OS	ES	OS	IS
PC	.68 (.82)			.53 (.73)	.46 (.68)		.54 (.73) ^a		
TS	.70 (.84)			.62 (.79)	.59 (.77)		.80 (.90)		
RAP	.40 (.63)			.36 (.60)	.39 (.63)				
BM		.86 (.93)		.35 (.59)		.85 (.92)		.89 (.95)	
P		.74 (.86)		.34 (.58)		.75 (.87)		.72 (.85)	
NC ^r		.52 (.72)		.33 (.58)		.51 (.72)		.47 (.69) ^a	
ILF			.65 (.81)	.78 (.88)	.73 (.83)				
CU			.87 (.93)	.76 (.87)	.83 (.91)				.84 (.92)
AI			.64 (.80)	.55 (.74)	.62 (.79)				.66 (.81)
QF			.66 (.81)	.56 (.75)	.62 (.79)				.72 (.85)
ID			.55 (.74)	.47 (.69)	.55 (.74)				.59 (.77)
χ^2	151.18(41)			218.60(44)	158.76(43)		28.58(23)		
SRMR	.10			.12	.10		.10		
RMSEA	.24			.29	.24		.07		
CFI	.75			.60	.73		.98		
TLI	.66			.50	.66		.97		

Note. ET = Effective Teaching; ES = Emotional Support; OS = Organizational Support; IS = Instructional Support. ^rReversed Negative Climate. ^aResiduals of Positive Climate and Negative Climate were allowed to correlate. ^bRAP and ILF excluded. PC = Positive Climate; TS = Teacher Sensitivity; RAP = Regard for Adolescent Perspectives; BM = Behavior Management; P = Productivity; NC = Negative Climate; ILF = Instructional Learning Formats; CU = Content Understanding; AI = Analysis and Inquiry; QF = Quality of Feedback; ID = Instructional Dialogue.

small sample size of the present study ($n = 46$), this is not surprising because it is known that RMSEA is inflated even for properly specified models when the sample size is less than 200 (Curran, Bollen, Chen, Paxton, & Kirby, 2003).

The correlations between the factors in the revised three-factor model were .57 between Emotional Support and Organizational Support, .45 between Organizational Support and Instructional Support, and .76 between Emotional Support and Instructional Support. Figure 2 presents the revised CLASS-S three-factor model.

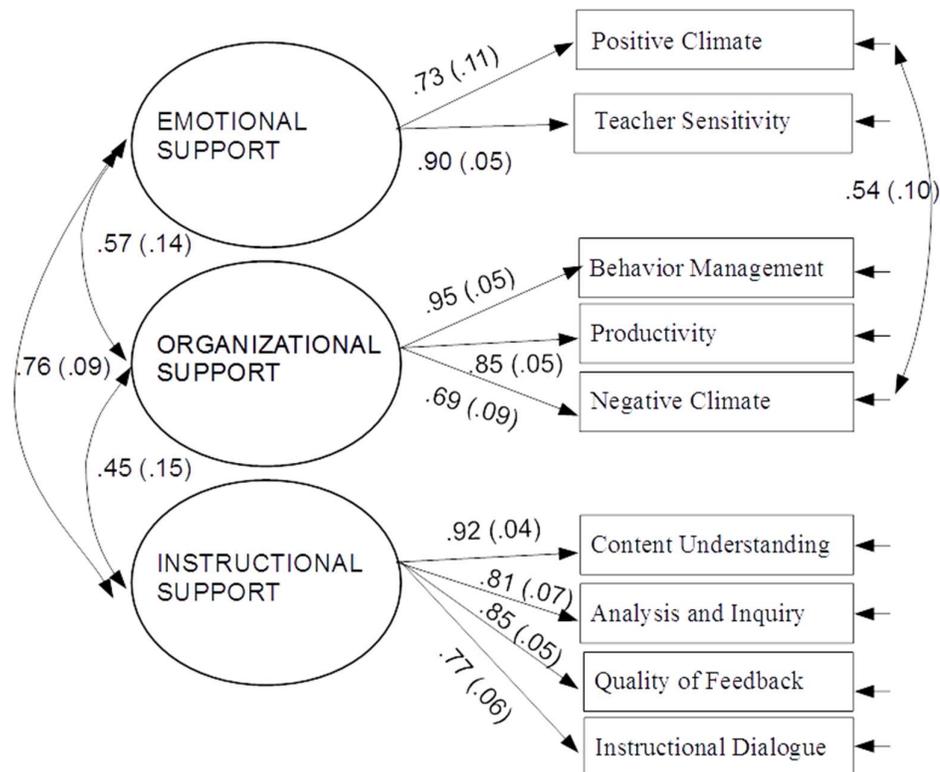


Figure 2. The revised CLASS-S three-factor model with standardized estimates.

