THE PRESCHOOL BEHAVIOR SCREENING SYSTEM:

PARENT, SPANISH FORM

A DISSERTATION

SUBMITTED TO THE FACULTY

OF

THE GRADUATE SCHOOL OF APPLIED AND PROFESSIONAL PSYCHOLOGY

OF

RUTGERS,

THE STATE UNIVERSITY OF NEW JERSEY

BY

JENNIFER C. RODRIGUEZ

IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE

OF

DOCTOR OF PSYCHOLOGY

NEW BRUNSWICK, NEW JERSEY   OCTOBER 2015

APPROVED: _____
                     Ryan J. Kettler, PhD

                     _____
                     Kelly A. Feeney-Kettler, PhD

DEAN: _____
                     Stanley Messer, PhD

ABSTRACT

Hispanic-American youth have been identified as a high-risk group for developing behavioral and emotional difficulties. Currently, there exist few behavioral and emotional screening measures targeting young children, with even fewer accessible to the Spanish speaking populations in the U.S. The purpose of the present study was to determine whether the Preschool Behavior Screening System (PBSS) in Spanish was able to reliably provide data leading to valid inferences regarding the behaviors and emotions of preschool children. The Preschool Behavior Screening System, Parent Spanish Form (PBSS-PSF) is a two-phase screening tool used for identifying preschool children who may be at-risk for developing emotional and behavioral difficulties. The first phase of the measure includes two nomination rubrics measuring internalizing and externalizing behaviors. The second phase is a 59-item rating scale, yielding three composite scores and one total score. Following a rigorous translation process of the English version of the measure, the PBSS-PSF was field tested on 49 Spanish-speaking parents of preschool age children from four schools in New Jersey. Analyses were used to assess the relationships between PBSS-PSF Phase 1 (PBSS-PSF P1) and PBSS-PSF Phase 2 (PBSS-PSF P2), as well as prediction to a published screening measure, the BASC-2 Behavioral and Emotional Screening System (BESS), Spanish Parent form. Results indicated that the PBSS-PSF P1 nomination rubrics worked well together in identifying at-risk children. PBSS-PSF P1 was found to be a highly sensitive, but not very specific tool when predicting PBSS-PSF P2 and the BASC-2 Behavioral and Emotional Screening System, Spanish (BESS, Spanish Parent). However, PBSS-PSF P1 did not work as effectively as was expected with PBSS-PSF P2 in identifying those individuals who may be at-risk. This was due to significant difficulties parents experienced in completing the PBSS-PSF P1 nomination rubrics. Reliability coefficients for the

PBSS, Spanish Parent PBSS-PSF P2 scales were found to be in the moderate, high, and excellent ranges. PBSS, Spanish PBSS-PSF P2 also yielded acceptable correlations among the Internalizing Symptoms Scale, Externalizing Symptoms Scale, Prosocial Behavior Scale, and Total Score Scale and with the BESS, Spanish Parent form. Future studies will address the observed difficulties parents experienced in completing the PBSS-PSF P1 nomination rubrics and determine the generalizability and validity of these results with different Spanish-speaking populations.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

PAGE

LIST OF TABLES

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). 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). Advances in emotional and behavioral screening systems have aided in the identification of at-risk children. 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 and Carter (2008) found that over 50% of children who were identified, by their parents, as having emotional/ behavioral problems or low competence at around age one, were also later identified by parents and teachers as having behavioral/ emotional issues in early elementary school. Early screening measures are critical in identifying and treating young children at-risk for developing psychological difficulties. 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.

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

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). Further, attempted suicide among Latino youth have been reported at higher rates (10.7%) than

those for African American youth (7.3%) and European American, non-Latino youth (6.3%; National Alliance on Mental Illness, 2006). For these reasons, screening for behavioral and emotional problems at a young age is critical among this particular population of children.

Currently, approximately 17% of the population identifies as Hispanic (U.S. Census Bureau, 2011), with these numbers rising. 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. Behavioral and emotional 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. While there are few preschool behavioral and emotional 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 behavioral and emotional rating scales available to assess the functioning of preschool children. Research in the translation of these instruments from English to Spanish is minimal. Further research in this area is crucial in allowing these measures to become more accessible to the Spanish- speaking population in the U.S.

Issues in Translation and Cross-Cultural Research

Several issues should be considered when translating a psychological measure from one language to another. There is an ongoing discussion on whether research should be conducted in an "emic" manner or in "etic" terms. Emic refers to research conducted in a manner which preserves the individual characteristics of a particular culture. For example, an emic researcher studying the construct of "happiness" would use the beliefs and norms of a particular culture or population in order to create an operational definition of the construct of interest. "Etic" refers to research which attempts to generate universal hypotheses and theories from aspects of many different cultures (Behling & Law, 2000). For example, an etic researcher studying the construct of "happiness" would use the views and beliefs of several cultures and population to develop a universal definition of the construct. While psychological research has geared itself toward an etic perspective, the issue remains that not all measures standardized and normed in one population may be administered cross-culturally. Berry (1969) outlines general suggestions when 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 behavioral and emotional 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). It is important, however, that these categories are appropriate to each culture and adaptable to explain the cultural practices of a particular culture.

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, cultural and societal norms, etc.,

must be considered in conducting quality cross-cultural research. Behling and Law (2000)

outline three main issues that arise when translating measures for the use in cross-cultural

research: (a) lack of semantic equivalence between the original and translated measure, (b) lack

of conceptual equivalence, and (c) lack of normative equivalence. Semantic equivalence refers to

identifying words and phrases in the target language that are meaningfully equivalent to words in

the original language. For example, researchers may encounter difficulties in determining

whether a word or set of words in Spanish has the same meaning as a word or set of words in

English. Conceptual equivalence refers to whether constructs in one culture are being

operationally defined the same way in another. Lastly, normative equivalence looks at how

similarly the social norms and conventions, in two distinct societies, influence the behaviors of

society members. Societal norms often dictate an individual's willingness to discuss certain

personal topics, the way in which ideas are expressed, and responsiveness to strangers (i.e.,

researchers, clinicians). These societal norms may dictate the way individuals view and respond

to psychological measures and should be taken into consideration when adapting measures for

cross-cultural research.

Regarding rating scales, which seek to gain unobservable information about an individual

through observable behavioral characteristics, Behling and Law (2000) state that the risk of

semantic and conceptual inequality is generally low, as it is often simpler to semantically

translate concrete behaviors than more abstract concepts. Broader conceptual constructs are not

typically the norm in behavioral questionnaires, as the focus of items lies in more specific target

behaviors. In terms of normative equivalence, however, the risk of inequality in this area is

considered to be high. There is great variety among cultures in the level of disclosure regarding

negative behaviors to "strangers." Individuals may, therefore, complete behavioral

questionnaires and rating scales in a way that does not disclose equivalent information compared

to the culture for which the questionnaire was initially intended and in which it was initially

used. Researchers should seek to create translated measures which possess adequate semantic

and conceptual equivalence, while attempting to reduce the potential for normative inequality.

There are several considerations in addressing and reducing issues in semantic,

conceptual, and normative equivalence. In terms of semantic equivalence, the type of translation

method should be carefully considered, as each may increase or decrease the potential for

inequality between the original and translated measure (Brislin, 1980). In addressing issues of

conceptual equivalence, Behling and Law (2000) propose logical and empirical tests of

conceptual equivalence. Logical tests of conceptual equivalence include an exploration of the

theory behind a particular construct, the operational definition of the studied construct in both

cultures, and differences that may exist between the two. Empirical tests of conceptual

equivalence involve exploratory and confirmatory factor analyses to assess whether the factorial

structure of a scale is comparable in both cultures.

As there is typically a high chance for normative inequality between the source and target

cultures, when considering the translation of behavioral rating scales, it is critical to assess the

factors which may exacerbate the inequality. Sources of normative inequalities may result from

differences in societal norms which dictate the openness with which individuals discuss certain

topics, political views, willingness to discuss personal and family matters, ways in which ideas

are expressed, conformity versus assertiveness, directness versus indirectness, response biases,

level of reticence towards strangers, and hospitality norms (Behling & Law, 2000; Pareek &

Rao, 1980). One way of closing normative gaps may be to develop close relations and positive

rapport with respondents, as well as to assure respondents that responses will remain confidential or anonymous (Behling & Law, 2000). Further, researchers should consider involving translators in more than just the translation process, such as part of a "multicultural team" to assess potential issues and solutions of scales (Douglas & Craig, 1983). Finally, pilot or field testing and appropriate statistical analyses should be conducted to assess for possible normative inequalities (Geisinger, 1994).

Concerns about the Reliability and Validity of Translated Measures

Concerns regarding the reliability of translated measures arise, as certain response patterns have been observed in several Hispanic cultures. These differential response styles may reduce the observed variance of scores in the target culture. An extreme response style (Chronbach, 1950) is the propensity of individuals to choose upper or lower limits of rating scales, irrespective of the content of the scale. High rates of extreme response styles have been found among Hispanic, African American, and Mediterranean populations (Clarke, 2000; Marin & Marin, 1995). Marin, Gamba, & Marin (1992) suggest that level of education and acculturation may influence the observation of extreme response styles in Hispanic respondents. More specifically, Hispanic individuals who are more acculturated to the U.S. and have more than twelve years of formal education are less likely to make more extreme choices. However, it should be noted that there are several limitations in these studies and the research in this area is variable, as not all studies have been able to replicate similar findings (Arce-Ferrer, 2006).

