Behind the numbers. A mixed-methods study of the CLASS tool application in Italy

Quello che i numeri non dicono. Uno studio mixed-methods sull'applicazione del CLASS in Italia

Valentina Pagani
University of Milan-Bicocca, Dep. of Human Sciences for Education "Riccardo Massa", Milan (Italy)

OPEN ACCESS

Abstract
The increasing globalization of tools used to assess ECEC quality requires reflection on their cultural coherence and ecological validity when applied in cultural contexts other than those of origin, as well as on the meaning of inconsistencies at the psychometric level that sometimes emerge - inconsistencies that quantitative methods alone often struggle to explain adequately. This paper proposes an integrated framework that combines the statistical analysis of quality assessment tools with a qualitative exploration of the meanings and interpretations that accompany their use in other cultures. This mixed-methods study, analyzing the implications of the CLASS Pre-K tool application in Italian preschools, offers a broader reflection on the importance and the need to adopt a critical, culturally-sensitive approach when using assessment tools outside their cultural cradle.

Keywords: quality assessment; mixed-methods; standardized instruments; ECEC; cross-cultural research.

Riassunto
La crescente globalizzazione degli strumenti per valutare la qualità dei servizi educativi per l'infanzia richiede una riflessione sulla coerenza culturale e la validità ecologica di questi strumenti quando applicati in contesti culturali diversi da quelli d'origine, nonché sul significato delle incongruenze a livello psicometrico che talora emergono – incongruenze che i soli metodi quantitativi spesso faticano a spiegare adeguatamente. Il presente lavoro propone un framework integrato che combina l'analisi statistica degli strumenti di valutazione con un'esplorazione qualitativa dei significati e delle interpretazioni che accompagnano il loro utilizzo in un'altra cultura. Nello specifico, lo studio mixed-methods, analizzando le implicazioni dell'applicazione del CLASS Pre-K nelle scuole dell'infanzia italiane, offre una riflessione più ampia sull'importanza e la necessità di adottare un approccio critico e culturalmente sensibile quando si utilizzano strumenti di valutazione al di fuori della loro culla culturale.

Parole chiave: valutazione della qualità; mixed-methods; strumenti standardizzati; servizi educativi per l'infanzia; ricerca cross-culturale.

1. Introduction

In recent years, the acknowledged key role of quality and the consequent growing public investment in early childhood education and care (ECEC) has drawn the attention of educational researchers towards developing evaluation tools to monitor ECEC quality (Ishimine & Tayler, 2014; Fenech, 2011; Grammatikopoulos, Gregoriadis & Zachopoulou, 2015). Not surprisingly, as governments increasingly demand means to assess quality objectively, most of these evaluation tools are quantitative and standardized – many of them developed in a U.S. context (e.g., ECERS/ECERS-R, ITERS/ITERS-R, CLASS) and widely used at the international level (Pastori & Pagani, 2017).

The adoption of existing tools offers multiple advantages compared to developing new instruments, including time and cost savings as well as a common ground for cross-country comparisons (Grammatikopoulos, Gregoriadis & Zachopoulou, 2015; Limlingan 2011). Nonetheless, the cultural and methodological complexities of the cross-cultural use of these tools are often not taken into account and problematized (Pastori & Pagani, 2017).

Indeed, these instruments are often used outside their *cultural cradle* without a rigorous critical reflection on their cultural consistency and ecological validity (Pastori & Pagani, 2017). When applying these instruments in cultural contexts that are different from the original ones, many studies rely only on statistical techniques (i.e., mainly factor-analytic procedures) to evaluate the construct equivalence of the imported assessment tools. Rarely is content validity questioned, despite being crucial in cross-cultural application and validation (Geisinger, 1994; van de Vijver & Poortinga, 1997). It is implicitly assumed that the underlying values and conceptualizations of quality conveyed by the tool overlap entirely across cultures (van de Vijver & Poortinga, 1997) without taking the culture-bound nature of quality into account (Dahlberg, Moss & Pence, 2007; Tobin, 2005). Moreover, in order to overcome possible psychometric inconsistencies within the original statistical model, at times the choice is made to retain certain elements of the model and discard others which do not fit in with the local context. This choice raises issues in terms of comparability and may undermine the reliability and validity of the instrument (Mathers et al., 2007).

Overall, a unified framework that integrates a statistical analysis of the tools with a qualitative exploration of the meanings and interpretations that accompany their use in other cultures appears to be lacking. Yet this can contribute to revealing the reasons behind psychometric inconsistencies, responding to questions that cannot be answered by quantitative methods alone (Fenech, Sweller & Harrison, 2010).