Reliability

Our next aim was to examine the reliability of the CLASS-S at the level of items, scales, and raters. As shown in Table 3, except for the single-factor model, item reliabilities in the models typically exceeded .50 (see Kline, 2011). Cronbach’s alpha estimates for the revised three-factor

structure scales were .83 for Emotional Support, .82 for Organizational Support, and .90 for Instructional Support, indicating good internal consistency for the scales at the item level. Factor score reliabilities for Emotional Support, Organizational Support, and Instructional Support, .83, .92, and .91, respectively, indicated similarly good internal consistency at the scale level.

Interrater reliabilities (IRR) for the revised three-factor model were calculated as both percentages of agreement and ICCs at the cycle level (each lesson consisted of three observation cycles). As can be seen in Table 4, IRR were generally acceptable. Percentages of perfect cycle-level interrater agreement varied between 21.7 (Content Understanding) and 81.4 (Negative Climate). The majority of IRR exceeded .80 when calculated using cycle-level agreement within 1 point (Pianta et al., 2012). The three dimensions within the domain of Organizational Support, in particular,

Classroom Observation Interrater Reliabilities for the Revised Three-factor Model

	^a Cycles with perfect agreement (%)	^b Cycles with agreement within 1 point (%)	^c Cycle intraclass correlations (ICC)
Domains			
Emotional Support			
Positive Climate	30.0	88.6	.57
Teacher Sensitivity	31.4	82.9	.25
Organizational Support			
Behavior Management	42.9	91.4	.73
Productivity	52.9	94.3	.75
Negative Climate	81.4	97.1	.26
Instructional Support			
Analysis and Inquiry	30.0	81.5	.67
Quality of Feedback	31.9	72.5	.67
Instructional Dialogue	28.6	71.4	.64
Content Understanding	21.7	73.8	.69

Note. ^aExact agreement. ^bAdjacent agreement. ^cIntraclass correlations. Teacher Sensitivity and Negative Climate intraclass correlation estimates were statistically nonsignificant. The other intraclass correlation estimates were statistically significant at $p < .001$.

reached high agreement among the independent pairs of coders. ICCs were typically greater than .60, indicating good reliability (Cicchetti & Sparrow, 1981). However, there were two exceptions, namely Teacher Sensitivity and Negative Climate, which showed poor levels of interrater consistency.

Concurrent Associations

Our third aim was to investigate concurrent associations of the CLASS-S with respect to teachers' self-ratings of efficacy beliefs (classroom management), psychological control (teaching style), teaching-related stress, exhaustion at work, work experience, and class size. Correlations between the CLASS domain scores and teachers' self-ratings are shown in Table 5.

Table 5

Correlations of Latent Factors with Concurrent Teacher Ratings and Control Variables (n = 42)

Three-factor model ^a					
	ES	OS	IS	Mean	Variance
Predictors					
Efficacy for classroom management	.24	.43***	.06	4.11	0.24
Psychological control	-.25+	-.17	-.24	2.68	0.66
Teaching stress	-.13	-.42***	-.30+	4.46	0.61
Exhaustion at work	-.02	-.33*	-.13	2.88	1.10
Control variables					
Work experience	-.07	.41*	.03	3.93	1.48
Class size	-.27*	-.17	-.06	20.64	3.60

Note. ^aRevised three-factor model; ES = Emotional Support; OS = Organizational Support; IS = Instructional Support. + $p < .07$; * $p < .05$; ** $p < .01$; *** $p < .001$.

First, teachers' efficacy beliefs for classroom management were positively related to high-quality Organizational Support. Second, teachers' self-report of psychological control showed a marginal ($p < .07$) negative association with Emotional Support. Third, teaching-related stress and exhaustion at work were negatively related to Organizational Support. In addition, teaching-related stress showed a marginal ($p < .07$) negative association with Instructional Support. Moreover, the

extent of teacher work experience was positively linked to Organizational Support, with longer teaching experience being associated with higher organizational support in the classroom. Finally, class size was negatively associated with Emotional Support, showing that the smaller the class size the higher the observed Emotional Support in the classroom.

