

The factor structures of the Classroom Assessment Scoring System Pre-K and mature play observation tool in multi-ethnic Norwegian early childhood centers

Journal of Early Childhood Research

1–16

© The Author(s) 2023



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/1476718X231195708

journals.sagepub.com/home/ecr

Lisa Karlsen 
Veslemøy Rydland 
University of Oslo, Norway

Elisabet Solheim Buøen

Regional Centre for Child and Adolescent Mental Health, Eastern and Southern Norway, Norway

Deborah Lowe Vandell

University of California, USA

Ratib Lekhal

University of Oslo, Norway

BI Norwegian Business School, Norway

Abstract

The use of standardized assessment tools for the evaluation of quality in early childhood education and care (ECEC) is on the rise, yet a greater understanding of the applicability of these tools across contexts is still needed. This study investigates the factor structure of two assessment tools, the Classroom Assessment Scoring System Pre-K (CLASS) and Mature Play Observation Tool (MPOT) in a free-play focused context serving high numbers of children with diverse language backgrounds in Norway. The study also evaluates the extent to which these tools complement each other to create a more comprehensive understanding of children's experiences in ECEC in this context. Using confirmatory factor analyses, our results from a sample of 125 multi-ethnic ECEC groups in Norway show a good fit for the two-factor (i.e., adult- and child-focused) model proposed by the authors of MPOT. In line with previous research, the three-factor (i.e., emotional support, classroom organization, and support for learning) model of CLASS required post hoc modifications, resulting in a marginally acceptable model fit. Overall, our findings provide evidence that the original factor structures of these tools can be modeled in urban ECEC centers in Norway, and using these tools provides different insights into children's ECEC experiences.

Corresponding author:

Lisa Karlsen, Department of Education, University of Oslo, Sem Sælunds vei 7, Helga Engshus, Oslo 0317, Norway.

Email: l.m.karlsen@iped.uio.no

Keywords

classroom assessment scoring system pre-k, confirmatory factor analysis, free play, mature play observation tool, Norway

Introduction

Due to the well-documented benefits of high-quality early childhood education and care (ECEC) provisions (Burchinal et al., 2015; Melhuish et al., 2015; Sibley et al., 2015), there has been an increased focus on ECEC quality over the last few years (Li et al., 2020; OECD, 2022). The Nordic context is no exception to this trend (Furenes et al., 2021). Process quality plays an especially central role in research into ECEC quality, leading to the development of tools to assess the quality of teacher–child interactions. Quality assessment tools are frequently used, aiming to monitor or improve the quality of ECEC (CARE (European Early Childhood Education and Care), 2016). However, review studies focusing on widely used quality assessment tools find there is still a need “to capture more fine-grained information on children’s experiences, including peer experiences [and] child-level interactions . . .” (Cadima et al., 2010: 38), perhaps using several tools with different foci.

The Classroom Assessment Scoring System Pre-K (CLASS) is one tool used to measure process quality in ECEC that has gained widespread use by researchers as well as ECEC staff. CLASS has been used to measure teacher–child interactions and quality in ECEC in the United States (Tout et al., 2010) as well as other countries such as Finland (Pakarinen et al., 2010), Chile (Leyva et al., 2015) and China (Hu et al., 2016). However, additional research is needed to validate CLASS in ECEC settings in countries, such as Norway, focusing on child-initiated free play.

Although CLASS is frequently used to measure the quality of teacher–child interactions, it was not designed specifically for, or in, a free-play focused context. It also does not focus on play or peer interactions. Because peer interactions often take a central role in play (Ridgway et al., 2020), this means some important aspects of children’s experiences may be lost when using CLASS.

Both adult–child and peer interactions have been found to be especially important for child outcomes (e.g., Rao et al., 2021; Ribeiro et al., 2017), but there are limited observation tools available to assess these types of interactions during play (Germeroth et al., 2019). One measure, The Mature Play Observation Tool (MPOT), was intended to fill this gap, focusing on the maturity of children’s pretend play, peer interactions, and adult facilitation (Germeroth et al., 2019). In contrast to other quality assessment tools, MPOT was designed to assess children’s behaviors during pretend play and is based on theory and research supported in settings that value free play.

Like CLASS, the use of MPOT and its applicability to other contexts offers valuable information about the generalizability of what is considered high-quality provision across regions. The present study considers the applicability of CLASS and MPOT in Norwegian ECEC centers serving children from diverse language backgrounds. Although these tools were designed in the United States, they reflect the Nordic ECEC pedagogy. They also offer systematic methods of assessing process quality and play—something currently missing from tools designed in this region. The present study aims to answer recent calls for more cross-cultural investigations of quality assessment tools (Hu et al., 2016; Li et al., 2020), as well as examine how much CLASS and MPOT overlap in their assessment of ECEC quality. Determining the extent to which these tools offer different perspectives on ECEC quality may add a more differentiated view of quality.

Table 1. CLASS domains and dimensions.

Emotional support	Description	Classroom organization	Description	Instructional support	Description
Positive climate	Degree of emotional connection and respect in the group.	Behavior management	The extent to which clear behavioral expectations are provided and reinforced.	Concept development	Degree to which staff promote thinking and reflection.
Negative climate	Frequency and intensity of expressed negativity in the group.	Productivity	The extent to which staff manage and facilitate instructional time and routines.	Quality of feedback	Content and extent to which children receive feedback.
Staff sensitivity	Extent to which staff notice and follow up on children’s cues.	Instructional learning formats	The extent to which staff maximize children’s interest and engagement.	Language modeling	Quality and amount of language stimulation and facilitation.
Regard for child perspectives	Extent to which staff emphasize and encourage children’s interests and autonomy.				