This paper attempts to narrow the gap by adopting a mixed-methods design to analyze the implications of CLASS Pre-K (*Classroom Assessment Scoring System*; Pianta, La Paro & Hamre, 2008) tool application in Italian preschools.

1.1 The CLASS tool

The CLASS tool (Pianta, La Paro & Hamre, 2008) is an observation system designed to provide a research-based framework to assess teacher-child interactions. It is based on developmental theory and focuses on process quality, including both emotional and instructional aspects in the classroom. The tool is available in several versions (from infant-toddler centers to secondary schools), each providing context-specific and developmentally sensitive parameters for the corresponding age level. It offers a common metric and vocabulary across grades, addressing the need for continuity and coherence in education.

Since its publication, the CLASS has been extensively used in evaluation and research in the U.S. as well as internationally (Pastori & Pagani, 2017). Recent studies (Ishimine & Tayler, 2014; Pastori & Pagani, 2017; Sandilos et al., 2014; Lemay, Leher & Naud, 2017) have pointed out some issues that question the validity of the CLASS framework when applied in cultural contexts different from that of origin. On the one hand, when exporting the tool into different cultural contexts, several researchers (e.g., Bouchard et al., 2014; Declercq & Laevers, 2015; Dessus, Cosnefroy & Joët, 2014; Leyva et al., 2015; Pakarinen et al., 2010) reported psychometric inconsistencies with regard to the original factorial model (i.e., *Teaching through Interactions* framework) and consequently they sometimes altered the scales. On the other hand, the CLASS acknowledgment of cultural differences is limited to the behavioral level, which does not affect the overall structure of the tool (Hamre et al., 2013). However, since culture shapes

the way that adults and children interact (Rogoff, 2003; Tobin, 2005), conclusions about the supposed universality of any conceptualizations of quality should be drawn with caution.

Despite the international diffusion of the CLASS tool, these issues have not yet been properly investigated.

1.2 Purpose and research questions

The present study presents a critical, mixed-methods approach to the application of standardized assessment tools outside their *cultural cradle*, with specific regard to the CLASS Pre-K.

The structural validity of the CLASS tool in the Italian ECEC context was examined. However, this step, rather than being the only and ultimate goal of the study, also included a qualitative analysis of the instrument that further developed the *critical-cultural discussion* about the CLASS that initiated within the CARE project (*Curriculum Quality Analysis and Impact Review of European ECEC* – http://ecec-care.org/; see Pastori & Pagani, 2017).

The mixed-methods study addressed four main research questions:

1 Is the three-factor model postulated by the *Teaching through Interactions* framework consistent with the Italian data?
2 Is there continuity between the conceptualization of effective teaching rooted in Italian pedagogical theory and practice and the conceptualization embedded in the tool?
3 Are there any differences, discrepancies or missing elements in the framework provided by the tool that are nonetheless crucial to fully capture the quality of teacher-child interactions in the Italian context?
4 What do the combined quantitative and qualitative results tell us about the applicability and generalizability of the CLASS framework in relation to Italian ECEC settings?

2. Procedure, Materials and Methods

2.1 Research design

A *mixed-methods convergent parallel design* (Creswell & Plano Clark, 2011) was adopted. Quantitative data (i.e., 162 preschool classroom observation cycles coded using the CLASS Pre-K) were used to describe classroom quality as postulated by the CLASS in Italian preschools and to examine the applicability and generalizability of the *Teaching through Interactions* framework in the Italian ECEC context. A qualitative approach was adopted in order to explore Italian practitioners' cultural values and beliefs concerning effective teacher-child relationships and ECEC quality, as well as to compare them with the perspective proposed by the tool.

Results from both the quantitative and qualitative data analyses were then compared, so as to offer a more comprehensive picture and bring greater insight to the problem.

2.2 Participants

Data was gathered from 27 preschool classrooms. Centers were selected from six provinces to cover different geographical areas and types of provision (including both state-run and municipal preschools).

Firstly, the procedures and the primary goals of the study were explained in detail to teachers and parents. Teachers were selected to participate on a voluntary basis and were asked for their written consent. Similarly, parents were asked for consent for their children's participation.

Fifty preschool teachers (48 females, 2 male), two of whom were special education teachers, participated in the study. The average age of participants was 46.9 years (SD = 10.3) and they had an average teaching experience of 22.4 years (SD = 12.3).

