

BRIEF REPORT

Reliability and Validity of a Spanish Language Assessment of Children's Social-Emotional Learning Skills

Jaclyn M. Russo
University of Virginia

Clark McKown, Nicole M. Russo-Ponsaran,
and Adelaide Allen
Rush University Medical Center

Few Spanish language tools are available for assessing important social-emotional learning (SEL) skills. The present study presents evidence of the psychometric properties of a Spanish-language version of SELweb (SELweb-S), a web-based system for assessing children's ability to recognize others' emotions and perspectives, solve social problems, and engage in self-control. With a sample of 524 students in Grades K to 3, we examined the reliability and validity of SELweb-S. This study provided evidence that (a) individual assessment modules exhibited moderate to high internal consistency and moderate 6-month temporal stability, (b) composite assessment scores exhibited high reliability, (c) assessment module scores fit a theoretically coherent factor structure, and (d) performance on SELweb-S assessment modules was positively related to teacher-reported SEL skills. Findings are discussed in terms of the importance of direct assessments of SEL skills in languages other than English. In addition, we highlight the importance of abiding by rigorous recommendations in the literature for the translation and cultural adaptation of assessments.

Public Significance Statement

This study advances the direct assessment of social-emotional learning (SEL) skills in school-age children by presenting evidence of reliability and validity of a Spanish-language version of SELweb (SELweb-S), a web-based direct assessment of SEL skills for children in Grades K to 3. This study highlights the importance of developing direct assessments of SEL skills in languages other than English.

Keywords: direct assessment, social-emotional comprehension, transadaptation, reliability, validity

Social-emotional learning (SEL) skills include the thinking skills and behaviors needed to create positive connections. A large body of research has shown that SEL skills in childhood are related

to later academic, behavioral, mental health, and even economic outcomes (e.g., McClelland, Acock, & Morrison, 2006; Sabol & Pianta, 2012). Reflecting the increased focus on SEL, four states (Illinois, Kansas, Pennsylvania, and West Virginia) have adopted comprehensive SEL standards, and more than 10 states have incorporated aspects of SEL into their K to 12 standards (CASEL, 2015). For educators committed to nurturing children's SEL skills, many evidence-based SEL programs have been created (Dusenbury, Calin, Domitrovich, & Weissberg, 2015). However, one area lags behind. Although standards and programs are proliferating, there are few usable, scalable, and informative systems for measuring important SEL skills. Even fewer assessment systems have been created for non-English-speaking students. The purpose of this study is to describe the psychometric properties of a Spanish-language translation of SELweb (SELweb-S), a web-based system for directly assessing important SEL skills (McKown, Russo-Ponsaran, Johnson, Russo, & Allen, 2016).

SEL skills include mental processes involved in the comprehension of social-emotional information and behaviors involved in the execution of social interactions (Lipton & Nowicki, 2009). Social-emotional (SE) comprehension includes the ability to recognize others' emotions ("social awareness") and perspectives ("social meaning"), solve social problems ("social reasoning"), and some

This article was published Online First June 19, 2017.

Jaclyn M. Russo, Center for Advanced Study of Teaching and Learning, University of Virginia; Clark McKown, Nicole M. Russo-Ponsaran, and Adelaide Allen, Department of Behavioral Sciences, Rush University Medical Center.

This research was supported by the Institute of Education Sciences through Grant R305A110143 to Rush University Medical Center. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. Clark McKown is the founder and president of xSEL Labs, a company that licenses and sells SELweb. This financial interest began after this article was accepted for publication. The authors are grateful for the help of the students, teachers, administrators, and field staff who participated in this study; the members of the translation committee; and voice actor Johnny Arena. We also thank Jason K. Johnson for his significant contributions to this project.

