

Running Head: PSYCHOMETRICS OF THE SDQ IN SPANISH

PSYCHOMETRICS OF THE SDQ IN SPANISH

A DISSERTATION

SUBMITTED TO THE FACULTY

OF

THE GRADUATE SCHOOL OF APPLIED AND PROFESSIONAL PSYCHOLOGY

OF

RUTGERS,

THE STATE UNIVERSITY OF NEW JERSEY

BY

HEATHER MARIE BLUMERT

IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE

OF

DOCTOR OF PSYCHOLOGY

NEW BRUNSWICK, NEW JERSEY

OCTOBER 2013

APPROVED:

\_\_\_\_\_  
Ryan Kettler, PhD.

\_\_\_\_\_  
Nancy Fagley, PhD.

DEAN:

\_\_\_\_\_  
Stanley Messer, PhD.

Copyright 2012 by Heather Blumert

## ABSTRACT

With the recent influx of Latinos into the United States, it is essential to understand how their backgrounds and cultures will affect the way they view their children's emotional, social, and educational development. Researchers continue to evaluate the psychometrics of various screening instruments in order to ensure a reliable and valid Spanish-language instrument is being used to measure children's behaviors. The purpose of this study was to compare the psychometric properties of the Spanish version with the English version of the Strengths and Difficulties Questionnaire (SDQ), a 25-item behavioral screener. Participants included in this study were 488 English-speaking parents and 435 Spanish-speaking parents of preschool age children (ages 3-5) that took part in the California University (Irvine) Initiative for the Development of Attention and Readiness (CUIDAR) program from 2004-2008. This study used data from the CUIDAR program to explore mean rating differences between the English and Spanish versions of the SDQ, along with coefficient alpha as an indicator of reliability at the scale and composite level, and factor analytic evidence of score validity. Confirmatory factor analysis was used to compare the relative fit of multiple models, including the Five First Order Factor (5F) Model that is prevalent in research on the SDQ. Results indicated mean ratings of the individual scales and the Total Difficulties scales were very similar across both language forms. Reliability coefficients indicated alphas were higher for the English forms compared to the Spanish forms at the scale and composite level, although neither form had adequate reliability at the scale level. Finally, the 5F Model was the best-fitting and most valid representation of all 25 items of the SDQ, despite the language of the form.

The English models also fit the proposed factor structure better than the Spanish models did.

## ACKNOWLEDGEMENTS

This study would not have been possible without the support of many educators, researchers, and family members. I want to extend my gratitude to the children, parents, and educators who participated in the study, including everybody involved in the California University (Irvine) Initiative for the Development of Attention and Readiness (CUIDAR) program from 2004-2008. In particular, I want to thank Dr. Kimberley Lakes for making this data so readily available and helping me with the many questions I had along the way.

I would also like to thank a wonderful dissertation committee, including Drs. Ryan Kettler and Nancy Fagley. I especially want to thank Ryan for making this study possible, answering countless emails, and scheduling last minute meetings. I sincerely appreciate your help throughout this process.

Of course, none of this would be possible without the support of my wonderful family. I am grateful to Judy, Steve, and Jessica for welcoming me with open arms into their family. In addition, I will forever be grateful to my parents, Mark and Shari, and my brother Bradly, who have loved, supported, and encouraged me throughout schooling and all life's endeavors.

Finally, I want to thank my husband Matthew, who is always there to motivate me and express his pride in being my spouse. I love you.

## TABLE OF CONTENTS

	PAGE
ABSTRACT .....	ii
ACKNOWLEDGEMENTS .....	iv
LIST OF TABLES .....	vii
CHAPTER	
I. INTRODUCTION.....	1
Criteria for Evaluating Rating Forms.....	1
Psychometrics of Spanish-translated Rating Forms.....	3
The Behavioral and Emotional Rating Scale-2.....	3
The Social Anxiety Scale for Adolescents (SAS-A).....	4
The Child Behavior Checklist (CBCL).....	5
The Strengths and Difficulties Questionnaire.....	6
Internal Structure of the SDQ.....	11
Summary.....	14
Current Research Questions.....	15
II. METHODS.....	16
Participants.....	16
Measures.....	18
Procedures.....	19
Treatment of Data/Plan of Analysis.....	19
III. RESULTS.....	21

## TABLE OF CONTENTS

	PAGE
Differences in Mean SDQ Ratings.....	22
Reliability.....	25
Factor Analysis of English Models.....	28
Factor Analysis of Spanish Models.....	32
Factor Analytic Comparison of English and Spanish Models.....	37
IV. DISCUSSION.....	38
Mean Ratings of English and Spanish SDQ Forms.....	39
Reliability of Spanish and English Forms of the SDQ.....	42
Factor Analysis of the SDQ in English and Spanish.....	44
Suggested Spanish Language Instruments.....	46
Limitations.....	48
Implications for Future Research and Practice.....	48
Conclusion.....	50
REFERENCES.....	52

## LIST OF TABLES

Table # 1 Reliability Coefficients (Coefficient alpha) Across Parent-Informant SDQ Studies.....	pg 15
Table # 2 Demographic Information.....	pg 17
Table # 3 Mean Parent Ratings for All Participants.....	pg 22
Table # 4 Mean Parent Ratings Between Spanish and English forms of the SDQ.....	pg 23
Table # 5 Mean Parent Ratings of SDQ Scales Across Spanish and English forms of the SDQ based on Mother's Country of Origin.....	pg 25
Table # 6 Reliability Coefficients (Coefficient alpha) Across Spanish and English forms of the SDQ.....	pg 26
Table # 7 Reliability Coefficients (Coefficient alpha) of Total Difficulties Scales Across Spanish and English forms of the SDQ based on Mother's Country of Origin.....	pg 27
Table # 8 Factor Loadings for English Models.....	pg 29
Table # 9 Goodness of Fit Indices for English and Spanish Factor Models.....	pg 32
Table # 10 Factor Loadings for Spanish Models.....	pg 34



## **Introduction**

Latinos live in countries all over the world including Mexico, Cuba, Puerto Rico, El Salvador, Guatemala, Colombia, and Costa Rica (Sharkey, You, Morrison, & Griffiths, 2009).. They represent the largest ethnic minority group in the United States (Pedrotti & Edwards, 2010). They are overrepresented in terms of families afflicted by poverty, behavioral disorders, and mental health disorders (Smokowski, Reynolds, & Bezruczko, 1999). In addition, Latino children are at greater risk of failing in school as well as dropping out of school (Tinkler, 2002).

Due to the Latino population increasing in numbers and the number of Latino school students rising, it is essential to have a reliable and valid instrument for Spanish-speaking individuals to measure their behavioral strengths and weaknesses. It is particularly important to understand the Latino parent perspective when they are asked to rate their child's behaviors. Therefore, researchers are continuing to evaluate the psychometrics of various screening instruments in order to ensure a reliable and valid instrument is being used to measure children's behaviors and create accurate and effective intervention and treatment plans to utilize in the classroom (Jimerson, Sharkey, Nyborg, & Furlong, 2004).

### **Criteria for Evaluating Rating Forms**

Exploring the psychometric properties of original rating scales that have been translated into Spanish is essential for this population. Key aspects in exploring the psychometric properties of a test or scale entail evaluating how reliable and valid its scores are.

Reliability refers to the how consistent a measure is when the assessment is repeated on a population (American Education Research Association, 1999). Coefficient alpha, or Cronbach's alpha, is an indicator of reliability and the internal consistency of items within a test (Cortina, 1993). Coefficient alpha is equal to the mean of all split-half reliabilities when the standard deviations are equal. If a test has a large coefficient alpha, then, "it can be determined that a large portion of the variance in the test is attributable to general and group factors" (Cortina, 1993).

Validity refers to the degree to which theory and evidence provide backing for the interpretations of test scores entailed by the designed use of tests (American Educational Research Association [AERA], American Psychological Association [APA], and National Council on Measurement in Education [NCME], 1999). Factor analysis is one type of evidence for validity that is included in the *Standards for Educational and Psychological Testing* (the *Standards*: AERA et al., 1999). Over sixty years ago, in the article *New Standards for Test Evaluation*, Guilford (1946) discussed the term "factorial validity," referring to the loading of meaningful, common, reference factors that indicate whether in fact a test is measuring what it is theoretically intended to measure. There are two types of factor analysis: exploratory and confirmatory factor analysis. While exploratory factor analysis is not based on any proposed theory and addresses the underlying structure of a large set of variables, confirmatory factor analysis confirms or disconfirms pre-established factors and subsequent loadings onto each based on theory (Thompson & Daniel, 1996). In confirmatory factor analysis, the fit of each proposed model is tested to determine the best structure of a test (Sharkey et al., 2009). Subsequent links between validity and factor analysis lie in the theory of falsification, that a theory should not be considered credible

until efforts to disconfirm that theory have taken place (Thompson & Daniel, 1996). A strong program of construct validation requires that rival hypotheses be tested which may suggest alternative explanations for the meanings of test scores. Similarly, in confirmatory factor analysis, rival models can and should be tested because multiple models may fit the same data. Indeed, testing multiple plausible models with confirmatory factor analysis provides stronger evidence of validity.