Preschool class size (i.e., total number of children enrolled in the class) ranged from 19 to 29 children (M = 24.7, SD = 2.4). All of the classes hosted children from ages three to six, provided full-time service, and were Italian speaking. On average, 16.11 (SD = 6.1) children were present during the observation cycles (min = 5, max = 27), depending on the type of activity (routines, small group activities, large group activities, meals/snacks…) taking place. Children-teacher ratio ranged from 3 to 27 (M = 11.8, SD = 6.0).

2.3   Data collection and analysis

Quantitative data. In each participating class, video observations began in the morning when the educational activity started and lasted approximately 3 hours.

From each class video recording, six 15-20 minute video segments (cycles) were selected for coding, consistent with the selection criteria presented in the CLASS manual (Pianta, La Paro & Hamre, 2008) and adopted in previous studies that used the tool for video observations outside the U.S. (Araujo et al., 2014; Hu et al., 2016; Leyva et al., 2015), in order to ensure that the selected segments were representative of the average experience of children in the observed classrooms.

The resulting 162 selected video segments were coded by a certified CLASS observer (the author) who rated each dimension on the 7-point scale.

Prior to the analysis, univariate data screening was conducted using the computer software package SPSS 25.0, following the procedures outlined by Kline (2011) and Tabachnick and Fidell (2001). Subsequently, the descriptive statistics for the individual CLASS Pre-K dimensions were examined and compared with those collected in previous studies conducted at the international level. Following this, structural validity investigations were carried out using the SPSS 25.0 and Amos software packages.

Qualitative data. In each preschool, teachers participated in three reflective seminars aimed at familiarizing them with the CLASS tool (for a more detailed description of the procedure adopted, see Pagani, 2016; Pastori & Pagani, 2017; Pagani & Pastori, 2019). In each reflective seminar, teachers were involved in focus groups to elicit their opinions about the tool and discuss its framework. The CLASS was assumed to be a powerful trigger for exploring the pedagogical values and teaching choices of the participants, making them explicit through comparison with the cultural perspective embedded in the tool itself (Pastori & Pagani, 2017; Pagani & Pastori, 2019). The focus groups with teachers were audiotaped and transcribed *verbatim*. Finally, the entire qualitative data set was analyzed through a thematic analysis (Boyatzis, 1998), adopting a semantic, data-driven inductive approach according to the guidelines proposed by Braun and Clarke (2006).

3.   Results

3.1   Quantitative results

Descriptive statistics. The range and distribution of scores for each of the 10 dimensions were examined (Table 1). At the domain-level, the overall level of Emotional Support (M= 5.70, SD = 0.62) and Classroom Organization (M = 5.57, SD = 0.70) were moderately high. In contrast, the overall level of Instructional Support was rather low (M = 2.67, SD = 0.61), with two dimensions (Concept Development and Quality of Feedback) in the low range and one dimension (Language Modeling) in the middle range.

The standard deviations for the dimensions ranged from 0.60 to 1.05 and for most dimensions they were approximately 1 scale point.

A closer inspection of the data reveals that the Negative Climate dimension poorly differentiated classroom quality among Italian preschool classes, as already pointed out by Pakarinen and colleagues (2010) and by Cadima, Aguiar and Barata (2018) with regard, respectively, to the Finnish and Portuguese samples.

The general picture of classroom quality depicted by the CLASS seems to suggest that: a) Italian preschool teachers generally had positive, warm and supportive interactions with children; b) moderate language stimulation and facilitation were provided; c) classrooms were rather effectively organized in terms of behavior and instructional time management. Conversely, Concept Development and Quality

of Feedback means in the low range seem to suggest that, for the most part, the observed activities focused more on basic skills rather than promoting children's reasoning and thinking.

This pattern – higher levels of Emotional Support and Classroom Organization, with lower scores in Instructional Support dimensions – is similar to those reported in previous studies conducted internationally (e.g., Pianta, La Paro & Hamre, 2008; Pakarinen et al., 2010; Leyva et al., 2015; Hu et al., 2016; Cadima, Aguiar & Barata, 2018).

Exploratory factor analysis. Structural validity was investigated to examine whether the Italian data indicated a similar three-factor structure as postulated by the *Teaching through Interactions* framework. The 162 observation cycles were used for structural analysis.