Correspondence concerning this article should be addressed to Jaclyn M. Russo, Center for Advanced Study of Teaching and Learning, University of Virginia, 405 Emmet Street South, Ruffner Hall, Room 221, Charlottesville, VA 22904. E-mail: jr4ym@virginia.edu

aspects of self-control. SEL skills develop early in childhood (Blair & Raver, 2015) and are strongly and positively correlated with age (McKown, Allen, Russo-Ponsaran, & Johnson, 2013). Furthermore, SEL skills are vital in building a child's foundation for later learning in school (Blair & Raver, 2015). Specifically, self-control and social competence significantly predict initial levels of math and reading scores between kindergarten and second grade, and between third and sixth grade after controlling for child IQ, age, ethnicity, and maternal education level (McClelland et al., 2006).

Further, more prosocial engagement in first grade is associated with higher math achievement in fourth grade (Luo, Hughes, Liew, & Kwok, 2009). Given the strong link between SEL skills and children's academic achievement, it is surprising that there are few tools to directly assess SE comprehension in schools. Most existing tools for assessing SEL skills that are appropriate for use in schools are teacher-report measures that are well suited to assessing observable social-emotional behavior. Those tools are not well suited to measure the mental processes that make up SE comprehension because they require observers to make a high level of inference (Denham, 2006; Ringwalt, 2008). Direct assessments, in which children solve tasks that reflect comprehension skills, are better suited to assess the mental processes that make up SE comprehension (McKown et al., 2013).

Addressing this gap, SELweb is a web-based system designed to assess children's SE comprehension from kindergarten through third grade. Five assessment modules directly assess children's ability to recognize others' emotions and perspectives, solve social problems, and engage in self-control. To measure emotion recognition skill, for example, children view pictures of faces and indicate what each child is feeling from their facial expression. Two large field trials of the English version of SELweb found that (a) SELweb scores exhibited high reliability, (b) SELweb modules demonstrated a theoretically coherent factor structure, (c) factor scores demonstrated convergent and discriminant validity, and (d) controlling for IQ and demographic characteristics, performance on the assessment modules was positively associated with peer acceptance, teacher report of social skills, and multiple indicators of academic achievement, and negatively related to teacher report of problem behaviors (McKown et al., 2016). SELweb provides a feasible and psychometrically sound assessment of SE comprehension in children.

However, the absence of direct assessments of SEL in languages other than English means that little is known about non-English-speaking children's SE comprehension skills and reflects a limitation in the area of SEL assessment. To the best of the authors' knowledge, no research exists describing the development of assessments of SE comprehension skills in non-English-speaking populations. However, during the 2013–2014 school year, 76.5% (approximately 3.8 million students) of all English-language learners spoke Spanish, accounting for 7.7% of all public K–12 students. Thus, Spanish is the most prevalent language next to English in American schools today (Kena et al., 2016). To expand its reach and benefit, SEL direct assessments should therefore be available in Spanish.

Translating an assessment from one language to another requires a multistep process (Bracken & Barona, 1991). Simply translating and back translating an assessment may result in accurately translated instructions, but it does not ensure the translation

will be culturally appropriate (Peña, 2007). Instead, translation should be accompanied by a cultural adaptation process so that questions are familiar and appropriate for the students completing the assessment (Peña, 2007). To achieve an appropriate translation, a committee of fluent speakers should assess the cultural relevance of the translated assessment items, recommend modifications, and come to consensus about the wording of translated items; this process is called *transadaptation* (van Widenfelt, Treffers, de Beurs, Siebelink, & Koudijs, 2005).

In light of the rapidly growing Spanish-speaking population in the United States and increased awareness of the need for strong SEL assessment in young children, we translated, culturally adapted, and validated a Spanish version of SELweb ("SELweb-S"). The purpose of this article is to present evidence of the psychometric properties of SELweb-S. Confirmatory factor analyses (CFAs; Arbuckle, 2012) were used to test the hypothesis that scores from SELweb-S would reflect four moderately correlated factors. Cronbach's alphas were calculated to test the hypothesis that factor score reliabilities would exceed .80 for all SELweb-S modules. Finally, two-level hierarchical linear models (HLMs; Raudenbush, Bryk, & Congdon, 2004) were run to show that students' scores on SELweb-S exhibited criterion-related validity, as demonstrated by a positive association between age and SELweb-S scores, and between performance on SELweb-S and teacher report of social skills.

