This rejoinder was authored by members of the CQEE and included the collaboration of Dr. Marlene Zepeda, Dr. Zulmara Cline, Dr. Janet Oh, Dr. Carola Matera, Dr. Linda Espinosa, Dr. Lisa Lopez, Antonia Lopez, Rachelle Pastor, Diane Ujiuye, Elizabeth Crocker, Ofelia Medina, Fran Chasen and Dawn Kurtz.
The Campaign for Quality Early Education (CQEE) is a group of California-based organizations and individuals who have successfully impacted California’s early learning public policies (including child and program assessments, frameworks, curricula, professional development, family engagement, data and research, evaluation, and workforce development). Over the past two years, the CQEE has monitored and raised concerns about the implementation of the early education observation tool CLASS (Classroom Assessment Scoring System) in early care and learning environments where children do not understand or speak English or who are at varying levels of English proficiency (Dual Language Learners – DLLs).

The Campaign for Quality Early Education (CQEE) recognizes that there are many positive aspects of the CLASS that may make it a valid classroom rating tool in early care and learning environments where the children and teachers are monolingual English speakers. However, the CQEE is very concerned with the underrepresentation of behavioral indicators that reflect the language and culture of DLLs. If we are striving to improve the quality of pedagogical practice for this growing population of children, classroom assessment must incorporate the specific linguistic and cultural dimensions of instruction, assessment and teacher-child interaction that are important for the learning of DLLs.

Recently Teachstone has issued a brief, Dual Language Learners and the CLASS™ Measure: Research and Recommendations, by Virginia E. Vitiello, Ph.D. CQEE applauds Teachstone for attempting to address issues and concerns raised by DLL advocates, however, their responses are not adequate as there are still several major concerns regarding the appropriateness of CLASS for DLL.

The Teachstone brief lays out a rationale for the use of the CLASS in classroom observations that include young DLLs. While there are a number of areas where we agree with the author, there are many where the justification for using the CLASS to evaluate the quality of teacher interaction with DLL children falls short. Our primary concerns focus on: 1) the contradictory argument that human behavior is universal while also stating that culture influences how individuals interact in the teaching-learning context; and 2) discussion of research cited to support the utility of the CLASS with DLL populations which fails to note important aspects of the research that would lead to a more nuanced and qualified set of conclusions with respect to DLL populations. What follows is a rejoinder to this brief.

---

**IS GOOD FOR ALL, GOOD FOR DLLS?**

The author states that there is an emerging agreement in early childhood that DLLs profit from a set of teacher behaviors that are good for all children regardless of their language and culture. Specifically, the brief states:
However, there is a growing consensus that dual language learners also benefit from the same fundamental experiences as monolingual students, including warm, supportive, and stimulating interactions with their teachers. Looking at social and emotional development, a study of diverse kindergartners found that the quality of teacher-child relationships was more strongly predictive of children’s skills than was their status as dual language learners (Howes et al., 2011).

Although we agree that there is a basic set of universal adult behaviors that are important for healthy child development, there is also a growing recognition in developmental psychology and early childhood education that adult behaviors and interactions with children are highly influenced by culture, which, in turn, are represented/manifested in language interactions (Eisenberg, 1999; Rogoff, 1990, Vygotsky, 1978). To be clear, the Howes et al. (2011) research cited in the brief focused on one aspect of social emotional development: peer interaction. In this study, the researchers found that the closeness of teacher-child relationships was positively associated with positive peer group relations regardless of ethnic background. However, the status of children as DLLs was not clearly distinguished; the sample included 143 Latino origin children, 57% of whose mothers reported that their children spoke Spanish in the home. The Teachstone brief suggests that language does not matter in this case because language proficiency of the child in either Spanish or English did not contribute to the relationship between teacher closeness and competence in peer relations. However, the authors themselves caution against making strong generalizations from this work because the sample and effect size were small. In fact, the authors say that the relationship was not strongly predictive; what is strongly predictive was the teacher closeness measure with specific child characteristics such as being anxious-withdrawn, aggressive or victims of aggression. Furthermore, Howes et al., (2011) point to the importance of having a differentiated view of social development that considers aspects of the social environment. Specifically, they say,