Domain and dimension	M	SD	Min	Max
Emotional Support	**5.70**	**0.62**		
Positive Climate	5.50	0.93	3	7
Negative Climate (Reversed Negative Climate)	1.25 (6.75)	0.60	1	6
Teacher Sensitivity	5.59	0.71	4	7
Regard for Student Perspectives	4.98	1.05	3	7
Classroom Organization	**5.57**	**0.70**		
Behavior Management	5.97	0.85	3	7
Productivity	5.64	0.88	3	7
Instructional Learning Formats	5.09	0.91	3	7
Instructional Support	**2.67**	**0.61**		
Concept Development	1.92	0.75	1	6
Quality of Feedback	2.40	0.63	1	6
Language Modeling	3.69	0.92	2	6

Note: Each scale ranges from 1 to 7 points.
Table 1: Means, standard deviations and ranges for teacher-child interactions in preschool classrooms

Prior to the exploratory factor analysis (EFA), data screening was conducted to examine the inter-correlation between variables, verifying the absence of singularity or extreme multicollinearity and to test the assumption of univariate normality. First, the magnitude of the correlations between dimensions and between domains was examined (Table 2). The results suggested that the presence of multicollinearity or singularity was not a significant concern as there was a significant correlation among individual CLASS dimensions ranging from .16 to .59. Overall, the results suggested modest to moderate intra-domain correlations and weak inter-domain correlations. The only exceptions were correlations among Emotional Support dimensions and Classroom Organization dimensions that presented some moderate convergence values. This result paralleled the high correlation (.67) registered between these two domains, compared to the weak correlations between Emotional Support and Instructional Support (.34) and between Classroom Organization and Instructional Support (.34). Then, the skewness and kurtosis of individual CLASS dimensions were examined. Only one dimension, Negative Climate, demonstrated severe skewness (3.87)

and kurtosis (22.79), resulting as non-normally distributed (Fabrigar et al., 1999). Due to these extreme values, Negative Climate was excluded from the factor analysis.

Afterwards, an EFA was carried out to identify the factor structure underlying the dataset without imposing any restrictions. A principal component analysis (PCA) was conducted on the CLASS dimensions (excluding Negative Climate) with oblique rotation (direct oblimin). The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis, KMO = .79, which was above the acceptable limit of .50 (Field, 2009). Bartlett's test of sphericity χ^2 (36) = 470.58, p < .001, indicated that correlations between items were sufficiently large for factor analysis.

	CO	IS	1	2	3	4	5	6	7	8	9
Emotional Support (ES)	.67**	.34**									
1. Positive Climate			–								
2. Negative Climate			-.43**	–							
3. Teacher Sensitivity			.52**	-.40**	–						
4. Regard for Student Perspectives			.44**	-.33**	.34**	–					
Classroom Organization (CO)	–	.34**									
5. Behavior Management			.42**	-.42**	.30**	.40**	–				
6. Productivity			.39**	-.23**	.39**	.35**	.27**	–			
7. Instructional Learning Formats			.59**	-.36**	.44**	.41**	.45**	.59**	–		
Instructional Support (IS)	–	–									
8. Concept Development			.24**	-.09	.17*	.23**	.16*	.26**	.38**	–	
9. Quality of Feedback			.11	-.09	.16*	.11	-.06	.09	.10	.45**	–
10. Language Modeling			.40**	-.19**	.18*	.27**	.20**	.34**	.33**	.48**	.41**

*Note: *p < .05   **p < .01*
Table 2: Correlations among the CLASS Pre-K domains and dimensions

According to Horn's (1965) parallel analysis, two components were retained. In combination, they explained 56.83% of the variance. Table 3 shows the factor loadings after rotation. The value of .40 was used as cut-off for acceptable factor loadings (Field, 2009).

All the CLASS dimensions from the Emotional Support and Classroom Organization domains loaded together on the first factor. All the CLASS dimensions from Instructional Support loaded together on the second factor. Both two factors had high scale reliabilities as measured by Cronbach's alpha: Factor 1 = .81, Factor 2 = .70.

Thus, the EFA results do not seem to provide preliminary support for the three-factor model posited in the *Teaching through Interactions* framework.

Confirmatory factor analysis. Next, a confirmatory factor analysis (CFA) was conducted to test the current sample against the theoretical model proposed by the *Teaching through Interactions* framework, assuming three positively correlated factors (Emotional Support, Classroom Organization, Instructional

Support). Furthermore, three alternative solutions were tested (the first two were based on the work of Hamre and colleagues, 2013):

a. a model with 10 dimensions loading on one global domain (Effective Teaching);
b. a model with 10 dimensions loading on two domains (Social Support and Instructional Support);
c. the two-factor solution excluding the Negative Climate dimension indicated in the EFA.