Method

Participants

Data were collected in the fall of 2013 from three urban elementary schools in a large Midwestern metropolitan area; the Rush University Institutional Review Board approved all study procedures. Participants included 524 native Spanish-speaking students in bilingual classrooms in kindergarten through third grade (32.3%, 24.6%, 20.0%, and 22.7%, respectively). Of the 524 students, 282 were boys (53.8%) and children were, on average, 7.16 years old ($SD = 1.09$ years).

Measures

SELweb. SELweb modules are described briefly in the following section. A detailed description of the administration protocol and scoring procedures can be found in McKown et al. (2016). Group administration occurred in school settings, either in a computer lab or in classrooms on laptops. Children completed the assessment independently, on average taking 40 to 45 min to complete.

Social awareness. Six photographs of child faces with neutral facial expressions, including three girls and two ethnic minorities, were used to create facial emotion recognition stimuli. With FaceGen software (Singular Inversions, 2005), digitized photographs were altered into high-intensity displays of happy, sad, angry, and frightened. For each face and emotion, 10 faces were created, ranging from low- to high-intensity affect displays, forming a pool of 240 images, or items. From this item pool, four test forms were created, each with 40 items, and children randomly received one of the four test forms. Item difficulties were determined from an initial item tryout with 1,239 children using Rasch modeling (Item

difficulty range = -3.2 to 5.1 logits). Faces were assigned to test forms to ensure a balance of emotions, item difficulties, and child faces within a given form. After each face was presented, children clicked to indicate whether the face reflected happy, sad, angry, scared, or just okay. To adjust for differences in test form difficulties and thereby equate scores, item scores were scored, summed, and standardized within form.

Social meaning. Consistent with existing measures (e.g., Happé, 1994), we created 12 illustrated and narrated vignettes in which a character is disappointed, scared, sarcastic, lying, hiding feelings, or harboring a false belief. After each story, children were asked a question whose correct answer required accurate inferences about the story character's mental state. Item scores were summed across vignettes and standardized within form to place scores from the different forms on a common metric.

Social reasoning. We created five illustrated and narrated vignettes involving ambiguous provocation situations and five involving peer entry. Each form also included three vignettes with girls and three with boys as protagonists. After each vignette, children selected from multiple-choice options: (a) an attribution about the problem's origin (benign or hostile intent), (b) a social goal, and (c) solution preference. Scores for each question were summed across vignettes and standardized within test form to equate scores. To reduce respondent fatigue, we created five test forms with six vignettes each. Each form included three ambiguous provocation vignettes and three peer entry vignettes. Each vignette was included on three forms. Children randomly received one of five test forms.

Self-control. We developed a choice-delay task (Kuntsi, Stevenson, Oosterlaan, & Sonuga-Barke, 2001) and a frustration-tolerance task (Bitsakou, Antrop, Wiersema, & Sonuga-Barke, 2006). For the choice-delay task, children earned points for sending rocket ships to space; the slower and therefore more tedious the rocket ship, the more points the child earned. Children had 10 opportunities to score as many points as possible. For the frustration-tolerance task, children were told to identify whether or not two shapes were identical and to get as many correct as they could in 90 s. For several items, the computer was programmed to get "stuck" before moving on to the next item, thereby inducing mild frustration.

The Devereux Student Strengths Assessment (DESSA)-Mini. Teachers completed one of four equivalent forms of the DESSA-mini (Naglieri, LeBuffe, & Shapiro, 2011), an 8-item questionnaire in which the frequency of behaviors reflecting SEL skills are rated. Items are scored on a 5-point scale (i.e., never to very frequently). The DESSA-mini provides one overall social-emotional competence score, the Social-Emotional Total (Naglieri et al., 2011). The average internal consistency across forms was $\alpha = .88$ (range $\alpha = .85$ to $.91$).