### **Psychometrics of Spanish-translated Forms**

Research regarding the effect of translating instruments into Spanish, or other languages, has found different results. It appears as though the effect of translation varies by the measure that is being analyzed.

**The Behavioral and Emotional Rating Scale-2.** The Behavioral and Emotional Rating Scale-2 Parent Report (BERS-2) is a school-based scale, which measures the strengths of a student (Sharkey et al., 2009). It is used primarily with children that have significant mental health concerns, including Attention Deficit Hyperactivity Disorder (ADHD), Oppositional Defiant Disorder (ODD), and mood disorders.

Buckley, Ryser, Reid, and Epstein (2006) performed an exploratory factor analysis of the original English version of the BERS-2. They assessed various factor structures, including a 3-factor model and found the intended 5-factor structure, which included 25 items loading onto: Interpersonal Strengths, Family Involvement, Intrapersonal Strengths, School Functioning, and Affective Strengths to be the best-fitting model (Buckley, 2006). Sharkey et al. (2009) then explored the factor structure of the BERS-2 with Spanish-Speaking parents of at-risk youths. There were two samples included in this study. The first consisted of parents of students in fourth through seventh grade from low

socioeconomic status neighborhoods in two school districts in Central California. The second sample consisted of parents of youths enrolled in a community program providing services to criminally involved families. Exploratory factor analysis indicated that a three-factor model was a better fit than the original five-factor model of the English version for the latter sample. Intercorrelations among factors for the 3- factor model ranged from .68 to .81, while they ranged from .65 to .94 for the five-factor model, indicating some redundancy.

Further analysis revealed overlap among items on the Family Involvement and Interpersonal Strength factors. In other words, items tended to co-load on both of the factors. Authors proposed that the strong value of familismo, the importance of immediate and extended familial ties in the Mexican-American population, caused this overlap, as a parent's basis for understanding their child's interpersonal functioning is within the family system (Sharkey, et al., 2009). Thus, Mexican American parents may be less likely to identify personal strengths distinct from family strengths. Furthermore, authors noted that the BERS-2 Spanish version might measure culturally sensitive constructs. However, because the BERS-2 factor structure has not been assessed thoroughly in prior studies, the authors note that they are unsure if a three-factor model is unique to the Spanish-speaking population. They note that further reliability and validity studies would benefit users of the BERS-2.

**The Social Anxiety Scale for Adolescents (SAS-A).** Although there is evidence to support alternate structures for some rating scales, there is research to support that some structures remain psychometrically sound even when used in Spanish with a Spanish-speaking population. The Social Anxiety Scale for Adolescents (SAS-A) is an instrument

designed to measure social anxiety responses (Olivares, Ruiz, Hidalgo, Garcia-Lopez, Rosa, & Piqueras, 2004). Confirmatory factor analysis of the SAS-A by LaGreca and Lopez (as cited in Olivares et al., 2004) supported the original three-factor structure consisting of Fear of Negative Evaluation (FNE), Social Avoidance and Distress specific to new situations or unfamiliar peers (SAD-New) and Social Avoidance and Distress that is experienced more generally in the company of peers (SAD-General) in an English-speaking sample.

Olivares et al., (2009) assessed alternative models to the original three-factor model of the SAS-A: a null or independent model, a one-factor model in which all 18 items loaded onto a general social anxiety factor, a two-factor model of FNE and SAD combined, and the original model with a Spanish-speaking adolescent population in Spain. Results indicated that the three-factor model was confirmed and was a better fit compared to the alternative structures proposed. The three-factor model had the highest Goodness of Fit index (.89) and Comparative Fit Index (.89) compared to the other models but didn't quite reach recommended levels. In addition, the Standardized Root Mean Square Residual of .06 indicated a good fit. Compared to the English version of the SAS-A, the factor model was only slightly less of a better fit, with Goodness of Fit and Comparative Fit Indices (.90) reaching the goodness of fit mark. Coefficient alphas were similar to those obtained in prior analyses by LaGreca and Lopez (1998) of the English form of the SAS-A ranging from .87 to .94 for the scales and total scales. Authors suggested that this measurement study provides support for the SAS-A to be used in a Spanish-speaking population.

**The Child Behavior Checklist (CBCL).** Another instrument that's structure proved to be valid despite translation is the Child Behavior Checklist. This measure

can be used to assess emotional problems as well as attention and social related concerns (Goodman & Scott, 1999). A study by Gross, Fogg, Young, Ridge, Cowell, Richardson, and Silvan (2006) was completed in which the Child Behavior Checklist (CBCL) was examined among parents who represented different races, ethnicities, incomes, and language backgrounds. Overall model fit was assessed through Confirmatory Factor Analysis based on the relative chi-square (ratio of the chi-square to the degrees of freedom) and the root-mean-square error of approximation (RMSEA). The authors noted that in confirmatory factor analysis, utilizing this method, a relative chi-square less than 2.0 and a RMSEA less than .05 were indications of an adequate fit model. Furthermore, the authors found that despite language, racial, and socio-economic differences, the model was considered a good fit when it was translated to Spanish. The RMSEA statistics were both at .03 and the relative chi-square was 1.66 for the English form and 1.67 for the Spanish form indicating good fit, regardless of the language of the form.

### **The Strengths and Difficulties Questionnaire**

The Strengths and Difficulties Questionnaire was developed in the United Kingdom by Robert Goodman as a rating instrument to assess youth ages 3 through 16 regarding twenty-five attributes, some positive and some negative, using a three-point Likert scale (Goodman, 2001). The SDQ is used in over 40 languages in an array of settings (Goodman, 1997). There are forms for parents, teachers, and self-raters above age eleven to complete. There are five scales consisting of five items each, “generating scores for Emotional Symptoms ES, Conduct Problems CP, hyperactivity-inattention HI, peer problems PP, and prosocial behavior PB; all but the last are summed to generate a Total

Difficulties score TD” (Goodman, 2001). Higher scores on scales indicate difficulties, with the exception of the PB scale, on which higher scores reflect strengths (Muris, Meesters, & van den Berg, 2002).

There are many uses of the SDQ including clinical assessment, evaluating outcomes, epidemiology, research, and screening for child psychiatric disorders in a community, school, or mental health clinic (Youth in Mind Organization, 2009). Much of the research produced regarding the SDQ has focused on its comparability to similar instruments for assessing youth behavior. The SDQ was developed based on the respected Rutter Questionnaire and Child Behavior Checklist (Goodman, 1997). Compared to the other two instruments, the SDQ is a much shorter rating scale and more strength-based. In addition, unlike the SDQ, the Rutter Questionnaire does not have a self-report version. In a study by Goodman (1997), the validity of the SDQ was assessed by comparing it to the Rutter Questionnaire. The SDQ was administered along with the Rutter Questionnaire to parents and teachers of 403 children ages four through 16, who were receiving services at dental and psychiatric clinics in London, England. Goodman found evidence for concurrent validity of the SDQ, as it was highly correlated with the Rutter Questionnaire, with correlations ranging from .78 to .92 on similar scales. As Hopkins (2002) suggested, correlations are considered very large when Pearson’s  $r$  exceeds .70, and nearly perfect correlations occur when  $r$  equals .90 and above. The correlations between parent and teacher ratings on each instrument showed similar correlations.

Goodman (1997) also looked at whether or not the Rutter and SDQ differed in their ability to differentiate between the high-risk psychiatric sample and the low risk dental clinic sample. Utilizing receiver operating characteristic (ROC) curves, Goodman found

there were equal area intervals under each of the curves, indicating each measure possessed good predictive validity and was equally able to discriminate among high and low risk samples. Goodman concluded that the SDQ appears to be a brief yet valid child behavioral screener, which measures strengths as well as difficulties, and possesses forms for parents as well as teachers.

Goodman and Scott (1999) conducted additional research comparing the SDQ to the Child Behavior Checklist (Achenbach, 1991). The CBCL is much longer than the one-page SDQ. It contains 113 items and possesses eight syndrome scales including Withdrawn, Somatic Complaints, Anxious/Depressed, Social Problems, Thought Problems, Attention Problems, Delinquent Behavior, and Aggressive Behavior along with two broader syndrome scales, Internalizing and Externalizing problems. The SDQ and CBCL differ not only in their length but also on their item selection rationale. The SDQ items were included based on nosological concepts as well as factor analysis (Goodman & Scott, 1999). Relevant concepts were taken from the DSM-IV and ICD-10 to develop questions linking answers to diagnostic criteria. According to Goodman and Scott (1999), the CBCL includes several items that contain no conceptual link to diagnostic criteria for disorders. Like the Rutter Questionnaire, the CBCL is entirely composed of negatively worded items. Goodman and Scott (1999) demonstrate an example of the comparison of two questions aimed at assessing attention in youth. Whereas the CBCL includes an item stating, "The child can't concentrate", the SDQ frames it as "The Child has a good attention span."

