Behavioral Screening with a Translated Measure: Reliability and Validity Evidence for the Preschool Behavior Screening System – Parent Spanish Form

Jennifer C. Rodriguez[1], Ryan J. Kettler[1], and Kelly A. Feeney-Kettler[2]

[1]*Rutgers, The State University of New Jersey, USA*
[2]*FK&K Consulting, USE*

ABSTRACT: *The present study evaluated whether the Preschool Behavior Screening System (PBSS) translated into Spanish was able to reliably provide information regarding the behaviors and emotions of preschool children. The PBSS Parent Spanish Form (PBSS-PSF) could fill a void as the first multi-stage emotional and behavioral screening system in Spanish. The PBSS-PSF was field tested on 49 Spanish-speaking parents of preschool age children. Reliability coefficients for PBSS-PSF Phase 2 were in the moderate to excellent ranges. Results indicated that PBSS-PSF Phase 1did not yield the expected and predicted correlations with PBSS-PSF Phase 2. Correlations were acceptable between PBSS-PSF Phase 2 and the Behavioral and Emotional Screening System, Spanish Parent form. Collectively, the results are promising for the PBSS-PSF and provide guidance for future refinement of the tool.*
Keywords: *behavior rating scale, cross-cultural, preschool, Spanish, translation, universal screening*

I. INTRODUCTION

In the U.S., over four million children and adolescents suffer from serious mental illness that impairs the ability to succeed academically, socially, and in general, daily living (National Alliance on Mental Illness, 2010). Studies have revealed Latino youth in the U.S. to be a high-risk group, particularly for depression and anxiety (National Alliance on Mental Illness, 2006), as well as for general and pervasive feelings of hopelessness and sadness (American Psychiatric Association, 2010). Identified in the early childhood years, however, children at risk for developing serious mental illness may be successfully treated to prevent the long-term, disabling effects of these conditions (President's New Freedom Commission on Mental Health, 2003). Early screening of emotional and behavioral issues has been shown to effectively enhance the detection rate of children who later develop disabling conditions (Briggs- Gowan & Carter, 2008). Nonetheless, there currently exist few behavioral and emotional screening measures targeting young children, with even fewer tailored for Spanish speaking populations (Carney & Merrell, 2002). The purpose of the present study was to determine the reliability of scores from a parent-rating screening measure in Spanish to draw valid inferences about the behaviors and emotions of preschool children.

1.1 Need for More Accessible Measures among the U.S. Hispanic Population

Emotional and behavioral rating scales for children rely on input from primary caregivers and teachers. It is critical to ensure the accessibility of these scales to all parents, particularly including those whose native language is not English. A study by the Pew Hispanic Center found that 38% of Hispanic respondents are Spanish dominant and 38% report to speak both Spanish and English (Taylor, Hugo Lopez, Hamar Martinez, Velasco, 2012). With many Spanish-speaking individuals reporting that they speak English "not well" or "not at all," (American Psychiatric Association, 2010) it is evident that research must now focus on creating measures more accessible to this rapidly growing population. While there are several preschool emotional and behavioral screening measures available for English speaking individuals, there are fewer available for monolingual-Spanish speaking populations. Currently, the BASC-2 Behavioral and Emotional Screening System (BESS; Kamphaus & Reynolds, 2007), the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001), and the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) are among the few emotional and behavioral rating scales available to assess the functioning of preschool children in Spanish. Research on the translation of these instruments from English to Spanish is minimal. Further research in this area is crucial for allowing these measures to become more accessible to the Spanish-speaking population in the U.S.

1.2 Spanish Screening Measures for Preschool Children

Currently, there are a limited number of widely accepted and published behavioral and emotional screening measures for preschool-aged children. The BASC-2 Behavioral and Emotional Screening System (BESS; Kamphaus & Reynolds, 2007) is a brief, universal screening version of the Behavior Assessment

System for Children (2nd Edition; BASC-2; Reynolds & Kamphaus, 2004). The BESS yields one total problems score. The Spanish version of the BESS has lower but acceptable internal consistency compared to the English version.

Another measure, the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997), is a social-emotional and behavioral screener consisting of five subscales (i.e., Emotional Symptoms, Behavioral Problems, Hyperactivity, Peer Relationship Problems, and Prosocial Behaviors), with a total of 25 items. Analyses revealed that the Spanish version of the SDQ demonstrated moderate internal consistency (Cronbach's alpha = .73), also lower than the English version of the scale (Cronbach's alpha = .81; Blumert, Kettler, & Lakes, 2015).

1.3 Rational for a New Screening Tool

The Preschool Behavior Screening System (PBSS; Feeney-Kettler, Kratochwill, & Kettler, 2009) is a two-phase behavioral and emotional measure. The instrument contains nomination rubrics (Phase 1; PBSS P1) and rating scales (Phase 2; PBSS P2), used to screen for children at-risk of developing internalizing and/or externalizing disorders. PBSS P1 consists of two nomination rubrics: one for internalizing behaviors and one for externalizing behaviors. Those children who receive high scores on PBSS P1 are further assessed on PBSS P2. Four scores are generated following the completion of PBSS P2: Total Score, Internalizing Symptoms Scale score, Externalizing Symptoms Scale score, and Prosocial Behavior Scale score (Feeney-Kettler, Kratochwill, & Kettler, 2011).

The PBSS improves upon the aforementioned rating scales by incorporating multiple phases of screening. By only using PBSS P2 with those students identified in PBSS P1, users of the measure spend the most time considering students with the most severe needs, and obtain the most information on this group. This method of screening increases efficiency in identifying children who need intervention services. The most popular multi-stage measure of preschool behavioral and emotional difficulties, the Early Screening Project (ESP; Walker, Severson, & Feil, 1995), is not available in Spanish. Creating a Spanish version of the PBSS extends the benefit of multi-stage screening to Spanish-speaking populations.