Discussion

The present study reports the first results outside the US concerning the structure and reliability of the secondary school version of the widely employed observation tool, the Classroom Assessment Scoring System (CLASS-S; Pianta et al., 2012). The hypothesized three-factor structure of the CLASS-S – Emotional Support, Organizational Support, and Instructional Support – provided a better fit for the data than the one- and two-factor structures, but some modifications were required with respect to certain dimensions. Factor loadings were high in all three domains, indicating good CLASS-S structural validity. Moreover, except for two interrater ICCs, all item, scale, and IRR were either acceptable or good. There was also some evidence for concurrent associations between the three CLASS-S factors and teacher self-ratings of self-efficacy, teaching-related stress, teacher work experience, and class size. The results also provide evidence for the applicability of the CLASS-S instrument in educational contexts outside the US.

Our first aim was to test several factor structures of the CLASS-S (i.e., structural validity) in the Finnish school context. In line with Hypothesis 1, the present data supported a model of three positively intercorrelated factors: Emotional Support, Organizational Support, and Instructional Support (Hafen et al., 2015; Pianta et al., 2012). However, the three-factor model required some modifications to the constellation of dimensions found in previous studies (Bell et al., 2012; Hafen et al., 2015). Two dimensions were excluded from the model because they showed poor discriminant validity, leading to poor fit of the preliminary estimated models. The first excluded dimension was Regard for Adolescent Perspectives. It is intended to assess the extent to which students are provided opportunities for autonomy and leadership during lessons by following their own ideas, encouraging

peer interaction, and flexibly adapting the lesson plan. The present data suggested that this dimension captures a wide range of aspects within classroom interactions, which are only partly aligned with emotional support, as it also encapsulates teacher practices of organizing the classroom work (e.g., allowing students to choose assignments, giving students responsibility and leadership through tasks), and instructional support (e.g., prompting students for their thoughts, ideas, and sharing). Thus, the behavioral indicators of the Regard for Adolescent Perspectives dimension appear to represent a mix of emotional, instructional, and organizational support where the emphasis is divided between the teacher's ability to provide a flexible structure that allows students' autonomous choices and freedom of movement, and the teachers' sensitivity to student initiative, valuing student input, conveying interest, and making connections to students' experiences.

Another possible explanation for the low discriminant validity of Regard for Adolescent Perspectives is that in the Finnish school system, although Grade 6 students are between 12 and 13 years of age, they attend the last grade of primary school (unlike most students in the US school system). Therefore, they are still mainly taught by classroom teachers rather than subject teachers. Thus, at this stage, Finnish teachers may be working from the perspective that a strong hand is required to direct the interactions and learning content, which leaves less room for learning formats that emphasize students' autonomy and leadership during lessons. The result may also reflect the culturally specific way of interpreting and evaluating this dimension. The coders may have underestimated the extent of regard for the students' perspectives due to challenges in interpreting individual work (e.g., solving workbook problems), which is a strong feature of the Finnish school context in Grade 6.

The other dimension excluded from our three-factor solution, Instructional Learning Formats, focuses on the teacher's provision of instruction to maximize student engagement by communicating clear learning objectives, using a variety of modalities, strategies, and materials, and involving students through questioning, encouragement, and appropriate pacing. One possible

explanation for poor discriminant validity of this dimension is small item variance (the second smallest after Negative Climate). Finnish Grade 6 classroom teachers are not subject specialists but rather pedagogical experts. When teaching core academic subjects, such as mathematics and Finnish and literacy, classroom teachers often rely on the national core curriculum guidelines (Finnish National Board of Education, 2014) and the textbooks following them (see Atjonen et al., 2008; Moate, 2017). Since a major proportion (88%) of the video recorded lessons involved these two academic subjects, this may have led to a narrower range of instructional strategies used by the teachers, as opposed to targeting those subjects that lend themselves more easily to interactional dialogue or project work, such as arts-and-crafts, religion, or music. This dimension also seems to have been somewhat problematic in the developmental phase of the CLASS-S instrument, with respect to its linkage to its respective domain, Instructional Support. The original CLASS-S manual (Pianta et al., 2010) presents Instructional Learning Formats as an item measuring Organizational Support, while the more recent version used in the present study (Pianta et al., 2012) specifies it as a dimension belonging to Instructional Support. There appears to be a need to refine the defining qualities and clarify the key behavioral indicators of the Instructional Learning Formats dimension in order to improve its discriminant validity in different educational contexts.