Goodman and Scott (1999) conducted another study involving raters of children from psychiatric clinics and dental clinics in London. For this study, only mothers were used as raters of their children ages four to seven years old. Mothers of 132 children



completed SDQ and CBCL questionnaires. The mothers of the low-risk dental clinic sample chose which of the two questionnaires they preferred completing. The mothers in the high-risk sample participated in semi-structured interviews called the Parental Account of Child Symptoms. Trained interviewers questioned the respondents on factors linked to inattention, hyperactivity, externalizing, and internalizing behaviors. They were unable to see the questionnaires the parents had previously completed.

Results from the Goodman and Scott study (1999) indicated that the SDQ and CBCL were comparable in many ways. The scores derived from each were correlated highly with one another, and ROC curves demonstrated nearly equal ability to discriminate between low and high-risk populations. As a guide to interpretation, Goodman and Scott note that the area under the curve would be 1.0 in order to be considered a perfect discrimination and .5 for a measure that did not discriminate any better than chance accuracy. In this instance, both questionnaires discriminated between the populations well, with areas under the curve equaling .95 for Total and Externalizing scales.

Goodman and Scott also examined how well the questionnaire predicted interview ratings of internalizing symptoms, inattention-hyperactivity, and externalizing problems. The CBCL Internalizing score, CBCL Hyperactivity-Inattention score, CBCL Conduct score, SDQ CP score, SDQ EP score, and SDQ HI score were analyzed along with investigator-based ratings on these dimensions. Results indicated that for internalizing and externalizing problems, the interview ratings correlated similarly with the two questionnaires, with correlations ranging from .44 to .64. However, there was a significant difference for the HI rating, with correlations significantly higher with the SDQ factor ( $r = .43$ ) than the CBCL factor ( $r = .15$ ).

Goodman (1999) notes that the SDQ and CBCL possess different strengths. Goodman states that the SDQ may be more useful when screening for attention-related problems or as a community-wide screening measure, because of the brevity and strength-based questions on the instrument. However, Goodman also explains that the CBCL covers a broader range of problems, which may make it more useful for clinical assessments.

Goodman, Ford, Simmons, Gatward, and Meltzer (2000) used parent, teacher, and self-report versions of the SDQ to compare SDQ algorithms and independent psychiatric diagnosis from the 1999 British Child Mental Health Survey of over 7,000 5 to 15-year olds. The aim of this study was to assess if the SDQ could be used as a tool to improve the detection of child psychiatric disorders in a community. The SDQ prediction equations determined if it were unlikely, possible, or probable for an individual to fit into any of the following categories: conduct-oppositional, hyperactivity-inattention, and anxiety-depressive. Results indicated that the SDQ identified over 70 percent of individuals with conduct, hyperactivity, depressive, and some anxiety disorders. However, the SDQ only identified around 50 percent of those with specific phobias, separation anxiety, and eating disorders. The authors note that this is consistent with their hypothesis as the SDQ does not contain any questions regarding dieting or panic attacks, and only has one question related to separation anxiety. The SDQ algorithm appears to capitalize on the co-morbidity of disorders, as it is able to detect three-quarters of those with depressive or obsessive compulsive disorders by querying about anxious (Emotional Problems Scale) and conduct related problems. In addition, it is able to detect three quarters of children with pervasive developmental disorder (PDD) as the algorithm looks at the emotional and hyperactive

problems associated with PDD (Goodman et al., 2000). The authors indicate that the SDQ is a useful tool in community screening programs, as it may increase the detection of child psychiatric disorders. Since the SDQ could be used on such a large scale, it is essential that the psychometric properties of the various forms as well as its translated versions be researched intensively.

**Internal structure of the SDQ.** The psychometric properties of the SDQ have been examined in various studies of samples throughout the world. Internal consistency studies have also found the SDQ to be adequately reliable. Factor analysis of the three versions of the SDQ has mostly confirmed that the questionnaire contains five factors corresponding with hypothesized domains of psychopathology and personal strengths (Murriss et al., 2002). However, many studies have been limited by small sample size (Murriss et al., 2002).

In 2001, Robert Goodman, developer of the SDQ, conducted a study to assess its psychometric properties. SDQ's were collected from parents, teachers, and self-informants of a nationwide epidemiological sample of over 10,000 British students ages 5 to 15. Ninety-six percent of the informants were parents (Goodman, 2001). Internal consistency was assessed and Cronbach's alpha coefficients were generally satisfactory for scales representing the five factors with a mean of .73 across all forms. The internal consistency of the TD category was also deemed sufficient with a Cronbach's alpha of .82.

Factor analyses were conducted to examine the factor structure for each category of informant. Results for each form indicated that all 25 items loaded onto their intended factors, while a few items loaded onto additional factors as well. In the case of the parent-informant form, all of the 25 items loaded more heavily onto their respective factors than

any of the additional factors. The same was true for the other informant forms with the exception of one item on each form.

Goodman (2001) noted many items on the HI scale and PP scale on the teacher and self-informant form also substantially loaded (.34 to .52) onto the PB scale. These items were all positively worded indicating a general tendency for positive statements to load onto the PB scale. In addition, the predicted five-factor structure, which consists of the five scales of the SDQ, was confirmed, especially in the case of the parent-informant form. Although there was a slight divergence from this structure with positively worded items loading onto the PB scale, the five-factor structure was mostly confirmed for the self-informant and teacher-rater form as well.

Hawes and Dadds (2004) conducted a similar psychometric study to analyze the parent form of the SDQ given to a large Australian community sample of parents of young children ages 4 through 9. Hawes and Dadds (2004) noted that the appropriateness of the SDQ for use in different cultures is a relevant topic as the SDQ is available in over 40 languages. The five-factor structure was examined separately for males and females using SPSS principal component analyses with oblimin rotation. Results indicated that the five-factor structure was consistent with its design, with factor loadings generally stronger for boys than for girls. Consistent with Goodman's study (2001) and a study examining the SDQ in a Swedish population (Smedje, Broman, & Knorrning, 1999), cross loading occurred with a conduct scale item relating to obedience. In the case of the boys, this item loaded more strongly onto the PB scale, albeit negatively and the hyperactivity scale, while only loading more heavily onto the PB scale for girls. Hawes and Dadds note that perhaps the utility of this item as an indicator of conduct problems is unreasonable. They reason

that using a more negatively worded statement may produce a better indicator of conduct problems (i.e. “generally disobedient” rather than “generally obedient”).

Hawes and Dadds (2004) also looked at the correlations among the various scales of the SDQ and found that there were higher correlations among the scales that both measured more externalizing factors, such as hyperactivity-inattention and conduct problems, than those measuring internalizing and externalizing factors. Further, there was a correlation of .52 between the CP scale and HI scale, while there was a correlation of .27 and .28 respectively with ES and CP/HI scales. This finding may indicate that the scales are uncontaminated; they are measuring different constructs. Coefficient alpha statistics ranged from .59 to .80.

A similar study was conducted in a community sample of Dutch children and adolescents (Muris, Meesters, & van den Berg, 2002). Over 500 parents and self-informants were asked to complete the SDQ along with other behavioral rating measures. Analyses of the parent-informant SDQs showed that the first five factors: ES, CP, HI, PP, and PB all had Eigen-values greater than 1.0 (i.e. 4.8, 2.5, 2.0, 1.3, and 1.2). They also accounted for 47.6 percent of the total variance (Muris et al., 2002). In addition, all of the items loaded strongly onto their respective factors, with the exception of one item having a substantial secondary loading (Muris et al., 2002). In this case, the PB item, “considerate of other people’s feelings” loaded negatively (-.49) onto the CP factor. Consistent with the findings of Goodman (2001) and Hawes et al., (2004), the CP scale item on the self-report form on obedience also loaded negatively (-.55) onto the HI scale.

Analyses of the SDQ scales indicated higher mean problem scores for males in terms of hyperactivity-inattention and peer problems as well as lower levels of prosocial behavior when compared to their female counterparts (Muris et al., 2002). Internal consistency was generally satisfactory with a mean coefficient alpha of .70 for the parent version and .64 for the self-report form (Muris et al., 2002). However, Cronbach's alpha for the CP scale was notably lower ( $\alpha = .55$ ) compared to the rest of the scales and TD scale, ( $\alpha$  ranging from .66 to .80). Finally, cross scale correlations indicated that scales were measuring independent domains of difficulties with low to moderate correlations between .17 and .41. Consistent with what one would expect, the PB scale correlated inversely with the other scales, as higher scores indicate strengths as opposed to difficulties.

### **Summary**

Research appears to support the originally developed five-factor structure of the SDQ. Factor analytic studies have shown this to be true utilizing various translations of the SDQ in many parts of the world. There have been slight but consistent deviations regarding the composition of the factors, as some items tend to co-load onto a second factor. Cross loading occurred consistently with a conduct scale item relating to obedience across various studies. In addition, there was a trend for more positively worded statements to co-load onto the PB Scale along with their intended scale. Across the various studies, parent-informant SDQ's generally have a satisfactory and higher mean coefficient alpha around .70 as compared to self and teacher-informant forms. Average scale reliabilities, however, are lower, ranging from .61 to .81. Internal consistency reliabilities of English, Australian, and Dutch parent-informant SDQs are reported and summarized below in Table 1. Further, the parent-informant version of the SDQ has been found to be a

reliable and valid instrument in measuring a child's psychological adjustment, regardless of the sample or language translation that is being utilized. However, there exists a lack of research directly comparing the SDQ's factor structure, reliability, and mean scores across forms in two languages.