1.4 Issues in Translation and Cross-Cultural Research

Several issues should be considered when translating a psychological measure from one language to another. The first consideration is that not all measures standardized and normed in one population may be administered cross-culturally. Berry (1969) outlines general suggestions for conducting cross-cultural research. He suggests that only behaviors which are functionally equivalent between the two cultures should be studied. For example, infants who cry when they are hungry in one culture do so in another, as the behavior serves the same function. This is relevant to the use of emotional and behavioral rating scales, as such instruments measure systematic and observable behaviors across cultures. As a result, similar descriptive categories may be used for functionally equivalent behaviors, allowing these scales to be applied cross-culturally (Berry, 1969).

When conducting cross-cultural research, it is not sufficient for researchers to simply translate study materials from the original to the target language. Several other considerations, such as operational definition of the construct in both cultures, as well as cultural and societal norms, must be considered in conducting quality cross-cultural research. Behling and Law (2000) outline three main issues that arise when translating measures for use in cross-cultural research: (a) lack of semantic equivalence between the original and translated measures (i.e., meaningfully equivalent words between original and target language), (b) lack of conceptual equivalence (i.e., constructs in one culture are being operationally defined the same way in another), and (c) lack of normative equivalence (i.e., how similarly the social norms and conventions, in two distinct societies, influence the behaviors of society members.). Societal norms often dictate the way individuals view and respond to psychological measures and should be taken into consideration when adapting measures for cross-cultural research. Further, while constructs in two distinct cultures may be operationally defined in the same way, a comparison of scores from each of the measures may not be valid, as each culture may measure those constructs in different ways (Hambleton & Li, 2005).

1.5 Cross-Cultural Research and Translation Methods

To date, there is no universally accepted procedure for the translation of psychological measures. While there are suggested methods (e.g., simple direct translation, back translation, etc.) and suggestions (e.g., semantic equivalence, cultural considerations, etc.), there is no "gold standard" procedural guideline for translating psychological instruments. There are, however, several suggested methods of translation, which are outlined in Table 1. One method is "simple direct translation," which involves a bilingual individual translating the measure from the original language to the target language. While this technique may be highly practical, there is no additional information provided (e.g., from a second or multiple translators) regarding the quality of the translation, potential cultural issues (e.g., the way in which a construct is defined in the target population), etc. (Behling & Law, 2000). A second type of translation is called "modified direct translation," in which the

work of the individual translator is periodically checked by a panel of experts (Geisinger, 1994). While this may be a less practical translation method than direct translation (additional costs include time, human investment, and financial resources), this method provides the researcher with more information and raises the potential of an accurate translation. A third type of translation is the "ultimate test" method. "Ultimate test" asks subjects to complete a measure in the original language, having only read the directions in the target language (e.g., if translating a measure from English to Spanish, have participants complete the English measure with Spanish instructions; Brislin, 1970). The second step of this method then has various participants complete different "split" versions (i.e., completing only certain items or parts) of the measure. There are several limitations of this method, including its use in only specific situations, the need for many translators, and relatively little information provided about the quality of the translation and individual items. Another translation method is called "parallel blind technique," and involves two translators completing independent translations of the measure, then meeting to compare and discuss (Werner & Campbell, 1970). One important consideration in using this method is that translators must be comfortable criticizing each other's work. "Random probe technique," in which a draft of the translated measure is given to a target group to complete, is another procedure. Participants are asked to explain why they responded in the manner they did. This method, however, provides limited information and should only be used as a supplement to other translation techniques (Guthrey & Lowe, 1992).

To assess the accuracy of a translated measure, van der Vijver & Hambleton (1996) suggest using both systematic judgmental evidence and statistical evidence. Judgmental evidence involves using standardized translation methods (e.g., translation, back-translation). However, judgmental evidence alone does not indicate an appropriate translation. Statistical analyses are also encouraged to assess for similarity between the original and translated measures, as well as to identify potential issues with the translated tool. Often, factor analysis is used to assess for test equivalence (Vijver & Hambleton, 1996). Differential item functioning, or DIF, is another method used to assess for equivalence. DIF assesses for differences between items on the original and translated measures, by looking for significant group differences on a single item (Ericikan & Lyons-Thomas, 2013). These group differences on a single item can often be associated with linguistic or cultural differences between the two translations. Methods used for DIF detection include the Mantel-Haenszel method, the delta plot method, the standardization index, and item response theory (Ericikan & Lyons-Thomas, 2013).

1.6 Translation Methods Used in the Current Project

Translation/back translation is a work intensive method. In this process: (1) an initial translation is completed, (2) this version is then translated back to the original language of the measure by a second translator, (3) the original and back-translated versions are compared for equivalency, and (4) if there are substantial discrepancies between the two versions of the measure, this process is completed iteratively until an adequate translation is reached (Werner and Campbell, 1970). This comprehensive process allows for more information to be provided, compared to the previous methods discussed. Further, the back translation method allows for the involvement of monolingual test developers in the translation process.

One concern that exists with back- translation is that there is no universally accepted method for identifying the level of similarity that should exist between items on the original measure and items on the corresponding translated measure (Behling & Law, 2000). Brislin (1970) identified three factors that may lead researchers to falsely conclude that the original and back-translated items are equivalent when they are not. One factor is that individuals who are translating and back-translating the measure may share the same biases in their translation methods. The second is that individuals completing the back-translation may be able to identify the essential meaning of the item, even though the original translation was poor. The third is that the grammatical structure of the items may make it possible for the back-translator to guess the item correctly. A fourth factor, identified by Hambleton (1993), involved translators using wording that would be easily translated by a back-translator, rather than the optimal language. While there are limitations to this method of translation, back-translation has been widely used and successfully applied in various research studies (Achenbach & Rescorla, 2001; Arce-Ferrer, 2006; Dumas, Martinez, LaFreniere, 1998; Goodman, 1997; Rubio-Stipec, Bird, Canino, & Gould, 1990).