A similar shift from one domain to another took place in the development of the CLASS-S with respect to Negative Climate. In the original CLASS-S version (Pianta et al., 2010), Negative Climate belonged to the Emotional Support factor, whereas in the updated manual (Pianta et al., 2012), it is part of the Organizational Support factor. In the present study, modeling the residual of Negative Climate was allowed to cross-correlate with the residual of the Positive Climate item. Our results indicated that Negative Climate and Positive Climate share a common variance, suggesting that they may not be best described as two separate dimensions in the Finnish context, but perhaps as one classroom emotional climate continuum, ranging from low to high. In a study conducted in Finnish kindergartens (Pakarinen et al., 2010) using the CLASS Pre-K instrument, the dimension of

Negative Climate was excluded from the final CLASS measurement model because of low variation between classrooms.

Our second aim was to examine multiple indicators of the reliability of the CLASS-S. In line with Hypothesis 2, our findings indicated that almost all reliability coefficients were either acceptable or good. Item reliabilities matched those observed in earlier studies (Hafen et al., 2015) in that squared standardized factor loadings (Bollen, 1989) were high. All but one (Negative Climate) of the reliabilities exceeded the optimal level of .50, indicating that more than half of the variance was explained by the corresponding latent factor (Kline, 2011). Taken together, the nine items included in the Finnish version of the CLASS-S measured the three latent factors in a reliable manner. In the same vein, Cronbach's alpha estimates and factor score reliabilities were high. Alphas and factor score reliabilities exceeded .80 in all domains, with two domains (Instructional Support and Organizational Support) reaching a level above .90. These results indicated good internal consistencies for the three domains (high item correlations between the items measuring a given domain). In addition, IRR were generally high: six out of nine dimensions reached the threshold of more than 80% agreement (within 1 point) between two independent coders (Pianta et al., 2012). All three dimensions belonging to the Organizational Support domain exceeded 90% agreement.

With respect to the IRR measured by ICCs at the dimension level, those for Teacher Sensitivity and Negative Climate were below .30. Because an intraclass correlation coefficient (ICC) is highly dependent on the heterogeneity of the study sample (low variance corresponds to low intraclass correlation), the low value for Negative Climate, in particular, may be linked to low between-rater variance on this dimension. For Teacher Sensitivity, the low ICC may be due to difficulties in judging teacher responsiveness in cases where few opportunities exist for teachers to demonstrate their strategies of responding to challenges. For example, students may be actively engaged in independent work either in pairs or on their own, which is a typical learning activity in the Finnish school context in Grade 6. Overall, applying multiple methods to assess interrater

reliability may be a useful approach in studies using the CLASS-S. The current study showed that even though interrater agreement is within 1 point, the ICC may produce a relatively low interrater reliability estimate. This may warrant a closer look at a dimension, especially when used in studies outside the culture in which the measures originated.

Our third aim was to analyze the associations between the observed teacher–student interactions as assessed by the CLASS-S with teachers’ self-ratings of efficacy beliefs for classroom management, psychological control, teaching-related stress, exhaustion at work, work experience, and class size. The results provided some evidence for associations between the self-reports and the CLASS-S domain scores. As expected (Hypothesis 3a), teachers’ self-reported efficacy for classroom management was positively associated with observed organizational support, suggesting that teachers’ high efficacy beliefs about their capability to control disruptive behavior in the classroom are reflected in their actual classroom practices in the domain of Organizational Support. This finding fits well with the literature, which shows that teachers’ higher efficacy beliefs are associated with higher quality classroom interactions (Guo et al., 2012; Justice et al., 2008). This result suggests that interventions aimed at improving teachers’ self-efficacy may potentially improve the quality of classroom interactions by promoting teacher competencies in utilizing effective methods to prevent and redirect misbehavior, and to organize time and routines that facilitate maximal student involvement. In line with our expectations (Hypothesis 3b), we found a trend suggesting a negative association between teacher-reported psychological control and observed Emotional Support. This trend aligns with findings by Pakarinen and colleagues (2010), indicating that a high extent of teacher control through guilt is also associated with lower warmth and sensitivity to students’ needs and low responsiveness in classroom situations (see also Soenens et al., 2012).

Teachers’ self-reported teaching-related stress and exhaustion at work were negatively associated with observed Organizational Support and also with Instructional Support. These results

are in line with our expectations (Hypothesis 3c) and the increasing evidence showing that teacher well-being is associated with the quality of their teaching (Hamre et al., 2008; Hoglund et al., 2015; McLean, & Connor, 2015). According to Jennings and Greenberg (2009), for example, teacher well-being is important for the development and maintenance of supportive teacher–student relationships and effective classroom management. Findings by Li Grining et al. (2010) indicated that teacher stress was moderately predictive of a lower use of effective strategies of behavior management in the classroom.