Procedures

The English script, including all spoken dialogue and any print text, was translated into Spanish and then back to English by the Rush University Medical Center translation service. Next, the study authors reviewed the English script and its back-translation for accuracy. Specifically, study authors independently reviewed the entire script and made detailed notes where they believed the translation deviated from the original intent of the script. Having

no bilingual members of the research team, a translation committee was assembled, as described in the following section.

Following the recommendations of van Widenfelt et al. (2005), the authors convened a translation committee of five members. Committee members were bilingual in English and Spanish and held positions as psychologists, school administrators, or teachers. Because they occupied different professional roles and had different experiences as native Spanish speakers, each committee member brought an important perspective to the discussion of translations.

Each committee member reviewed the translated script prior to a group meeting. Members were asked to rank each translated line on a 4-point Likert scale of how well the line maintained the meaning and intent of the original line ($4 = \text{good with no reservations}$, $3 = \text{good, but alternatives are possible}$, $2 = \text{okay, but could be improved}$, $1 = \text{needs improvement}$). Any line receiving an average score of 2 or lower was flagged as unacceptable and prioritized to discuss with the group during the meeting. In addition, any line rated lower than a 4 was also edited and approved by the committee. Discrepant ratings were resolved in two 3-hr group meetings, starting with lines rated as a "1" or "2" and finishing with those rated a "3." The project team edited and finalized the script according to the decisions made by the translation committee.

Finally, a bilingual voice actor recorded the script. Project team members were present for all recording sessions. After all lines were recorded, they were processed and incorporated into the system. The translation committee reviewed the completed Spanish version and all members agreed that the assessment was ready for administration.

Results

Missing Data

Teachers completed DESSA-mini questionnaires for those children whose parent(s) provided consent to participate in an add-on study (i.e., 129 of the 524 children who participated in the study). Children with and without DESSA-mini scores did not differ significantly in sex, age, perspective-taking skills, problem-solving skills, or self-control. Children whose teachers completed the DESSA-mini scored .19 standard deviation higher on the emotion recognition task than children without DESSA-mini scores ($p < .05$).

Reliability

Internal consistency. To test reliability hypotheses, we calculated the internal consistency of scores. For Social Meaning, a single score measure, we computed Cronbach's alpha. Social Awareness, Social Reasoning, and Self-Control scores were derived from multiple forms and scores. Internal consistency coefficients for each score were calculated and averaged across forms. Factor score reliabilities were estimated using procedures described by Nunnally and Bernstein (1994, p. 271). Composite score reliabilities were deemed acceptable and ranged from .67 to .94. Comprehensive score reliabilities are summarized in Table 1 (i.e., r_{yy}).

Table 1
Score Reliabilities

SELweb composite score	r_{yy}^a	r_{12}^b
Social-Emotional (SE) Comprehension	.94	.73
Social Awareness	.91	.56
Social Meaning	.67	.68
Social Reasoning	.85	.64
Self-Control	.81	.51
SELweb module and score	α	r_{12}
Social Awareness		
Happy	.71	.36
Sad	.75	.43
Angry	.68	.45
Scared	.81	.49
Social-Meaning/Perspective-Taking	.67	.68
Social Reasoning		
Positive Attribution	.66	.54
Positive Social Goal	.68	.53
Positive Solution Selection	.77	.54
Self-Control		
Delay of Gratification	.63	.42
Frustration Tolerance	.84	.42

Note. SEL = social-emotional learning.

^a Internal consistency of SELweb composite and module scores. ^b 6-month stability of SELweb composite and module scores.

Six-month stability. From the total sample, 209 children completed SELweb a second time in the spring. Mean interval between testing was 176 days (range = 183 to 186 days). We used these data to evaluate 6-month measurement stability, and findings are presented in Table 1 (i.e., r_{12}). Because children were randomly assigned to Social Awareness and Social Reasoning test forms, for those assessment modules, temporal-stability estimates reflected a mix of alternate forms and test-retest reliability. Six-month stability scores were moderate with the Social Reasoning composite score demonstrating the strongest estimate of temporal stability (i.e., .68).