Before conducting the analysis, five univariate outliers were identified and deleted listwise (Field, 2009). The three-factor model provided the best relative fit to the data, $\chi^2(32) = 61.795$, p = .001, CFI = .930, TLI = .902, RMSEA = .075, SRMSR = .0599, compared to the one-factor and two-factor models – respectively, $\chi^2(35) = 95.437$, p < .001, CFI = .858, TLI = .817, RMSEA = .102, SRMSR = .0749, and $\chi^2(34) = 70.993$, p < .001, CFI = .913, TLI = .885, RMSEA = .081, SRMSR = .0610. Based on the cut-off points suggested by Browne and Cudeck (1993), it nonetheless did not fit the Italian data well. Moreover, the high correlation between Emotional Support and Classroom Organization (.85) indicated that there was still a notable overlap among domains.

Subsequently, the solution suggested by the EFA was tested. Fit statistics suggested that the model had less than adequate fit: $\chi^2(26) = 56.122$, p = .001, CFI = .919, TLI = .888, RMSEA = .084, SRMSR = .0596. Modification indexes suggested that the fit of the model would increase, introducing some minor modifications. Specifically, the residuals of the observed variables Positive Climate and Productivity, Regard for Student Perspective and Behavioral Management, and Instructional Learning Formats and Concept Development were allowed to correlate, resulting in a revised two-factor model (Figure 1) that fit the data well: $\chi^2(23) = 31.577$, p = .109, CFI = .977, TLI = .964, RMSEA = .047, SRMSR = .0505. The present model provided statistically significant parameter estimates.

The results seem to show that the two-factor structure, excluding the Negative Climate dimension, is a better solution to fit the current sample than the original three-factor model.

Domain and dimension	Factor 1	Factor 2
Emotional Support		
Positive Climate	**.78**	.06
Teacher Sensitivity	**.67**	.00
Regard for Student Perspectives	**.66**	.04
Classroom Organization		
Behavior Management	**.73**	-.20
Productivity	**.65**	.12
Instructional Learning Formats	**.79**	.10
Instructional Support		
Concept Development	.13	**.76**
Quality of Feedback	-.18	**.87**
Language Modeling	.23	**.69**

Note: Factor loadings above the cut-off value of .40 are evidenced in bold.
Table 3: Factor loadings after rotation for the two-factor solution

3.2 Qualitative results

Italian practitioners used the CLASS dimensions, indicators and behavioral markers as a lens and frame to observe and compare the perspective of the tool to their local-cultural and pedagogical perspective. This discussion – not only *on* the instrument, but also *with* it – allowed Italian teachers to reflect on and become more aware of their conceptualization of quality in ECEC. At the content level[1], several noteworthy elements emerged (for a more detailed presentation see Pagani 2016; Pastori & Pagani, 2017; Pagani & Pastori, 2019):

- *Continuities.* Italian teachers appreciated the key role assigned to teacher-child interactions in defining classroom quality and the dual, comprehensive concept of relationships that the CLASS considers, addressing both the emotional and the instructional features of the classroom;
- *Differences.* A relevant difference regards the concept of children's learning. While the CLASS tool focused solely on cognitive and linguistic facets, Italian practitioners shared a broader vision of learning (e.g., also embracing socio-emotional learning, learning from each other among peers, learning to respect and value diversity…). At the same time, most teachers recognized that in their classrooms many occasions to foster children's reasoning were lost and they agreed with the feedback provided by the tool on this matter. Thus, they appreciated the cognitive declination of the CLASS lens precisely because it drew their attention to this often undervalued aspect and prompted them to reflect on how to more effectively foster children's cognitive skills throughout the different activities and moments of the school day (see Pagani & Pastori, 2019);
- *Missing elements.* Italian teachers mentioned various key features of teacher-child relationships not captured by the CLASS: the lack of attention paid to the space and materials, considered only as structural quality features rather than a 'third educator' (Malaguzzi, 1993) that scaffolds educational and relational processes; the far too marginal role assigned to peer relationships, considered almost exclusively from the socio-emotional point of view and not as a key factor in promoting children's learning and socio-cognitive development; the lack of emphasis placed on teachers' efforts to foster inclusive competences and to encourage children to be aware of and respect all forms of diversity;
- *Disagreements.* Two aspects raised criticism among participants. Firstly, Italian teachers, rather than stressing the productivity of the classroom as the tool did, were more concerned with the significance of the activities provided and ensuring the children a holistic and rich experience, free from pressure. Secondly, the research participants did not think that their immediate response to a child's request for help or attention was always necessary. For them, effective teachers were reflective professionals who, albeit aware of children's signals, took enough time to observe the reactions of the children and their peers before deciding if and how to intervene.
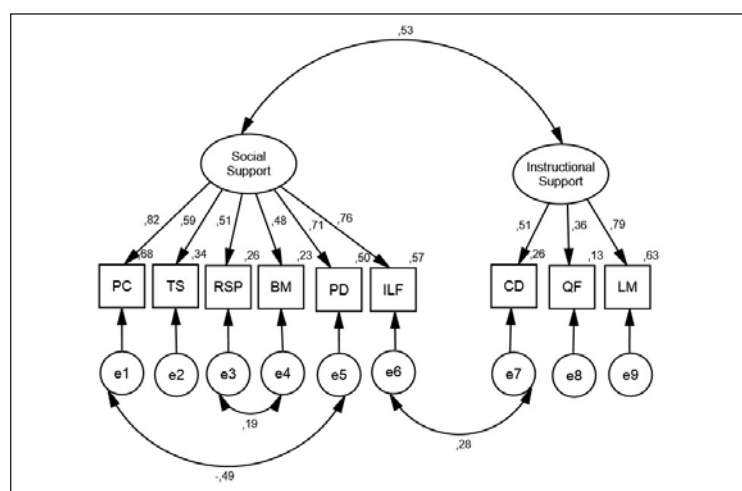