CLASS

CLASS is an assessment tool used to rate instructional quality, emotional support, and classroom organization in ECEC (Pianta et al., 2008). It is intended for use in ECEC groups serving children between 36 and 60 months. Observers score 10 dimensions (see *Table 1*) using seven-point scales with 1 and 2 representing low scores, 3 to 5 representing medium scores and 6 and 7 representing high scores.

CLASS was developed in the United States in collaboration with various experts in the field, and is based on extensive literature review, practice observations, effective teaching practices, and piloting (Pianta et al., 2008). Studies assessing the factor structure of CLASS in the United States conclude that a three-factor model consisting of emotional support, classroom organization and instructional support fits the tool best, although not all indices show adequate fit depending on threshold values used (see Hamre et al., 2013). The tool has been used extensively outside of the United States and validation studies of its applicability in other countries, although limited, do exist. European studies support the three-factor structure of CLASS though various adjustments to the model were needed to reach a good model fit (e.g., Finland: Pakarinen et al., 2010; Germany: Stuck et al., 2016). In Chile, Leyva et al.’s (2015) findings support construct and predictive validity of CLASS. Hu et al. (2016) report similar findings in the Chinese context. Few of these studies outside of the United States used CLASS in free-play focused centers serving linguistically diverse families.

The Mature Play Observation Tool (MPOT)

MPOT (Germeroth et al., 2019) was developed by a research team in the United States, based on research and theory on pretend play. This observation tool is used to assess pretend play maturity

Table 2. MPOT domains and dimensions.

Child-focused	Description	Adult-focused	Description
Child-created props	Number of props created by children or instances of creative use of ready-made props (including behaviors and gestures) in the play.	Center management	Extent to which there is a system for managing and facilitating children's play.
Child metaplay	Children's ability to discuss play scenarios and their role.	Make-believe play time	The amount of time dedicated to uninterrupted play.
Play interactions	Level of children's collaboration in play.	Teacher intervention	The degree and content of staff involvement in play.
Children's role- playing	Instances of sequenced actions related to a role.		
Child role speech and communication during play	Children's use of theme-related words or gestures to communicate in their role.		

and adult support for pretend play, not specifically ECEC quality. It is intended for use with children between 36 and 60 months at group level. The tool assesses adult-child and peer interactions on eight dimensions, divided between two domains; adult- and child-focused (see *Table 2*). These domains focus on children's use of props and language, creativity and real-world experiences, as well as how pretend play is supported, group organization and play interactions. Play sequences are rated on a four-point scale indicating either immature (score 1 and 2) or mature (score 3 and 4) play. Development of the tool involved multiple trials and piloting, as well as careful consideration about dimensions included (Germeroth et al., 2019).

Previous studies have not looked specifically at the factor structure of MPOT, instead focusing on other aspects of validity in the United States (see Germeroth et al., 2019). Our study expands on Germeroth et al.'s (2019) by investigating the factor structure of MPOT in Norwegian ECEC serving linguistically diverse children.

ECEC in Norway

The availability of universal ECEC from age one is a political priority in Norway (Ministry of Education and Research [MER], 2020a). ECEC is heavily subsidized as part of the national welfare system (MER, 2020b). All children have access to ECEC irrespective of socioeconomic background, as a way to ensure equal opportunities for all families (MER, 2020b; Official Norwegian Records [NOU], 2009, 2010, 2012). To achieve this, additional subsidies are provided to families needing extra support. All children, regardless of subsidy status, attend the same ECEC provision. ECEC settings have an attendance rate of 93.4% of 1- to 5-year-olds (Statistics Norway, 2022), with 20% considered minority language speakers (Norwegian Directorate for Education, 2022). In Oslo, an average of 30% of children are placed within this category, and up to 75% in some city districts (Norwegian Directorate for Education, 2022). Structural quality standards are regulated by law and supervised by the Ministry of Education, and practice is guided by a common framework plan (Kindergarten Act, 2005; MER, 2017). Play, as a general term, is placed at the core of this framework plan and ECEC's daily activities. Children spend most of their day in child-led free play, with staff prioritizing this over planned adult-led activities (Lekhal et al., 2013). Daily

activities are organized in an informal way, with children moving freely between play themes and activities, often without a staff member nearby (Karlsen & Lekhal, 2019; Kleppe, 2017; Sandseter et al., 2020). With its universally accessible provision and heavy focus on play, Norwegian ECEC offers a setting for the exploration of the applicability of both CLASS and MPOT in a free-play focused context.

The present study

This study was conducted in ECEC centers across different city districts in Norway's capital, Oslo. The study's aim was two-fold—to examine the factor structure of CLASS and MPOT and explore the use of these tools together as assessments of process quality in a free-play focused context. To do this we aimed to answer two questions: (1) Are the original factor structures of CLASS and MPOT supported in the urban ECEC context in Norway? and (2) Is there an association between CLASS and MPOT domains?

Methods

Participants and procedure

Data for this study were collected in ECEC centers participating in The Oslo Early Education Study (OEES), an ongoing longitudinal cluster randomized control trial intervention conducted in 214 ECEC groups across five city districts in the municipality of Oslo. OEES is a researcher–sector intervention intended to train multi-ethnic ECEC centers in utilizing their potential to support children's language development. All center managers in five multi-ethnic city districts in the municipality of Oslo were invited to participate in OEES. Groups not able to recruit a minimum of two children and two staff members were excluded. A total of 56 centers and 214 groups were included in the first round of data collection, and the study was approved by the Norwegian Centre for Research Data.