Validity

Factor structure. We hypothesized that SELweb-S scores would fit a four-factor model of social-emotional comprehension in which Social Awareness, Social Meaning, Social Reasoning, and Self-Control reflect correlated latent variables (see McKown et al., 2016). To test this hypothesis, we used Amos 21.0.0 (Arbuckle, 2012) to run confirmatory factor analyses (CFAs). For the single-indicator latent variable reflecting perspective-taking, the error variance was modeled as 1 minus the internal consistency of the indicator variable score (Kline, 2005). We compared our hypothesized model with three plausible alternatives to evaluate the extent to which our hypothesized model was superior to those alternatives. Because simpler CFA models were nested within more complex models, the change in chi-square goodness of fit was used to compare the fit of the models with one another (Kline, 2005).

The fit of a one-factor model, in which all scores loaded on a single factor, was poor (comparative fit index [CFI] = .78, root mean square error of approximation [RMSEA] = .130, 90% confidence interval [CI] [.117, .142]). Because emotion recogni-

tion and perspective-taking both reflect understanding others, we next tested a two-factor model with Emotion Recognition and Perspective-Taking scores loading on one factor, and the Social Problem-Solving and Self-Control scores on a second factor. Model fit was again poor (CFI = .76, RMSEA = .136, 90% CI [.124, .149]). Because problem solving and self-control may be distinct from one another, we next tested a three-factor model, with Emotion Recognition and Perspective-Taking scores loading on one factor, Self-Control scores on a second factor, and Social Problem-Solving scores on a third factor, and the fit was good (CFI = .91, RMSEA = .078, 90% CI [.064, .092]). The fit of the four-factor model was excellent (CFI = .99, RMSEA = .036, 90% CI [.017, .053]). That model, depicted in Figure 1, was a significantly better fit to the data than the alternatives (all comparisons with the four-factor model, $\Delta\chi^2/df > 40$, $p < .05$).

Social Awareness, Social Meaning, Social Reasoning, and Self-Control loaded on a second-order Social-Emotional Comprehension latent variable (CFI = .97, RMSEA = .034, 90% CI [.015, .051]). The Social-Emotional Comprehension factor score from this analysis ("SE composite") was computed as a weighted sum, with factor loadings as the weighting value. This score was used in criterion-related validity analyses. Along with factor scores reflecting Social Awareness, Social Meaning, Social Reasoning, and Self-Control, we use the SE composite as an overall indicator of social-emotional comprehension in criterion-related validity analyses.