“Theoretical formulations to explain and understand children’s development must be sufficiently complex to account for influences of individual social competence, dyadic and group processes, and dimensions of the social context.” (p.399)

It is important to note that there is some empirical evidence that places the assertion that a child’s primary language does not matter in relation to the quality of the teacher-child relationship in doubt. In a study examining teacher’s use of English and Spanish with DLL children, Chang et al. (2007) found that teachers who spoke some Spanish rated Spanish-speaking children in their classrooms as having better social skills when compared with teachers who spoke only English in the classroom. This study also showed that when the teacher spoke Spanish in the classroom, DLL children experienced less bullying by peers when compared to teachers who only spoke English.

One of the points derived from a universalistic perspective of child development is the belief that good quality for all children will result in good results regardless of the background characteristics of the children. It is stated:

In reviewing evidence on teaching dual language learners, one researcher writes, “good instruction for students in general tends to be good instruction for ELLs [English language learners] in particular” (Goldenberg, 2008, p. 17).
Although we agree that good instruction for children in general tends to be good instruction for DLLs, the statement attributed to Goldenberg is incomplete and, as a result, misleading. The brief neglects to mention the second part of the statement where Goldenberg specifically says that accommodations must be made to address the needs of DLLs. Goldenberg writes,

“What we know about good instruction and curriculum in general holds true for English learners as well; but when instructing English learners in English, teachers must modify instruction to take into account students’ language limitations” p. 18 (emphasis added)

**TECHNICAL ADEQUACY OF THE CLASS FOR DLLS**

Teachstone points to two investigations that suggest that the CLASS is a reliable and valid measure for use in classrooms with DLL children. The first is a preliminary study of reliability comparing reliability scores from English-only videos and scores from presumably Spanish-only videos. Researchers were interested in whether CLASS observations can be conducted reliably in dual language classrooms (Vitiello, 2012). They looked at CLASS scores from 2,000 videotapes that had been coded by more than one observer. Overall, there was no significant difference in reliability on the English-only videos versus the videos that used a language other than English, most of which used English and Spanish. This provides preliminary evidence that the CLASS measure can be used reliably to observe interactions in linguistically diverse classroom.

With respect to reliability, tests can achieve acceptable levels of reliability by training assessors to utilize one set of expectations from which to make interpretations. In other words, a test can be reliable but not necessarily valid (i.e., measuring what it is intended to measure). There is a history in this country of assessments and measures demonstrating adequate reliability that turned out to be culturally and/or racially biased because they utilized a mainstream or middle-class view of what constituted an appropriate response (Olmedo, 1981; Samuda, 1998). Asserting that the CLASS is reliable in DLL classrooms is problematic because its indicators, based on universals regarding teacher behavior, do not account for and may even penalize teacher behaviors that are culturally responsive, let alone linguistically responsive. For example, in the domain of Emotional Support there is an item labeled “regard for students’ perspectives” where, ideally, a teacher is supposed to respect the child’s perspective and demonstrate flexibility, incorporate the child’s ideas and follow their lead. However, if the teacher does not comprehend or speak the language of the child, how can they understand the child so that they can incorporate the child’s ideas?

The second study mentioned in the brief to support the reliability and validity of the CLASS with DLLs is the work of Downer, Lopez, Grimm, Hamaguri, Pianta et al., (2012). Although CLASS was not developed or normed specifically for DLLs, the developers contend that the structure and predictive validity of the CLASS applies equally well across preschool classrooms with different DLL compositions at the end of preschool.
Results showed that the scores predicted gains in children’s language/literacy, math, social competence, and letter naming, as has been reported in previous studies (e.g., Mashburn et al., 2008). Effects were the same for dual language learners and monolingual English speakers. These results suggest that the CLASS measure can be used validly in dual language classrooms. The structure of the measure held true across diverse groups of classrooms, average scores did not differ across the classrooms, and CLASS scores predicted outcomes for dual language learners and non dual language learners alike. These findings support the use of the CLASS tool in linguistically diverse preschool settings.