The validity of newly translated measures is also of concern to cross-cultural researchers. Low quality translation and cultural variables may negatively influence the validity of a translated measure (Behling & Law, 2000). Construct validity, or the extent to which an instrument measures what it is intended to measure (Chronbach & Meehl, 1955), is assessed

through various types of validity evidences, such as content and predictive validity. Content

validity is the degree to which the items of a particular instrument adequately measure the

operational definition of construct of interest (Chronbach & Meehl, 1955). Consideration must

be taken in the way in which a construct is operationalized (Behling & Law, 2000). For example,

an operational definition of the construct "happiness" may vary greatly from culture to culture

and may contain varying dimensions. Therefore, a particular measure may not have adequate

content validity in the target culture, even though it has been shown in the original. Predictive

validity, which may also be compromised by a low quality translation, is the extent to which a

measure accurately predicts the criterion of interest (Chronbach & Meehl, 1955). Predictive

validity is the extent to which a behavioral screener can accurately identify those children who

are actually at-risk for a future difficulty. The predictive validity of a measure may vary across

cultures and societies. Other considerations in translating measures for cross-cultural use include

the utility of a measure, such as practical utility and economic utility (Behling & Law, 2000).

Some research on the translation of psychological measures from English to Spanish has

found that the internal structure of the measure changes when translated and used with the target

group. For example, in one study assessing a Spanish translation of Rotter's Locus of Control

scale, factor analyses revealed that different factor structures were created when the measure was

translated from English to Spanish (Garza, 1977). In other words, this instrument was measuring

distinct constructs when used with English versus Spanish speaking participants. Cultural biases

may also affect the internal structure of a translated psychological measure. Azocar, et al. (2001)

found that four items on the Spanish version of the Beck Depression Inventory are biased, in that

Latino participants were equally likely to endorse or not endorse these items, regardless of level

of depression. Other research, however, has found that with proper translation techniques and

considerations (i.e., selecting appropriate and studied translation methods, considering cultural

norms and beliefs, etc.), translated behavioral and emotional measures demonstrate psychometric

properties similar to those of their English counterparts (Achenbach & Rescorla, 2001;

Goodman, 1997; Reynolds & Kamphaus, 2004).

Considerations in Cross-Cultural Research with Hispanic Populations

While there are general areas of concern which apply across cultures and languages, there

are specific issues in conducting cross-cultural research and in using translated measures with

Hispanic populations. Hispanic individuals tend to be more cautious of researchers than

individuals of other cultures. One theory is that there is fear that immigration status information

of the individual may be used against them (Marin & Marin, 1991). Further, it has been found

that Hispanic individuals fear the possibility of experimental exploitation by those researchers

that are not Hispanic (Hirsch, 1973; Moore, 1973). It should be noted that the findings in this

area have been mixed, with some researchers finding little to no differences in perception of

scientific research (Marin, Perez-Sable, Marin, 1989; Wendler et al., 2005). With regard to

potential issues with normative equivalence, it has been found that Hispanic individuals,

particularly men, prefer lower levels of self-disclosure to strangers (Franco, Malloy, & Gonzalez,

1984). This factor may limit the ability to recruit Hispanic participants, subsequently restricting

the range of responses provided. Methods of remediating these issues include ensuring

confidentiality of responses, legitimacy of the researchers and study, and providing study

information, in detail, when recruiting participants (Marin & Marin, 1991). Additionally, it has

been found in some studies that Hispanic individuals are less likely to complete study materials

than other populations (Marin, Perez-Sable, Marin, 1989). Possible causes of this may be that:

(a) researchers fail to interact with participants in their preferred language; (b) regional Spanish,

rather than standard Spanish, is used for study materials; (c) there is a lack of researchers who are ethnically similar to the target group, (d), there is a lack of compensation for participation; (d) rating scales and study materials include complex and advanced language; (e) and data collection methods are inappropriate (Marin & Marin, 1991).

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, 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 if 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 may have difficulty "criticizing" one

another's translation, regardless if the criticism is intended to be constructive and/or useful to the

translation process. "Random probe technique," in which a draft of the translated measure is

given to a target group to complete, is another technique. Subjects are then 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).

An alternate type of translation is "translation/back translation." 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, (4) if there are substantial discrepancies between the two versions of the

measure, this process is completed again until an adequate translation is reached (Werner and

Campbell, 1970). This comprehensive process allows for more information to be provided than

the previous methods discussed. Further, the back translation method allows for the involvement

of monolingual researcher in the translation process.

Table 1. *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.

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 issues with this method of translation, back-translation method 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).

The final discussed 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 all translation methods, there are

strengths, as well as limitations. Determining the best method for a given study will help reduce

inequalities and inaccuracies in translation. For this particular study a combination of the back-

translation, committee translation methods, and field testing was used. This process of

combining methods has been successfully used in several studies and increases the potential for

an accurate and appropriate translation (Dumas, Martinez, & LaFreniere, 1998; Jacobsen, 1954).

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.

Existing Measures and Methods of Translation

Various widely accepted and published behavioral and emotional screening measures

have been translated from English to Spanish via the aforementioned methods. Table 2 outlines

the procedures and psychometric properties of such measures.

Table 2. *Methods of Translation and Psychometric Properties of Published Scales*

Measure	Method of Translation	Psychometric Properties	
		English	**Spanish**
Strengths and Difficulties Questionnaire (SDQ)	• Initial translation from English to Spanish. • Translated measure administered to group of parents and teachers to assess for cross-cultural issues. • Measure back-translated into English by independent expert.	• Cronbach's alpha *English* = .73	• Cronbach's alpha *Spanish* = .76 • Some differential factor loading of items onto different scales than original.
Behavior Assessment System for Children (BASC-2)	• Initial translation by translation company. • English and Spanish versions compared by nine bilingual psychologists. • Reviewer comments and suggestions given to translation company for final edits.	• *English* Parent Preschool (PRS): Externalizing Problems= .87 (ages 2-3), .90 (ages 4-5);. Internalizing Problems = .85 - .87. Adaptive Skills = .93 - .91. Behavioral Symptoms Index: .93 - .93.	• *Spanish* PRS, Preschool: Externalizing Problems = .80 (ages 2-5). Internalizing Problems = .79 (ages 2-5). Adaptive Skills = .89 (ages 2-5). Behavioral Symptoms Index: .85 (ages 2-5).
BASC-2 Behavioral and Emotional Screening System (BESS)	• Initial translation by translation company. • English and Spanish versions compared by nine bilingual psychologists. • Reviewer comments and suggestions given to translation company for final edits.	• *English* Parent, Preschool Rating Scale (PRS): .91 (age 3), .93 (ages 4-5).	• *Spanish* PRS, Preschool: preschool = .86
Child Behavior Checklist (CBCL)	• Translation/ back translation method.	• *English* alpha levels for broad band internalizing and externalizing composites = .78 to .97	• *Spanish* alpha levels for broad band internalizing and externalizing = .89 to .94

Strengths and Difficulties Questionnaire. The Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) is a social-emotional and behavioral screening measure consisting of five subscales (i.e., Emotional symptoms, Behavioral problems, Hyperactivity, Peer Relationship problems and Prosocial behaviors), with a total of 25 items. The English version of this measure has moderate psychometric properties and reliability (Cronbach's alpha = .73; Goodman, 2001). It has also been translated into several different languages, including Spanish. In order to translate the SDQ into Spanish, a bilingual child psychiatrist conducted the initial translation. Following an initial translation, the measure was administered to a group of parents and teachers, in Spain, to assess for cross-cultural issues such as semantic incompatibilities, scale equivalence, and unclear or misleading translations. Following a review from this focus group, the measure was then back-translated into English by an independent expert in translation. Analyses revealed that the Spanish version of the SDQ demonstrated moderate internal reliability values similar (Cronbach's alpha = .76) to those found in several European studies following translation of the measure into various languages, and to the original English version of the SDQ. Factor analyses, however, revealed that certain items on the Spanish version loaded onto different composite scales than were predicted (Rodriguez-Hernandez, et al., 2012).

Behavior Assessment System for Children, Second Edition. The Behavior Assessment System for Children, Second Edition (BASC-2; Reynolds & Kamphaus, 2004) is a widely used behavioral and emotional rating scale for children and adolescents between the ages of two and eighteen. The Spanish forms of the Parent Rating Scales (PRS; i.e., preschool, child, and adolescent forms) were created following the completion and standardization of the English version of the instrument. In order to create an appropriate translation, a translation company with expertise in psychological measures was tasked with reviewing translated Spanish items on

the first edition of the measure as well as to translate the newly created BASC-2 items.

Translations were conducted in a manner which would be understandable to various dialects and

cultures among the Spanish speaking population in the United States. Following an initial

translation by the translation services company, the English and Spanish BASC-2 measures were

compared by nine bilingual psychologists in the United States who had familiarity with the

measure. The compilation of reviewer comments and suggestions were given to the translation

company to review and consider.

While the English and Spanish versions of the BASC-2 Parent Rating Scales

demonstrated comparable psychometric properties, the Spanish versions of the measures

demonstrated lower internal consistency reliabilities. Internal consistency reliabilities for the

English PRS scales for the Externalizing Problems composite ranged from .87 - .94 on the

preschool, child, and adolescent forms. For the Internalizing Problems composite, internal

consistency reliabilities were .85 - .91. Coefficient alphas for the Adaptive Skills composite were

.91 - .95. The total Behavioral Symptoms Index yielded the coefficient alphas in the .93 - .95

range. On the Spanish PRS forms, analyses revealed that internal consistency reliabilities for the

Externalizing Problems composite were .80 - .90 on the preschool, child, and adolescent forms.

Coefficient alphas for the Internalizing Problems composite ranged from .78 - .82. Internal

consistency reliabilities for the Adaptive Skills composite were: .89 - .92. Lastly, coefficient

alphas for the total Behavioral Symptoms Index were .85 - .93.

BASC-2 Behavioral and Emotional Screening System. The BASC-2 Behavioral and

Emotional Screening System (BESS; Kamphaus & Reynolds, 2007) is a brief, universal

screening version of the BASC-2. A Spanish version of the BESS has also been created, using

the same procedural methods as described for the BASC-2. Again, higher levels of internal

consistency have been found for the English versions of the parent measures (preschool, age 3 = .91, age 4-5 = .93; child/adolescent, age 5-9 = .94, age 10-14 = .95, age 15-18 = .94) than the Spanish versions of the BESS, parent forms (preschool = .86, child = .90, adolescent = .86), as compared to the English and Spanish BASC-2 measures.