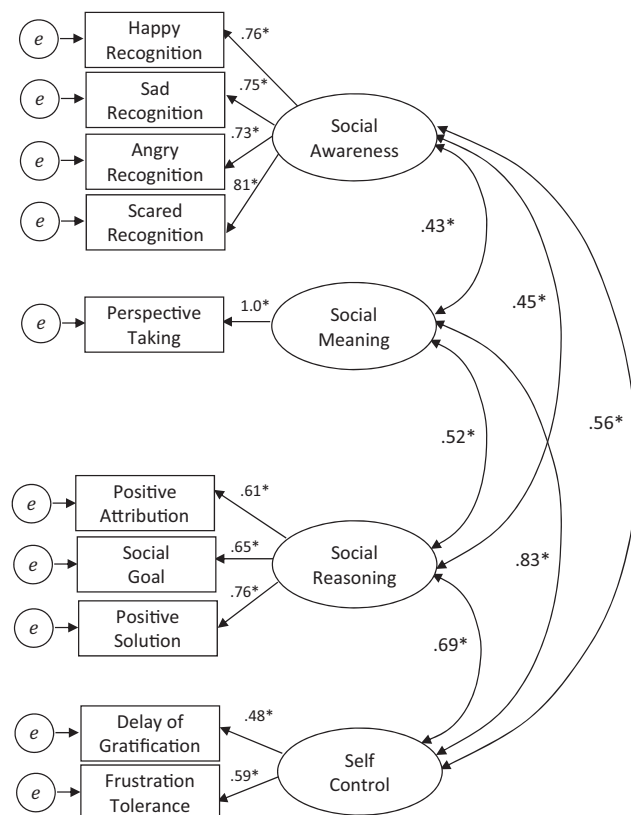


Figure 1. Confirmatory model of social-emotional comprehension. CFI = comparative fit index; RMSEA = root mean square error of approximation; CI = confidence interval. * $p < .05$.

Criterion-related validity. We ran two kinds of analyses to test criterion-related validity hypotheses. First, we ran two-level HLMs (Raudenbush et al., 2004), with classroom as the Level 2 unit, evaluating the relationship between age and each social-emotional comprehension score. We found that age demonstrated a curvilinear relationship related to social-emotional comprehension scores as shown in Table 2. Specifically, children's scores in social awareness, social meaning, social reasoning, self-control, and overall SELweb-S performance increased with age. In addition, the age-related increase in these skills declined with age. Figure 2 shows that each of the skills measured by SELweb-S improved with age, but the slope of the relationship between age and performance also leveled off with age.

To further evaluate criterion-related validity, we ran two-level HLMs evaluating the relationship between each social-emotional comprehension score and the DESSA-mini *T* score, controlling for age and sex. Social-emotional comprehension scores and the DESSA-mini *T* score were standardized for these analyses so that coefficients reflecting the association between social-emotional comprehension and DESSA-mini scores can be interpreted as standardized regression coefficients. As shown in Table 2, each of the social-emotional comprehension scores was significantly associated with teacher report of social-emotional comprehension.

Discussion

SELweb addresses a need for direct assessments of SEL skills appropriate for use in schools. In consideration of the fast-growing population of native Spanish-speaking students in the United States, the authors developed and field tested a Spanish-language version of SELweb. As such, the purpose of this study was to describe the process by which SELweb was transadapted into Spanish and to present evidence of the reliability and validity of SELweb-S.

Using the transadaptation process, we translated and culturally adapted a complex developmental assessment. Study findings highlight that SELweb-S yields reliable scores that are valid for their intended purpose. Specifically, results supported our hypotheses that (a) individual assessment modules exhibited moderate to high internal consistency and moderate 6-month temporal stability, (b) composite assessment scores exhibited high reliability, (c) assessment module scores fit a theoretically coherent factor struc-

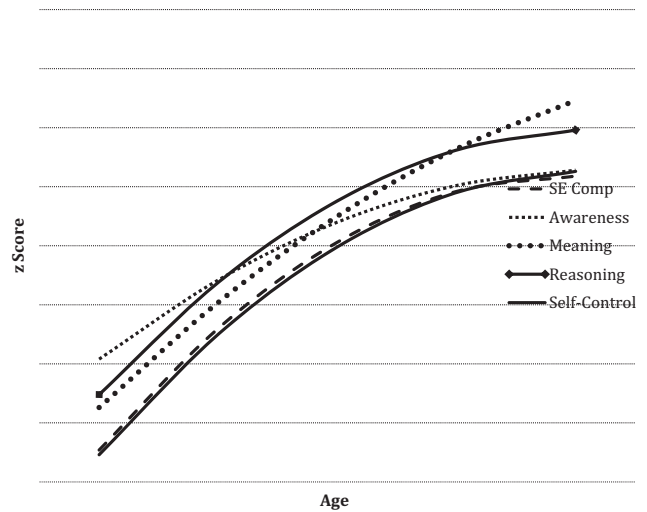


Figure 2. Model-inferred relationship between age and social-emotional comprehension. SE = social-emotional.