The present study uses data from the first round of data collection, prior to randomization in 2021. Of the 214 groups participating in OEES, 125 were observed with MPOT. One of these groups was not able to be observed with CLASS, resulting in 124 CLASS observations. All 125 groups were included in the present study. The remaining 89 groups catered exclusively to children under 36 months and were not observed with these tools. The included groups consisted of 7 to 24 children ($M=17.5$; $SD=3.4$), up to 60 months. Based on data provided by the Norwegian Directorate for Education (2022), between 37 and 75% of children in ECEC centers within the included city districts are considered minority language speakers. Consent forms submitted by parents in this study indicate that an average of 68% ($SD=27$) of children were from families with Norwegian as an additional language, with over 60 different languages represented by children and staff.

Following several information meetings and consent from parents and staff, adult–child and peer interactions were observed live by 18 certified CLASS and/or MPOT research assistants. Observations using CLASS and MPOT were completed during two observation sessions usually by two observers on two different days. As a means of calibrating and ensuring reliable coding practice, an expert observer accompanied certified observers during their first observation.

Measures

CLASS observations. Observations of adult–child interactions in different situations were observed using CLASS between 09:00 and 12:00. ECEC groups were observed and scored for four 15-minute cycles, before an average score for the group, on each of the 10 dimensions (see *Table 1* for these dimensions), was calculated. This deviated slightly from the observation protocol prescribed by the tool developers. Due to the informal organization of Norwegian ECEC, a specified activity was observed for two of the four observation cycles. These specified activities were shared reading and water play.

MPOT observations. The maturity of pretend play in the groups was assessed during free play using MPOT. Observations were conducted during two 30-minute cycles, which were then averaged. This deviated slightly from the observation protocol prescribed by the developers of MPOT, where 60 minutes of uninterrupted observation is advised. Due to the informal organization of play activities in Norwegian ECEC, prior to observations, observers chose one child for each of the two cycles to follow if the group dissolved. Criteria for choosing each child were: a child with Norwegian as an additional language, between 3 and 4 years, already in play. In order to ensure gender diversity, a girl was followed for the first observation and a boy for the second. This did not mean that children were observed and scored individually, but a child needed to be followed if the play group changed during the observation. Observations were still conducted at the group level as prescribed by Germeroth et al. (2019).

Seven of the eight MPOT dimensions (see *Table 2*) were used in the present study. Since free play takes up the majority of the day in Norwegian ECEC, the “planned play time” dimension was not rated because the original dimension required an hour of uninterrupted play time, something most, if not all, Norwegian centers exceed. In addition, the “center management” dimension was adjusted to better reflect the Norwegian context, where little planning or management is usually involved in free play. Norwegian ECEC centers generally have few restrictions as to what or where children play during free play. This means there is little, if any, use of visual aids or assigned roles. ECEC staff generally only organize play groups when necessary and not in advance of play (Alvestad et al., 2019). In OEES, the center management dimension therefore no longer requires proof of management of play through with visual aids, as needed in the original dimension, to receive a mature score (score 3 or 4). Instead, other evidence of facilitation of play could be demonstrated. Examples include preparing play centers, lack of wandering by children, and practitioner awareness of children not in play—actions not specifically emphasized in the original dimension.

Data analytic plan

To examine the factor structure of the tools, we used a confirmatory factor analysis (CFA). The CFA was based on ECEC groups’ mean scores on each dimension of CLASS and MPOT. All analyses were conducted with Stata 17. CLASS’ three-factor model and MPOT’s two-factor model were evaluated using four commonly reported indices (Kline, 2005): comparative fit index (CFI), root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR) and Tucker-Lewis index (TLI). For CFI and TLI, values ≥ 0.95 commonly indicate a good model fit (Hu and Bentler, 1999). However, this cut-off has been criticized for its conservativeness (see, Marsh et al., 2004). As recommended by Browne and Cudeck (1993), a threshold of >0.90 was used in the present study to indicate a good model fit, while 0.80–0.90 indicated

an acceptable fit. For RMSEA and SRMR, a good model fit was indicated as ≤ 0.05 , and acceptable as 0.06–0.10 (Hu and Bentler, 1999; MacCallum et al., 1996; Schermelleh-Engel et al., 2003). To evaluate factor loadings, we used R^2 estimates ≥ 0.25 and standardized factor loadings ≥ 0.40 . Factor loadings of 0.32 were rated as poor, 0.45 as fair, 0.55 as good, 0.63 as very good, and ≥ 0.71 as excellent (Comrey and Lee, 1992).

Further, we investigated the extent to which the tools measured overlapping aspects of quality. To do this, we assessed correlation patterns between the three domains of CLASS (i.e., emotional support, classroom organization and instructional support domains) with the two domains of MPOT (i.e., adult- and child-focused domains) using a correlation analysis.

Findings

Three-factor structure of CLASS

Examining the three-factor model of CLASS, initial findings indicated an acceptable to poor fit (see *Table 3*). As previous studies that have made similar findings have done (e.g., Pakarinen et al., 2010; Slot et al., 2018), we adjusted the original model by allowing some error variances to correlate based on modification indices (MI) provided by Stata. We first allowed for the measurement error of negative climate and behavior management to be correlated (MI=17.02, standard EPC=0.39). These modifications did not result in a significantly improved model (CFI=0.82; RMSEA=0.13; SRMR=0.1; TLI=0.84). We therefore ran the modification indices again and allowed productivity and instructional learning formats to also correlate (MI=21.33, standard EPC=2.15). This resulted in a somewhat improved and overall acceptable fit, although RMSEA remained borderline poor (see *Table 3*). Because values based on a third adjustment did not reach more than MI 13 (lower than the initial MI value of 17.02), we decided not to proceed with adjusting the model further.

For the sake of thoroughness, we also tested our initial model without the negative climate and language modeling dimensions because they violated the cutoff value for R^2 estimates. These dimensions were removed one at a time, and the model was analyzed individually. The model was not significantly improved by these changes so both dimensions were retained in the final model. The first two modification indices suggested by Stata were the only changes made to the final model.