The Child Behavior Checklist. The Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001) is an established, broad measure of a child or adolescent's emotional and behavioral difficulties. The CBCL has been translated from English to various different languages, including Spanish. In order to obtain a Spanish version of the measure, the translation methods proposed by Brislin, Lonner, and Thorndike (1973) were used. Researches used the translation and back translation procedures to ensure a quality translation of the measure. Analyses revealed high internal consistency for many of the scales on the CBCL, with alpha levels ranging from .89 to .94 on the broadband internalizing and externalizing composites of the measure (Rubio-Stipec, M., Bird, H., Canino, G., & Gould, M, 1990).

Preschool Behavior Screening System

The Preschool Behavior Screening System (PBSS; Feeney-Kettler, Kratochwill, & Kettler, 2009) is a two-phase behavioral and emotional screening measure. This instrument contains nomination rubrics (Phase 1; PBSS-PSF P1) and rating scales (Phase 2; PBSS-PSF P2), in order to screen for children at-risk of developing internalizing and/or externalizing disorders. In PBSS-PSF P1, there are two nomination rubrics, one for internalizing behaviors and one for externalizing behaviors. Those children who receive high scores in PBSS-PSF P1 will be further assessed in PBSS-PSF P2 of the scale. In other words, only those children exhibiting elevated scores on either or both of the nomination rubrics in PBSS-PSF P1 will proceed to PBSS-PSF P2 of the instrument. Four scores are generated following the completion of PBSS-PSF P2, the Total

Score, the Internalizing Symptoms Scale score, the Externalizing Symptoms Scale score, and the Prosocial Behavior Scale score (Feeney-Kettler, Kratochwill, & Kettler, 2011).

A comprehensive procedure of item analysis and expert review was used to create the final versions of the English parent and teacher PBSS forms. Several analyses were conducted to determine the reliability and validity of the PBSS-Parent form. Cronbach's alpha revealed high to excellent internal consistency on the PBSS PBSS-PSF P2 (Internalizing Symptoms Scale = .81, Externalizing Symptoms Scale = .96, Prosocial Behavior Scale = .90, PBSS-PSF P2 Total Score = .95). Concurrent validity analyses demonstrated that the PBSS-PSF P1 nomination rubrics were highly sensitive predictors of a parent's Total Score (.96); however, PBSS-PSF P1 had low specificity (.49). The PBSS-PSF P1 nomination rubrics were also highly sensitive predictors of the Internalizing Symptoms Scale score (.80) and the Externalizing Symptoms Scale score (.90) (Feeney-Kettler, Kratochwill, & Kettler, 2011).

Conditional Probability Analyses

Conditional probability analyses are important in assessing the validity of screening measures, as they take into account the dichotomies which exist in screening measures (Kettler & Feeney-Kettler, 2011). Conditional probability analyses yield various indicators of the accuracy of a screening measure, which include sensitivity (the probability that a screening measure will correctly identify a child who actually has behavioral difficulties), specificity (the probability that a screening measure will correctly not identify a child who does not have behavioral difficulties), positive predictive value (PPV; the probability that a child identified as at-risk actually has behavioral difficulties), and negative predictive value (NPV; the probability that a child who is not identified as at-risk actually has no behavioral difficulties; Kettler & Feeney-Kettler, 2011). Kettler and Feeny-Kettler (2011) argue that low values for specificity and

PPV may be appropriate for screening measures in which making a false negative is much more costly than making a false positive. For the purposes of the PBSS, identifying a child as at-risk who does not actually have behavioral difficulties would be considered far less costly than not identifying a child who is actually at-risk. For this reason, the observed low specificity and PPV values are considered to be acceptable.

Research Questions

The purpose of the present study was to collect reliability evidence for the Preschool Behavior Screening System, Parent Spanish Form (PBSS-PSF) scale to draw valid inferences about the behaviors and emotions of preschool children. Concurrent validity analyses were conducted to determine comparability of this measure to the Spanish BESS. Further, comparative analyses were conducted to assess whether the PBSS-PSF demonstrates similar psychometric properties as its English counterpart. The following research questions were asked:

1. What is the internal structure validity evidence for PBSS-PSF P1 of the PBSS-PSF?

2. What is the relationship between the PBSS-PSF P1 nomination rubrics and the PBSS-PSF P2 scales?

3. How well do the two PBSS-PSF phases work together in identifying children who are at-risk?

4. How well does the PBSS-PSF P1 compare to and predict the BESS, Spanish Parent Total Score?

5. What is the internal consistency of PBSS-PSF P2 of the PBSS-PSF?

6. What is the internal structure validity evidence for the PBSS-PSF P2?

7. How well does the PBSS-PSF P2 compare to and predict the BESS, Spanish Parent Total Score?

8. What is the subjective experience of the participants in completing the PBSS-Parent,

    Spanish? Is the measure clearly written and easy to use?

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, it was predicted that the PBSS-PSF would, as

well. Table 3 outlines the predicted value for the correlations between PBSS-PSF P1 nomination

rubrics and indices on PBSS-PSF P2 of the PBSS-PSF.

Table 3. *Correlation Predictions for the PBSS-PSF P1 Nomination Rubrics and the PBSS-PSF*

P2 Indices on the PBSS-PSF

Variable	INR	ENR	ISS	ESS	PBS	Total
PBSS-PSF P1 Nomination Rubrics						
Internalizing Nomination Rubric (INR)		.30 to .60	.30 to .60	.30 to .60	-.40 to 0.00	.30 to .60
Externalizing Nomination Rubric (ENR)			.00 to .40	.60 to .80	-.40 to 0.00	.30 to .60
PBSS-PSF P2 Rating Scale						
Internalizing Symptom Scale (ISS)				.00 to .40	-.40 to 0.00	.60 to .80
Externalizing Symptom Scale (ESS)					-.40 to 0.00	.80 to 1.00
Prosocial Behavior Scale (PBS)						-.80 to -1.00
Total Scale (Total)						

Note. Predictions based on findings from Feeney-Kettler, Kratochwill, & Kettler, 2011.

It was also hypothesized that the PBSS-PSF would be comparable to the BESS, Spanish

Parent, as both screening instruments measure similar constructs. As the BESS, Spanish Parent

provides only one total score (preschool = .86), it was predicted that the internal consistency for

the PBSS-PSF Total Score would lie within a range of .85 to .95. Lastly, Table 4 delineates the

predicted values for conditional probability of the PBSS-Parent, Spanish.

Table 4. *Predicted Conditional Probability Analyses of the PBSS-PSF.*

	Descriptions	Predicted Value Ranges
Internal Consistency	• Assesses whether items intended to measure the same construct produce similar scores.	• PBSS-PSF P2 Internalizing Symptoms Scale = .80-1.00 • PBSS-PSF P2 Externalizing Symptoms Scale = .80- 1.00 • PBSS-PSF P2 Prosocial Behavior Scale = .80-1.00 • PBSS-PSF P2 Total Score = .80- 1.00
Sensitivity	• Probability that the PBSS-PSF will correctly identify those children who actually have behavioral difficulties.	• PBSS-PSF P1 nomination rubrics will be sensitive predictors of PBSS-PSF P2 Total score (.80-1.00) and of the BESS, Spanish Parent Total Score (.80-1.00). • PBSS-PSF P1 nomination rubrics will be sensitive predictors of PBSS-PSF P2 Internalizing Symptoms Scale score (.80- 1.00) and Externalizing Symptoms Scale score (.80-1.00).
Specificity	• Probability that the PBSS-PSF will correctly not identify those children who do not have behavioral difficulties.	• PBSS-PSF P1 nomination rubrics will not be specific predictors of PBSS-PSF P2 Total score (.40-.60) and of the BESS, Spanish Parent Total Score (.40-.60). • PBSS-PSF P1 nomination rubrics will not be specific predictors of PBSS-PSF P2 Internalizing Symptoms Scale score (.40-.60) and Externalizing Symptoms Scale score (.40-.60).
Positive Predictive Value	• Probability that a child identified as at-risk actually has difficulties.	• Positive predictive value of PBSS-PSF P2 Total Score (.40-.60), Internalizing Symptom Scale score (.40-.60), Externalizing Symptoms Scale score (.40-.60), and BESS, Spanish Parent Total Score (.40-.60) will be low.
Negative Predictive Value	• Probability that a child who is not identified as at-risk actually has no difficulties.	• Negative predictive value of PBSS-PSF P2 Total Score (.80- 1.00), Internalizing Symptom Scale score (.80- 1.00), Externalizing Symptoms Scale score (.80- 1.00), and BESS, Spanish Parent Total Score (.80-1.00) will be high.

Note. Descriptions obtained from Kettler et al., 2013. Predictions based on findings from Feeney-Kettler, Kratochwill, & Kettler, 2011.

Method

Participants

The sample included 49 Spanish-speaking 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). The mean age of the children who were rated was 52.04 months (4.33 years; SD= 7.01).

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 largest ethnic representation in the sample. More

specifically, the country of origin for participants included Mexico, Dominican Republic,

Honduras, Ecuador, and Bolivia. Table 5 delineates further demographic information of the

aforementioned sample.

Table 5. *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%

Measures

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

Preschool Behavior Screening System. The Preschool Behavior Screening System, Parent version (PBSS; Feeney-Kettler, Kratochwill, & Kettler, 2009) is a two-phase, universal screening instrument, shown to have acceptable reliability and validity. It includes two nomination rubrics (PBSS-PSF P1), which assess a child's internalizing and externalizing behaviors, and a 60-item rating scale (PBSS-PSF P2), assessing internalizing, externalizing, and prosocial behaviors. The PBSS Parent version yields four index scores: Internalizing Symptoms Scale (alpha= .81), Externalizing Symptoms Scale (alpha= .96), Prosocial Behavior Scale (alpha= .90), and PBSS-PSF P2 Total Score (alpha= .95).

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 for every child, 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 will be completed for all children. Parents were asked to complete the newly created Spanish, parent version of this measure. Field testing was used to determine the reliability and validity of this scale.