Years of teaching experience (Hypothesis 3d) were positively related to Organizational Support, suggesting that longer work experience helps teachers in setting rules and expectations, as well as providing activities that maximize the time students spend in learning activities. This finding corroborates the NICHD (2002) results, which indicate that teacher experience is related to higher-quality classroom interactions.

Finally, partially in line with our expectations (Hypothesis 3e), our results showed that class size was negatively related to observed Emotional Support (but not Instructional Support). The finding suggests that small classrooms provide a more optimal context for building emotional connections between teachers and students, facilitating positive peer interactions, and promoting higher teacher responsiveness to students' needs (see also Graue et al., 2009). Relatively small classes are likely to provide children with more opportunities to interact with the teacher and with each other (Graue et al., 2009), than in classes with a high number of students. For example, a previous study by Lerkkanen et al. (2016) in Finnish first grade classrooms indicated that teachers deployed more child-centered practices in smaller classrooms.

The mean values of each domain – Emotional Support (Ms = 3.1 to 5.1), Organizational Support (Ms = 5.8 to 6.8), and Instructional Support (Ms = 2.5 to 4.8) – were generally at the same level or higher in the present study than in the study conducted in the US (Ms Emotional Support = 3.4 to 4.7, Ms Organizational Support = 5.3 to 6.8, Ms Instructional Support = 3.2 to 3.9; Allen et al.,

2013). There are at least three possible explanations for this result. First, in Finland, class sizes are relatively small (OECD, 2014), a factor that is associated with high-quality classroom interactions (Allen et al., 2013; Graue et al., 2009). Second, the teaching profession is highly valued in Finland: only 7% of applicants gain acceptance to classroom teacher-training programs. This stiff competition leads to a situation whereby the most suitable candidates are selected to master's level programs. Teacher training includes multiple practical periods in schools for the students during their university studies, and regular in-service training is provided for teachers throughout their professional careers.

Three dimensions, however, were found to be rated higher in the US sample (Allen et al., 2013) than in the present Finnish sample: Analysis and Inquiry, Quality of Feedback, and Regard for Adolescent Perspectives. One possible reason for this is that the present study was conducted among sixth graders (12 years of age) in their last year of primary school, unlike the students in the study by Allen et al. (2013), who were attending secondary school. Secondary school academic requirements, where the curriculum is taught by subject teachers, may call for higher levels of student autonomy and study strategies, which provide more explicit feedback on cognitive learning processes. Moreover, secondary school teachers may view their students as more mature and independent learners than primary school students, which may lead to instructional strategies where regard for students' perspectives is more prevalent. This study also showed smaller variances than in the US sample (Allen et al., 2013) in almost all CLASS-S dimensions. Only the dimensions of Behavior Management and Analysis and Inquiry had higher variances in the Finnish sample. The smaller variances may be a reflection of the homogeneous Finnish teacher training, the principles of equity for all students and free public education (Finnish National Board of Education, 2014; Sahlberg, 2015), and small differences between schools as attested in the PISA studies.

Limitations

The present study also has some limitations that need to be taken into account. First, although we found some evidence on concurrent associations of the observations based on using

teacher self-ratings as reference variables, another observational measure would be needed in the future to provide documentation on concurrent validity of the CLASS-S in the Finnish context. Accordingly, future studies would benefit from measuring criterion variables, such as teacher stress and burnout, using more refined measures (e.g., MBI: Maslach, Jackson, & Leiter, 1996) than the one used in the present study. Second, it would be relevant to analyze the predictive validity of the CLASS-S scores with respect to student outcomes as well. Third, variations in the educational context should be taken into account to a greater extent. Our sample consisted of 42 mainstream education classes and four special education classes. Possible moderation effects of the student composition on the CLASS-S scores should be investigated with larger samples. Fourth, teachers participated in the study on a voluntary basis, presenting the possibility of a selection effect. For instance, the observed teachers may have been more active than average teachers and, therefore, more likely to volunteer. Finally, the study was conducted in one specific educational and cultural context – Finland – which is somewhat different than the US context. A recent study in seven European countries (Slot, Cadima, Salminen, Pastori, & Lerkkanen, 2016) indicated that cultural sensitivity is needed when adapting and validating classroom observation instruments, such as the CLASS, and making generalizations from one cultural and educational context to another.

Conclusion

Overall, this study supports the overall structure of the domains of classroom interactions captured by the CLASS-S in an educational context other than the US. However, the results concerning the modifications needed at the dimension level suggest that cultural and educational contexts need to be taken into account when training observers and applying the CLASS-S instrument in another cultural context. When compared to US secondary school students, the current study also showed that the quality of teacher–student interactions as assessed by CLASS-S in the Finnish Grade 6 classrooms were generally at a slightly higher level. Moreover, small variances

indicated that the quality of Finnish students' classroom interactions is relatively uniform compared to those of US students.