One final translation technique is "committee translation." Committee translation is a process by which two or more individuals translate a measure from the original to the target language and compare results (Brislin, Lonner, Thorndike, 1973). This method of translation may involve all members completing an individual translation and then comparing translations. Individuals may also complete a translation together or review and provide feedback to the translation of another member of the committee.

With each translation method there are strengths and limitations. The current study incorporated a combination of the back-translation and committee translation methods followed by field testing. This process of combining methods has been successfully employed in several studies and increases the potential for an accurate and appropriate translation (Jacobsen, 1954; Dumas, Martinez, & LaFreniere, 1998). Further,

considerations were taken to reduce the limitations and inaccuracies associated with these two methods of translation. For example, a stringent system for identifying the level of similarity between items on the original measure and items on the translated measure was created. Further, the committee translation technique was used to assess the level of similarity between items on the original English version of the PBSS and on the back-translated version, therefore eliminating the need for bilingual experts.

II.   RESEARCH QUESTIONS

The purpose of the present study was to collect reliability and validity evidence regarding the Preschool Behavior Screening System, Parent Spanish Form (PBSS-PSF). The following research questions were asked:
1.   What is the internal consistency of the PBSS-PSF?
2.   What are the intercorrelations among scores from the PBSS-PSF?
3.   How well does the PBSS-PSF relate to the BESS, Spanish Parent Total Score?
4.   What is the subjective experience of the participants in completing the PBSS-PSF?

It was hypothesized that the PBSS-PSF would demonstrate adequate psychometric properties, similar to those of the PBSS-English. While translated rating scales may vary considerably from the original measure, brief screening measures that are translated are often psychometrically similar to the measure in the original language. As the PBSS- English demonstrated adequate reliability and validity (Feeney-Kettler, Kratochwill, & Kettler, 2011), it was predicted that the PBSS-PSF would, as well.

III.   METHOD

3.1 Participants

Participants in the current study emanated from a convenience sample that responded to phone recruitment in the area targeted by the funding agent for the study. The sample included 49 monolingual Spanish and bilingual Spanish-dominant parents of preschool-aged children from four schools in central New Jersey. The mean age of the parent participants was 32.91 years (SD= 6.86 years). The mean age of the children who were rated was 4.33 years; (SD= 7.01 years). Among the sample, 34.7% of the participants reported having two children, with the remaining reporting one (12.2%), three (24.5%), four (12.2%), five (8.2%), and eight (2.2%) children. The Latino/a American ethnicity was the most represented in the sample. More specifically, the countries of origin for participants included Mexico, Dominican Republic, Honduras, Ecuador, and Bolivia. Table 2 delineates further demographic information for the aforementioned sample.

3.2 Measures

Participants were asked to complete the PBSS-PSF and the BASC-2 Sistema de Evaluación Emocional y de Conducta (BESS, Spanish; Kamphaus & Reynolds, 2007). Further, participants completed a demographic questionnaire, along with an evaluation survey.

3.2.1 Preschool Behavior Screening System – Parent Spanish form.

The PBSS-PSF is a two-phase universal screening instrument. Phase 1 (P1) includes two nomination rubrics which assess a child's internalizing (internalizing nomination rubric [INR]) and externalizing (externalizing nomination rubric [ENR]) behaviors. Phase 2 (P2) is a 60-item rating scale that yields four index scores: Internalizing Symptoms Scale (ISS), Externalizing Symptoms Scale (ESS), Prosocial Behavior Scale (PBS), and Total Score (TS).

Both phases are completed by a parent or caregiver that knows the child well. In practice, the brief nomination rubrics of PBSS-PSF P1 are completed on every student, while the longer PBSS-PSF P2 is only completed on students identified as at-risk in PBSS-PSF P1. Children who are identified as at-risk in PBSS-PSF P2 are candidates for further assessment and intervention. In this study, both phases were completed for all children. Parents were asked to complete the newly created Spanish, parent version of this measure (i.e., the PBSS-PSF). Field testing was used to determine the reliability and validity of this scale.

Evidence has been collected to establish the reliability of PBSS Parent English form scores, as well as the validity of inferences drawn from the instrument (Feeney-Kettler, Kratochwill, & Kettler, 2011). The PBSS-P2 had good to excellent internal consistency for the ISS (alpha = .81), ESS (alpha = .96), PBS (alpha = .90), and TS (alpha = .95) rating scales. Content validity for the measure was ensured through extensive expert review. Intercorrelations among the PBSS scores were in the acceptable range for related but non-redundant variables. Correlations among scores that shared a construct but differed on method tended to be higher than correlations among scores that shared a method but differed on construct. Scores from PBSS P1 were also very sensitive predictors of scores from PBSS P2, as would be expected for a multi-stage screening system.

3.2.2 BASC-2 Sistema de Evaluación Emocional y de Conducta.

The BASC-2 Sistema de Evaluación Emocional y de Conducta (BESS, Spanish; Kamphaus & Reynolds, 2007) is a brief, universal screening system, used to identify the emotional and behavioral strengths and weaknesses of young children, along with more specific information on each child's internalizing and externalizing behaviors, school problems, and daily adaptive skills. For this study, the parent, preschool, Spanish version of the BESS was used. This measure consists of 30 items, taking the rater approximately five to ten minutes to complete. The BESS, Spanish Parent yields a total score, with no additional subscale scores.

The reliability and validity argument for the BESS Spanish is primarily founded on research on the English language version of the measure. Comparable albeit lower levels of internal consistency have been found for the Spanish versions of the BESS parent forms (preschool = .86, child = .90, adolescent = .86) compared to the English versions of BESS parent forms (preschool, age 3 = .91, age 4-5 = .93; child/adolescent, age 5-9 = .94, age 10-14 = .95, age 15-18 = .94). The BASC-2 Sistema de Evaluación Emocional y de Conducta was selected for the current study because the BESS is one of the most established behavioral screeners available, and because it was brief enough to be completed within the meeting time allotted for this research.