Final parameter estimates based on the modifications showed standardized factor loadings ranging from 0.44 to 0.95, and R^2 values between 0.19 and 0.90 (see *Table 4* for full description). Initial and final models are presented in *Figures 1* and *2*.

We also examined descriptive statistics of the three-factor model of CLASS. Means on the three domains varied, with the emotional support domain showing the highest values ($M=6.06$), and the instructional support domain showing the lowest ($M=2.38$). More descriptive information can be found in *Table 5*.

Table 3. Initial and final fit indices for the three-factor model of CLASS.

Model of CLASS	CFI	RMSEA	SRMR	TLI
Initial three-factor model	0.86**	0.15	0.10**	0.81**
Final modified three-factor model	0.93*	0.11	0.07**	0.89**

*Good fit CFI = > 0.90 RMSEA = ≤ 0.05 SRMR = ≤ 0.05 TLI = > 0.90 .

**Acceptable fit CFI = 0.80–0.90 RMSEA = 0.06– 0.10 SRMR = ≤ 0.05 TLI = 0.80–0.90.

Table 4. Final standardized factor loadings and R^2 for three-factor model of CLASS.

CLASS dimension	Standardized factor loading (≥ 0.40)	R^2 (≥ 0.25)
Positive climate	0.81	0.65
Negative climate	0.44	0.19
Teacher sensitivity	0.95	0.90
Regard for child perspectives	0.74	0.54
Behavior management	0.77	0.60
Productivity	0.58	0.34
Instructional learning formats	0.73	0.54
Concept development	0.72	0.52
Quality of feedback	0.93	0.86
Language modeling	0.44	0.19

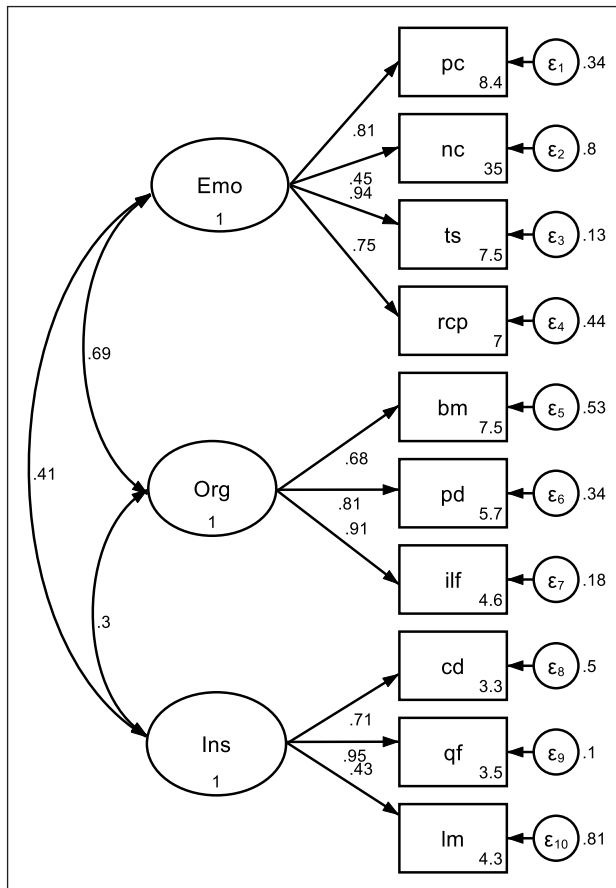


Figure 1. Initial three-factor model of CLASS.

pc: positive climate; nc: negative climate; ts: teacher sensitivity; rcp: regard for child perspectives; bm: behavior management; pd: productivity; ilf: instructional learning formats; cd: concept development; qf: quality of feedback; lm: language modeling; emo: emotional support; org: classroom organization; ins: instructional support.

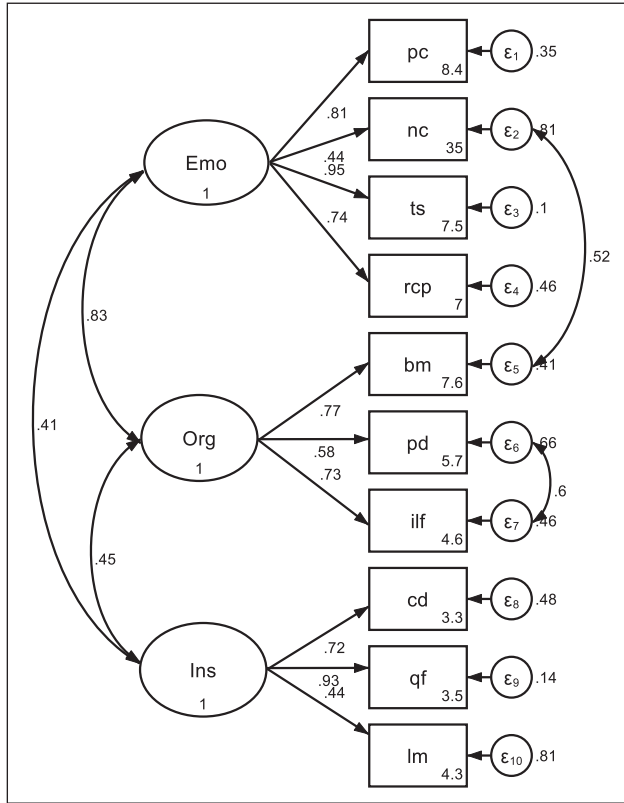


Figure 2. Adjusted three-factor model of CLASS.

pc: positive climate; nc: negative climate; ts: teacher sensitivity; rcp: regard for child perspectives; bm: behavior management; pd: productivity; ilf: instructional learning formats; cd: concept development; qf: quality of feedback; lm: language modeling; emo: emotional support; org: classroom organization; ins: instructional support.