References

- Allen, J., Gregory, A., Mikami, A., Lun, J., Hamre, B., & Pianta, R. (2013). Observations of effective teacher–student interactions in secondary school classrooms: Predicting student achievement with the classroom assessment scoring system—secondary. *School Psychology Review, 42*(1), 76–98.
- Andrews, P., Ryve, A., Hemmi, K., & Sayers, J. (2014). PISA, TIMSS and Finnish mathematics teaching: An enigma in search of an explanation. *Educational Studies in Mathematics, 87*(1), 7–26.
- Atjonen, P., Halinen, I., Hämäläinen, S., Korkeakoski, E., Knubb-Manninen, G., Kupari, P., Mehtäläinen, J. Risku, A-M., Salonen, M. & Wikman, T. (2008). *Tavoitteista vuorovaikutukseen. Perusopetuksen pedagogiikan arviointi* [From goals to interaction. Evaluation of pedagogy in Finnish basic education]. Publications by the Education Evaluation Council 30. Jyväskylä.
- Aunola, K., Lerkkanen, M., Poikkeus, A., & Nurmi, J. (2005). *Teacher interactional styles scale*. Jyväskylä, Finland: University of Jyväskylä.
- AUTHORS. (2006-2016).
- AUTHORS. (2015).
- Bell, C. A., Gitomer, D. H., McCaffrey, D. F., Hamre, B. K., Pianta, R. C., & Qi, Y. (2012). An argument approach to observation protocol validity. *Educational Assessment, 17*(2–3), 62–87.

- Blatchford, P., Bassett, P., Goldstein, H., & Martin, C. (2003). Are class size differences related to pupils' educational progress and classroom processes? Findings from the Institute of Education Class Size Study of Children Aged 5–7 Years. *British Educational Research Journal*, 29, 709–730
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley Interscience.
- Bowlby, J. (1979). The Bowlby-Ainsworth attachment theory. *Behavioral and Brain Sciences*, 2(4), 637–638.
- Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Vol. 1: Theoretical models of human development* (5th ed., pp. 993-1028). New York: Wiley.
- Byrne, B. M. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. New York: Routledge.
- Casabianca, J. M., McCaffrey, D. F., Gitomer, D. H., Bell, C. A., Hamre, B. K., & Pianta, R. C. (2013). Effect of observation mode on measures of secondary mathematics teaching. *Educational and Psychological Measurement*, 73(5), 757–783.
- Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, 86(2), 127–137.
- Connell, J. P., & Wellborn, J. G. (1991). Competence, autonomy, and relatedness: A motivational analysis of self-system processes. In M. R. Gunnar, & L. A. Sroufe (Eds.), *Self processes and development: The Minnesota symposia on child psychology* (pp. 43–77). Hillsdale (N.J.): Erlbaum.

- Costello, E., Mustillo, S., Erkanli, A., Keeler, G., & Angold, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of General Psychiatry*, *60*(8), 837–844.
- Curran, P. J., Bollen, K. A., Chen, F., Paxton, P., & Kirby, J. B. (2003). Finite sampling properties of the point estimates and confidence intervals of the RMSEA. *Sociological Methods & Research*, *32*(2), 208–252.
- Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*(4), 227–268.
- Douglas, K. (2009). Sharpening our focus in measuring classroom instruction. *Educational Researcher*, *38*(7), 518–521.
- Emmer, E. T., & Stough, L. M. (2001). Classroom management: A critical part of educational psychology, with implications for teacher education. *Educational Psychologist*, *36*(2), 103–112.
- Evertson, C. M., & Weinstein, C. S. (2006). Classroom management as a field of inquiry. In C. M. Evertson, & C. S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 3–15). Mahwah, N.J: Lawrence Erlbaum.
- Finnish National Board of Education. (2014). National core curriculum for basic education 2014. National core curriculum for basic education intended for pupils in compulsory education. Retrieved from http://www.oph.fi/english/curricula_and_qualifications/basic_education
- Gettinger, M., & Walter, M. J. (2012). Classroom strategies to enhance academic engaged time. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 653–673). New York: Springer.

