May 2021

NeMTSS Research Brief

The Classroom Assessment Scoring System for PreK (CLASS-PreK) and Child Outcomes

Abril Rangel-Pacheco, M.A. & Amanda Witte, Ph.D.

NEBRASKA CENTER FOR RESEARCH ON CHILDREN, YOUTH, FAMILIES & SCHOOLS

NeMTSS Implementation Support Team, University of Nebraska–Lincoln is funded by the Nebraska Department of Education (USDE IDEA, Part B-611 Grant). Contents do not necessarily represent the policy of NDE or USDE, and no endorsement should be assumed. Developed by UNL-CYFS MTSS Implementation Support Team.

The Classroom Assessment Scoring System for PreK (CLASS-PreK) and Child Outcomes: An NeMTSS Research Brief

Key Points:

- There has been mixed evidence regarding the use of the Classroom Assessment Scoring System for PreK (CLASS-PreK) in predicting child outcomes. Two studies have found significant relationships between CLASS-PreK domain scores and outcomes in language, literacy, social skills, math, and behavior at the end of the PreK year. However, effect sizes were small to modest in magnitude (Zaslow et al., 2010).

- More recent studies have highlighted the weak relationships between domain scores and child outcomes, as well as issues with the instruments interrater consistency and ability to capture accurate measurements throughout different times of the day and content (Styck et al., 2020; Thorpe et al., 2021)

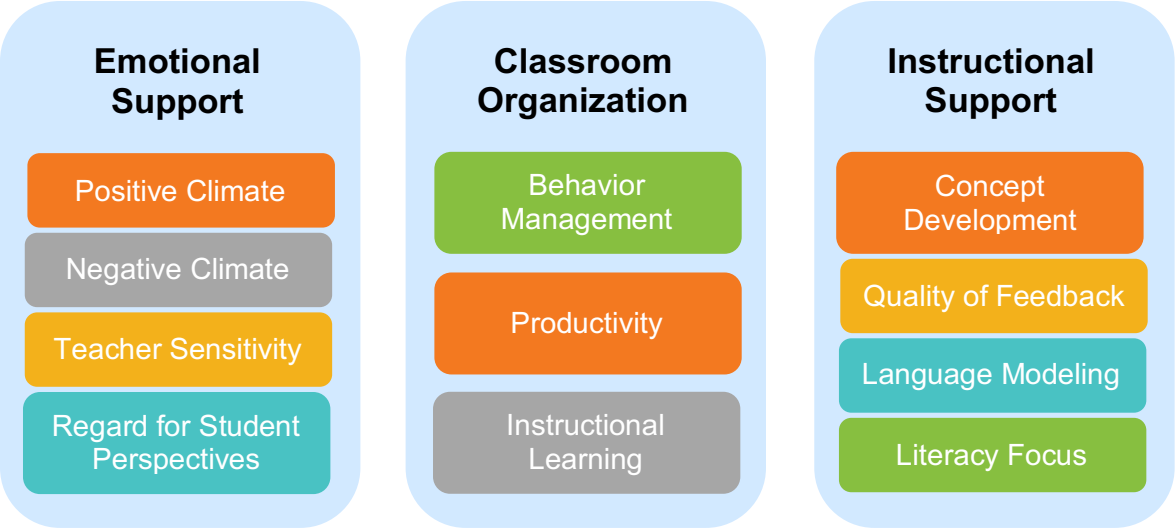
- Overall, key limitations of the CLASS PreK concern the cultural alignment of the assessment to diverse cultures and countries outside of the US, consistency of ratings across observers and time of day, and predictive validity in child outcomes (Bihler et al., 2018; Barnes-Najor et al., 2017)

- It is recommended that users of the CLASS-PreK consider the impact of rater effects (and other extraneous variables) on scores from teacher–child interaction variables. Further, it is recommended that CLASS-PreK scores not be used for individual high stakes decisions (e.g., teacher evaluations) due to the issues with interobserver consistency and predictive validity. However, a standardized approach where observations focus on specific activities of interest including both care and educational activities may reduce some biases and enhance comparability (Styck et al., 2020; Thorpe et al., 2021).