BASC-2 Sistema de Evaluación Emocional y de Conducta. The BASC-2 Sistema de Evaluación Emocional y de Conducta (BESS; Kamphaus & Reynolds, 2007) is a brief, universal screening system, used to identify the behavioral and emotional 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 BESS Spanish, parent form has demonstrated high internal consistency (preschool= .86).

Demographic Questionnaire. Participants completed a demographic questionnaire, in Spanish, which 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 *(Appendix A)*.

Evaluation Survey. The evaluation survey was used to gain qualitative and quantitative information regarding the subjective experience of the participants in completing the measure. Questions include 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 *(Appendix B)*.

Procedures

Stage 1: Translation. To attain an accurate and valid translation of the Preschool Behavior Screening System (PBSS; 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 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 had received their Master's degrees in Spanish Translation and are native speakers of Spanish. The head translator, who is

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 is a university professor of Spanish, as well as a certified member of the American Translators Association. The third reviewer holds 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 of the PBSS-Parent form back into English. This method was chosen, as it has been widely studied and allows for a direction comparison of the original and translated versions of the measure. Further, it allows for the involvement of monolingual, English speaking researchers in the translation process (Behling & Law, 2000). The back-translations were completed by two individuals with Master's level degrees in Spanish Translation. One translator has obtained a Master's degree in Spanish Translation and is currently a school teacher. The second translator has been a Notary Public for over ten years, specializing in the translation of legal immigration papers. She is also a native Spanish speaker and has training in Spanish translation. Each translator was assigned a subset of items on the PBSS-Spanish, Parent to translate from Spanish to English. Following the completion of the back-translated version of the PBSS-Parent, Spanish, a second committee compared the original, English version of the PBSS-Parent to the back-translated measure. This committee consisted of two Ph.D. level researchers in Educational Psychology, along with a doctoral student in school psychology, the principle investigator of this project. 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 among corresponding items on

the original PBSS- Parent and the PBSS- Parent English, back translated version of the measure.

Committee members were asked to rate each corresponding item on their level of semantic

similarity. In order to reduce the previously discussed limitations of this method, a stringent

definition of unacceptable similarity was adopted. More specifically, those items which 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 holds a doctoral degree in second language education.

      Stage 2: Field Testing. The PBSS-PSF was field tested on a sample (N= 49) of bilingual

and monolingual-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 ensured the completion, and also ensured that all parents were fully informed of

their participation and all questions were answered. Approval to conduct this study was obtained

from the Institutional Review Board at Rutgers, The State University of New Jersey prior to data

collection at the preschools.

Data Analysis

Several quantitative techniques were used to analyze the collected data. Table 6 outlines

the types of evidence that were addressed and the analyses that were conducted.

Table 6. *Data Analyses*

Type of Evidence	Variables	Analyses
Accessibility	• Evaluation surveys.	• Qualitative/ Quantitative analyses: 　○ Assessing for themes. 　○ Percentage(s) of yes/no responses.
Reliability	• Items and scales on the PBSS-PSF P2. • Items on BESS, Spanish Parent form.	• Cronbach's alpha for: 　○ PBSS, Spanish Total score. 　○ PBSS, Spanish Internalizing, Externalizing, Prosocial scales. 　○ BESS, Spanish Parent total score.
Internal Structure	• PBSS-PSF P2.	• An exploratory comparisons of means was used to assess whether those who were identified as "at risk" on PBSS-PSF P2 scored higher than those who were not. • Correlation matrix to assess relation between the phases and scales. • Conditional probability analyses.
Concurrent Validity	• PBSS-PSF P1 and PBSS-PSF P2. • BESS, Spanish Parent form.	• Pearson correlations between: 　○ BESS, Spanish Parent Total score and PBSS-PSF P1 nomination rubrics. 　○ BESS, Spanish Parent Total score and PBSS PBSS-PSF P2 Total score. 　○ BESS, Spanish Parent Total score and each PBSS PBSS-PSF P2 index. • Exploratory comparisons of means between PBSS-PSF P1 scores and BESS, Spanish Parent Total score. • Conditional probability analyses.

To assess the internal consistency of the PBSS-PSF, coefficient alpha was calculated. The

same was done for the BESS, Spanish Parent 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 (i.e., $\geq$ .90 = excellent, .80 = high, .70- .79 = moderate, .60-

.70 = low, < .60 = unacceptably low). A correlation matrix was also created to assess the

relationships between the PBSS-PSF P1 nomination rubrics, the PBSS-PSF P2 scales, and the

BESS, Spanish Parent. Cohen's (1992) classification system, with Hopkin's (2001) extensions,

were used to describe the Pearson correlations. This system classifies r = .00 as a nonexistent

effect, $r = .10$ or $r = -.10$ as a small effect, $r = .30$ or $r = -.30$ as a medium effect, $r = .50$ or $r = -$

.50 as a large effect, $r = .70$ or $r = -.70$ as a very large effect, and $r = .90$ or $r = -.90$ as a nearly

perfect correlation. These are inner boundary ranges, such that .10 to .30 is considered to be

small. Exploratory comparisons of means were used to assess the relationship between PBSS-

PSF P1 and PBSS-PSF P2 of the PBSS- Spanish. Concurrent validity analyses were conducted to

compare the PBSS-PSF and the BESS, Spanish Parent form. These analyses exploratory

comparisons of means and Pearson correlations. Conditional probability analyses were also used

to assess the sensitivity, specificity, and predictive value of this screening measure.

Classification categories from Kettler and Feeney- Kettler (2011) were used (i.e., .80- 1.00 =

high; .60- .79 = moderate; .40- .60 = low). One-tailed significance tests were used throughout the

analyses. Lastly, descriptive statistics will be used to analyze the qualitative information

collected from the questionnaires.

Prior to running the proposed analyses, data was systematically analyzed and

cleaned. All data was double entered by two data clerks. Differences among the two entries were

calculated to assess for errors in data entry. Further, the data was checked for values which were

outside of a given range (e.g., a score of 6 inputted for a 5-item scale). To account for missing

data on the PBSS-PSF nomination rubrics, scores were imputed by taking the middle score on

each of the two questions which comprise the nomination rubrics. If parents indicated a certain

number of symptoms for their child at the top of the rubric, however, left the corresponding

question on the rubric blank, that question was scored according to the number of symptoms a

parent endorsed. For missing data on the rating scales, scale scores were computed by taking the average scores of items and then multiplying that average by the number of items on the given scale. Separate data sets were kept for imputed and missing data.

Results

To assess how parents rated their children on the constructs of interest, scale means were calculated for the sample. On the PBSS-Spanish, the mean rating for the Internalizing Symptoms Scale (ISS) was 32.21 (SD = 8.68; Possible range = 20 - 100; Observed range = 21- 56) and the mean rating for the Externalizing Symptoms Scale (ESS) was 33.97 (SD = 12.38; Possible range = 22 - 110; Observed range= 22 - 70). The Prosocial Behavior Scale (PBS) had a mean rating of 40.16 (SD = 12.58; Possible range= 17 - 85; Observed range = 38 - 85). 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 (Total Score) had a mean rating of 108.98 (SD = 22.36). On the BESS, Spanish Parent Scale, the mean rating for the Total Score was 48.74 (SD = 10.95).

Internal Consistency

To answer the first research question, the internal consistency coefficient (Cronbach's alpha) was calculated for each of the four scale scores on the PBSS-PSF P2 rating scale. Table 7 outlines the Cronbach's alpha values for the English and Spanish versions of the PBSS-PSF form. 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 in the moderate range, although still acceptable for screening measures (Murphy & Davidshofer, 2004). Overall, reliability coefficients for the PBSS-PSF were comparable to those of the PBSS-English Parent form.

Table 7. *Reliability Coefficients (Cronbach's alpha) of PBSS-PSF P2 Scales*

PBSS-PSF P2 Scale	PBSS- English	PBSS-Spanish
Internalizing Symptoms Scale	.81	.78
Externalizing Symptoms Scale	.96	.92
Prosocial Behavior Scale	.90	.89
Total Score	.95	.89

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

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.

Internal Structure

Correlation Matrix. To answer the second question, Pearson product-moment correlation coefficients were computed to assess the relationships between the PBSS-PSF P1 nomination rubrics and PBSS-PSF P2 composite scales. Cohen's (1992) classification system for Pearson correlations with Hopkin's (2001) extensions of the system were used to qualitatively describe the relationships. One-tailed significance tests were used. Table 8 outlines the Pearson correlation values.

Table 8. *Correlations for the PBSS-PSF P1 Nomination Rubrics and the PBSS-PSF P2 Indices on the PBSS-PSF.*

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
PBSS-PSF P1 (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 (PBSS-PSF P2) Rating Scale							
Internalizing Symptom Scale (ISS)				.40*	.09	.55*	.10
Externalizing Symptom Scale (ESS)					-.17	.80*	.61*
Prosocial Behavior Scale (PBS)						-.61*	.51*
Total Score (Total)							.67*

*Correlation is significant at the 0.05 level (1-tailed).

When assessing the relationship between the PBSS-PSF P1 INR and ENR, analyses revealed a large and significant positive relationship between the two nomination rubrics ($r = .67$, $p < .05$). In assessing PBSS-PSF P1 and PBSS-PSF P2, the PBS yielded medium and significant negative relationships with the INR ($r = -.30$, $p < .05$) and the ENR ($r = -.30$, $p < .05$). There was a medium and significant positive correlation between the PBSS-PSF P2 Total Score and the ENR ($r = .35$, $p < .05$).

There was a medium positive correlation between the ISS and the ESS on PBSS-PSF P2 of the PBSS-PSF ($r = .40$, $p < .05$). The ISS, ESS, and PBS are subscales that contribute to the

PBSS-PSF Total Score. There was a large positive correlation between the Total Score scale and the ISS ($r = .55$, p < .05) and a very large positive relationship between the Total Score and the ESS ($r = .80$, p < .05). The PBS yielded a large negative relationship with the Total Score scale ($r = -.61$, p < .05).