ture, and (d) performance on SELweb-S assessment modules was positively related to teacher-reported SEL skills. Additionally, our HLM results provide users of SELweb-S with confidence that the translated version of this assessment is measuring what it was designed to measure (i.e., SE comprehension) within a population of native Spanish-speaking children. Additionally, and aligned with results from our English-speaking population of children, the association between the teacher report (i.e., DESSA-mini) and SELweb-S was moderate indicating that the SEL skills measured through SELweb-S are distinct from the behavioral components of SEL as measured through teacher report. Finally, our results highlighted a strong relationship between age and SELweb-S performance across SEL skills with their age-mates. Overall, our study produced a psychometrically sound Spanish language direct assessment of social-emotional comprehension, the first to our knowledge that is available for use at scale in schools. Importantly, SELweb-S allows teachers to directly, feasibly, and validity assess their Spanish-speaking students' SEL skills without needing to speak Spanish themselves.

Although there were no significant differences in SEL skills as measured by SELweb-S between the 129 children whose parents consented and those children whose parents did not consent, a low consent rate reflects a limitation of our study. Additionally, our study design did not include a version of the assessment that was translated by other means, so we cannot say definitively that the transadaptation process is superior to other translation methods. However, given the strong evidence of score reliability and validity, we are confident that this method of translation is highly effective. The rigorous requirements of the transadaptation process may deter many test developers from using it to translate tests into another language. However, we believe that following these recommendations is likely to result in better measurement systems and, as such, we strongly suggest that other test developers adhere to this procedure.

In addition, our analyses did not account for the nesting of children in classrooms and schools. Because the degrees of free-

Table 2

Hierarchical Linear Models of the Association Between Social-Emotional Comprehension and DESSA-mini T Score, Controlling for Age and Sex

Variable	SELweb predictors				
	SELweb overall	Social awareness	Social meaning	Social reasoning	Self-control
Intercept	-.12	-.14	-.13	-.10	-.11
Age	-.08	.04	-.07	.01	-.10
Sex	.14	.19	.20	.18	.15
SELweb score	.36*	.25*	.28*	.20*	.37*

Note. SEL = social-emotional learning; DESSA-mini = Devereux Student Strengths Assessment Mini. Coefficients are standardized.

* $p < .05$.

dom in the models exceeded the number of classrooms, we could not use a more advanced modeling strategy such as the Complex Samples Facility in *Mplus* (Muthén & Muthén, 2012). In the future, it will be important to replicate this work with either larger samples of classrooms or independent samples of children.

Conclusion

The push for reliable and valid direct assessments of social-emotional comprehension that are appropriate for use in schools continues to grow. As a result, educational researchers aim to create easily implementable direct assessments. SELweb was created to address this need. To the best of our knowledge, this study is the first to translate and culturally adapt a computer-automated, scalable, direct assessment of social-emotional comprehension into Spanish. As such, we were able to provide important information to teachers for a group of students who otherwise would not have been able to complete the assessment.