CLASS Assessment: Overview

The Classroom Assessment Scoring System for PreK (CLASS-PreK) is an observation instrument that assesses the quality of teacher-child interactions in center-based preschool classrooms and settings (ECLKC, 2020). As part of the assessment procedures, four cycles of 15 to 20-minute observations of teachers and students are conducted by a certified CLASS observer. Observations are then rated on a 7-point scale (7 being the highest score) using a manual of behaviors and responses (Teachstone, n.d.). The CLASS-PreK assesses three broad domains of teacher-child interactions that support children's learning and development: Emotional Support, Classroom Organization, and Instructional Support (Hamre, 2009; Teachstone, n.d.). Within each domain are dimensions that capture more specific details about teachers' interactions with children (Figure 1). Collectively, these dimensions assess the extent to which teachers are effectively supporting children’s social and academic development (Hamre, 2009).

Figure 1. Organization of the CLASS-PreK Framework



Researchers, teachers, principals, school psychologists, and educational consultants can use the CLASS for research, program planning and evaluation, for professional development, and to support supervision (Hamre, 2009; Rodriguez & Garza, 2014). Additionally, the CLASS can be used for accountability as a monitoring system to address the quality of early childhood classrooms (Rodriguez & Garza, 2014). The CLASS-PreK is used widely in the United States (US) as a tool for monitoring early childhood education programs, and has gained increasing use internationally (Bihler et al., 2018).

Psychometric Evidence

The CLASS-PreK has been compared to other classroom quality measurements in order to validate the instrument. Several studies have shown the CLASS-PreK to possess adequate reliability and validity in the US (Buell et al., 2016; La Paro et al., 2004), and has been validated in international samples of early childhood education students as well (Bihler et al., 2018; Pakarinen et al., 2010; Cloney et al., 2017). Regarding factor structure, there has been diverging evidence supporting one (Hamre et al., 2013), two (La Paro et al., 2004; Hamre et al., 2014), and three factor models of the instrument, with most of the extant literature supporting the existing domains of the three-factor model (Figure 1) (See Bihler et al., 2018 for a summary).

The CLASS-PreK has also been utilized with diverse samples of students in order to examine the generalizability of its psychometric properties. In a study of the CLASS and Dual Language Learners (DLL) Downer et al. (2012) examined whether the CLASS domains were consistent for classrooms with varying proportions of DLLs and if the class was as predictive of DLL outcomes compared to those of monolingual English speakers. The researchers concluded that the CLASS was reliable and valid for both samples and equally related to DLL children's school success. Similarly, in a sample of American Indian and Alaska Native Head Start programs, Barnes-Najor et al. (2017) found the instrument to provide a reliable measure targeting the quality of child-teacher interactions, albeit with limitations (see Limitations section).

Relationship to Child Outcomes

In a large study of state funded PreK programs, Mashburn et al., (2008) examined associations between different measures of PreK quality and children's development. The CLASS-PreK was used to assess the quality of teacher-child interactions. To measure development, the researchers administered a battery consisting of receptive language, expressive language, Rhyming, applied problem solving, and letter naming tasks (Peabody Picture Vocabulary Test-Revised (PPVT-R), Oral Expression Scale from the Oral and Written Language Scales (OWLS), and The Woodcock-Johnson III Test of Achievement (WJIII ACH) subscales. Social skills of children were also measured using the Teacher-Child Rating Scale (TCRS).

Overall, results indicated that observed quality of emotional interactions measured by the CLASS were positively related to children's development of social competence and negatively related to children's development of problem behaviors during PreK. Specifically, higher scores in the instructional support domain were significantly and positively associated with all measures of academic or language skills (*B* coefficients ranged from .33 to 1.07). Higher scores in the Emotional Support domain were associated with teacher's ratings of higher social competence and lower problem behaviors by the end of the academic year (*B* = .06 and -.05, respectively). Lastly, the authors found that in comparison to other measures of PreK quality (ECERS-R & NIEER), the CLASS-PreK was most consistently and strongly associated with children's development (Mashburn et al., 2008).