3.2.3 Demographic Questionnaire.

Participants completed a demographic questionnaire in Spanish. The demographic questionnaire included information such as role in family (i.e., mother, father, or other caregiver), gender, marital status, number of children, education level, age, and ethnic/racial background.

3.2.4 Evaluation Survey.

The evaluation survey was used to gain qualitative and quantitative information regarding the experience of the participants in completing the PBSS-PSF. Questions address the length of time to complete different phases of the measure, organization and clarity of the measure, and whether the measure can provide useful information regarding preschool age children. Responses are informative regarding the accessibility of the translated scale for its raters, as well as the usability or utility of the scale. Rating scales can only be useful if they are accessible to the raters; information from the evaluation survey provides evidence in this area and informs future iterations of the scale.

3.3 Procedures
3.3.1 Stage 1: Translation.

To attain an accurate and valid translation of the Preschool Behavior Screening System (PBSS-PSF; Feeney-Kettler, et al., 2009), Brislin's (1970) and Werner and Campbell's (1970) translation models for cross-cultural research were used. An initial English to Spanish translation of the PBSS-PSF was completed by a professional translation service. This service aligned with Brislin's (1970) committee translation, as there were three individuals who translated and reviewed the measure. All of the translators held Master's degrees in Spanish Translation and were native speakers of Spanish. The head translator, who was also federally certified as a court interpreter/translator, completed the initial translation of the measure. This translation was then given to a second reviewer, who was a university professor of Spanish, as well as a certified member of the American Translators Association. The third reviewer held a specialty in Education, and ensured that the content of the translation was appropriate for use by a school psychologist, school personnel, and parents. As suggested by Marin & Marin (1991), translations were done in a "contemporary" or "standard" Spanish, appropriate for use by Hispanic individuals in the United States.

Following an initial translation, back translation was used to translate the Spanish version back into English. The back-translations were completed by two individuals with Master's level degrees in Spanish Translation. One translator obtained a Master's degree in Spanish Translation and was also a school teacher. The second translator served as a Notary Public for over ten years, specializing in the translation of legal immigration papers. She was also a native Spanish speaker and had training in Spanish translation. Each translator was assigned a subset of items from the translated form to back-translate from Spanish to English.

Following the completion of the back translation, a second committee compared it with the original, English PBSS-PSF. This committee consisted of two Ph.D. level researchers in Educational Psychology, along with the primary author. A 4-point Likert scale (1= not at all similar, 2= somewhat similar, 3= very similar, 4= exactly the same) was created and used to determine similarity between corresponding items on the original and back translated versions of the measure. Committee members were asked to rate each corresponding item on their level of semantic similarity. To reduce the previously discussed limitations of this method, a stringent definition of unacceptable similarity was adopted. More specifically, those items that did not meet criteria for acceptable similarity (i.e., yielding a score below 3 from any rater) were reassessed. The translations and suggestions of a third translation expert were used to revise unacceptable items. This expert was a university professor in Spanish with over twenty-five years of experience in Spanish translation. She was also employed by the World Language Institute and held a doctoral degree in second language education.

3.3.2 Stage 2: Field Testing.

The PBSS-PSF was field tested on a sample of monolingual Spanish and bilingual Spanish-dominant parents of preschool children in two central New Jersey preschools. Participating schools were initially contacted via telephone and provided information regarding the study. The schools who agreed to participate were provided all study materials by the principal investigator. Each school held a meeting for parents to complete the measures. Each parent was given a PBSS-PSF; a BESS, Spanish Parent form; a demographic questionnaire; and an evaluation survey. Parents who attended the meeting were first given an explanation of the study and procedures by the school director. Parents were then given time to ask questions and complete all study materials. Having parents complete the measures during a group meeting helped to ensure measure completion. Approval to conduct this study was obtained from the Institutional Review Board at the institution of the lead author, prior to data collection at the preschools.

3.4 Data Analysis

Prior to analysis, data was systematically checked and cleaned. All data was double entered by two data clerks. To account for omitted items on the PBSS-PSF nomination rubrics, pre-established rules for obtaining scores were used. One or both nomination rubrics were entirely skipped by 20% of respondents. The percentage is higher for those who indicated specific symptoms at the top of the rubric, but did not complete one or both of the corresponding questions. The rules for omitted items on the PBSS-PSF P1 nomination rubrics are conservative in the sense that students are more likely to advance to PBSS-PSF P2 rating scales, and cases are highly unlikely to become false negatives due to skipped items. Specifically, nomination scores were calculated by taking the middle score on each of the two questions which comprise the rubrics. If parents indicated a certain number of symptoms for their child at the top of the rubric, but left the corresponding question on the rubric blank, that question was scored according to the number of symptoms a parent endorsed. For omitted items on the rating scales, scale scores were calculated by taking the average scores of completed items and multiplying that average by the number of items on the scale.

To assess the internal consistency of the PBSS-PSF P2, coefficient alpha was calculated. The same was done for the BESS, Spanish Parent form to compare the internal consistency of both measures with the current sample. Murphy and Davidshofer's (2004) qualitative descriptions of coefficient alpha levels were used. A correlation matrix was also created to assess the relationships between the PBSS-PSF P1 nomination rubrics, the PBSS-PSF P2 rating scales, and the BESS, Spanish Parent. Cohen's (1992) classification system, with Hopkin's (2001) extensions, was used to describe the Pearson correlations. Pearson correlations were conducted to compare the PBSS-PSF and the BESS, Spanish Parent form. One-tailed significance tests, consistent with directional predictions, were used throughout the analyses. Specifically, only correlations with the Prosocial Behavior Scale were predicted to be negative. Lastly, descriptive statistics were used to analyze the information from the evaluation survey.