Figure 1: Revised two-factor model

---

1   For a discussion on differences at the methodological level, see Pagani and Pastori (2019).

3.3 Integrating qualitative and quantitative results

Combining and comparing the two strands of data raised questions about the assumption, suggested by the CLASS authors (Hamre, Goffin & Kaft-Sayre, 2009; Hamre et al., 2013), that the dimensions of classroom quality assessed by the tool were relevant across cultures.

The quantitative findings suggest that a two-factor solution, which unifies classroom socio-emotional features within a single domain, better describe the Italian data compared to the original three-factor model posited by the *Teaching through interactions* framework (Hamre et al., 2013). However, no explanation is offered for this evidence. In this regard, the qualitative data could help to offer a possible interpretation regarding the two-fold structure that emerged from the EFA and CFA.

The participating Italian practitioners recognized that, while the CLASS framework contained many elements of interest and continuity, it also presented noteworthy discrepancies with regard to their idea of what ECEC quality and effective teaching were. Moreover, some points they highlighted seem to suggest that the demarcation line between Emotional Support and Classroom Organization, valid at the theoretical level, actually faded in their daily educational practices. There were several connections that Italian teachers identified between dimensions belonging to these two domains that the CLASS Pre-K kept separate (Pagani, 2016). For example: the attention Italian teachers paid to recognizing and valuing children as competent subjects, involving them in discussing and defining – rather than predetermining and imposing – classrooms rules; the promotion of self-regulatory skills that allowed children to enforce rules and learn to accept peer/mutual correction; the active role assigned to children in co-constructing projects, activities, routines and actively shaping the classroom curriculum; the effort to offer children a significant, pressure-free experience in which even time spent waiting could become an opportunity to assimilate the rules of social life and mutual respect (Pastori & Pagani, 2017; Pagani, 2016). In this light, the emotional support provided by teachers ceased to be conceived of only in terms of how effectively they created a warm, secure, supportive climate in the classroom, but was broadened to embrace how teachers deliberately supported children in learning emotional skills – a kind of learning that necessarily takes place within the social life of the class through the definition of the underlying rules, times and routines.

Overall, this combined qualitative and quantitative evidence strongly challenges the otherwise taken-for-granted universality of the CLASS framework and suggests the need to reconceptualize the CLASS model in order to properly assess relational quality in Italian preschools.

4. Discussion and Conclusions

The data presented here – based on the application of the CLASS Pre-K in Italian preschools, but likely to result in a broader reflection on the international use of standard-based assessment instruments – show that even tools with a solid theoretical and empirical background, such as the CLASS, cannot be considered culture-free. As *children* of the context where they were developed (Pastori & Pagani, 2017), they are vessels of culturally-bound conceptualizations regarding quality and pedagogy that may not fully reflect the viewpoints and interpretations of different groups in different places.

Despite these findings, the overall purpose of the present study is not to criticize *tout court* the cross-cultural use of standardized assessment measures, nor to deny the valuable advantages that they may confer. Rather, it intends to highlight the importance and the need to adopt a critical stance in the cross-cultural application of these tools (*Ibidem*), integrating qualitative and quantitative methods in order to be attentive to the cultural and methodological complexities when these instruments are used in cultural contexts that are different from the original ones.