Table 5. CLASS descriptive statistics.

Domains and dimensions	Mean	Range
Emotional support	6.06 (0.53)	4.3–6.9
Positive climate	6.06 (0.72)	3.8–7
Negative climate (reversed)	6.89 (0.20)	6.0–7
Teacher sensitivity	5.74 (0.76)	3.5–7
Regard for child perspectives	5.56 (0.80)	3.3–7
Classroom organization	5.32 (0.81)	2.6–6.8
Behavior management	5.71 (0.76)	3.3–7
Productivity	5.59 (0.99)	2.5–7
Instructional learning formats	4.67 (1.02)	1.8–6.8
Instructional support	2.38 (0.51)	1.3–3.8
Concept development	1.76 (0.53)	1.0–3.8
Quality feedback	2.25 (0.64)	1.0–4.3
Language modeling	3.14 (0.74)	1.0–5

Table 6. Fit indices for the two-factor model of MPOT.

Model of MPOT	CFI	RMSEA	SRMR	TLI
Two-factor model	0.99*	0.05*	0.04*	0.98*

*Good fit CFI = >0.90 RMSEA = ≤0.05 SRMR = ≤0.05 TLI = >0.90.

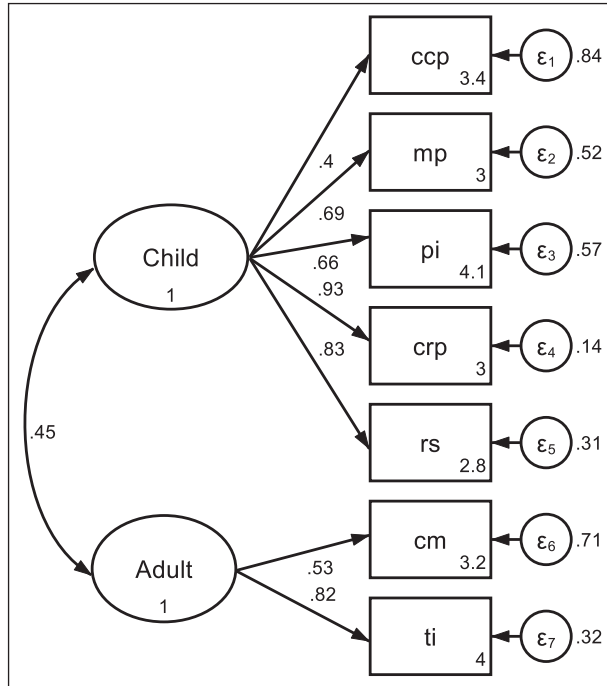


Figure 3. Two-factor model of MPOT.

ccp: child-created props; mp: metaplay; pi: play interactions; crp: child role play; rs: child role speech; cm: center management; ti: teacher intervention; Child: child-focused; Adult: adult-focused.

Two-factor structure of MPOT

Examining the two-factor model of MPOT proposed by Germeroth et al. (2019), the CFA indicated a good model fit on all indices as shown in Table 6, and all factor loadings were significant (see Figure 3). Parameter estimates revealed standardized factor loadings ranging from 0.40 to 0.93, and R^2 values between 0.16 and 0.86 (see Table 7). Children’s role-playing had the highest value on these estimates (0.93 and 0.86), while child-created props had the lowest (0.40 and 0.16). Because the child-created props dimension violated the cutoff value for R^2 estimates, the model was run again without this dimension. The model fit was not significantly improved by this adjustment nor was it conceptually meaningful to remove this dimension so no further exploration of this was undertaken and the original model was retained.

We also examined descriptive statistics of the two-factor model of MPOT. Findings revealed that mean scores tended toward the lower end of the scale, with the adult-focused domain producing slightly higher results (child-focused: $M=1.68$; adult-focused: $M=2.15$). See Table 8 for more information about the descriptive statistics.

Table 7. Standardized factor loadings and R^2 for the two-factor model of MPOT.

MPOT dimension	Standardized factor loading (≥ 0.40)	R^2 (≥ 0.25)
Child-created props	0.40	0.16
Metaplay	0.69	0.48
Play interactions	0.66	0.43
Children’s role playing	0.93	0.86
Role speech	0.83	0.69
Center management	0.53	0.29
Teacher intervention	0.82	0.68

Table 8. MPOT descriptive statistics.

Domains and dimensions	Mean (SD)	Range
Child-focused	1.68 (0.40)	1–2.8
Child-created props	1.54 (0.46)	1–3
Child metaplay	1.54 (0.51)	1–3
Play interaction	2.10 (5.1)	1–3.5
Children’s role-play	1.50 (0.49)	1–3
Child role speech and communication	1.72 (0.61)	1–3.5
Adult-focused	2.15 (0.52)	1–3.5
Center management	2.25 (0.70)	1–4
Teacher intervention	2.06 (0.52)	1–4

Correlation patterns

Correlation analysis showed few significant correlations between CLASS and MPOT domains with values between -0.01 and 0.16 . However, albeit weak, significant positive correlation between adult- and child-focused domains of MPOT and the classroom organization domain of CLASS were found (0.14 and 0.16 respectively). We also found a weak positive correlation ($p < 0.2$) between CLASS instructional support domain and MPOT’s child-focused domain as shown in *Table 9*.

Table 9. Correlations between CLASS and MPOT domains.

Domains	1	2	3	4	5
1. Adult-focused domain (MPOT)					
2. Child-focused domain (MPOT)	0.31				
3. Instructional support domain (CLASS)	0.05	0.12			
4. Classroom organization domain (CLASS)	0.14 [^]	0.16 [*]	0.29		
5. Emotional support domain (CLASS)	-0.02	-0.01	0.34	0.53	

* $p < 0.1$; [^] $p < 0.15$.