IV. RESULTS

To assess how parents rated their children on the constructs of interest, scale means were calculated for the sample. On the PBSS-PSF P2 rating scales, the mean rating for the Internalizing Symptoms Scale (ISS) was 32.21 (SD = 8.68) and the mean rating for the Externalizing Symptoms Scale (ESS) was 33.97 (SD = 12.38). The Prosocial Behavior Scale (PBS) had a mean rating of 40.16 (SD = 12.58). It should be noted that items on the PBSS-PSF P2 PBS were reverse coded for the purposes of the analyses. The PBSS-PSF P2 Total Score had a mean rating of 108.98 (SD = 22.36). On the BESS, Spanish Parent the mean rating for the Total Score was 48.74 (SD = 10.95).

4.1 Internal Consistency

The internal consistency coefficient (Cronbach's alpha) was calculated for each of the four scale scores on the PBSS-PSF P2. Table 3 outlines the Cronbach's alpha values for the English and Spanish versions. The ESS (.92), PBS (.89), and Total Score (.89) all fell within the excellent and high ranges, indicating that the items on each of the scales fit together. As expected, the ISS (.78) yielded an alpha that was in the moderate range, although still acceptable for screening measures (Murphy & Davidshofer, 2004). Overall, reliability coefficients for the PBSS-PSF P2 were comparable to those of the English form. Cronbach's alpha was also calculated for the BESS, Spanish Parent form. Results indicated an alpha of .88, similar to the published alpha level of .86 and comparable to the PBSS-PSF P2 Total Score.

4.2 Intercorrelations

Pearsonproduct-moment correlation coefficients were computed to assess the relationships among the scores from the PBSS-PSF P1 and the PBSS-PSF P2 rating scales. The correlation between the two PBSS-PSF P1 nomination rubrics (r = .67, p < .05) was large. There was a medium positive correlation between the ENR

and the PBSS-PSF P2 Total Score (r = .35, p < .05). The PBS shared medium negative relationships with the INR (r = -.30, p < .05) and the ENR (r = -.30, p < .05). There was a medium positive correlation between the ISS and the ESS on PBSS-PSF P2 (r = .40, p < .05).Table 4 outlines the Pearson correlation values.

4.3 Concurrent Validity
Pearson correlations were also used to compare the PBSS-PSF with the BESS, Spanish Parent form. The relationship between the INR and the BESS, Spanish Parent Total Score (r = .26, p < .05) was small and the relationship between the ENR and the BESS, Spanish Parent Total Score (r = .35, p < .05) was medium. There was a large positive relationship between the PBSS-PSF P2 Total Score and the BESS, Spanish Parent Total Score (r = .67, p < .05). The ESS (r = .61, p < .05) and the PBS (r = -.51, p < .05) yielded large and significant correlations with the BESS, Spanish Parent Total.

4.4 Evaluation Survey
An evaluation survey was provided to each parent participant to gain more information on the experience of parents completing the PBSS-PSF. Overall, 78% of parents did not have concerns about the way their child behaved, while another 11% reported they had "a little" concern, and 9% reported they did have concern. Of the children whose parents endorsed concern, two were also identified as at-risk on the INR and the ENR, one was not identified on either nomination rubric, and one was identified on the ENR only. Of the children whose parent endorsed "a little" concern, three were also identified as at-risk on the INR and the ENR, one was not identified on either nomination rubric, and one was identified on the ENR only. Ninety-five percent of parents believed that the PBSS-PSF could provide useful information about preschool-age children and 84% reported that the form was clearly organized and written. Further, it took parents an average of 18 min (SD = 13 min) to complete both of the PBSS-PSF P1 nomination rubrics and an average of 20 min (SD = 12 min) to complete the PBSS-PSF P2 rating scales.

Regarding whether or not the PBSS-PSF provides valuable information about preschool children, two overarching themes were identified. The first theme is that the PBSS-PSF helps to identify children with emotional and behavioral difficulties. One parent stated, "…ayuda a la gente capacitada para conocer las necesidades que tienen los niños y así ayudarlos a descubrir cual es la problema a través de su comportamiento./…it helps trained people to identify the needs of children and help them to discover what the problem is through their behaviors." The second theme among parent responses is that the information provided by the scale may lead to appropriate interventions for children identified to be at-risk. One parent remarked that the scale may be useful "porque tanto el maestro y como los padres podríamos ayudar mas a nuestros hijos/ because both teachers and parents could help our children more."

Regarding whether the PBSS-PSF is clearly written and organized, many parents reported some confusion in completing the measure. It was further revealed through data analysis that many parents had difficulty in understanding how to adequately complete the nomination rubrics, with several parents completing PBSS-PSF P1 incorrectly. One parent noted, "Era un poco confuso porque no entendí muy bien lo que tenia que hacer/ It was a little confusing because I did not understand very well what I had to do." A second parent shared, "Mas o menos. Era un poco difícil de entender./ More or less. It was a little difficult to understand." Overall, parents demonstrated difficulty in completing the PBSS-PSF P1 nomination rubrics.

V.   DISCUSSION
Following a rigorous translation process, the PBSS-PSF was created to screen for emotional and behavioral difficulties in preschool children. This translation process took into account various issues of cross-cultural research and translation procedures discussed in the literature. To increase semantic equivalence between the PBSS-PSF and the PBSS-PSF various translation methods were studied and considered. The use of a professional translation service, back-translation (Werner and Campbell, 1970), and committee translation (Brislin, Lonner, Thorndike, 1973) were selected due to their appropriateness for this particular measure, as well as their accepted use in cross-cultural research. The completed translation was then field tested on a sample of monolingual Spanish and bilingual Spanish-dominant parents of preschool children.