TABLE 1

*Reliability Coefficients (Coefficient alpha) Across Parent-Informant SDQ Studies*

	Goodman (2001)	Hawes et al., (2004)	Muris et al., (2002)	
Language of Form	English	Australian-English	Dutch	
Child's Age (years)	5-15	4-9	9-15	
SDQ Scale				Scale Average
Emotional Symptoms (ES)	.67	.66	.70	.68
Conduct Problems (CP)	.63	.66	.55	.61
Hyperactivity-Inattention (HI)	.77	.80	.78	.78
Peer Problems (PP)	.57	.59	.66	.61
Prosocial Behavior (PB)	.65	.70	.68	.68
Total Difficulties (TD)	.82	.82	.80	.81

*Note:* SDQ= Strengths and Difficulties Questionnaire.

**Current research questions.** The current study utilized existing data from the SDQ that was collected as part of the California University (Irvine) Initiative for the Development of Attention and Readiness (CUIDAR) Program. Data that have been collected from English-speaking parents and Spanish-speaking parents using either the English and Spanish form of the SDQ will allow for psychometric comparison of the two parent-informant forms. The objective of this study is to compare the psychometrics of the English and Spanish versions of the SDQ. This study will contribute to existing literature regarding the SDQ by allowing direct comparison of the instrument in two languages. In addition, it will look to confirm or disconfirm previous literature regarding the five-factor structure of the SDQ as

well as the established reliability of the parent-informant SDQ. Finally, it will provide information on whether the Spanish version of the SDQ is psychometrically sound and suitable to be used in a Spanish-speaking United States population. The study will address the following research questions:

1. Are there mean differences in SDQ scores based on the language of forms (English versus Spanish)?
2. Are there reliability differences in SDQ scores based on the language of forms (English versus Spanish)?
3. Is the internal structure validity evidence of SDQ scores different based on the language of forms (English versus Spanish)?

## **Method**

### **Participants**

Participants included in this study included 488 English-speaking parents and 435 Spanish-speaking parents of preschool age children (ages 3-5) that took part in the California University (Irvine) Initiative for the Development of Attention and Readiness (CUIDAR) program over a four-year period, from 2004-2008. The Mexican-American ethnicity represented the highest percentage of children, regardless of whether the forms were completed in English or Spanish. The mean age of the children who were rated in English was 3.79 years (SD= 1.64), while the mean age of the children who were rated Spanish was 4.11 years (SD= 1.70). Many of the children's parents originated from Mexico. Further demographic information representing the aforementioned sample is located in Table 2.



TABLE 2

*Demographic Information*

	English Form (n=488)	Spanish Form (n=435)
Percentage of each Gender (%)	44 Female 56 Male	52 Female 47 Male
Child's Ethnicity (%)	43 Mexican-American 19 European American 14 African American 10 Biracial 7 Other Hispanic	86 Mexican-American 12 Other Hispanic 2 Biracial
Parent's Education Level (%)	21 Did Not Complete HS 27 HS Diploma/GED 44 Some College 4 Bachelor's Degree 4 Advanced Degree	51 Did Not Complete HS 28 HS Diploma/GED 12 Some College 7 Bachelor's Degree 2 Advanced Degree

CUIDAR was designed to address potential barriers, such as lack of knowledge, lack of insurance, and cultural discrepancies that may affect the screening for and intervention with behavioral disorders affecting low-socioeconomic status and minority families (Lakes, Kettler, Schmdet, Haynes, Feeney-Kettler, Kamptner, Swanson, & Tamm, 2009). The goal of CUIDAR is to identify children with attention and behavioral difficulties prior to entering the school system so they will have a more successful educational experience (Lakes et al., 2009). The Children and Families Commission of Orange County and First 5 San Bernardino provide funding for this program. The program provides meals to attendees as well as parent training/education and childcare (Lakes, et al., 2009). The parent education model used in this program is a modified version of the original Creating Opportunities for Parent Empowerment (COPE) program (Cunningham,

Bremner, & Boyle, 1995) which focuses on parent-child interactions, building self-efficacy, and identifying common parenting errors.

### **Measures**

The Strengths and Difficulties Questionnaire was developed in the United Kingdom by Robert Goodman (1997) and is used to assess youth ages 3 through 16 based on 25 items related to positive and negative characteristics, using a 3-point Likert scale (0=*Not True*, 1=*Somewhat True*, 2=*Certainly True*; Goodman, 2001). There are forms for parents, teachers, and self-raters above age 11 to complete. In addition, there is a separate form used for preschool age children, 3 to 4, which has only 2 different questions than the form used for older children. These two questions are worded so that they are more developmentally appropriate for younger children, but are measuring the same behaviors as the form used for older children. There are 5 scales consisting of 5 items each, generating scores for Emotional Symptoms (ES), Conduct Problems (CP), Hyperactivity-Inattention (HI), Peer Problems (PP), and Prosocial Behavior (PB). A Total Difficulties score is also computed from the four problem scales (TD). The theoretical structure of the SDQ developed by Goodman (1997) consists of five individual factors representing the five scales of the SDQ.

The theoretical five-factor structure of the SDQ is the same in English and in Spanish. Further, the Spanish and English versions of the SDQ all contain 25 questions, with five questions assessing each factor. There is an SDQ computer algorithm, which contains prediction equations determining if it is unlikely, possible, or probable for an individual to fit into any of the following categories: conduct-oppositional, hyperactivity-inattention, and anxiety-depressive (Goodman, et al., 2000).

**Procedures**

Analyses were conducted using an extant database from the CUIDAR program. During the introduction of the CUIDAR San Bernardino program, parents were invited to participate in a research study, designed to evaluate the effectiveness of a 10-week intervention protocol. As part of their entrance into the research study, participants completed the Strengths and Difficulties Questionnaire (SDQ). In their study, only parent report forms of the SDQ were used to rate their preschooler's behavior. Parents of children ages three and four used the preschool form while parents of children ages five and above used the form for children ages five through 16. In addition, participants initially completed a demographic questionnaire, which included questions regarding race, ethnicity, country of origin, and parent education level. Participants were given an SDQ form in either Spanish or English based on whether they enrolled in a Spanish or English-speaking parenting group.

**Treatment of Data/Plan of Analysis**

The study will determine if there are any significant mean score differences between the Spanish and English forms of the SDQ. Comparisons will be made between scale and TD scores to determine whether English or Spanish raters identified different degrees of concerns about their children's psychological adjustment.

Data analyzed in this study indicate how well the items from both the Spanish and English version of the SDQ fit together to generate a TD Score and four scale scores indicative of problem behaviors. Reliability will be estimated using Cronbach's alpha at both the composite and scale levels.

In addition, confirmatory factor analysis will be used to examine the internal

structure validity. Several indicators will be calculated including the normed fit index (NFI), goodness of fit index (GFI), and the comparative fit index (CFI), indicating how well the specific data is structured in relation to the proposed model. Further, the CFI indicates the fit of a target model to the fit of an independent model, which assumes all variables are uncorrelated (Bentler, 1990). The NFI compares the null model and target model and indicates how well the proposed model improves the fit relative to the independent model (Bentler, 1990). The GFI index involves the variance and covariances jointly explained by the model (Joreskog and Sorbom, 1986). It is somewhat difficult to determine whether a model is considered a good or bad fit but it is much easier to decipher that one model is better than another (Hair Jr., Black, Babin, Anderson, 2010). For instance, a model that has a CFI of .95 is considered a better fit than a model, which has a CFI of .85. In addition, simpler models with fewer observed variables and less subjects should have stricter criteria in regards to goodness of fit as compared to models, which contain more variables and more subjects (Hair Jr., et al., 2010). All of the indices mentioned require a statistic of .92 or more in order to be considered acceptable regarding how well the data fit the structured model (Hair Jr., et al., 2010). None of these tests is affected by sample size and normal distribution.

Other goodness-of fit- statistics that are often used in confirmatory factor analysis and will be used in this study include the Standardized Root Mean Square Residual (SRMR) and the Root Mean Square Error of Approximation (RMSEA). Utilizing Hair Jr. et. al (2010) characteristics of goodness of fit indices, and our sample size and number of variables, it is determined that an SRMR of .08 or less is considered good fit and an RMSEA of .07 is considered good fit. Because there are many ways to interpret the

findings from confirmatory factor analysis, the various multiple fit statistics will be considered to represent various perspectives (Campbell, Gillaspay, and Thompson, 1995). In addition, several factor models will be assessed in order to ensure that the original 5-factor model designed by Goodman is the best fit for data from the Spanish form.

The current study will compare the relative fit of multiple models, including a Five First Order Factor (5F) Model consisting all five scales, a Five First Order within One Second Order Factor (5F1S) model consisting of all five scales scores nested within a second order TD score, and a Four First Order Factors within One Second Order Factor (4F1S) model consisting of the four problem behavior scales nested within the second order TD score as well as the non-nested PB scale.

Although CFA has been previously conducted on English versions of the SDQ, this study will include CFA on the English form to allow for comparison of the two forms on similar samples.