When assessing the yielded correlations, it is observed that the constructs within each phase correlate more highly to one another than the same construct across phases, specifically in regards to internalizing and externalizing behaviors. Similar to the INR and ENR in PBSS-PSF P1, which yielded a large positive relationship ($r = .67$, p < .05), the ISS and ESS in PBSS-PSF P2 demonstrated a medium correlation ($r = .40$, p < .05). This differs from relationship seen within the same construct across phases. The INR and ISS ($r = .02$, p > .05) yielded a small and nonsignificant correlation, comparable to the small relationship between the ENR and ESS ($r = .24$, p > .05).

Exploratory Comparisons of Means. Exploratory comparisons of means were conducted to assess whether individuals who are identified as at-risk on PBSS-PSF P1 of the PBSS-PSF score higher on PBSS-PSF P2 than those individuals who were not identified as at-risk. Total scores for the Internalizing and Externalizing Nomination Rubrics were dichotomized (i.e., identified as at-risk versus not identified) and compared to the PBSS-PSF P2 Total Score. Thirty-six participants were identified to be "at-risk" on the INR and thirty were identified on the ENR. Consistent with Feeney-Kettler, Kratochwill, & Kettler (2011), scores of five and above were considered to be at-risk on the INR, while scores of six and above were considered at-risk on the ENR. The relationship between being identified on either or both of the nomination rubrics in relation to the PBSS-PSF P2 Total Score was also assessed. Table 9 outlines the group

means of PBSS-PSF P2 Total Scale Scores for those identified or not identified as at-risk on the

PBSS-PSF P1 nomination rubrics.

Table 9. *Mean Parent Ratings on the PBSS-PSF PBSS-PSF P2 Total Score Scale*

		PBSS-PSF P2 Total Score		
		N	**Mean**	**Standard Deviation**
Internalizing Nomination Rubric	*Identified*	34	111.12	20.79
	Not Identified	10	103.70	27.64
Externalizing Nomination Rubric	*Identified*	30	112.57	22.51
	Not Identified	14	103.94	21.81
Either/Or Nomination Rubric	*Identified*	36	111.75	22.50
	Not Identified	8	99.00	21.31

Analyses revealed those who were rated as at-risk on the INR (M = 111.12, SD = 20.79)

yielded a higher mean score on the PBSS-PSF P2 Total Score than those who were not identified

(M = 103.70, SD = 27.64). On the ENR, those who were identified to be at-risk (M = 112.57, SD

= 22.51) had higher means on the PBSS-PSF P2 Total Score than those who were not identified

(M = 103.94, SD = 21.81). Lastly, the relationship between whether participants rated their

children in the at-risk range on either or both of the nomination rubrics and the PBSS-PSF P2

Total Score was assessed. As expected, those who scored in the at-risk range on one or both of

the nomination rubrics (M = 111.75, SD = 22.50) scored higher on the PBSS-PSF PBSS-PSF P2

Total Score than those who did not (M = 99.00, SD = 21.31).

Conditional Probability Analyses. Conditional probability analyses were computed to

assess how well PBSS-PSF P1 and PBSS-PSF P2 of the PBSS-PSF work together in identifying

those children who are at-risk and those who are not. Both the nomination rubric scores and the

PBSS-PSF PBSS-PSF P2 Total Score were dichotomized into two categories: those who were

identified as at-risk and those who were not identified. More specifically, the nomination rubric

scores were dichotomized into those who were identified on either or both of the nomination rubrics and those who were not identified at all. The dichotomization of the PBSS-PSF P2 rating scale was done by creating new variables, using a cut score of 145. Any child with a score of 145 or higher was considered to be at-risk. Kettler and Feeney-Kettler's (2011) qualitative descriptions for the conditional probability indices were used (i.e., .00 - .40 = very low; .40 - .60 = low; .60 - .80 = moderate; .80 or greater = high). Findings revealed that the PBSS-PSF P1 nomination rubrics were highly sensitive (1.0) in predicting the PBSS-PSF P2 Total Score, but had very low specificity (.19). PPV was very low (.05), while NPV was high (1.0).

Concurrent Validity

Exploratory Comparisons of Means. Exploratory comparisons of means were computed to compare the PBSS-PSF P1 nomination rubric scores to the BESS, Spanish Parent total T-scores. Once again, participant scores for the PBSS-PSF Internalizing and Externalizing Nomination Rubrics were dichotomized (i.e., identified as at-risk versus not identified) and compared to the BESS, Spanish Parent total score. Table 10 delineates the mean scores on the BESS, Spanish Parent for those participants identified or not identified as "at risk" on the INR and ENR. Further, the relationship between being identified on either one or both of the nomination rubrics and the BESS, Spanish Parent total score was assessed.

Table 10. *Mean Parent Ratings on the BESS, Spanish Parent Total Score*

		N	**Mean**	**Standard Deviation**
		BESS, Spanish Parent Total Score		
Internalizing Nomination Rubric (INR)	*Identified*	35	46.03	10.40
	Not Identified	13	42.15	7.60
Externalizing Nomination Rubric (ENR)	*Identified*	29	46.83	10.17
	Not Identified	19	42.16	8.70
Either/Or Nomination Rubrics	*Identified*	37	45.81	10.47
	Not Identified	11	42.18	6.70

Group means for the BESS, Spanish Parent were higher for those individuals who were identified to be at-risk (M= 46.06, SD= 10.40) than those who were not identified on the INR (M= 42.15, SD= 7.60). Group means were also higher on the BESS, Spanish Parent for those identified as at-risk (M= 46.83, SD= 10.17) on the ENR. When assessing the relationship between being identified on one or both of the nomination rubrics and total scores on the BESS, Spanish Parent, findings revealed that those who identified on either one or both of the rubrics yielded a higher mean score (M = 45.81, SD = 10.47) on the BESS, Spanish Parent than those who were not (M = 42.18, SD = 6.70).

Conditional Probability Analyses. Conditional probability analyses were calculated to assess the accuracy of the PBSS-PSF in correctly identifying those children who are truly at-risk and those who are not. For these analyses, the PBSS-PSF PBSS-PSF P1 nomination rubrics, the PBSS-PSF PBSS-PSF P2 Total Score, and the total score from the BESS, Spanish Parent form were used. Again, all of the scores were dichotomized into those who were identified as at-risk and those who were not. Table 11 depicts the sensitivity, specificity, PPV, and NPV of each PBSS-PSF P1 nomination rubric in predicting the BESS, Spanish Parent Total Score. A score of 61 or above on the BESS, Spanish Parent was considered to be at-risk.

Table 11. *Conditional Probability Analyses of PBSS-PSF P1 and PBSS-PSF P2 in Predicting BESS, Spanish Parent Total Score*

Variable	Sensitivity	Specificity	PPV	NPV
Internalizing Nomination Rubric (INR)	1.0	.31	.17	1.0
Externalizing Nomination Rubric (ENR)	.83	.43	.17	.95
Either/or Nomination Rubrics	1.0	.26	.16	1.0
PBSS-PSF P2 Total Score	.20	.98	.50	.91

Findings revealed that the INR was a highly sensitive (1.0) predictor of the BESS total score, but had low specificity (.31). PPV was very low (.17) and NPV was high (1.0). The ENR was a highly sensitive (.83) predictor of the BESS, Spanish Parent total score, but had low specificity (.43). PPV was very low (.17), while NPV was high (.95).

Next, conditional probability analyses were computed to assess how well being identified as at-risk or not at-risk on either one or both of the nomination rubrics predicted being identified or not on the BESS, Spanish Parent. The PBSS-PSF PBSS-PSF P1 nomination rubrics demonstrated high sensitivity (1.0) in predicting the BESS, Spanish Parent total score and very low specificity (.26). PPV was very low (.16), while NPV was high (1.0).

Sensitivity, specificity, PPV, and NPV were calculated for the PBSS-PSF PBSS-PSF P2 Total Score in predicting the BESS, Spanish Parent total score. Results revealed that the PBSS-PSF PBSS-PSF P2 Total Score has very low sensitivity (.20) and high specificity (.98) in predicting the BESS, Spanish Parent total score. PPV was low (.50) and NPV was high (.91).

Pearson Product-Moment Correlation Coefficients. Pearson correlations were used to compare the PBSS-PSF form with the BESS, Spanish Parent form. There was a small and significant relationship between the INR and the BESS, Spanish Parent Total Score. (r = .26, p < .05) and a medium and significant correlation between the ENR and the BESS, Spanish Parent Total Score (r = .35, p < .05). There was a large and significant positive relationship between PBSS-PSF 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.

Evaluation Survey

Evaluation survey data was systemically collected and analyzed to assess for emergent themes and patterns among participant responses. A process by which convergent and divergent data (i.e., both evidence that supported and evidence that disagreed with recurring themes) was used to assess for the pervasiveness of themes. When exploring evidence for the overarching themes, negative case examples were identified and considered. Presenting opposing findings provided a neutral and complete evaluation of the parents' experiences using the PBSS-PSF (Patton, 2002).

An evaluation survey was provided to each parent participant to gain more information on the subjective experience of parents in completing the PBSS-Parent, Spanish. Information of interest included the amount of time taken by each parent to complete the form, whether or not the parent believed this scale would provide useful information about his or her child, and the clarity and structure of the scale. It should be noted 44 of the participants completed at least one item on the evaluation survey, with not all of the parents responding to all of the evaluation questions. Five parents did not complete any of the evaluation survey questions. 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 parent endorsed concern, two of these children 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 of these children 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 minutes (SD =

13) to complete both of the PBSS-PSF P1 nomination rubrics and an average of 20 minutes (SD

= 12) to complete the PBSS-PSF P2 rating scale.

When asked to elaborate on the aforementioned yes/no responses provided, only 22 of the

parents in the sample elaborated on one or more of the survey questions. In regards to whether or

not the PBSS-PSF provides valuable information about preschool children, two overarching

themes were identified from 15 of the parents who provided further information. 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." Another parent reported, "*Si, porque así uno puede saber si los niños*

están desarrollando algún problema emocional./ Yes, because that way one can know if the

children are developing an emotional problem."