5.1 PBSS-PSF Phase 1
The first phase of the PBSS-PSF consists of two nomination rubrics, screening for the presence of internalizing and externalizing symptoms. The INR and ENR shared a large relationship slightly above the predicted range. This correlation between the INR and ENR is similar to that of the English version of the measure. This degree of correlation suggests that the two nomination rubrics measure symptoms of constructs which may overlap, however, the value is not high enough to suggest that they are measuring the same construct. As a result, it can be deduced that there is an appropriate relationship between the nomination rubrics of the PBSS- PSF.

Because the PBSS-PSF is a two-phase screening measure, it was important to assess the relationship between the first and second phases of the tool. Pearson correlations were used to assess the relationships between the rubrics and scales in both phases. The PBSS-PSF P1 nomination rubrics did not yield the predicted and expected correlations with most of the PBSS-PSF P2 rating scales. There was no significant relationship between the INR and the ISS, nor between the ENR and ESS. An almost nonexistent relationship between the two pairs is an unexpected outcome, because they are intended to measure similar symptoms. Further, the INR and ENR yielded small and nonsignificant relationships with the remaining PBSS-PSF P2 scales, much lower than expected.

Pearson correlations were conducted to assess the relationship between the PBSS-PSF P1 nomination rubrics and the BESS, Spanish Parent. While correlations were significant, they were in the small and medium ranges. The magnitudes of the correlations between the PBSS-PSF P1 nomination rubrics and the criterion measures are lower than anticipated and do not provide strong evidence of concurrent validity.

5.2 PBSS-PSF Phase 2

PBSS-PSF P2 is composed of 59 items organized into 3 rating scales. Analyses revealed that PBSS-PSF P2 yields reliable scores from which valid inferences may be drawn, comparable to the English version of the scale, as well as to the BESS, Spanish Parent. This is expected, as PBSS-PSF P2 is more comprehensive than PBSS-PSF P1.

The PBSS-PSF P2 rating scales demonstrated reliability coefficients in the excellent and high ranges for the ESS, PBS, and Total Score. These findings indicate that the items on each scale fit together well, as the yielded alpha levels are appropriate for a screening tool. The ISS yielded an alpha level in the moderate range. This finding is consistent with other published measures, which show lower alpha coefficients for scales which measure internalizing behaviors (Achenbach & Rescorla, 2001; Reynolds & Kamphaus, 2004). The PBSS-PSF P2 rating scales also demonstrated lower, yet comparable reliability coefficients to those on the English version of the measure. Scores from Spanish versions of several other emotional and behavioral measures, including the BESS (Kamphaus & Reynolds, 2007) and the SDQ (Goodman, 1997), are likewise less reliable.

The PBSS-PSF P2 rating scales also appear to have acceptable internal structure. The correlations between the PBSS-PSF P2 rating scales scores all fell within or close to the predicted ranges, as well as to the ranges on the English version of the measure (Feeney-Kettler, Kratochwill, Kettler, 2011). Further, the ISS and ESS shared a medium correlation, suggesting that they are measuring somewhat related but separate constructs.

PBSS-PSF P2 was also compared to the BESS, Spanish Parent. There was a large, positive correlation between PBSS-PSF P2 Total Score and the BESS, Spanish Parent Total Score. The correlation was lower than originally predicted. However, it should be noted that apart from measuring internalizing and externalizing symptoms, the BESS, Spanish Parent also includes items which measure adaptive functioning (e.g., social skills and functional communication) and school problems (e.g., attention and learning problems). This difference may explain the lower correlation, because the PBSS-PSF does not incorporate these two additional areas of functioning.

5.2 Limitations and Implications for Future Research

The primary limitation of the current study is regarding the size and homogeneity of the sample. In order to create a scale that can be used by parents from a wide range of Spanish-speaking cultures, the PBSS-PSF was written in standard Spanish. The convenience sample was drawn from two schools in one region of New Jersey. The sample was homogenous with regard to ethnicity and education level. Future studies for the PBSS-PSF may focus on incorporating a larger, more diverse sample of participants. Such research will be necessary to draw conclusions about how the scale works for a diverse population of Spanish speaking persons.

A second limitation involves the ability to assess the accuracy of the translated PBSS. While a combination of judgmental and statistical evidence is encouraged in establishing equivalence between the original and translated tools, not all relevant statistical procedures could be conducted for this purpose. The limited number of participants included in this study prevented the suggested analyses from being conducted (i.e., factor analysis). Future studies should include a higher number of participants, in order to gain further statistical evidence of measure equivalence.

A third limitation is that the BESS, Spanish Parent form was used as the sole criterion measure. The BESS is based on parent report, as is the PBSS, so some of the shared variance is due to common methodology. The BESS also yields a single global risk score for behavioral difficulties. Future research could incorporate scores from different methodologies, as well as from Spanish language measures that yield scores for internalizing symptoms, externalizing symptoms, and prosocial behaviors, aligned more directly with the PBSS.

VI. CONCLUSION

The PBSS-PSF is a two-stage screening measure for identifying preschool children who may be at-risk for developing emotional and behavioral difficulties. The multiple stage format of the screening measure was designed to more efficiently and effectively identify those children who may be at-risk for later developing serious and disabling emotional and behavioral conditions. Following a rigorous translation process of the English version of the measure, the PBSS-PSF was field tested on a small group of monolingual Spanish and bilingual Spanish-dominant parents of preschool children.

The PBSS-PSF P2 may be a useful tool for practitioners and teachers to identify preschool children who are in need of additional services and early intervention programs. Results indicated that PBSS-PSF P1 requires further revision before it functions similarly to the original English version of the measure. The development of the PBSS in Spanish increases its accessibility for a larger population of children and parents who are in need of emotional and behavioral screening. Additional research is needed to address the presented concerns and determine the generalizability of these results to the broader Spanish-speaking population.