The data presented here offers an effective case in point to support this claim. In fact, a mere statistical investigation of the applicability of the CLASS framework in the Italian context did not confirm the original three-factor model, leading to two possible scenarios:

a. trying to *tame disobedient* dimensions, 'dropping or amending elements which [did] not fit with the local context' (Mathers et al., 2007, p. 268). However, the choice of altering the original model, albeit already done other times (Cadima, Aguiar & Barata, 2018; Leyva et al., 2015; Pakarinen et

al., 2010), would have raised issues in terms of comparability and validity (Sandilos et al., 2014; Mathers et al., 2007; Lemay, Leher & Naud, 2017);

b. accepting the psychometric inconsistencies as an indicator of cross-cultural differences, without, however, being able to provide more than a tentative explanation of the results.

Instead, the choice to adopt a mixed methodology allowed us to explore some possible reasons behind the statistical inconsistencies in the CLASS framework when applied to Italian ECEC settings. Thus, incorporating structural elements of quality with qualitative explorations has the potential to provide more nuanced understandings of ECEC quality, as well as creating a stronger platform to inform policy and practice (Fenech, Sweller & Harrison, 2010, p. 294). It also permits the application of assessment instruments in different cultural contexts in a more critical and culturally sensitive way, accounting for cultural peculiarities while respecting local values and pedagogical interpretations.

This notwithstanding, some potential limitations need to be considered when interpreting the results of the current study. First, the limited sample size, although adequate for performing factor analysis according to MacCallum and colleagues' (1999) criteria, may have affected the accuracy of structural equation modelling (Gagné & Hancock, 2006). Secondly, although data was collected in different Italian provinces, the sample cannot be considered representative of Italian ECEC services at the national level. Thirdly, the presence of a single coder, albeit certified as a reliable CLASS Pre-K observer, did not allow for double-coding and, consequently, for monitoring observer drift and corroborating the reliability of the ratings assigned.

In this regard, further research could potentially be directed at extending the existing study, by involving a larger, more representative sample group and a greater number of certified coders. Such an extension could provide a broader, sounder basis for verifying the findings obtained. This could also provide indications regarding the feasibility and usefulness of working on adapting the tool for the Italian ECEC context, or whether it might be more profitable to develop a distinct instrument to assess the quality of teacher-child interactions, better adapted to incorporate Italian peculiarities.

Conflict of interests

The authors declare no conflict of interest.

References

Araujo, M. C., Carneiro, P., Cruz-Aguayo, Y., & Shady, N. (2014). *A Helping Hand? Teacher Quality and Learning Outcomes in Kindergarten*. Working paper, version: October 4, 2014. Retrieved from http://-economics.yale.edu/sites/default/files/carneiropedroahelpinghand_final.pdf (accessed on 26th August 2020).

Bouchard, C., Cantin, G., Charron, A., Crépeau, H., & Lemire, J. (2014). *Validation du Classroom Assessment Scoring System (CLASS) en maternelle 4 ans à mi-temps au Québec.* Paper presented at the 82nd Congres de l'Acfas, Colloque 519, in Montréal, Canada, May 15.

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development.* Thousand Oaks, CA: Sage.

Braun, V., & Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in Psychology*, *3*(2), 77-101.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models*. Newbury Park, CA: Sage.

Cadima, J., Aguiar, C., & Barata, M. C. (2018). Process quality in Portuguese preschool classrooms serving children at-risk of poverty and social exclusion and children with disabilities. *Early Childhood Research Quarterly*, *45*, 93-105.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and Conducting Mixed Methods Research* (2nd ed.). Thousand Oaks, CA: Sage.

Dahlberg, G., Moss, P., & Pence, A. (2007). *Beyond Quality in Early Childhood Education and Care: Languages of Evaluation* (2nd ed.). London: Falmer Press.

Declercq, B., & Laevers, F. (2015). *Measuring 'Pedagogical' Quality with an Observation Tool: Insights from the Pilot Study.* Paper presented at the 25th EECERA conference, in Barcelona, Spain, September 7-10.

Dessus, P., Cosnefroy, O., & Joët, G. (2014). *Evaluer la qualité des interactions enseignant-élèves en début de scolarité:*

Qualités psychométriques d'une adaptation française du CLASS. Paper presented at the 82nd Congres de l'Acfas, Colloque 519, in Montréal, Canada, May 15.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological methods*, *4*, 272-299.