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." Another

stated that it can help "*para corregir los errores de cada niño*/ to correct the errors of each

child."

When asked to elaborate on whether the PBSS-PSF is clearly written and organized,

sixteen parents provided responses. Results were mixed for this area of questioning. As

mentioned, 84% of parents believed that the PBSS-PSF is clear and easy to use. One parent

noted, "*Las preguntas fueron claras y fácil de contestar*/ The questions were clear and easy to answer." It should be noted, however, that 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 PBSS-PSF P1 of the measure incorrectly. In total, 30 out of the 49 participants completed the PBSS-PSF PBSS-PSF P1 nomination rubrics incorrectly. Of the 30 participants, 29% of the participants left the nomination rubrics partially or entirely blank, 12% endorsed a number of symptoms different than what was previously endorsed, and 24% percent used the Likert scale at the bottom of the nomination rubrics to rate each symptom at the top. 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." Another parent offered a suggestion on how to improve the clarity of the nomination rubrics; "*La primera fase fue un poco confusa ya que no esta bien explicado como marcar. Yo creo que si se ponen dos columnas enfrente de cada síntoma para marcar cuantos síntomas y la frecuencia de cada uno, seria mas sencillo.*/ The first phase was a little confusing because it was not well explained how to mark it. I think that if you put two columns in front of each symptom to mark how many symptoms and the frequency of each, it would be simpler." Overall, parents demonstrated considerable difficulty in completing the PBSS-Parent, Spanish, particularly the PBSS-PSF P1 nomination rubrics.

Discussion

Following a rigorous translation process, the PBSS-PSF version was created to screen for the presence of behavioral and emotional difficulties in preschool-age 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-English 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 bilingual and monolingual-Spanish speaking parents of preschool children.

Key Findings

PBSS-PSF P1 was found to be a highly sensitive, but not very specific tool when predicting PBSS-PSF P2. Similarly, PBSS-PSF P1 was found to be a highly sensitive, but not very specific tool when predicting to the BESS, Spanish Parent. PBSS-PSF P1 did not work as well as expected with the PBSS-PSF P2 in identifying those individuals who may be at-risk. This was due to inconsistencies in the way in which parents completed PBSS-PSF P1 and PBSS-PSF P2. Many of the surveyed parents reported that the nomination rubrics were difficult to complete, with a large portion of the sample leaving the rubrics blank or completing them incorrectly. PBSS-PSF P2 demonstrated internal consistency in the moderate, high, and excellent ranges. All of the correlations among the PBSS-PSF P2 scales were found to be within the predicted and expected ranges. Further, the PBSS-PSF PBSS-PSF P2 scales yielded large and significant correlations with the BESS, Spanish Parent.

PBSS-PSF P1 Nomination Rubrics

The first phase of the PBSS-PSF consists of two nomination rubrics, screening for the presence of internalizing and externalizing symptoms. Analyses were conducted to assess the

psychometric properties of PBSS-PSF P1, which included the comparison of PBSS-PSF P1 to PBSS-PSF P2 of the PBSS, as well as to the BESS, Spanish Parent. Results indicated that the nomination rubrics present with certain strengths, as well as certain weaknesses.

Internal Structure Validity. The INR and ENR shared a large and significant relationship ($r = .67$, $p < .05$), just above the predicted range of .30 - .60. This correlation between the INR and ENR is similar to that of the English version of the measure ($r = .55$, $p < .05$). Kettler and Feeney-Kettler (2011) indicate that 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 which comprise

PBSS-PSF Phase 1 and Phase 2

As 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. While many of the PBSS-PSF P2 scales correlated with one another as predicted, the PBSS-PSF P1 nomination rubrics did not yield the predicted and expected correlations with most of the PBSS-PSF P2 scales. There was a small and non-significant negative relationship between the INR and the ISS, as well as between the ENR and ESS. An almost nonexistent relationship between the two pairs is an unexpected outcome, as they are intended to measure similar symptoms. Further, the INR and ENR yielded small and non- significant relationships with the remaining PBSS-PSF P2 ISS, ESS, and Total Score scales, much lower than expected. It is hypothesized that the reason for the relatively low magnitudes of these relationships may be due to participants' difficulties in appropriately completing the nomination rubrics. Incorrect and blank nomination rubrics from

most of the sample may have affected the way in which the nomination rubrics and scales correlated with one another in this study.

Also of interest was whether or not PBSS-PSF P1 and PBSS-PSF P2 of the PBSS-PSFwork well together in identifying those children who are at-risk for developing behavioral and emotional difficulties. More specifically, it was of interest to see whether those who were identified as at-risk on PBSS-PSF P1 also scored higher on PBSS-PSF P2 than those who were not identified. It was found that those who were identified to be at-risk on either or both of the nomination rubrics also yielded higher means on the PBSS-PSF P2 rating scale. While this is an encouraging finding, there is considerable evidence to suggest that the two phases of the PBSS-PSF may not work together as predicted. The PBSS-PSF P1 and PBSS-PSF P2 were created to specifically work with one another to identify at-risk children. Inaccurate completion of the first phase of the scale would affect the interconnected relationship with the second phase of the screening tool. Further, the low sample sized used for this study, along with the low percentage of children with reported behavioral and emotional difficulties may have restricted the observed outcomes.

In further assessing the accuracy of the PBSS-PSF P1 in identifying those children who are and are not at-risk, conditional probability analyses were conducted. According to Kettler and Feeney-Kettler (2011), indices on conditional probability analyses between .80 and 1.00 suggest that the screening measure is working appropriately. However, in cases in which false negatives are more costly than false positives, as in this case, it is acceptable to have lower specificity and PPV and higher sensitivity and NPV. As the nomination rubrics work together in identifying at-risk children who will move on to PBSS-PSF P2, it is more useful to assess the rubrics as a unified phase. When considering individuals who were unidentified on either one or

both of the nomination rubrics, it was found that PBSS-PSF P1 is highly sensitive, but not very

specific in predicting PBSS-PSF P2. The frequency count for false positives was 34, while the

count of false negatives was zero. While 34 false positives may not be ideal, zero false negatives

is what would be expected for the first phase of the measure and provides evidence to support its

efficacy. The cost of falsely identifying a child results in a parent having to take approximately

15-20 extra minutes to complete PBSS-PSF P2. This is a fairly innocuous consequence, as

compared to a child in need who is not identified and will, therefore, not receive early

intervention services at the time of the screening. With a base rate of 4.5% and only 2 individuals

truly having behavioral and emotional problems in the population, the very low PPV and high

NPV is acceptable.

Concurrent Validity. Pearson correlations were conducted to assess the relationship

between the nomination rubrics and the BESS, Spanish Parent. While small and medium

correlations were found, they were higher than those correlations between the nomination rubrics

and PBSS-PSF P2 of the PBSS-PSF. Further, the correlations between PBSS-PSF P1 and the

BESS, Spanish Parent were significant, while those with PBSS-PSF P2 were not. An exploratory

comparison of means showed group means were higher on the BESS, Spanish Parent for those

who were identified on either or both of the nomination rubrics than for those who were

identified as at-risk. This is an encouraging finding and provides evidence that the nomination

rubrics work well together as one phase.

It was found that PBSS-PSF P1 demonstrated high sensitivity and low specificity in

predicting the BESS, Spanish Parent. Further, PBSS-PSF P1 yielded a high number of false

positives (31) and a no false negatives. A low rate of false negatives is an important feature of

PBSS-PSF P1, as it is more costly to miss a child who is truly at-risk than to identify a child who

is not truly at-risk. PBSS-PSF P1 demonstrated very low PPV (.16), and high NPV (1.0) in

predicting the BESS, Spanish Parent total score. This is an acceptable outcome, as the base rate

for those who actually have issues in the population was 12.5%. When the base rate is low it is

difficult to obtain a high PPV, as the rate of those with true issues in the population is low

(Kettler & Feeney-Kettler, 2011).

Summary. Table 12 outlines the overall strengths and weaknesses demonstrated by the

PBSS-PSF P1 nomination rubrics.

Table 12. *Strengths and Weaknesses of the PBSS-PSF P1 Nomination Rubrics*

Strengths	*Weaknesses*
• Large and expected correlation between INR and ENR. • PBSS-PSF P1 is a highly sensitive predictor of PBSS-PSF P2 and BESS, Spanish Parent. Very low rate of false negatives. • Acceptable PPV and NPV for PBSS-PSF P1 predicting PBSS-PSF P2 Total Score and NPV for predicting BESS Total Score. • Significant correlations between PBSS-PSF P1 and the BESS Total Score. • Higher means on PBSS, Spanish PBSS-PSF P2 Total Score and BESS Total Score for those identified as at-risk on PBSS-PSF P1.	• Small, non-significant correlations between the INR and ISS and the ENR and ESS. • Nomination rubrics yielded weak and unexpected correlations with other PBSS-PSF P2 scales. • PBSS-PSF P1 demonstrated very low specificity in predicting PBSS-PSF P2 Total Score.

Overall, the PBSS-PSF P1 nomination rubrics work well as one phase. The PBSS-PSF P1

nomination rubrics correlated with one another as predicted. PBSS-PSF P1 is a good predictor in

identifying those who were at-risk on PBSS-PSF P2 and on the BESS, Spanish Parent. PBSS-

PSF P1 also yielded higher scores on the PBSS-PSF P2 Total Score and on the BESS, Spanish

Parent for those who were at-risk. The INR and the ENR demonstrated significant relationships

with the BESS, Spanish Parent Total Score. However, the INR and the ENR yielded weak and

unexpected correlations with the PBSS-PSF P2 scales.