6.1 Compliance of Ethical Standards

6.2 Acknowledgements

REFERENCES

[1]. Achenbach, T. M., & Rescorla, L. (2001). *ASEBA School-Age Forms & Profiles.* ASEBA.
[2]. American Psychiatric Association. (2010). Mental health disparities: Hispanics/ Latinos [Data file]. Retrieved from http://www.nami.org
[3]. Arce-Ferrer, A. J. (2006). An investigation into the factors influencing extreme response style: Improving meaning of translated and culturally adapted rating scales. *Educational and Psychological Measurement*, 66, 374-392.
[4]. Behling, O., & Law, K. S. (2000). *Translating questionnaires and other research instruments: Problems and solutions.* Thousand Oaks, CA: Sage Publications, Inc.
[5]. Berry, J. W. (1969) On cross-cultural comparability. *International Journal of Psychology*, 4, 110-128.
[6]. Blumert, H., Kettler, R.J., & Lakes, K.D. (2015). The impact of language translation on the internal structure of a rating scale: The strengths and difficulties questionnaire in Spanish. *Global Journal of Human Social Science*, 15(1), 57-68.
[7]. Briggs-Gowan, M. J., & Careter, A. S. (2008). Social-emotional screening status in early childhood predicts elementary school outcomes. *Pediatrics*, 121(5), pg. 957-962.
[8]. Brislin, R., W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1(3), 185-216.
[9]. Brislin, R. W., Lonner, W. J., & Thorndike, R. M. (1973). *Cross-cultural research methods* (pp. 32-58). New York: J. Wiley.
[10]. Carney, A. G., & Merrell, K. W. (2002). Reliability and comparability of a Spanish-language form of the preschool and kindergarten behavior scales. *Psychology in the Schools*, 39(4), 367-373.
[11]. Cohen, J. (1992). A power primer. Psychological Bulletin, 112, 155-159.
[12]. Dumas, J. E., Martinez, A., LaFreniere, P. J. (1998). The Spanish version of the Social Competence and Behavior Evaluation (SCBE)- Preschool edition: Translation and field testing. *Hispanic Journal of Behavioral Sciences*, 20(2), 255-269.
[13]. Ercikan, K., & Lyons-Thomas, J. (2013). Adapting tests for use in other languages and cultures. In K. F. Geisinger et al. (Eds.), *APA handbook of testing and assessment in psychology, Vol. 3: Testing and assessment in school psychology and education* (pp. 545–569). Washington, DC: American Psychological Association.
[14]. Feeney-Kettler, K. A., Kratochwill, T. R., & Kettler, R. J. (2009). *Preschool Behavior Screening System*. Unpublished instrument, Vanderbilt University.
[15]. Feeney-Kettler, K. A., Kratochwill, T. R., & Kettler, R. J. (2011). Identification of preschool children at risk for emotional and behavioral disorders: Development and validation of a universal screening system. *Journal of School Psychology*, 49, 197-216.
[16]. 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.
[17]. Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry*, 38(5), 581-586.
[18]. Guthrey, D., & Lowe, B. A. (1992). Translation problems in international marketing research. *Journal of Language for International Business*, 4, 1-14.
[19]. Hambleton, R. K. (1993). Technical standards for translating test and establishing test score equivalence. Symposium presented at the 101 Annual Convention of the American Psychological Association, Toronto, Ontario, Canada.
[20]. Hambleton, R. K., & Li, S. (2005). Translation and adaption issues and methods for educational and psychological tests. In C. L. Frisby & C. R. Reynolds (Eds.), *Comprehensive Handbook of multicultural school psychology* (881-903). New York, NY: John Wiley.

[21]. Hopkins, W. G. (2001). A scale of magnitudes for effect statistics. Retrieved from the New View of Statistics Website: http://eee.sportsci.org/resource/stats/effectmag.html

[22]. Jacobsen, E. (1954). Methods used for producing comparable data in the OSCR seven-nation attitude survey. *Journal of Social Issues*, 10(4), 40-51.

[23]. Kamphaus, R. W., & Reynolds, C. R. (2007). *Behavioral & Emotional Screening System*. NCS Pearson.

[24]. Marin, G., & Marin, B. V. (1991). *Research with Hispanic populations*. Thousand Oaks, CA: Sage Publications, Inc.

[25]. Murphy, K. R., & Davidshofer, C. O. (2004). *Psychological testing: Principles and applications* (6[th]ed.). Upper Saddle River, New Jersey: Pearson.

[26]. National Alliance on Mental Illness (2010). Facts on children's mental health in America. Retrieved from http://www.nami.org.

[27]. National Alliance on Mental Illness (2006). Latino community mental health fact sheet. Retrieved from http://www.nami.org

[28]. President's New Freedom Commission on Mental Health (2003). Achieving the promise: Transforming mentalhealth care in America. Final report. DHHS pub. no. SMA-03-3832. Rockville, MD: U.S. Department of Health and Human Services.

[29]. Reynolds, C. R., & Kamphaus, R.W. (2004). *Behavior Assessment System for Children (2nd Edition)*. Circle Pines, MN: AGS Publishing.

[30]. Rubio-Stipec, M., Bird, H., Canino, G., & Gould, M. (1990). The internal consistency and concurrent calidity of a Spanish translation of the Child Behavior Checklist. *Journal of Abnormal Child Psychology*, 18(4), 393-406.

[31]. Taylor, P., Hugo Lopez, M., Hamar Martinez, J., & Velasco, G. (2012). When label don't fit: Hispanic and their views of identity. Washington, D. C.,: Pew Hispanic Center, April. http://www.pewhispanic.org/files/2012/04/PHC-Hispanic-Identity.pdf

[32]. Van der Vijver, F. J. R., Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, 1(2), 88-99.