Fenech, M. (2011). An Analysis of the Conceptualisation of 'Quality' in Early Childhood Education and Care Empirical Research: Promoting 'Blind Spots' as Foci for Future Research. *Contemporary Issues in Early Childhood*, *12*(2), 102-117.

Fenech, M., Sweller, N., & Harrison, L. (2010). Identifying high-quality centre-based childcare using quantitative data-sets: what the numbers do and don't tell us. *International Journal of Early Years Education*, *18*(4), 283-296.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage publications.

Gagné, P., & Hancock, G. R. (2006). Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivariate Behavioral Research*, *41*(1), 65-83.

Geisinger, K. F. (1994). Cross-cultural normative assessment: translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological assessment*, *6*(4), 304-312.

Grammatikopoulos, V., Gregoriadis, A. & Zachopoulou, E. (2015). Evaluation of Early Childhood Evaluation Environments and Professional Development: Current Practices and Implications. In O. N. Saracho (Ed.), *Contemporary Perspectives on Research in Testing and Evaluation in Early Childhood Education*, 153-172. Maryland: Information Age Publishing.

Hamre, B. K., Goffin, S. & Kraft-Sayre, M. (2009). *Classroom Assessment Scoring System (CLASS) Implementation Guide: Measuring and Improving Classroom Interactions in Early Childhood Settings*. Charlottesville: Teachstone.

Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., Brown, J. L. et al. (2013). Teaching Through Interactions. *The Elementary School Journal*, *113*(4), 461-487.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179-185.

Hu, B. Y., Fan, X., Gu, C., & Yang, N. (2016). Applicability of the Classroom Assessment Scoring System in Chinese preschools based on psychometric evidence. *Early Education and Development*, *27*(5), 714-734.

Ishimine, K., & Tayler, C. (2014). Assessing Quality in Early Childhood Education and Care. *European Journal of Education*, *49*(2), 272-290.

Kline, R. B. (2011). *Principles and practice of structural equation modelling* (3rd ed.). New York, NY: Guilford Press.

Lemay, L., Lehrer, J., & Naud, M. (2017). Le CLASS pour mesurer la qualité des interactions en contextes culturels variés. *Les dossiers des sciences de l'éducation*, *37*, 15-34.

Leyva, D., Weiland, C., Barata, M., Yoshikawa, H., Snow, C., Treviño, E., & Rolla, A. (2015). Teacher-child Interactions in Chile and Their Associations with Prekindergarten Outcomes. *Child development*, *86*(3), 781-799.

Limlingan, M. C. (2011). On the Right Track: Measuring Early Childhood Development Program Quality Internationally. *Current Issues in Comparative Education*, *14*(1), 38-47.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological methods*, *4*(1), 84-99.

Malaguzzi, L. (1993). For an Education Based on Relationships. *Young Children*, *49*, 9-12.

Mathers, S., Linskey, F., Seddon, J., & Sylva, K. (2007). Using quality rating scales for professional development: Experiences from the UK. *International Journal of Early Years Education*, *15*(3), 261-274.

Pakarinen, E., Lerkkanen, M. K., Poikkeus, A. M., Kiuru, N., Siekkinen, M., Rasku-Puttonen, H., & Nurmi, J. E. (2010). A Validation of the Classroom Assessment Scoring System in Finnish Kindergartens. *Early Education and Development*, *21*(1), 95-124.

Pagani, V. (2016). *Teacher-child relationship quality in ECEC settings. A mixed-methods study of the CLASS tool application in Italy*. PhD diss., University of Milano-Bicocca.

Pagani, V., & Pastori, G. (2019). Coniugare valutazione standardizzata e approccio partecipativo-riflessivo. Un'esperienza di ricerca collaborativa nei servizi per l'infanzia italiani. *RicercAzione*, *11*(2), 103-120.

Pastori, G., & Pagani, V. (2017). Is validation always valid? Cross-cultural complexities of standard-based instruments migrating out of their context. *European Early Childhood Education Research Journal*, *25*(5), 682-697.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System Pre-K Manual*. Baltimore: Brookes Publishing.

Rogoff, B. (2003). *The Cultural Nature of Human Development*. Oxford: University Press.

Sandilos, L. E., Di Perna, J. C. & The Family Life Project Key Investigators (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.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston, MA: Allyn and Bacon.

Tobin, J. (2005). Quality in Early Childhood Education: An Anthropologist's Perspective. *Early Education and Development*, *16*(4), 421-434.

van de Vijver, F., & Poortinga, Y. H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European journal of psychological assessment*, *13*(1), 29-37.