Analysis of this data has been approved by the Institutional Review Board of Rutgers, The State University of New Jersey.

## **Results**

First, scale means were computed for the entire sample (see Table 3). Parents rated their children highest on the HI scale with a mean rating of 4.66 (SD=2.42). Mean ratings for the CP Scale were 3.58 (SD=2.28), followed by a rating of 2.52 (SD=1.82) for the PP Scale. Parents rated their children lowest in terms of difficulties related to emotionality with a mean score of 2.20 (SD=2.03) on the ES Scale. Higher ratings for the PB Scale

were indicative of less difficulty in the area of prosocial behavior. Mean ratings for this scale were 7.12 (SD=2.09). Finally, the mean rating for the TD Scale was 13.07 (SD=6.07).

TABLE 3

*Mean Parent Ratings for All Participants (N=923)*

SDQ Scale	Mean Ratings (SD)
Emotional Symptoms (ES)	2.20 (2.03)
Conduct Problems (CP)	3.58 (2.28)
Hyperactivity-Inattention (HI)	4.66 (2.42)
Peer Problems (PP)	2.52 (1.82)
Prosocial Behavior (PB)	7.12* (2.09)
Total Difficulties (TD)	13.07 (6.07)

*Note:* SDQ= Strengths and Difficulties Questionnaire. Range of possible ratings is (0-10) on Emotional Symptoms, Conduct Problems, Hyperactivity-Inattention, Peer Problems, and Prosocial Behavior. Range of possible ratings for Total Difficulties is (0-40). \*Higher Ratings signify less difficulty/problem on the Prosocial Behavior Scale.

### **Differences in Mean SDQ Ratings**

In order to answer our first research question and to examine whether there were mean differences in SDQ scores based on the language of the forms, mean ratings were compared at both the individual scale level and at the composite level (see Table 4). Mean ratings of the individual scales and the TD Scales were very similar across the two forms. Parents who completed the form in either English or Spanish both rated hyperactivity-inattention to be the difficulty their child endured the most while emotional symptoms were rated least problematic.

TABLE 4

*Mean Parent Ratings Between Spanish and English forms of the SDQ*

SDQ Scale	English Form (N=245)	Spanish Form (N=240)
Emotional Symptoms (ES)	2.19 (2.05)	2.16 (1.94)
Conduct Problems (CP)	3.63 (2.45)	3.46 (2.01)
Hyperactivity-Inattention (HI)	4.95* (2.54)	4.33 (2.20)
Peer Problems (PP)	2.56 (1.90)	2.51 (1.73)
Prosocial Behavior (PB)	7.25 <sup>a</sup> (2.15)	7.01 <sup>a</sup> (2.00)
Total Difficulties (TD)	13.46* (6.43)	12.52 (5.45)

*Note:* SDQ= Strengths and Difficulties Questionnaire. Range of possible ratings is (0-10) on Emotional Symptoms, Conduct Problems, Hyperactivity-Inattention, Peer Problems, and Prosocial Behavior. Range of possible ratings for Total Difficulties is (0-40). <sup>a</sup>=Higher Ratings signify less difficulty/problem on the Prosocial Behavior Scale. ( )=Standard Deviation. \*= Significantly higher mean rating on English Form compared to Spanish Form ( $p<.05$ ).

Independent samples t-tests were conducted in order to compare the means of the two samples. Mean ratings were significantly higher on the HI scale,  $t(3.47) = 12.04$ ,  $p<.01$ , and the TD scale,  $t(1.98) = 3.92$ ,  $p<.05$ , when the SDQ was completed in English. Although the difference in ratings was significant, the effect sizes of the difference between the two forms of the HI scale ( $d=.14$ ) and TD scale were small ( $d=.24$ ).

In order to further analyze mean differences among parent ratings, the language of the form as well as the mother's country of origin were assessed to see if country of origin influenced how parents rated their children. To determine which countries of origin to analyze, the two countries with the highest representation in the sample (United States and Mexico) were used. As shown in Table 5, raters that were from the United States and completed the form in English rated their children's total difficulties the highest out of the 4 subgroups. Independent samples t-tests were run in order to compare the means of the four subgroups. There was only a significant difference between raters from the United

States who completed the form in English as compared to the Spanish form,  $t(1.97)=6.16$ ,  $p<.05$ . Those who completed the form in English rated their children's total difficulties significantly higher than those who completed the form in Spanish. Although the difference was significant, there effect size was somewhere in between a small and medium effect ( $d=.41$ ).

Because the sample size was so small ( $n=24$ ) of those from the United States who completed the form in Spanish, it is difficult to demonstrate that country of origin actually affected mean ratings. For raters from Mexico, it did not appear that language of the form affected total difficulty ratings, as scores were similar across English and Spanish forms. In summary, English-speaking parents from the United States rated their children's total difficulties higher than did either language group from Mexico. Regardless of whether the form was completed in Spanish or English or whether they were born in Mexico or the United States, parents rated hyperactivity-inattention to be the most problematic area. Further breakdowns of language of form and country or origin mean ratings by scale are listed below in Table 5.



TABLE 5

*Mean Parent Ratings of SDQ Scales Across Spanish and English forms of the SDQ based on Mother's Country of Origin*

SDQ Scale	English Form	Spanish Form
United States (Country of Origin)	n=198	n=24
Emotional Symptoms (ES)	2.25 (2.05)	1.61 (1.54)
Conduct Problems (CP)	3.89 (2.50)	3.42 (2.13)
Hyperactivity-Inattention (HI)	5.14 (2.57)	4.31 (2.49)
Peer Problems (PP)	2.61 (1.90)	2.48 (1.72)
Prosocial Behavior (PB)	7.19 <sup>a</sup> (2.16)	6.52 <sup>a</sup> (2.17)
Total Difficulties (TD)	14.11 (6.46)	11.54 (4.54)
Mexico (Country of Origin)	n=47	n=216
Emotional Symptoms (ES)	2.07 (1.88)	2.26 (1.97)
Conduct Problems (CP)	3.20 (2.07)	3.59 (2.05)
Hyperactivity-Inattention (HI)	4.02 (2.16)	4.42 (2.21)
Peer Problems (PP)	2.65 (2.08)	2.60 (1.77)
Prosocial Behavior (PB)	7.40 <sup>a</sup> (2.08)	7.00 <sup>a</sup> (1.96)
Total Difficulties (TD)	12.25 (6.11)	12.91 (5.60)

*Note:* SDQ= Strengths and Difficulties Questionnaire. Range of possible ratings is (0-10) on Emotional Symptoms, Conduct Problems, Hyperactivity-Inattention, Peer Problems, and Prosocial Behavior. Scores in parentheses signify standard deviation. Range of possible ratings for Total Difficulties is (0-40). <sup>a</sup>=Higher Ratings signify less difficulty/problem on the Prosocial Behavior Scale.

### **Reliability**

The second research question addressed whether reliability differences existed based upon language of the SDQ form. In order to answer this question, the internal consistency coefficients (Cronbach's alphas) for the various SDQ scales were compared at the composite and scale levels (see Table 6). On the SDQ English form the reliability estimates varied greatly: one of the five scales had an alpha in the very low range. Alphas

for two of the five scales were in the low range and two were in the moderate range. The alpha representing the TD Scale was in the good range ( $\alpha=.81$ )

On the Spanish version of the SDQ, alphas for all five scales were in the very low range. The TD Scale had an alpha in the moderate range.

TABLE 6

*Reliability Coefficients (Coefficient alpha) Across Spanish and English forms of the SDQ*

SDQ Scale	English Form	Spanish Form
Emotional Symptoms (ES)	.65	.57
Conduct Problems (CP)	.74	.59
Hyperactivity-Inattention (HI)	.73	.59
Peer Problems (PP)	.47	.35
Prosocial Behavior (PB)	.69	.59
Total Difficulties (TD)	.81	.73

*Note:* SDQ= Strengths and Difficulties Questionnaire.

The TD scale Cronbach's alphas were then examined based on mother's country of origin to research whether differences in alpha statistics were based on language of the form or where the rater originated from. In order to determine which countries of origin to analyze, the two countries with the highest representation in the sample (United States and Mexico) were used. Both coefficient alphas were in the good range when the form was completed in English regardless of whether the mother informant was born in the United States or Mexico. However, the coefficient alpha for the forms completed in Spanish was in the low range when the informant was born in the United States and the in the moderate range when the informant was from Mexico (see Table 7). In order to analyze whether there were any statistically significant differences between coefficient alphas, Feldt and Kim's (2006) statistical method was used. Results indicated that there were no

statistically significant differences between either the language of the form or the country of origin in terms of reliability. Further, results suggest that Cronbach alphas were similar when compared based on mother's country of origin and that more differences were observed based on the language of the form, albeit none of them statistically significant. It is important, however, to note the small sample of participants born in the United States who completed the form in Spanish (n=26), thus making it more difficult to achieve a higher coefficient alpha.