The two phases of the PBSS-PSF may have been heavily influenced by the errors that parents displayed in completing PBSS-PSF P1. As parents had difficulty completing the PBSS-PSF P1 nomination rubrics, but were able to complete the rating scale adequately, it is not surprising that the relationships between the PBSS-PSF P1 rubrics and the PBSS-PSF P2 scales are small. It is as though PBSS-PSF P1 and PBSS-PSF P2 are independent measures, and were completed as such by parents. Similarly, since parents completed each nomination rubric in the same manner, it is expected that the PBSS-PSF P1 rubrics would correlate better with each other than with the PBSS-PSF P2 scales. While the relationship between PBSS-PSF P1 and 2 was greatly impacted by the errors in PBSS-PSF P1, results indicate that the PBSS-PSF P1 nomination rubrics worked as expected as a single phase. PBSS-PSF P1 is also a good predictor of the BESS, Spanish Parent Total Score, as well as being able to accurately identify those who have higher means on the BESS, Spanish Parent Total Score as at-risk. The nomination rubrics are highly sensitive, with very low false negative rates. These findings may be attributed to the low number of identified individuals on PBSS-PSF P2, making it easier to avoid a false negative.

Observed Difficulties. Many of the parent participants struggled in completing the PBSS-PSF. More specifically, parents demonstrated difficulty in completing the two nomination rubrics in the first phase of the tool. In fact, one or more scores for the nomination rubrics were imputed for over 40% of the sample. While some parents left the nomination rubrics blank, many others completed the PBSS-PSF P1 rubrics incorrectly. For example, some parents indicated a higher or lower number of internalizing and/or externalizing symptoms than previously endorsed. Other parents were observed using the Likert scale at the bottom of each of the nomination rubrics (i.e., the Likert scale which asks parent to rate the frequency of the aggregate of symptoms) to rate each potential symptom listed at the top of the nomination rubrics. Some

parents also specifically identified the nomination rubrics as being problematic and unclear on the evaluation survey. While there is no clear explanation for the cause of these difficulties, several different factors were considered.

The first step in determining the potential cause of parents' difficulties in completing the PBSS-PSF P1 nomination rubrics was to assess the method of translation chosen for the rating scale. Two of the most common and widely used methods of translation, the use of a professional translation service (i.e., committee translation) and back-translation, were used. During each stage of the translation phase of this study, potential issues identified by the translators and researchers were considered and addressed. Through the use of back-translation, there was no indication that the translation provided by the professional translation service was inappropriate or incorrect. Further, all of the translations were completed in a standard, universally understood Spanish. While the translation may have contributed to the difficulties parents displayed in completing the nomination rubrics, it is highly unlikely due to the types of translation methods chosen, the rigor of the translation process, and the ease at which parents were able to complete the PBSS-PSF P2 rating scale. Should the translation have been an issue, it would be expected that parents would have had difficulties on all aspects of the PBSS-Parent, Spanish. While no other method is deemed to have been more appropriate for the PBSS-PSF , it may have been useful to add a third type of translation technique, the random probe technique. This method involves having a target group of parents complete a draft of the translated PBSS and subsequently share their experiences in completing the measure, along with a rationale for why he or she chose to respond in the manner he or she did (Guthrey & Lowe, 1992). This method may have provided a more detailed account of the difficulties parents experienced, as well as provided insight into the potential cause of the issue.

Second, it was important to assess for potential issues in cross-cultural research and translation, which may have affected the observed outcomes. Behling and Law's (2000) three main translation issues were heavily considered in this translation process. To increase semantic equivalence, a standard Spanish was used to increase the usability of the measure to various Spanish-speaking cultures. While the goal was to increase parent understanding of the various components, it cannot be definitely stated that all parents understood all translated material. Next, conceptual equivalence, or whether the target constructs were similar in both English and Spanish-speaking cultures, was considered. As the measured constructs dealt with observable behaviors present and identified cross-culturally, it was determined that conceptual inequality was low. Last, normative equivalence, or how similarly the social norms and conventions of both cultures influence the individuals within that society, was assessed. To reduce the potential for normative inequality, the researchers made themselves visible to the participants and increased positive relationships with the preschool partners. Further, participants were ensured that confidentiality would be maintained to increase parents' responsiveness in completing the scale. While the discussed issues may have affected the ability of parents to complete the nomination rubrics, there is no evidence to support this theory as the cause of the difficulties.

Third, the characteristics of the sample and of the PBSS-PSF were assessed and are believed to be the underlying causes of the observed difficulties. It may be that the rubrics are written and structured in a way that is too complex and ambiguous for individuals who have had no prior experience with the scale. It may also be that the novelty of completing a two phase measure caused some confusion for parents. Parents were able to appropriately complete the = PBSS-PSF P2 rating scale and the BESS, Spanish Parent and reported no difficulties with these scales. The rating scale format may be more familiar to parents, as this structure is used on a

variety of forms in everyday life (e.g., medical forms). As a result, level of experience in

completing a tool such as the PBSS-PSF may have had an effect. Another explanation may be

the specific characteristics of the study sample. Forty-nine percent of the parents in the sample

reported having a high school diploma and 33% reported no educational degree. Preschool

directors of participating schools also reported that many of the parents are unable to adequately

read and required assistance in completing the measures. The complex and novel nature of the

PBSS-PSF , along with parent level of ability may have resulted in the observed difficulties,

particularly with the PBSS-PSF P1 nomination rubrics.

PBSS-PSF P2 Rating Scale

The second phase of the PBSS-PSF is a 59-item rating scale. Analyses revealed that

PBSS-PSF PBSS-PSF P2 is a reliable and valid component of the PBSS-PSF and is 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 and was designed to provide more

information.

Reliability. The PBSS-PSF P2 rating scale 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 and that the yielded alpha levels are appropriate for a

screening tool. The ISS yielded an alpha level in the moderate range. Murphy & Davidshofer

(2005) suggest that while an alpha level of .80 is ideal, an alpha level above .70 is accepted for

screening tools. 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). It is noted in the literature that informants can more

reliably report observable behaviors, rather than unobservable behaviors (Achenbach,

McConaughy, Howell, 1987; Duhig, Renk, Epstein, & Pahres, 2000). As internalizing symptoms tend to be less observable than externalizing behaviors, the lower alpha level on the ISS may be due to parents' difficulties in reporting their children's internalizing problems. The PBSS-PSF P2 nomination rubric indices also demonstrated lower, yet comparable reliability coefficients to those on the English version of the measure. Several other measures, such as the BESS, the Behavior Assessment System for Children, Second Edition (BASC-2; Reynolds & Kamphaus, 2004), and the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001) have all demonstrated lower coefficient reliabilities on the Spanish versions of each measure.

Internal Structure Validity. Aside from demonstrating that the PBSS-PSF P2 rating scale possesses high internal consistency, it was also found to have acceptable internal structure. The correlations between the PBSS-PSF P2 scale scores all fell within or close to those on the English version of the measure and to the predicted ranges (Feeney-Kettler, Kratochwill, Kettler, 2011). Further, the ISS and ESS yielded a significant correlation of .40, suggesting that they are measuring somewhat related, but separate constructs (Ketter & Feeney-Kettler, 2011).

Concurrent Validity. PBSS-PSF P2 of the PBSS-PSF was assessed to see whether it performs comparably to a widely used and published screening measure, the BESS, Spanish Parent. There was a large and significant positive relationship between PBSS-PSF P2 Total Score and the BESS, Spanish Parent Total Score. The yielded 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). For this reason, the yielded correlation is appropriate and expected, as the PBSS-PSF does not incorporate these two additional areas of functioning. The

ESS and the PBS yielded large and significant correlations with the BESS, Spanish Parent Total

Score, while the ISS produced a small and non-significant positive correlation. In interpreting

these findings, it should be considered that many of the adaptive functioning and school

problems items overlap with prosocial behavior and externalizing problems, respectively. For

that reason, it may be that prosocial behavior and externalizing symptoms are highly represented

on the BESS, Spanish Parent, yielding a higher correlation with these scales on the PBSS-PSF.

When predicting the BESS, Spanish Parent Total Score, the PBSS-PSF P2 rating scale

was found to be a highly specific, but not very sensitive tool. This is not an unexpected finding,

as only 2 individuals were identified as being at-risk on both the BESS, Spanish Parent and the

PBSS-PSF and the frequency count for true negatives was 37. The PBSS-PSF also displayed a

low number of false negatives. This is a positive finding, as failing to identify a student who is

truly at-risk is considered to be very costly.

Summary. The PBSS-PSF P2 rating scale of the PBSS-PSF demonstrated several

strengths. Reliability coefficients for the composite scales were in the moderate to excellent

ranges. Internal structure was also a strength of the rating scale, as all of the intercorrelations

among the scales were in the expected ranges. When compared to a published screening

measure, the PBSS-PSF P2 Total Score and the total score on the BESS, Spanish Parent

correlated largely and positively, indicating that PBSS-PSF P2 is comparable to the BESS,

Spanish Parent rating scale. While findings indicated that PBSS-PSF P2 is not a very sensitive

measure, the low base rate of truly at-risk children in the population reduced the likelihood of

false negative cases.

Limitations

This study has several limitations that will need to be researched. The generalizability of the results may be limited. In order to create a scale that may be used by parents from a wide range of Spanish-speaking cultures, the scale was written in standard Spanish. The current sample, however, was only representative of a small range of Spanish-speaking cultures. Further, there were unequal distributions of ethnicities in the sample, as most of the parent participants were of Mexican-American origin. As a result, it is difficult to make generalizations regarding the usefulness and accessibility of this measure for all Spanish-speaking parents.

Parent educational level and abilities may have impacted the outcomes of the study. Parent participants reported some struggle in understanding how to complete the rubrics. Difficulties in completing the PBSS-PSF P1 nomination rubrics may have compromised the results, making it difficult to make conclusions about the efficacy of PBSS-PSF P1 of the PBSS.

Also, the small sample size of participants in the study and the low base rate yielded a small amount of children who were identified as at-risk on the PBSS-PSF scale. Further consideration will be needed in determining the most appropriate format and structure for PBSS-PSF P1 to reduce the observed difficulties.