In a similar study, Howes et al. (2008) also examined the relationship between CLASS measured teacher-child interactions and academic and behavioral outcomes of children in a large sample of state-funded PreK programs. Researchers used Woodcock-Johnson III Tests of Achievement, a language and literacy questionnaire, and the Social Skills and Behavior Problems Scale (Hightower et al., 1986) to measure student outcomes.

Results revealed the CLASS Instructional Support domain to be a significant predictor of gains in receptive language (d = .06) and expressive language (d = .07). Analysis also indicated that Instructional Support domain scores were significantly associated with the number of letters the children identified (d = .07) and the teacher's ratings of global language and literacy skills (d = .10) at the end of the year. Furthermore, there were significant associations between the CLASS Emotional Climate domain (d = .05) and math skills and behavior problems (d = −.04). Overall, Mashburn et al. (2008) and Howes et al. (2008) believe that gains in skills related to school readiness are supported both by sensitive, responsive interactions with teachers and by teacher–child interactions that focus on skill-development.

Limitations

In addition to the individual limitations of each study mentioned, there have been several articles highlighting the overall limitations of the CLASS-PreK. For example, Bihler et al. (2018) stated that because there have been several country-specific findings regarding the CLASS, it raises the question of whether it is appropriate to use outside of the US, because the scoring of the CLASS dimensions and indicators could be culturally bound. For example, a behavior like eye contact in teacher-child interactions may be interpreted differently in other contexts or may not occur in some cultures (Bihler et al., 2018). Barnes-Najor et al. (2017) affirmed this limitation in their study examining the cultural alignment of the CLASS in American Indian and Alaska Native head start programs. Specifically, their study provided evidence of cultural misalignment in the CLASS trainings, some of the CLASS behavioral markers, and the Instructional Support domain.

In examining the effects of teacher-child interactions as measured by the CLASS-PreK on academic and behavioral outcomes, effect sizes in Bihler et al. (2018) and Barnes-Najor et

al. (2017), were often small to modest, indicating weak relationships (Zaslow et al., 2010). Styck et al., (2020) theorized that this was due to systematic differences in observer ratings, and conducted a study examining rater effects on the CLASS and its relation to reading and math outcomes in 1st and 2nd grade students. Results indicated that observers differed significantly on severity for all three CLASS domains. Further, adjusting for these effects did not improve the statistical relationship between CLASS scores and child outcomes in their study. In other words, although rater effects influenced CLASS score variation, adjusting for those rater effects did not result in stronger associations between CLASS scores and theoretically related constructs (Styck et al., 2020). These results were in line with previous studies examining rater effects on the CLASS (see Styck et al., 2020 for a detailed summary). This indicates the CLASS may have limited consistency across raters which may in turn limit the accuracy of child outcome predictions.

Along a similar vein, Thorpe et al. (2021) examined two contextual factors that might affect assessment functioning: (1) When – the timing of assessment in the early childhood education (ECE) day and (2) What – the content and format of the activities observed. Their study showed a decline in instructional, organizational, and emotional support across the ECE day (8am to 4pm) with recovery in emotional support at the end of the day. Within-classroom analyses demonstrated that whole group and small group formats and science, math, and social science content inflated, while mealtimes, physical activity, and transitions constrained CLASS scores. Their results indicated that time of day and format and content matter when it comes to CLASS observation scores. They concluded that variations in content and format systematically biased CLASS scores, and assessments undertaken at different times of day were not equally comparable.

Lastly, in a review of the literature, there were no studies examining the impacts of PreK quality/teacher-child interactions in PreK and academic or behavioral outcomes past the end of the PreK year. In other words, all studies cited in this brief tested only relationships between CLASS domain scores and end-of-the-year student outcomes; none of the studies examined whether PreK CLASS scores had a relationship with student outcomes in later academic years. Therefore, information regarding the CLASS-PreK and its impact on long-term academic and behavioral outcomes of children must be interpreted with caution. Overall, the key limitations of the CLASS-PreK concern the cultural alignment of the assessment to diverse cultures and countries outside of the US, consistency of ratings across observers and time of day, and predictive validity in child outcomes.