TABLE 7

*Reliability Coefficients (Coefficient alpha) of Total Difficulties Scales Across Spanish and English forms of the SDQ based on Mother's Country of Origin*

SDQ Scale	English Form	Spanish Form
United States (Country of Origin)	n=200	n=26
Emotional Symptoms (ES)	.65	.41
Conduct Problems (CP)	.74	.67
Hyperactivity-Inattention (HI)	.75	.72
Peer Problems (PP)	.45	.47
Prosocial Behavior (PB)	.71	.72
Total Difficulties (TD)	.81	.64
Mexico (Country of Origin)	n=49	n=236
Emotional Symptoms (ES)	.58	.58
Conduct Problems (CP)	.66	.60
Hyperactivity-Inattention (HI)	.61	.59
Peer Problems (PP)	.53	.36
Prosocial Behavior (PB)	.63	.56
Total Difficulties (TD)	.81	.75

*Note:* SDQ= Strengths and Difficulties Questionnaire. United States and Mexico comprised the greatest percentage of countries of origin and were thus used as a base for comparison.

**Factor Analysis of English Models**

In order to answer our third research question as to whether the factor structure of the Spanish SDQ differed from the English SDQ, the internal structure validity of both forms was first measured via confirmatory factor analysis based on a 5F Model, including factors related to ES, CP, HI, PP, and PB. The model for the SDQ in English appeared to be a good fit, with the NFI (.88), the CFI (.91), and the GFI (.87) each at or approaching .92. The Standardized Root Mean Square Residual (Standardized RMR) of .07 also indicated good fit as it exceeded the level expected for a good fit (.08). The Root Mean Square Error of Approximation (RMSEA), at .07 also indicated a good fit. The 5F model accounted for between 5% and 52% of the variance in each individual item. The saturated model had a lower AIC (650.00) than did the five-factor model (996.12), indicating that the saturated model was a better fit, when not considering theory. The AIC of the Independence model (6385.36) was much higher than either. While factor loadings were high for the CP Factor with four out of five items exceeding .60 and moderately high for the ES Factor, HI Factor, and PB factor, loadings were lower and more difficult to interpret for the PP Factor. Three of the five items linked to this factor were below .30. Table 8 reports factor loading for each item of the 5F model.

TABLE 8

*Factor Loadings for English Models*

SDQ Items/Scale	5F Model	5F1S Model	4F1S Model
<b><i>Emotional Symptoms Scale (ES)</i></b>			
Somatic Complaints	.32	.32	.32
Worried	.59	.58	.59
Unhappy	.60	.61	.61
Nervous/Clingy	.46	.47	.46
Many fears	.61	.61	.61
<b><i>Conduct Problems Scale (CP)</i></b>			
Temper tantrums	.60	.60	.63
Obedient	.63	.62	.57
Fights w/children	.61	.63	.61
Lies/Cheats	.64	.62	.66
Steals	.52	.53	.54
<b><i>Hyperactivity-Inattention Scale (HI)</i></b>			
Restless/Overactive	.68	.67	.68
Fidgeting/Squirming	.72	.71	.72
Distracted	.63	.63	.64
Thinks before Acting	.40	.43	.40
Attention Span	.52	.53	.51
<b><i>Peer Problems Scale (PP)</i></b>			
Solitary	.26*	.25*	.28*
One good friend	.47	.45	.42
Liked by other children	.63	.63	.61
Bullied by other children	.23*	.23*	.29*
Gets along w/adults more than peers	.29*	.30*	.33
<b><i>Prosocial Behavior Scale (PB)</i></b>			
Consider of others	.62	.65	.58
Shares	.58	.58	.53
Helpful	.51	.51	.59
Kind	.60	.57	.53
Volunteers	.49	.50	.60

*Note:* SDQ= Strengths and Difficulties Questionnaire. (5F) Model=Five First Order Factor Model. (5F1S) Model= Five First Order within One Second Order Factor Model. (4F1S) Model= Four First Order Factors within One Second Order Factor Model. \*= at or below .30 considered low factor loading (N=488)

Next, the internal structure validity of the English form was measured via confirmatory factor analysis based on a theoretical 5F1S model including the factors of ES, CP, HI, PP, and PB nested within the TD Factor. The model for the SDQ in English appeared to be a good fit, with the NFI (.87), the CFI (.90), and the GFI (.86) each approaching .92. The Standardized RMR of .07 also indicated good fit, exceeding the level expected for a good fit (.08). The RMSEA at .08 indicated a moderate fit as it was slightly above what is considered to be an acceptable fit (.07). The 5F1S model accounted for between 5% and 51% of the variance in each individual item. The saturated model had a much lower AIC (650.00) than did the (5F1S) model (1077.60), indicating that the saturated model was a better fit, when not considering theory. The AIC of the Independence model (6385.36) was much higher than either. While factor loadings were high for the CP Factor with four out of five items exceeding .60 and moderately high for the ES Factor, HI Factor, and PB factor, loadings were lower and more difficult to interpret for the PP Factor. Three of the five items linked to this factor were at or below .30. Table 8 depicts factor loading for each item of the Five 5F1S model.

The next confirmatory factor analysis that was analyzed was a 4F1S model that included ES, CP, HI, and PP isolated from the single PB Factor. The model for the SDQ in English appeared to be of moderate fit, with the NFI (.84), the CFI (.88), and the GFI (.85) each exceeding .80. The Standardized RMR of .12 indicated moderate fit. The RMSEA, at .08, indicated a moderate fit as well. The 4F1S model accounted for between 8% and 53% of the variance in each individual item. The saturated model had a much lower AIC (650.00) than did the (4F1S) model (1141.45), indicating that the saturated model was a better fit, when not considering theory. The AIC of the Independence model

(6385.36) was much higher than either. While factor loadings were high for the CP Factor with three out of five items exceeding .60 and moderately high for the ES Factor, HI Factor, and PB factor, loadings were again lower and more difficult to interpret for the PP Factor. Two of the five items linked to this factor were at or below .30. Table 8 depicts factor loading for each item of the 4F1S model.

When comparing the three English models, it appears as though the 5F Model has the highest internal structural validity. Although the statistics are similar when comparing the 5F Model and the 5F1S Model, the Five First Order Factor 5F Model still has the highest NFI, CFI, and GFI. The Standardized RMR of .07 was the same in the 5F Model and the 5F1S Model, yet substantially better than the fit of the (4F1S) model (.12). In addition, the RMSEA (.07) of the 5F Model matched the level expected for a good fit while the other two models were slightly below what would be considered a good fit (.07). Also, the 5F Model had the lowest AIC of 996.12, in comparison to the 5F1S Model of 1077.60 and the 4F1S model of 1141.45, again indicating a better fit. The models all accounted for approximately 5 to 50 percent of the variance in each individual item. Factor loadings were strikingly similar across all three of the models, with items on the CP Scale loading most convincingly onto their respective factor. Items on the PP Scale had the weakest loadings onto their respective factor across all three models. Items related to being solitary, getting bullied by others, and relating better to adults than peers loaded the weakest onto the PP factor. Table 9 depicts goodness of fit statistics for all three of the factor models in English and in Spanish.

TABLE 9

*Goodness of Fit Indices for English and Spanish Factor Models*

Indices	Models					
	5F ENG	5F1S ENG	4F1S ENG	5F SP	5F1S SP	4F1S SP
NFI	.88	.87	.84	.74	.70	.68
CFI	.91	.90	.88	.80	.76	.74
GFI	.87	.86	.85	.85	.82	.83
SRMR	.07	.07	.12	.08	.09	.10
RMSEA	.07	.08	.08	.08	.09	.09
AIC	996.12	1077.60	1141.45	1103.85	1270.86	1259.50

*Note:*. 5F ENG=Five First Order Factor Model, English Form. 5F1S ENG= Five First Order within One Second Order Factor Model, English Form. 4F1S ENG= Four First Order Factors within One Second Order Factor Model, English Form. 5F SPA= Five First Order Factor Model, Spanish Form. 5F1S SPA= Five First Order within One Second Order Factor Model, Spanish Form. 4F1S SPA= Four First Order Factors within One Second Order Factor Model, Spanish Form

**Factor Analysis of Spanish Models**

In order to more directly compare whether the factor structure of the Spanish SDQ differed from the English SDQ, the three factorial models were assessed with the Spanish form of the SDQ. The Spanish form was first measured via confirmatory factor analysis based on a 5F Model including factors related to ES, CP, HI, PP, and PB. The model for the SDQ in Spanish appeared to be a moderate fit, with the NFI (.74), the CFI (.80), and the GFI (.85) each at or approaching .80. The Standardized RMR of .08 indicated good fit. The RMSEA, at .08, indicated a moderate fit as it exceeded the level expected for a good fit (.07). The 5F Model accounted for between 2% and 38% of the variance in each individual item. The saturated model had a significantly lower AIC (650.00) than did the (5F) Model (1103.85), indicating that the saturated model was a better fit, when not considering theory. The AIC of the Independence model (3330.00) was much higher than either. While factor loadings were moderate for the CP Factor, ES Factor, and PS Factor, loadings were lower



and more difficult to interpret for the PP Factor and HI Factor. Three of the five items linked to the PP Factor were below .30. In addition, although two items associated with the HI Factor were considered to be highly loading onto their factor, two items were recognizably low and below .30. Table 10 depicts factor loading for each item of the 5F Model.