- Gibbs, R., & Poskitt, J. (2010). Student engagement in the middle years of schooling (years 7-10): A literature review. Retrieved from http://www.educationcounts.govt.nz/_data/assets/pdf_file/0010/74935/940_Student-Engagement-19052010.pdf
- Gitomer, D., Bell, C., Qi, Y., McCaffrey, D., Hamre, B. K., & Pianta, R. C. (2014). The instructional challenge in improving teaching quality: Lessons from a classroom observation protocol. *Teachers College Record, 116*(6), 1–32.
- Graue, E., Rauscher, E., & Sherfinski, M. (2009). The synergy of class size reduction and classroom quality. *The Elementary School Journal, 110*, 178–201.
- Gregory, A., Allen, J. P., Mikami, A. Y., Hafen, C. A., & Pianta, R. C. (2014). Effects of a professional development program on behavioral engagement of students in middle and high school: Increasing Behavioral Engagement. *Psychology in the Schools, 51*(2), 143–163.
- Guo, Y., Connor, C. M., Yang, Y., Roehrig, A. D., & Morrison, F. J. (2012). The effects of teacher qualifications, teacher self-efficacy and classroom practices on fifth graders' reading outcomes. *The Elementary School Journal, 113*, 3–24.
- Hafen, C. A., Hamre, B. K., Allen, J. P., Bell, C. A., Gitomer, D. H., & Pianta, R. C. (2015). Teaching through interactions in secondary school classrooms: Revisiting the factor structure and practical application of the classroom assessment scoring system-secondary. *The Journal of Early Adolescence, 35*(5–6), 651–680.
- Hagelskamp, C., Brackett, M., Rivers, S., & Salovey, P. (2013). Improving classroom quality with the RULER approach to social and emotional learning: Proximal and distal outcomes. *American Journal of Community Psychology, 51*(3), 530–543.

- 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., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., . . . Hamagami, A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *The Elementary School Journal*, 111(4), 461–487.
- Hamre, B. K., Pianta, R. C., Downer, J. T., & Mashburn, A. J. (2008). Teachers' perceptions of conflict with young students: Looking beyond problem behaviors. *Social Development*, 17 (1), 115–136.
- Hamre, B. K., Pianta, R. C., Mashburn, A. J. & Downer, J. T. (2007). Building a science of classrooms: Application of the CLASS framework in over 4,000 U.S. early childhood and elementary classrooms. Retrieved from

- Kane, T., Staiger, D., McCaffrey, D., Cantrell, S., Archer, J., Buhayar, S., & Parker, D. (2012). Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains (Technical report). Seattle, WA: Bill & Melinda Gates Foundation, Measures of Effective Teaching Project.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.
- La Paro, K., M., Pianta, R. C., & Stuhlman, M. W. (2004). The classroom assessment scoring system: Findings from the prekindergarten year. *Elementary School Journal, 104*(5), 409–409.
- Leyva, D., Weiland, C., Barata, M., Yoshikawa, H., Snow, C., & Treviño, E. (2015). Teacher–child interactions in Chile and their associations with prekindergarten outcomes. *Child Development, 86*, 781–799.
- Li Grining, C., Raver, C. C., Champion, K., Sardin, L., Metzger, M., & Jones, S. M. (2010). Understanding and improving classroom emotional climate and behavior management in the “real world”: The role of Head Start teachers’ psychosocial stressors. *Early Education and Development, 21*, 65–94.
- Malmberg, L., Hagger, H., Burn, K., Mutton, T., & Colls, H. (2010). Observed classroom quality during teacher education and two years of professional practice. *Journal of Educational Psychology, 102*(4), 916–932.
- Mashburn, A. J., Hamre, B. K., Downer, J. T., & Pianta, R. C. (2006). Teacher and classroom characteristics associated with teachers’ ratings of prekindergartners’ relationships and behaviors. *Journal of Psychoeducational Assessment, 24*, 367–380.

Maslach, C, Jackson, S.E. and Leiter, M.P. (1996). *Maslach Burnout Inventory Manual* (3rd edn.). Palo Alto, CA: Consulting psychologists Press.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, Electronic; 1*(1), 30–46.

McLean, L., & Connor, C. M. (2015). Depressive symptoms in third-grade teachers: Relations to classroom quality and student achievement. *Child Development, 86*(3), 945–954.

Moate, J. (2017). *Living between two educational systems*. In A. Raiker, & M. Rautiainen (Eds.), *Educating for Democracy in England and Finland: Principles and Culture* (pp. 17-26). Routledge.

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

National Institute of Child Health and Human Development Early Child Care Research Network. (2002). The relation of global first-grade classroom environment to structural classroom features and teacher and student behaviors. *Elementary School Journal, 102*, 367–387.