[33]. Walker, H. M., Severson, H. H., & Feil, E. G. (1995). *Early screening project: User manual*. Longmont, CO: Sopris West.

[34]. Werner, O., & Campbell, D. (1970). Translating, working through interpreters, and the problem of decentering. In R. Naroll and R. Cohen (Eds.), *A handbook of method in cultural anthropology* (pp. 398-420). New York: American Museum of Natural History.

Tables

Table I. Methods of Translation in Cross-Cultural Research

Method	Technique	Strengths	Limitations
Simple Direct Translation	Bilingual individual translating a measure from the original language to the target language.	• Highly practical (i.e., translation obtained quickly and with minimal financial resources).	• Provides no additional information about the translations (i.e., quality of the translation, potential issues, etc). • Subject to biases of the translator.
Modified Direct Translation	The work of the individual translator is periodically checked by a panel of experts.	• Increases the chance of an accurate and reliable translation.	• Less practical than simple direct translation. • Panel members may be no more competent than principle translator. • Difficulty obtaining group consensus.
Ultimate Test	Step 1: subject completes a measure in the original language, having only read the directions in the target language. Step 2: various participants complete different "split" versions of the measure.	• Correlations between source and target measure scores provide objective information.	• Can only use in specific situations. • Lack of expert review. • Relatively little additional information provided. • Requires large number of bilingual participants.
Parallel Blind Technique	Two translators independently complete a translation of the measure, then meet to compare and discuss.	• High practicality (i.e., can complete translation quickly) • Increases potential quality of the measure when comparing two translations.	• Must have two translators fluent in the source and target language. • Translators may have difficulty "criticizing" one another. • Translators may share similar biases/ misinterpretations.
Random Probe Technique	A draft of the translated measure is given to a target group to complete. Subjects are then asked to explain why they responded in the manner they did.	• High practicality (i.e., simple technique, quick translation, and little financial resources needed).	• Mostly provides qualitative information through open ended questions. • Often used as a supplement to another technique.
Translation/ Back Translation*	An initial translation is completed, a second translator translates measure back to original language, original and back-translated versions are compared for equivalency, discrepancies are addressed.	• Can compare back-translated version to original measure. • Data allows involvement of and discussion among monolingual, researchers.	• No universally accepted method for identifying the level of similarity between items on the original measure and items on the translated measure.
Committee Translation*	Two or more individuals translate a measure from the original to the target language and compare results (independently or as a group).	• Allows for discussion among several experts.	• Several bilingual experts needed. • May take longer period of time to complete than other methods.

* *Note.* Methods used in the current study are marked with an asterisk.

Table II. Demographic Characteristics of Total Sample

Characteristic	Frequency	Percentage
Total N	49	100%
Gender		
Female	43	91.5%
Male	4	8.5%
Family Role		
Mother	42	89.4%
Father	4	8.5%
Other Caregiver	1	2.1%
Ethnicity		
African American	1	2.1%
Asian American/ Pacific Islander	0	0.0%
European American	0	0.0%
Latino/a American	46	97.9%
Native American	0	0.0%
Other	0	0.0%
Marital Status		
Single	20	46.5%
Married	22	51.2%
Divorced	1	2.3%
Highest Level of Education Completed		
No degree	16	34.0%
High School Diploma	24	51.1%
Associate's Degree	4	8.5%
Bachelor's Degree	1	2.1%
Master's Degree	2	4.3%
Doctoral Degree	0	0.0%

Table III. Reliability Coefficients of PBSS-PSF P2 Scales with 95% Confidence Intervals

PBSS-PSF P2 Scale	PBSS- English	PBSS-Spanish
Internalizing Symptoms Scale	.81 (.75 to .86)	.78 (.67 to .87)
Externalizing Symptoms Scale	.96 (.95 to .97)	.92 (.88 to .95)
Prosocial Behavior Scale	.90 (.87 to .93)	.89 (.84 to .93)
Total Score	.95 (.94 to .96)	.89 (.83 to .93)

Note: PBSS- English values were obtained from Feeney-Kettler, Kratochwill, & Kettler, (2011).

Table IV. Correlations among PBSS and BESS Scores

Variable	PBSS-PSF P1 INR	PBSS-PSF P1 ENR	PBSS-PSF P2 ISS	PBSS-PSF P2 ESS	PBSS-PSF P2 PBS	PBSS-PSF P2 Total	BESS Total
PBSS-PSF P1 Nomination Rubrics							
Internalizing Nomination Rubric (INR)	-	.67*	.02	-.06	-.30*	.17	.23
Externalizing Nomination Rubric (ENR)		-	.08	.24	-.30*	.35*	.35*
PBSS-PSF P2 Rating Scales							
Internalizing Symptom Scale (ISS)	.51 or higher	-.18 or higher	-	.40*	-.09	.55*	.10
Externalizing Symptom Scale (ESS)	-.31 or higher	-.02 or higher	.16 or higher	-	-.17	.80*	.61*
Prosocial Behavior Scale (PBS)	-.05 or lower	-.05 or lower	.17 or lower	.09 or lower	-	-.61*	.-51*
Total Score (Total)	-.09 or higher	.10 or higher	.35 or higher	.68 or higher	-.42 or lower	-	.67*
Behavioral and Emotional Screening System							
Total	-.01 or higher	.11 or higher	-.16 or higher	.41 or higher	-.29 or lower	.50 or higher	-

Note. Correlations are above the diagonal. One-tailed 95% confidence intervals are below the diagonal. PBSS = Preschool Behavior Screening System; PSF = Parent Spanish Form; P1 = Phase 1; P2 = Phase 2; INR = Internalizing Nomination Rubric; ENR = Externalizing Nomination Rubric; PBS = Prosocial Behavior Scale; BESS = Behavioral and Emotional Screening System.
*Correlation is significant at the 0.05 level (1-tailed).