TABLE 10

*Factor Loadings for Spanish Models*

SDQ Items/Scale	(5F) Model	(5F1S) Model	(4F1S) Model
<b><i>Emotional Symptoms Scale (ES)</i></b>			
Somatic Complaints	.38	.35	.36
Worried	.50	.48	.48
Unhappy	.55	.56	.57
Nervous/Clingy	.45	.46	.45
Many fears	.44	.47	.46
<b><i>Conduct Problems Scale (CP)</i></b>			
Temper tantrums	.39	.37	.41
Obedient	.53	.57	.46
Fights w/children	.56	.54	.58
Lies/Cheats	.48	.46	.50
Steals	.35	.34	.40
<b><i>Hyperactivity-Inattention Scale (HI)</i></b>			
Restless/Overactive	.64	.61	.64
Fidgeting/Squirming	.62	.55	.62
Distracted	.51	.51	.51
Thinks before Acting	.28*	.34	.27*
Attention Span	.28*	.36	.29*
<b><i>Peer Problems Scale (PP)</i></b>			
Solitary	.17*	.19*	.22*
One good friend	.46	.42	.36
Liked by other children	.58	.57	.54
Bullied by other children	.19*	.23*	.27*
Gets along w/adults more than peers	.13*	.20*	.24*
<b><i>Prosocial Behavior Scale(PB)</i></b>			
Consider of others	.41	.41	.42
Shares	.48	.43	.37
Helpful	.47	.48	.56
Kind	.54	.55	.48
Volunteers	.49	.53	.57

*Note:* SDQ= Strengths and Difficulties Questionnaire. (5F) Model=Five First Order Factor Model. (5F1S) Model= Five First Order within One Second Order Factor Model. (4F1S) Model= Four First Order Factors within One Second Order Factor Model. \*= at or below .30 considered low factor loading (N=436)

The internal structure validity of the Spanish form was measured via confirmatory factor analysis based on a 5F1S model including the factors of ES, CP, HI, PP, and PB nested within the TD Factor. The model for the SDQ in Spanish appeared to be a moderate fit, with the NFI (.70) slightly lower than the CFI (.76), and the GFI (.82). The Standardized RMR of .09 also indicated moderate fit. The RMSEA, at .09, indicated a moderate fit as well as it was slightly above the level expected for a good fit (.07). The 5F1S model accounted for between 2% and 46% of the variance in each individual item. The saturated model had a much lower AIC (650.00) than did the 5F1S model (1270.86), indicating that the saturated model was a better fit, when not considering theory. The AIC of the Independence model (3330.00) was much higher than either. While factor loadings were moderate for the CP Factor, ES Factor, and (PS) Factor, loadings were lower and more difficult to interpret for the PP Factor and HI Factor. Three of the five items linked to the PP Factor were below .30. In addition, although two items associated with the HI Factor were considered to be highly loading onto their factor, two items were recognizably low and only slightly above .30. Table 10 depicts factor loading for each item of the 5F1S model.

Finally, the last confirmatory factor analysis that was analyzed was a 4F1S model that included ES, CP, HI, and PP isolated from the single PB Factor. The model for the SDQ in Spanish appeared to be of poorer fit compared to the 5F Model in Spanish, with the NFI (.68), the CFI (.74) far below .92 and the GFI slightly approaching .92 (.83). The Standardized RMR of .10 indicated moderate fit. The RMSEA, at .09, indicated a moderate fit as well as it was slightly above the level expected for a good fit (.07). The 4F1S model accounted for between 5% and 41% of the variance in each individual item. The saturated

model had a much lower AIC (650.00) than did the (4F1S) model (1259.50), indicating that the saturated model was a better fit, when not considering theory. The AIC of the Independence model (3330.00) was much higher than either. While factor loadings were moderate for the CP Factor, ES Factor, and (PS) Factor, loadings were lower and more difficult to interpret for the PP Factor and HI Factor. Two of the five items linked to the PP Factor were below .30. In addition, although two items associated with the HI Factor were considered to be highly loading onto their factor, two items were below .30. Table 10 depicts factor loading for each item of the 4F1S model.

When comparing the three Spanish models, it appears similar to the English models in that the 5F Model has the highest internal structural validity. The 5F Model has the highest NFI, CFI, and GFI. The Standardized RMR at .08, met the level expected for a good fit, while the other two models were slightly above the criterion at (.09) and (.10). In addition, while the RMSEA of the 5F model (.08) was slightly above the level expected for a good fit (.07), the other two models were still less of a good fit (.09). It is not clear when comparing the 5F1S model and the 4F1S model which model attains better fit since their goodness of fit indices are similar. However, it is clear that the 5F model attains the best fit of the three models. The 5F Model also had the lowest AIC of 1103.85, in comparison to the 5F1S model of 1270.86 and the 4F1S model of 1259.50, again indicating a better fit. The models all accounted for approximately 2 to 50 percent of the variance in each individual item. Factor loadings were also very similar across all three of the Spanish models, with items on the (EP) Scale, CP Scale, and PB Scale loading moderately onto their respective factors. On the other hand, items on the PP Scale had the weakest loadings onto their respective factor across all three models. Items related to preferring to be

solitary, getting bullied, and relating better to adults than children loaded the weakest onto the PP Scale across all of the Spanish models. Finally, there were inconsistent loadings within the HI Scale across all 3 models of the SDQ Spanish form. In each model, two of the items loaded convincingly well while two loaded poorly, which were related to thinking before acting and having a good attention span. Table 9 depicts goodness of fit statistics for all factor models in English and Spanish.

### **Factor Analytic Comparison of English and Spanish Models**

After comparing the confirmatory factor analyses of the six models in English and Spanish, it is clear the English models fare better. The average NFI (.86) for the English models is significantly higher than average for the Spanish models (.71). In addition, the average CFI (.90) is also significantly higher than the average for the Spanish models (.77). The averages for the GFI indices are more similar than the other measures, yet they are still higher (.86) for the English models than the Spanish models (.83). The average of the Standardized RMR's of the Spanish and English models are identical and slightly below what is considered a good fit (.08) at .09. The RMSEA averages are also identical at .08 on both English and Spanish forms, which is considered slightly above a good fit (.07). Models in both English and Spanish accounted for approximately the same percentage (2% to 50%) of the variance in each individual item.

Finally, factor loadings were much higher across the English models than the Spanish models. Items on the English models loaded highly onto the CP Factor, moderately high onto the EP Scale, HP scale, and PB scale. Items did not load well on the PP scale. Factor loadings were moderate, at best, for the Spanish models. Similar to the English models, items related to being solitary, getting bullied, and getting along better

with adults than children loaded poorly onto the PP Scale. Furthermore, unlike the English factor models, loadings were inconsistent on the HI Scale across all of the Spanish models. Furthermore, the English models appear to have higher internal structural validity compared to the Spanish models, when evaluating the NFI, CFI, GFI, as well as factor loadings.

### **Discussion**

Parent raters who took part in the California University (Irvine) Initiative for the Development of Attention and Readiness (CUIDAR) program assessed their preschool age children's behaviors using the Strengths and Difficulties Questionnaire as part of their entrance into the intervention program. In this study, the psychometric properties of the instrument were assessed in order to explore mean rating differences between the English and Spanish versions of the SDQ, along with coefficient alpha indicators of reliability at the scale and composite level, and factor structure differences. Results indicated individual scale and TD scale means were very similar across both forms of the SDQ with the only significant difference occurring between raters who were born in the United States and completed the form in English as opposed to Spanish. Those who completed the form in English rated their children's total difficulties significantly higher than those who completed the form in Spanish. Reliability coefficients indicated alphas were higher for the English form compared to the Spanish form at the scale and composite level but these differences were not statistically significant. Also of importance is that neither form attained adequate reliability at the scale level. Finally, the Five First Order Factor (5F) Model was the best-fit and most valid representation of all 25 items of the SDQ, regardless

of the language of the form. When comparing Spanish and English factor models of the SDQ, the English models attain better fit.

### **Mean Ratings of English and Spanish SDQ Forms**

The first research question was whether there are mean differences in SDQ scores for students from Spanish-speaking families versus students from English-speaking families. First, the mean parent ratings were computed for the entire sample, and it was found that parents rated children highest on the HI scale and lowest on the ES scale. When the sample was then divided into those who completed the form in Spanish versus English, similar patterns were revealed. HI was still rated the most problematic area while ES was rated least problematic. Mean ratings were similar across English and Spanish forms, with significant but small mean differences on the HI scale and the TD scale.

The mean ratings for the HI scale for the English and Spanish forms in this study were 4.95 and 4.33 respectively. Previous research is also consistent with HI usually rated most problematic out of all the indices. Muris, Meesters, and van den Berg (2002) also found that mean ratings on the HI scale were the highest among the scales. Hawes and Dadds (2004) found that parents of young children in Australia rated Hyperactivity-Inattention problems the highest as well.