Implications for Practitioners

Generally, the studies reviewed agreed that for young children, much learning occurs via interactions, and high-quality emotional and instructional interactions are the mechanisms through which PreK programs transmit academic, language, and social competencies to children (La Paro et al., 2004; Howes et al., 2008; Mashburn et al., 2008; Hamre et al., 2009). Mashburn et al. (2008) argues that program policies and regulations aimed at improving the effectiveness of children's exposure to PreK should focus more directly on improving interactions that children experience in classrooms. Two avenues through which policies can directly improve the quality of emotional and instructional interactions in classrooms are teacher professional development and program monitoring systems (Mashburn et al., 2008). Observational data may then be used in feedback to centers about their areas of strengths and weaknesses, as well as linked to resources to implement changes (Mashburn et al., 2008)

However, there is a need to consider the impact of rater effects (and other extraneous variables) on scores from teacher–child interaction variables, in general, and on the CLASS (Styck, 2020). It is recommended that CLASS scores not be used for individual high stakes decisions (e.g., teacher evaluations) due to the issues with interobserver consistency and

predictive validity (Styck et al., 2020). According to Styck and colleagues, even lower stakes decisions, such as formative feedback, might be questionable as assigned scores have been found to be related to specific observers completing the assessment. Further, Styck et al., (2020) contend that if CLASS is used at the individual level, inter rater differences could lead to unfair, inefficient, or inaccurate decisions about teacher–student interaction quality. On the other hand, if analysis of CLASS results occurs at a group level, the impact of these rater effects will depend on how observers are assigned to teachers. Styck and colleagues argue that if the process is random, between-observer differences will add an extra layer of noise potentially attenuating statistical relations with important developmental outcomes. However, if there is any systematicity to observer assignment, the impact could mirror the individual case, with unfair or inaccurate group level decisions being made (Styck et al., 2020).

A key message in Thorpe et al., (2021) was that different formats and contents within PreK settings are not comparable – time of day and instructional content matter. Their study showed that sampling of an educationally focused format and content such as a whole group science lesson favored the assessment outcome, whereas sampling health and care content and formats such as outdoor physical activity and mealtimes presented a less-favorable quality assessment. To combat this, the authors suggested a standardized approach where observations would focus on specific activities of interest including both care and educational activities, which could reduce some biases and enhances comparability.

Regarding issues of cultural sensitivity, Barnes-Najor et al. (2017) recommend that trainers of the instrument be aware of and sensitive to cultural differences within classrooms. The authors also recommend further work be done in the training regimen to include discussions of how cultural differences manifest and how to address these within the coding system.

References

Barnes-Najor, J. V., Thompson, N. L., Cameron, A. F., Smith, T. M., Calac Verdugo, M., Brown, P. L., & Sarche, M. C. (2021). 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*(1), 162-183. https://doi.org/10.1080/02568543.2020.1723749

Bihler, L.-M., Agache, A., Kohl, K., Willard, J. A., & Leyendecker, B. (2018). Factor Analysis of the Classroom Assessment Scoring System Replicates the Three Domain Structure and Reveals no Support for the Bifactor Model in German Preschools. *Frontiers in Psychology*, 9. https://doi.org/10.3389/fpsyg.2018.01232

Buell, M., Han, M., & Vukelich, C. (2017). Factors affecting variance in classroom assessment scoring system scores: Season, context, and classroom composition. *Early Child Development and Care, 187*(11), 1635-1648. https://doi.org/10.1080/03004430.2016.1178245

Cloney, D., et al. (2017) Psychometric Properties of the Classroom Assessment Scoring System (Pre-K): Implications for Measuring Interaction Quality in Diverse Early Childhood Settings. *Journal of Applied Measurement*. 18(3):299-318. PMID: 29579740.