Discussion

This study used data from 125 multi-ethnic Norwegian ECEC centers to explore the factor structure of CLASS and MPOT using a CFA. The aim was to investigate whether the original factor structures of these tools would be supported in a free-play focused context. Correlation patterns between the tools were also analyzed to establish if there was an association between CLASS and MPOT domains.

Support for the factor structure of CLASS

To evaluate the factor structure of CLASS, and answer our first research question, we used results of the CFA based on CLASS data. Similar to previous studies (for overview see Li et al., 2020), initial results of the three-factor model indicated acceptable fit indices on all but one indicator. The fact that our initial findings needed adjustments, reflecting what others have found, is notable. Our study adds to previous findings by including a larger sample size than most as well as a multi-ethnic sample from outside of the United States while still producing similar results. Our multi-ethnic sample also adds to previous discussions about the applicability of CLASS with these samples and planned changes related to this (see Teachstone.com).

Our findings of lower factor loadings and R^2 values on the negative climate and language modeling dimensions compared with some previous studies are also noteworthy. Although it is not uncommon to find low factor loadings on the negative climate dimension (see, for example, Pakarinen et al., 2010; Stuck et al., 2016), factor loadings on this dimension did not violate the cutoff value (≥ 0.40) in our model, nor did removing this dimension improve the model. The negative climate dimension assesses the level of negativity in ECEC groups by focusing on, for example, negative affect and punitive control. Previous studies have found this dimension shows, in general, little variation across groups (La Paro et al., 2004; Pakarinen et al., 2010), and based on our findings of low factor loadings, our data supports this.

It is not as typical to find low factor loadings on the language modeling dimension, however. This dimension focuses on frequency, sophistication and quality of language stimulation through, for example, back-and-forth conversations, open-ended questions and language extension. Because diversity in language was not only represented by children but also staff, as others have emphasized, our findings here highlight the need for further research with diverse samples and a review of the tool's applicability in multi-ethnic settings (Barnes-Najor et al., 2020; Bichay-Awadalla and Bulotsky-Shearer, 2021).

Although we conclude that the three-factor model of CLASS was only marginally acceptable due to questionable RMSEA values, cut-off values are not universally agreed upon (e.g., Marsh et al., 2004). Nonetheless, one explanation for the slightly higher than desirable RMSEA value may be due to the fact that RMSEA measures absolute fit, not taking into consideration the complexity of the model (van Trijp et al., 2021). Interestingly, Li et al. (2020) report on an unpublished study of the three-factor structure of CLASS in the United States in which RMSEA values did not meet the acceptable criteria (0.13), indicating our results may not be so unusual.

Support for the factor structure of MPOT

To evaluate the factor structure of MPOT, and further answer our first research question, we used the results of the CFA based on MPOT data. These results supported the two-factor model of MPOT proposed by Germeroth et al. (2019) with good fit indices and factor loadings on all but one dimension. These findings indicate that our data fit the original model well despite slight deviation from the prescribed observation protocol and a different context. In addition, descriptive statistics were not dissimilar to those found in the United States (see Germeroth et al., 2019).

Low factor loadings related to child-created props, however, may be due to our particular sample. This dimension emphasizes props made by children or creative use of existing props. It is uncommon to find many child-made props in Norwegian ECEC, which may have led to only slight variation on this dimension between groups and having little influence on the model, thus producing low factor loadings (Ximénez, 2009). Previous studies have found that although Norwegian ECEC settings have a relatively large selection, materials are often stored out of children's reach (Rove Nilsen, 2021). If this is true for our sample, little variation in children's creative use of props

may have been seen due to only a selection of props being available for spontaneous use. Another aspect of this is the type of props accessible to children. If adults decide which props should be available at child level, children may be restricted to ‘ready-made’ or ‘appropriate’ props – imposing unimaginative views on prop use. With such restrictions, little variation in scores is likely. Nonetheless, keeping this dimension in the model is still conceptually and theoretically meaningful even in the Norwegian context, but further research into variation of scores on this dimension across groups may be beneficial to better understand the low factor loadings on this dimension.

Associations between the tools

To assess associations between CLASS and MPOT domains, we used results of the correlation analysis to answer our second research question. Our findings point to a weak correlation between CLASS and MPOT domains, indicating that dimensions assessed by the two tools are unrelated and that the use of several tools may better capture children’s experiences in ECEC. This is expected since, in general, CLASS focuses specifically on adult facilitation of support for learning and development, while MPOT focuses more heavily on child-initiated actions – therefore assessing two different types of interactions. Using CLASS alone may lead to a less refined assessment of the quality of peer interactions or child-initiated activities during play.

Due to weak yet significant correlations found between MPOT domains and organization and instructional support rated by CLASS, our findings also indicate that these domains are somewhat related. Focusing on how this finding indicates the complementary nature of these tools, improvements on CLASS scores could influence scores on MPOT domains in other areas. Moreover, the CLASS emotional support domain was not significantly correlated with either of the MPOT domains, indicating high emotional support in ECEC may not necessarily reveal high-quality support for peer interactions and facilitation of play. Further, high support for learning and organization may not be enough to indicate adequate support for peer interactions or play as measured by MPOT. Although MPOT was not designed to assess ECEC quality per se, because play takes up such a large part of the day in Nordic ECEC culture, assessment of pretend play maturity in this context may indicate important aspects of quality. By evaluating the extent to which CLASS and MPOT overlap in their assessment of ECEC quality, a more comprehensive exploration of quality may be possible. Overall, our findings suggest these tools may assess different aspects of children’s experiences in ECEC, confirming their use together as a way to assess quality in free-play focused contexts. This also indicates that using only one tool exclusively may result in insufficient information about the support young children receive in free-play focused settings.