References

- Arbuckle, J. L. (2012). Amos (Version 21.0.0). Chicago, IL: SPSS, Inc.
- Bitsakou, P., Antrop, I., Wiersma, J. R., & Sonuga-Barke, E. J. (2006). Probing the limits of delay intolerance: Preliminary young adult data from the Delay Frustration Task (DeFT). *Journal of Neuroscience Methods*, 151, 38–44. <http://dx.doi.org/10.1016/j.jneumeth.2005.06.031>
- Blair, C., & Raver, C. C. (2015). School readiness and self-regulation: A developmental psychobiological approach. *Annual Review of Psychology*, 66, 711–731. <http://dx.doi.org/10.1146/annurev-psych-010814-015221>
- Bracken, B. A., & Barona, A. (1991). State of the art procedures for translating, validating and using psycho-educational tests in cross-cultural assessment. *School Psychology International*, 12, 119–132. <http://dx.doi.org/10.1177/0143034391121010>
- CASEL. (2015). *State scan scorecard project*. Retrieved from <http://www.casel.org/state-scan-scorecard-project>
- Denham, S. (2006). Social-emotional competence as support for school readiness: What is it and how do we assess it? *Early Education and Development*, 17, 57–89. http://dx.doi.org/10.1207/s15566935eed1701_4
- Dusenbury, L., Calin, S., Domitrovich, C., & Weissberg, R. (2015). *What does evidence-based instruction in social and emotional learning actually look like in practice?* Retrieved from <http://www.casel.org/state-standards-for-social-and-emotional-learning/>
- Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, 24, 129–154. <http://dx.doi.org/10.1007/BF02172093>
- Kena, G., Hussar, W., McFarland, J., de Brey, C., Musu-Gillette, L., Wang, X., . . . Dunlop Velez, E. (2016). *The Condition of Education 2016 (NCES 2016-144)*. Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from <http://nces.ed.gov/pubsearch>
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford Press.
- Kuntsi, J., Stevenson, J., Oosterlaan, J., & Sonuga-Barke, E. (2001). Test-retest reliability of a new delay aversion task and executive function measures. *British Journal of Developmental Psychology*, 19, 339–348. <http://dx.doi.org/10.1348/026151001166137>
- Lipton, M., & Nowicki, S. (2009). The social-emotional learning framework (SELF): A guide for understanding brain-based social-emotional learning impairments. *The Journal of Developmental Processes*, 4, 99–115.
- Luo, W., Hughes, J. N., Liew, J., & Kwok, O. (2009). Classifying academically at-risk first graders into engagement types: Association with long-term achievement trajectories. *The Elementary School Journal*, 109, 380–405. <http://dx.doi.org/10.1086/593939>
- McClelland, M., Acock, A., & Morrison, F. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly*, 21, 471–490. <http://dx.doi.org/10.1016/j.ecresq.2006.09.003>
- McKown, C., Allen, A. M., Russo-Ponsaran, N. M., & Johnson, J. K. (2013). Direct assessment of children's social-emotional comprehension. *Psychological Assessment*, 25, 1154–1166. <http://dx.doi.org/10.1037/a0033435>
- McKown, C., Russo-Ponsaran, N. M., Johnson, J. K., Russo, J., & Allen, A. (2016). Web-based assessment of children's social-emotional comprehension. *Journal of Psychoeducational Assessment*, 34, 322–338. <http://dx.doi.org/10.1177/0734282915604564>
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus users guide* (7th ed.). Los Angeles, CA: Author.
- Naglieri, J. A., LeBuffe, P. A., & Shapiro, V. B. (2011). *Devereux Student Strengths Assessment-Mini (DESSA-mini)*. Lewisville, NC: Kaplan.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.
- Peña, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development*, 78, 1255–1264. <http://dx.doi.org/10.1111/j.1467-8624.2007.01064.x>
- Raudenbush, S. W., Bryk, A. S., & Congdon, R. (2004). HLM 6 for Windows [Computer software]. Skokie, IL: Scientific Software International.
- Ringwalt, S. (2008). *Developmental screening and assessment instruments with an emphasis on social and emotional development for young children ages birth through five*. Chapel Hill, NC: The University of North Carolina, FPG Child Development Institute, National Early Childhood Technical Assistance Center.
- Sabol, T. J., & Pianta, R. C. (2012). Patterns of school readiness forecast achievement and socioemotional development at the end of elementary school. *Child Development*, 83, 282–299. <http://dx.doi.org/10.1111/j.1467-8624.2011.01678.x>
- Singular Inversions. (2005). Facegen Main Software Development Kit (Version 3.1). Vancouver, BC: Author.
- van Widenfelt, B. M., Treffers, P. D., de Beurs, E., Siebelink, B. M., & Koudijs, E. (2005). Translation and cross-cultural adaptation of assessment instruments used in psychological research with children and families. *Clinical Child and Family Psychology Review*, 8, 135–147. <http://dx.doi.org/10.1007/s10567-005-4752-1>

Received August 22, 2016

Revision received April 20, 2017

Accepted May 4, 2017 ■