Early Childhood Learning & Knowledge Center (ECLKC) (2020). Use of Classroom Assessment Scoring System in Head Start. U.S. Department of Health and Human Services. https://eclkc.ohs.acf.hhs.gov/designation-renewal-system/article/use-classroom-assessment-scoring-system-class-head-start

Hamre, B.K., Goffin, S.G., Kraft-Sayre, M., (2009). Classroom Assessment Scoring System (CLASS) Implementation Guide: Measuring and Improving Classroom Interactions in Early Childhood Settings. https://www.boldgoals.org/wp-content/uploads/CLASSImplementationGuide.pdf

Hightower, A. D., Work, W. C., Cowen, E. L., Lotyczewski, B. S., Spinell, A. P., Guare, J. C., et al. (1986). The teacher–child rating scale: A brief objective measure of elementary children's school problem behaviors and competencies. School Psychology Review, 15, 393– 409. https://doi.org/10.1080/02796015.1986.12085242

Howes C., Burchinal M., Pianta R., Bryant D., Early D., Clifford R., et al. (2008). Ready to learn? children's pre-academic achievement in pre-kindergarten programs. Early Child. Res. Q. 23 27–50. https://doi.org/10.1016/j.ecresq.2007.05.002

La Paro, K. M., Pianta, R. C., & Stuhlman, M. (2004). The Classroom Assessment Scoring System: Findings from the Prekindergarten Year. *The Elementary School Journal, 104*(5), 409–426. https://doi.org/10.1086/499760

Mashburn A. J., Pianta R. C., Hamre B. K., Downer J. T., Barbarin O. A., Bryant D., et al. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. Child Dev. 79 732–749. https://doi.org/10.1111/j.1467-8624.2008.01154.x

Pakarinen, E., Lerkkanen, M., Polkkeus, A., Kluru, N., Slekklnen, M., Raskuputtonen, H., & Nurmi, J. (2010). A validation of the Classroom Assessment Scoring System in Finnish kindergartens. *Early Education and Development, 21*, 95–124. https://doi.org10.1080/10409280902858764

Rodriguez, B. J., & Garza, S. (2014). Test Review: Classroom Assessment Scoring System Manual: ToddlerLa ParoK. M.HamreB. K.PiantaR. C. (2012). Classroom Assessment Scoring System Manual: Toddler. Baltimore, MD: Paul H. Brookes. *Journal of Psychoeducational Assessment, 32*(6), 573–576. https://doi.org/10.1177/0734282914523085

Styck, K. M., Anthony, C. J., Sandilos, L. E., & DiPerna, J. C. (2020). Examining Rater Effects on the Classroom Assessment Scoring System. Child Development. https://doi.org/10.1111/cdev.13460

Teachstone (n.d.) The CLASS System: Birth through Secondary. http://info.teachstone.com/hubfs/CLASS-System-Paper-FINAL.pdf?t=1495735615702

Thorpe, K., Rankin, P., Beatton, T., Houen, S., Sandi, M., Siraj, I., & Staton, S. (2020). The when and what of measuring ECE quality: Analysis of variation in the Classroom Assessment Scoring System (CLASS) across the ECE day. *Early Childhood Research Quarterly*, *53*, 274–286. https://doi.org/10.1016/j.ecresq.2020.05.003

Zaslow, M.Anderson, R.Redd, Z.Wessel, J.Tarullo, L. & Burchinal, M. (2010). Quality dosage, thresholds, and features in early childhood settings: A review of the literature (OPRE 2011–5). https://www.acf.hhs.gov/sites/default/files/documents/opre/quality_review_0.pdf

Recommended Citation:

Rangel-Pacheco, A. & Witte, A. L. (2021). *The Classroom Assessment Scoring System for PreK (CLASS-PreK) and Child Outcomes: An NeMTSS Research Brief.* Nebraska Multi-tiered System of Support (NeMTSS).

Authorship Information:

Abril Rangel-Pacheco, M.A.
School Psychology Doctoral Student
Graduate Research Assistant
Nebraska Center for Research on Children, Youth, Families and Schools
University of Nebraska–Lincoln
arangel-pacheco2@huskers.unl.edu