Implications for Future Research

Future studies for the PBSS-PSF may focus on incorporating a larger, more diverse sample of participants. Areas on which to further diversify the sample may be on ethnic background and educational level. It will be important to incorporate participants of a variety of different cultural backgrounds, to assess the usability of the measure among Spanish-speaking parents of various cultural origins. Further, parent educational level may have played a crucial role in the observed outcomes of this study. It will be important to re-assess the psychometric

properties of the PBSS-PSF on a sample of parents with a wider range of educational levels. As both characteristics may have had an effect on the observed outcomes and generalizability of the PBSS-PSF , it is important the re-assess the functioning of the screening measure on these dimensions.

It will also be of great importance to consider whether the current format and structure of the PBSS-PSF P1 is the most appropriate format. Significant difficulties in correctly using the nomination rubrics may suggest that changes to the format are needed. To make PBSS-PSF P1 as "user-friendly" as possibly, it may be beneficial to use focus groups to assess potential areas of improvement. Further, different versions of the nomination rubrics may be field tested on a small sample of parents to assess the preferred and most effective version.

Conclusion

The PBSS-PSF is a two-phase screening measure for identifying preschool children who may be at-risk for developing emotional and behavioral difficulties. The multiple phase 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 behavioral and emotional conditions. Following a rigorous translation process of the English version of the measure, the PBSS-PSF was field tested on a small group of bilingual or monolingual Spanish-speaking parents of preschool children.

The PBSS-PSF PBSS-PSF P1 nomination rubrics worked well together as a unified phase. Consideration must be taken in the structure and format of PBSS-PSF P1 of the measure, as parents displayed significant difficulties in appropriately completing the nomination rubrics. These difficulties subsequently affected the relationship between the phases. PBSS-PSF P1 demonstrated weak relationships with the PBSS-PSF P2 rating scale. However, PBSS-PSF P1

was a good predictor of PBSS-PSF P2 and of the BESS, Spanish Parent. The PBSS-PSF P2 rating scale of the PBSS-PSF yielded reliable scores. Reliability coefficients for PBSS-PSF P2 indicated alpha levels in the moderate, high, and excellent ranges on all of the composite scales. Correlations between the PBSS-PSF P2 scales were within the expected ranges. The PBSS-PSF P2 rating scale also yielded large correlations with the published BESS, Spanish Parent form.

The PBSS-PSF may be a useful tool for practitioners and teachers to identify preschool children who are in need of additional services and early intervention programs. The creation of this tool in Spanish also increases its accessibility to a larger population of children and parents who are in need of such tools. Subsequent studies will be needed to address the presented concerns and determine the generalizability and validity of these results with different populations.

References

Achenbach, T. M., & Rescorla, L. (2001). *ASEBA School-Age Forms & Profiles*. ASEBA.

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/ adolescent behavioral and emotional problems: Implications of cross informant correlations for situational specificity. Psychological Bulletin, 101, 213–232.

American Psychiatric Association. (2010). *Mental health disparities: Hispanics/ Latinos* [Data file]. Retrieved from http://www.nami.org

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.

Azocar, F., Arean, P., Miranda, J., & Muñoz, R. F. (2001). Differential item functioning in a Spanish translation of the Beck Depression Inventory. *Journal of Clinical Psychology, 57*(3), 355-365.

Behling, O., & Law, K. S. (2000). *Translating questionnaires and other research instruments: Problems and solutions*. Thousand Oaks, CA: Sage Publications, Inc.

Berry, J. W. (1969) On cross-cultural comparability. *International Journal of Psychology*, 4, 110-128.

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.

Brislin, R., W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*(3), 185-216.

Brislin, R. W., Lonner, W. J., & Thorndike, R. M. (1973). *Cross-cultural research methods* (pp. 32-58). New York: J. Wiley.

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.

Clarke III, I. (2000). Extreme response style in cross-cultural research: An empirical

   investigation. *Journal of Social Behavior & Personality*, *15*(1), 137-152.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Cronbach, L. J. (1950). Further evidence on response sets and test design. *Educational and

   Psychological Measurement, 10*, 3-31.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological

   Bulletin*, 52, 281-302.

Douglas, S. P., & Craig, S. (1983). *International Marketing Research.* Englewood Cliffs, NJ:

   Prentice-Hall.

Dowdy, E., Chin, J. K., Twyford, J. M., & Dever, B. V. (2011). A factor analytic investigation of

   the BASC-2 Behavioral and Emotional Screening System Parent Form: Psychometric

   properties, practical implications, and future directions. *Journal of School Psychology,

   49*(3), 265-280.

Duhig, A. M., Renk, K., Epstein, M. K., & Phares, V. (2000). Inter parental agreement on

   internalizing, externalizing, and total behavior problems: A meta-analysis. Clinical

   Psychology: Science and Practice, 7, 435–453.

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.

Feeney-Kettler, K. A., Kratochwill, T. R., & Kettler, R. J. (2009). *Preschool Behavior Screening System*. Unpublished instrument, Vanderbilt University.

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.

Franco, J. N., Malloy, T., & Gonzalez, R. (1984). Ethnic and acculturation differences in self-disclosure. *The Journal of Social Psychology*, *122*(1), 21-32.

Garza, R. T. (1977) Personal control and fatalism in Chicanos and Angles: Conceptual and methodological issues. In J. L. Martinez (Eds.), *Chicanos psychology* (pp. 97-108). New York: Academic Press.

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.

Goodman, R. (2001). Psychometric Properties of the Strengths and Difficulties Questionnaire. Journal of the American Academy of Child and Adolescent Psychiatry, 40, 1337-1345.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry*, *38*(5), 581-586.

Guthrey, D., & Lowe, B. A. (1992). Translation problems in international marketing research. *Journal of Language for International Business, 4*, 1-14.

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.

Hirsch, H. (1973). Political scientists and other camaradas: Academic myth making and racial

    stereotypes. In R. O. de la Garza, Z. A. Kruswski, & T. A. Arcinego (Eds.), Chicanos and

    Native Americans. Englewood Cliffs, NJ: Prentice-Hall.

Hopkins, W. G. (2001). A Scale of magnitudes for effect statistics. *A New View of Statistics*.

    Retrieved from http://www.sportsci.org/resources/stats/effectmag.html.

Jacobsen, E. (1954). Methods used for producing comparable data in the OSCR seven-nation

    attitude survey. *Journal of Social Issues, 10*(4), 40-51.

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

    Pearson.

Kettler, R. J. & Feeny-Kettler, K. A. (2011). Screening systems and decision making at the

    preschool level: Application of a comprehensive validity framework. *Psychology in the*

    Schools, 48(5), 430-441.

Marin, G., Gamba, R. J., & Marin, B. V. (1992). Extreme Response Style and Acquiescence

    among Hispanics: The Role of Acculturation and Education. *Journal of Cross-Cultural*

    Psychology, *23*(4), 498-509

Marin, G., & Marin, B. V. (1991). *Research with Hispanic populations*. Thousand Oaks, CA:

    Sage Publications, Inc.

Marin, G., Perez-Sable, E. J., & Marin, B. V. (1989). Cigarette smoking among San Francisco

    Hispanics: The role of acculturation and gender. *American Journal of Public Health, 79*,

    196-198.

Moore, J. W. (1973). Social constraints on sociological knowledge: Academics and research

    concerning minorities. *Social Problems, 21*, 65-77.

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

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

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

Pareek, U., & Rao, T. V. (1980). Cross-cultural surveys and interviewing. In H. C. Triandis & J. E. Berry (Eds.), *Handbook of cross cultural psychology, Vol 2: Methodology* (pp. 127-180). Boston: Allyn & Bacon.

Patton, M. Q. (2002). *Qualitative research and evaluation methods, 3rd edition*. Thousand Oaks, CA: Sage. Excerpt from Chapter 9, Enhancing quality and credibility, pages 552-564.

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

Rodriguez-Hernandez, P. J., Betancort, M., Ramirez-Santana, G. M., Garica, R., Sanz-Alvarez, E. J., & De las Cuevas-Castresana, C. (2012). Psychometric properties of the parent and teacher versions of the Strength and Difficulties Questionnaire (SDQ) in a Spanish sample. *International Journal of Child and Health Psychology, 12*(2), 265-279.

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.

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

U.S. Census Bureau (2011). *The Hispanic population: 2010* [Data file]. Retrieved from

http://www.census.gov/prod/cen2010/briefs/c2010br-04.pdf

Wendler, D., Kington, R., Madans, J., Van Wye, G., Christ-Schmidt, H., Pratt, L. A., Brawley,

O. W., Gross, C. P. & Emanuel, E. (2005). Are racial and ethnic minorities less willing to

participate in health research?. *PLoS medicine*, *3*(2), e19.

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.

Appendix A

Demographic Questionnaire
Preschool Behavior Screening Project
Spanish Form

Please circle your answers to the following questions below.

I am a …..	Mother	**Completed is…..**	Bachelor's Degree
	Father		Master's Degree
	Other caregiver		Doctoral Degree
		My age is….. _____	
I am a …..	Female		
	Male	**My ethnicity American is…..**	African
			Asian/ Pacific
I am …..	Single	Islander	
	Married		European American
	Divorced		Latino/Latina
			Native
The number of children I have is…..		American	
	_____		Other

Highest Diploma Level of Education I	High School	**Thank you for completing this questionnaire!**	
	Associate's Degree		

Appendix B

Evaluation Survey- Parent Version
Preschool Behavior Screening Project- Spanish Form

1. Do you have any concerns about how your child behaves?

   Yes / No / A Little If Yes or A Little, please describe your concerns:

2. How long did it take to complete PBSS-PSF P1 (nomination rubrics) of the
   Preschool Behavior Screening System? _____

3. How long did it take to complete PBSS-PSF P2 (59 individual questions about
   children's behavior) of the Preschool Behavior Screening System?
   _____

4. Do you think that the Preschool Behavior Screening System could provide useful
   information about preschool age children?

   Yes / No Please explain:

5. Was the Preschool Behavior Screening System clearly organized and written?

   Yes / No Please explain:

6. Please share any other information that you think would be helpful to the
   Preschool Behavior Screening System, this project, or screening for social and
   emotional problems in general.