OECD. (2011), *Lessons from PISA for the United States, Strong Performers and Successful Reformers in Education*, OECD Publishing. <http://dx.doi.org/10.1787/9789264096660-en>

OECD. (2013). Programme for international student assessment (PISA). Retrieved from <http://www.oecd.org/pisa/>

OECD. (2014). Indicator D2: What is the student-teacher ratio and how big are classes? In *Education at a Glance 2014: OECD indicators*, OECD publishing. Retrieved from [http://www.oecd.org/edu/EAG2014-Indicator%20D2%20\(eng\).pdf](http://www.oecd.org/edu/EAG2014-Indicator%20D2%20(eng).pdf)

- Pakarinen, E., Kiuru, N., Lerkkanen, M.-K., Poikkeus, A.-M., Siekkinen, M., & Nurmi, J.-E. (2010). Classroom organization and teacher stress predict learning motivation in kindergarten children. *European Journal of Psychology of Education, 25*, 281–300.
- Pakarinen, E., Lerkkanen, M., Poikkeus, A., Kiuru, N., Siekkinen, M., Rasku-Puttonen, H., & Nurmi, J. (2010). A validation of the classroom assessment scoring system in Finnish kindergartens. *Early Education and Development, 21*(1), 95–124.
- Paris, S. G., & Paris, A. H. (2001). Classroom applications of research on self-regulated learning. *Educational Psychologist, 36*(2), 89–101.
- Parker, A., & Neuharth-Pritchett, S. (2006). Developmentally appropriate practice in kindergarten: Factors shaping teacher beliefs and practice. *Journal of Research in Childhood Education, 21*, 65–78.
- 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.
- Pianta, R. C., Hamre, B. K., & Mintz, S. L. (2010). *Classroom assessment scoring system—secondary (CLASS-S)*. Charlottesville, VA: University of Virginia.
- Pianta, R. C., Hamre, B. K., & Mintz, S. L. (2012). *The CLASS-secondary manual*. Charlottesville, VA: University of Virginia.
- Pressley, M., Dolezal, S. E., Raphael, L. M., Mohan, L., Roehrig, A. D., & Bogner, K. (2003). *Motivating primary-grade students*. New York: Guilford Press.

- Reyes, M. R., Brackett, M. A., Rivers, S. E., White, M., & Salovey, P. (2012). Classroom emotional climate, student engagement, and academic achievement. *Journal of Educational Psychology, 104*(3), 700–712.
- Rimm-Kaufman, S. E., & Sawyer, B. E. (2004). Primary-grade teachers' self-efficacy beliefs, attitudes towards teaching, and discipline and teaching practice priorities in relation to the responsive classroom approach. *The Elementary School Journal, 104*, 321–341.
- Sahlberg, P. (2015). *Finnish lessons 2.0: What can the world learn from educational change in Finland?* Forewords by Diane Ravitch and Andy Hargreaves. Afterwords by Ken Robinson (2nd ed.). New York: Teachers College Press.
- Sandilos, L. E., & DiPerna, J. C. (2014). Measuring quality in kindergarten classrooms: Structural analysis of the classroom assessment scoring system (CLASS K–3). *Early Education and Development, 25*(6), 894–914.
- Schacter, J., & Thum, Y. M. (2004). Paying for high- and low-quality teaching. *Economics of Education Review, 23*(4), 411–430.
- Schoorman, F., Mayer, R., & Davis, J. (2007). An integrative model of organizational trust: Past, present, and future. *Academy of Management Review, 32*(2), 344–354.
- Slot, P., Cadima, J., Salminen, J., Pastori, G., & Lerkkanen, M.-K. (2016). Multiple case study in seven European countries regarding culture-sensitive classroom quality assessment. Online publication: <http://ecec-care.org>.
- Soenens, B., Sierens, E., Vansteenkiste, M., Dochy, F., & Goossens, L. (2012). Psychologically controlling teaching: Examining outcomes, antecedents, and mediators. *Journal of Educational Psychology, 104*, 108–120.

- Stipek, D., & Byler, P. (1997). Early childhood education teachers: Do they practice what they preach? *Early Childhood Research Quarterly*, *12*(3), 305–325.
- Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, *17*(7), 783–805.
- Woolfolk, A. E., Rosoff, B., & Hoy, W. K. (1990). Teachers' sense of efficacy and their beliefs about managing students. *Teaching and Teacher Education*, *6*, 137–148.
- Yilmaz, K. (2011). The cognitive perspective on learning: Its theoretical underpinnings and implications for classroom practices. *Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *84*(5), 204–212.