Limitations and further studies

Because data was only collected in one city, further research assessing variation in scores across regions may be beneficial to better understand the use of these tools in other free-play focused contexts. Additionally, although care was taken to ensure centers represented the city districts they were situated within, they are part of an intervention study.

Focusing on our findings related to CLASS, our choice of predetermined situations for observations can be viewed as a limitation due to possible variation between scores. We argue, however, that this is also a strength as it enables comparison across situations. Although prescribed activities are not given by the CLASS manual, results did not reveal large differences from previous findings.

Finally, although it is common for researchers to make use of modification indices, research into their value has shown varying results (see Whittaker, 2012 for discussion). Nonetheless, it is not

unusual to adjust the CLASS three-factor model in validation studies both in the United States and internationally (Li et al., 2020). In addition to what others have found about the applicability of CLASS across contexts, our findings add to calls for the tool to be adapted to be more inclusive for multi-ethnic settings (Barnes-Najor et al., 2020; Bichay-Awadalla and Bulotsky-Shearer, 2021).

Conclusion

Our findings of support for the original two-factor model of MPOT and partial support for the three-factor model of CLASS add to limited validation studies in free-play focused contexts, while encouraging further research across contexts. Because ECEC is becoming increasingly multi-ethnic in many regions, including free-play focused contexts, our findings about the applicability of tools in these types of settings are important. This is further emphasized by the current review of CLASS underway (see Teachstone.com) and recent calls for research with these types of samples (Barnes-Najor et al., 2020; Bichay-Awadalla and Bulotsky-Shearer, 2021). Since correlation patterns revealed that CLASS and MPOT measure different aspects of quality, using these tools to complement each other when assessing quality in free-play focused contexts may provide a richer, more holistic representation of quality in these settings. Finally, our findings offer important considerations when using standardized tools across contexts.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was provided by The Research Council of Norway [grant number 320258].

ORCID iDs

Lisa Karlsen  <https://orcid.org/0000-0001-7606-9723>

Veslemøy Rydland  <https://orcid.org/0000-0002-2036-3920>

References

- Alvestad M, Gjems L, Myrvang E, et al. (2019) *Kvalitet i barnehagen* [Quality in Kindergarten]. Stavanger, Norway: Universitetet i Stavanger. <https://utdanningsforskning.no>
- Barnes-Najor JV, Thompson NL, Cameron AF, et al. (2020) Cultural and practice perspectives on the classroom assessment scoring system: Voices from American Indian and Alaska Native head start programs. *Journal of Research in Childhood Education* 35: 162–183.
- Bichay-Awadalla K and Bulotsky-Shearer RJ (2021) Examining the factor structure of the classroom assessment scoring system toddler (CLASS-T) in early head start and subsidized child care classrooms. *Early Education and Development* 33: 309–325.
- Browne MW and Cudeck R (1993) Alternative ways of assessing model fit. In: Bollen KA and Long JS (eds) *Testing Structural Equation Models* (pp. 136–162). Newbury Park, CA: Sage.
- Burchinal M, Magnuson K, Powell D, et al. (2015) Early childcare and education. In: Lerner RM, Bornstein M and Leventhal T (eds) *Handbook of Child Psychology and Developmental Science*, vol.4, 7th edn. Hoboken, NJ: Wiley, pp. 223–267.
- Cadima J, Leal T and Burchinal M (2010) The quality of teacher–student interactions: Associations with first graders’ academic and behavioral outcomes. *Journal of School Psychology* 48(6): 457–482.
- CARE (European Early Childhood Education and Care) (2016) D4.3: Overview of quality monitoring systems and results of moderator analysis. Available at: <https://ecec-care.org/resources/publications> (accessed 1 August 2022).
- Comrey AL and Lee HB (1992) *A First Course in Factor Analysis*, 2nd edn. Hillsdale, NJ: Lawrence Erlbaum Associates.