There are important limitations to this claim. This investigation is based on a large, multi-state study in which DLLs were not the main focus of the research. As a result, and as reported by the study, DLL status was not well defined, no information was reported regarding the bilingual capacity of the assessors and data are only reported for child outcomes at the end of preschool. When DLL children are the primary focus of the relationship of CLASS ratings to child outcomes, there is emerging data that suggests that there is no association between all subscales of the CLASS and child outcomes in Spanish (Lopez, Arango, & Ferron, 2012) and no association between any subscales of the CLASS and child outcomes in English for Latino DLLs in preschool (Lopez, personal communication). The concern about the use of the CLASS for DLL children is that the indicators are not reflective of specific strategies needed to optimize child outcomes for DLL children.

Another point made by the brief is that the CLASS can be used in culturally diverse classrooms because it has been used in many other countries.

The CLASS system has also been used in projects in Puerto Rico, Chile, and in Canada with French speaking teachers. These studies and projects suggest that the CLASS system can be used validly in culturally diverse classrooms.

This proposition is based on faulty reasoning. In order for this assertion to be valid, the definition of cultural diversity would have to be similar to that within US classrooms. How does the use of the CLASS in another country (primarily with linguistic majority populations) justify the use of the CLASS with a linguistic minority population residing in the US whose histories and current educational realities may be entirely different? Generalizing from a foreign country whose dynamics of cultural and language politics vary considerably from that of the US is simply not generalizable to linguistic minorities residing in the US.

**CONCLUSION**

However, if there are elements of instruction that are considered critical to the cultural group that are not included in the CLASS system, it may be important to supplement CLASS data with information from other sources.

Vitiello, p. 9
It is estimated that 36% of all entering kindergarten children in California come from homes where English is not the primary language (American Institute of Research, 2012). California educates 1/3 of all dual language learners in the US which is approximately 5.3 million children or 9 to 10% of the US school population (Migration Policy Institute, 2010). According to an analysis of data from the UCLA Center for Health Public Policy Institute by Children Now, a children’s policy organization in California, there are 1.6 million or 57% of California children under the age of five who live in the home where a language other than English is spoken (American Institute of Research, 2012).

In many parts of California, young DLLs constitute the majority of children in early care and learning environments. With the advent of Quality Rating Improvement Systems (QRIS) there is increasing interest in identifying program and classroom assessments that can be used to improve the quality of education for young DLLs and increase academic achievement. The CLASS has gained widespread support as a measure that can assist early educators in such important areas as classroom organization, emotional climate and instructional support. However, if the goal of QRIS efforts are to improve quality in early care and learning environments yet the needs of DLL children for linguistically and culturally appropriate pedagogy are not adequately considered, then QRIS efforts will be unsuccessful for a large proportion of the child population.

The primary criticism of CLASS for use with DLL populations is that it has incomplete construct validity (Messick, 1989). By overlooking how language and culture can affect teacher and child interactions the overall construct of classroom quality is not captured by the CLASS. This may be partially the reason for the lack of associations located by Lopez, Arango and Ferron (2012). In early childhood education there is a history of environmental measures demonstrating differential predictive validity for different population groups. A case in point is the Home Observation for Measurement of the Environment Inventory (HOME) that examined child outcomes in African-American, White and Mexican American children. It was found that the HOME significantly predicted cognitive performance in African-American and White children but for Mexican American children, some of whom were likely DLL, the results were not significant (Bradley et al., 1989). Clearly, the need for more empirical evidence regarding the use of the CLASS with DLL populations is needed, however, there is a pressing need to exercise caution when using the CLASS in culturally and linguistically diverse learning settings.

Given the present circumstances, CQEE proposes that if the CLASS is used, it be supplemented by an assessment that is developed specifically for DLL populations and incorporates the specific elements of quality important to the learning of young DLLs (Castro, Espinosa, & Paez, 2011). In other words, as noted by early childhood assessment experts, program assessments need to be multidimensional and appropriate for both the purpose of the evaluation as well as the population assessed (Espinosa, 2006;Meisels, & Atkins-Burnett, 2006; National Research Council, 2008). This assessment principle is even more critical in light of current efforts to improve quality for all and the urgent need for improved instruction for DLLs, a growing and significant proportion of young children.
References