- Furenes MI, Reikerås E, Moser T, et al. (2021) *Trender i empirisk barnehageforskning i de skandinaviske landene 2006-2019- en forskningskartlegging* [Trends in Scandinavian Kindergarten Research from 2006-2019]. Available at: www.uis.no/nb/forskning/ (accessed 10 February 2022).
- Germeroth C, Bodrova E, Day-Hess C, et al. (2019) Play it high, play it low: Examining the reliability and validity of a new observation tool to measure children's make-believe play. *American Journal of Play* 11(2): 183–221.
- Hamre BK, Pianta RC, Downer JT, et al. (2013) Teaching through Interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *Elementary School Journal* 113(4): 461–487.
- Hu BY, Dieker L, Yang Y, et al. (2016) The quality of classroom experiences in Chinese kindergarten classrooms across settings and learning activities: Implications for teacher preparation. *Teaching and Teacher Education* 57: 39–50.
- Hu LT and Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* 6(1): 1–55.
- Karlsen L and Lekhal R (2019) Practitioner involvement and support in children's learning during free play in two Norwegian kindergartens. *Journal of Early Childhood Research* 17(3): 233–246.
- Kindergarten Act (2005) *Barnehageloven*. Norway: Ministry of Education.
- Kleppe R (2017) Characteristics of staff-child interaction in 1–3-year-olds' risky play in early childhood education and care. *Early Child Development and Care* 188: 1487–1501.
- Kline RB (2005) *Principles and Practice of Structural Equation Modeling*. New York, NY: Guilford Press.
- La Paro KM, Pianta RC and Stuhlman M (2004) The classroom assessment scoring system: Findings from the Pre-Kindergarten year. *Elementary School Journal* 104(5): 409–426.
- Lekhal R, Vartun M, Gustavson K, et al. (2013). *Den norske mor og barn-undersøkelsen. Variasjon i barnehagekvalitet. Beskrivelser fra første datainnsamling fra barnehagene* [The Norwegian mother and child survey. Variations in kindergarten quality]. Folkehelseinstituttet. Available at: www.fhi.no (accessed 5 October 2022).
- Leyva D, Weiland C, Barata M, et al. (2015) Teacher-child interactions in Chile and their associations with pre-kindergarten outcomes. *Child Development* 86(3): 781–799.
- Li H, Liu J and Hunter CV (2020) A meta-analysis of the factor structure of the classroom assessment scoring system (CLASS). *Journal of Experimental Education* 88(2): 265–287.
- MacCallum RC, Browne MW and Sugawara HM (1996) Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods* 1(2): 130–149.
- Marsh HW, Hau KT and Wen Z (2004) In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal* 11(3): 320–341.
- Melhuish EC, Ereky-Stevens K, Petrogiannis K, et al. (2015) A review of research on the effects of early childhood Education and Care (ECEC) upon child development. In curriculum quality analysis and impact review of European early childhood education and care (ECEC). Available at: <https://ecec-care.org/resources/publications> (accessed 1 September 2022).
- Ministry of Education and Research (MER) (2017) Framework plan for the content and tasks of kindergartens. Available at: www.udir.no (accessed 7 July 2022).
- Ministry of Education and Research (MER) (2020a) Early Childhood Education and Care. Available at: <https://www.regjeringen.no/no/tema/familie-og-barn/barnehager/id1029/> (accessed 23 August 2023).
- Ministry of Education and Research (MER) (2020b) Finansiering av barnehager [Funding of kindergartens]. Available at: <https://www.regjeringen.no/no/tema/familie-og-barn/barnehager/id1029/> (accessed 23 August 2023).
- Norwegian Directorate for Education (2022) Fakta om barnehager 2021 [Facts about kindergarten 2021]. Available at: www.udir.no (accessed 1 September 2022).
- OECD (2022) Quality assurance and improvement in the early education and care sector. *OECD Education Policy Perspectives*, No. 55. Paris: OECD Publishing.
- Official Norwegian Records (NOU) (2009) *Fordelingsutvalget* [The Allocation Commission]. Oslo: Ministry of Finance.

- Official Norwegian Records (NOU) (2010) *Med forskertrang og lekelyst: Systematisk pedagogisk tilbud til alle førskolebarn* [Systematic Pedagogical Offer to All Preschool Age Children]. Oslo: Ministry of Education and Research.
- Official Norwegian Records (NOU) (2012) *Til barnas beste: Ny lovgivning for barnehagene* [In the Best Interest of Children: New Legislation for Kindergartens]. Oslo: Ministry of Education and Research.
- Pakarinen E, Lerkkanen M-K, Poikkeus A-M, et al. (2010) A Validation of the classroom assessment scoring system in Finnish kindergartens. *Early Education and Development* 21(1): 95–124.
- Pianta RC, La Paro KM and Hamre BK (2008) *Classroom Assessment Scoring System (CLASS) Pre-K Manual*. Baltimore, Maryland: Brookes, Publishing.
- Rao Z, Barker B, O'Farrelly C, et al. (2021) Maternal anxiety and depression and their associations with mother-child pretend play: A longitudinal observational study. *BMC Psychology* 9(1): 1–11.
- Ribeiro LA, Zachrisson HD and Dearing E (2017) Peer effects on the development of language skills in Norwegian childcare centers. *Early Childhood Research Quarterly* 41: 1–12.
- Ridgway A, Quiñones G and Li L (2020) *Peer Play and Relationships in Early Childhood*, vol.30. Cham: Springer International Publishing.
- Rove Nilsen T (2021) Pedagogical intentions or practical considerations when facilitating children's play? Teachers' beliefs about the availability of play materials in the indoor ECEC environment. *International Journal of Child Care and Education Policy* 15: 1.
- Sandseter EBH, Storli R and Sando OJ (2020) The dynamic relationship between outdoor environments and children's play. *Education* 3–13: 1–14.
- Schermelleh-Engel K, Moosbrugger H and Müller H (2003) Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online* 8: 23–74.
- Sibley E, Dearing E, Toppelberg CO, et al. (2015) Do increased availability and reduced cost of early childhood care and education narrow social inequality gaps in utilization? Evidence from Norway. *International Journal of Child Care and Education Policy* 9(1): 1.
- Slot PL, Bleses D, Justice LM, et al. (2018) Structural and process quality of Danish preschools: Direct and indirect associations with children's growth in language and preliteracy skills. *Early Education and Development* 29(4): 581–602.
- Statistics Norway (2022) *Barnehager* [Kindergartens]. Oslo, Norway: Statistisk Sentralbyrå. www.ssb.no
- Stuck A, Kammermeyer G and Roux S (2016) The reliability and structure of the classroom assessment scoring system in German pre-schools. *European Early Childhood Education Research Journal* 24(6): 873–894.
- Tout K, Starr R, Soli M, et al. (2010) *Compendium of Quality Rating Systems and Evaluations*. Available at: <https://files.eric.ed.gov/fulltext/ED566258.pdf> (accessed 23 August 2023).
- van Trijp CPJ, Lekhal R, Drugli MB, et al. (2021). Validation of the Leiden Inventory for the Child's Well-being in Daycare (LICW-D) Questionnaire in Norwegian early childhood education and care centers. *Frontiers in Psychology* 12: 1–10.
- Whittaker TA (2012) Using the modification index and standardized expected parameter change for model modification. *Source: The Journal of Experimental Education* 80(1): 26–44.
- Ximénez C (2009) Recovery of weak factor loadings in confirmatory factor analysis under conditions of model misspecification. *Behavior Research Methods* 41(4): 1038–1052.