One possible explanation for high ratings on the HI scale is that items on this scale measure more concrete observable behaviors. It is plausible that parents may not know as much with regards to their children's emotional well-being at a young age when children are not as self-expressive. It might be easier for a parent to recognize if their child is constantly fidgeting, squirming, or cannot sit still for long in comparison to whether or not they are worried, nervous in new situations, or liked by other children. Therefore, parents

may rate their children highest in terms of hyperactivity and inattention because they do not know as much yet with regard to their children's emotional well-being, peer interaction, or conduct-related behaviors. Furthermore, at a young age, children might not be able to articulate their feelings to express fears, concerns, and other emotions. In addition, they might not have the opportunity to build relationships with children, or the ability to lie. All of these behaviors are indicators of the ES, CP, PP, and PB scales in the SDQ. However, parents can easily observe how well their children can pay attention to things and sit still, which are characteristics that are measured in the HI Scale. This may also be something that is rated highest in comparison to other behaviors when children are younger, because children may not be developmentally able to maintain their attention for a long time.

Across studies, there does not appear to be any scale in which mean ratings were consistently the lowest. As indicated earlier, ratings were lowest on the Emotional Symptoms scale for both English and Spanish-speaking raters in this study, with mean scores of 2.19 and 2.16 respectively. However, in Muris et al. (2002), mean parent ratings were lowest on the Conduct Scale, with a mean score of .8. Hawes and Dadds (2004) found that Peer Problems were rated the lowest by both parents of boys and girls, with mean ratings of 1.52 and 1.27 respectively.

In order to further analyze mean differences among parent ratings, the language of the form, and mother's country of origin were assessed to see if any further differences occurred with regard to mean ratings. Mothers that were born in the United States and completed the form in English rated their children their children higher than those who were born in Mexico and completed the form in English. In fact, those from the United States, who completed the form in English, rated their children the highest out of all



subgroups with regard to Total Difficulties. There was a statistically significant difference between mean ratings for those who were born in the United States and completed the form in English as opposed to Spanish. In fact, the subgroup that completed the Spanish form and were from the United States, rated their children's Total Difficulties the lowest out of all four subgroups.. Prior studies have introduced the value of familisimo, the importance of immediate and extended familial ties in the Mexican-American population. Further, perhaps those who were born in the United States and utilized the Spanish form, viewed their child's interpersonal functioning as within the family system (Sharkey, et al., 2009). It is plausible that since these raters are more comfortable utilizing the Spanish language that they have embraced this Hispanic view on familial ties. Thus, parents may be less likely to identify personal strengths distinct from family strengths, causing United States born English raters to rate children higher. However, when making the claim that language of the form affects how United States born raters will rate their children, it is important to note that the Spanish speaking raters, who were born in the United States, are not equally represented in terms of sample size. There are only 24 participants in this group, which is much lower than the number of participants in the other subgroups. These results should thus be interpreted with caution. Furthermore, since there is no prior literature comparing the English form of the SDQ to the Spanish form of the SDQ, it is unknown whether these results are consistent.

Independent samples t-tests were run in order to compare mean ratings on the English and Spanish SDQ. Mean ratings were significantly higher on the HI Scale and the TD Scale when the SDQ was completed in English. Although the difference in ratings was significant, the effect sizes between the HI scales ( $d=.14$ ) and TD scales were small

( $d=.24$ ). Therefore, the difference in score between the English and Spanish forms may be of little practical importance. Further, since the effect sizes are small, it is difficult to draw conclusions from these results. What is important to note, however, is that despite small effect sizes, ratings on the HI scale scores were the highest of the scores on both forms.

Independent samples t-tests were also run in order to compare mean ratings on the English and Spanish SDQ while including country of origin as a variable. The only significant difference that was observed was between raters from the United States who completed the form in English as opposed to Spanish. Those from the United States who completed the form in English rated their children significantly higher on TD scales than those who completed the form in Spanish. The effect size of the significant relationship was between a small and medium effect ( $d=.41$ ).

### **Reliability of English and Spanish Forms of the SDQ**

The second research question asked how well the items from the Spanish version of the SDQ fit together to generate a TD score as compared to the English version of the SDQ. Cronbach's alphas for the SDQ scales were compared at the scale and total composite levels. At the scale level, the alphas are all inadequate. Although alpha levels are higher on the English form of the SDQ in comparison to the Spanish form, they are still not considered to represent adequate reliability. On neither the English nor the Spanish form was there a scale in which alpha was in the good range, except for the TD scale, on the English form. It appears as though prior research has found similar reliability coefficients at the scale and composite level. Goodman (2001) found Cronbach's alphas in the low to moderate range, with only the total composite in the good range. Hawes et al. (2004) and Muris et al., (2002) obtained similar Cronbach's alpha statistics, with alphas

ranging from the low to moderate range at the scale level, and above .80 and in the good range for the TD scale. Coupled with prior findings, this study indicates that there is a need for more items at the scale level. It is difficult to obtain alphas in the adequate or good range when there are only five items in each scale. Although a benefit of the SDQ is its brevity, increasing the number of items would make the scale more reliable, and probably increase the number of people utilizing the instrument.

The TD Cronbach's alphas were examined further to observe whether differences in alpha statistics were dependent on language of the form or mother's country of origin. Regardless of whether the mother was from the United States or Mexico, the two highly most represented countries of origin, the coefficient alphas were in the good range when the form was completed in English. However, when the form was completed in Spanish, regardless of country of origin, alphas were in the low and moderate ranges. However, there were no statistically significant differences between alpha ratings regardless of the language of the form or country of origin of the rater.

Again, it is important to note that the sample size of those who completed the form in Spanish and were from the United States was small, with only 24 people, so results should be interpreted with caution. Overall, when evaluating alphas at the composite level, the alphas appear somewhat more reliable, although not statistically, when the form is completed in English as opposed to Spanish. The Spanish TD scores were not shown to be reliable. This indicates that a more reliable measure is needed for Spanish-speaking parents. Recommendations for reliable measures for Spanish-speaking parents will be discussed in the conclusion.

### **Factor Analysis of the SDQ in English and Spanish**

The third research question involved whether the factor structure of the SDQ in Spanish differed from the SDQ in English. Three factor models were assessed through confirmatory factor analysis on both the English and Spanish form of the SDQ. The first was a 5F Model, which has been confirmed in prior literature to be the best-fit model. It consists of five factors related to ES, CP, HI, PP, and PB. The second model assessed was a 5F1S model with all factors nested within the TD factor. The third model that was assessed was a 4F1S model with four factors nested into the total, isolating the PB factor.

Prior research has indicated that the 5F Model was consistently found to be a good fit for all 25 items on the questionnaire. In this study, regardless of whether the form was completed in English or Spanish, the 5F Model was the best-fit and most valid representation of all 25 items in this study. Factor loadings were consistently lower on the Spanish model in comparison to the English model. However, on all models, whether they are in Spanish or English, loadings were consistently very low with items on the PP Scale. One can reason that this is so may be partly due to some items within this index needing to be reverse scored as well as scored normally. Perhaps, having a more uniform scoring system within an index would have enabled items to load higher within the index.

Similar to this study, Goodman (2001) confirmed the 5F Model, and indicated that all 25 items loaded onto their intended factors. He found that the model was the best fit in the case of the parent form as compared to the self-informant SDQ form, although the model was still a good fit for self-informant as well. Hawes and Dadds (2004) also confirmed the 5F Model with parents of Australian children, ages four through nine. They found that factor loadings were generally stronger for boys than for girls but that the design

was a good fit, regardless of gender. Further, the five-factor model, which included factors related to ES, CP, HI, PP, and PB had the highest internal structural validity in this study as well as other studies. On both the English and Spanish forms of the SDQ, the 5F Model had the highest NFI, CFI, and GFI.

In order to ensure adequate validity, a model is tested in comparison with other models. Prior studies, which assessed the factor structure of the SDQ, did not do this. Goodman (2001), Hawes and Dadds (2004), and Muris, Meesters, & van den Berg (2002) all confirmed the 5F model of the SDQ by assessing factor loading, and other statistical techniques that did not include comparison of hypothetical factor models. A benefit of this study was that the established (5F) Model of the SDQ was compared with other hypothetical models. Because the other two models, which include the TD factor, were considered to be a lesser fit in comparison to the 5F Model, which excludes the TD factor for both the English and Spanish forms, one must take a look at how SDQ scores are calculated and analyzed.

When assessing the 5F1S model and the 4F1S model, both of which include the second order factor, it is apparent that their structure does not fare as well in comparison to the highly established 5F Model. However, as discussed earlier, the TD Coefficient Alpha was by far the most reliable regardless of whether the form was completed in Spanish or English as compared to the other scale alphas. Furthermore, it appears as though, in the case of the second order factor models, the lower scale alphas are contributing to how well the factor model fares and not the inclusion of the second order TD Factor. Therefore, the inclusion of the TD Score as a measure of one's total problems across the domains measured appears to be appropriate.

When comparing the English factorial models of the SDQ to the Spanish models directly, it is apparent that the English models fare better. The English models have higher Normed Fit Index averages, Comparative Fit Index averages, and Goodness of Fit Index averages. Factor loadings were also much higher across English models as compared to Spanish models. Models in English and Spanish accounted for similar percentages of variance (2% to 50%) in each individual item.

### **Suggested Spanish Language Instruments**

The English form of the SDQ appears to have higher internal structural validity as well as higher coefficient alphas, at the scale and composite level, indicating better reliability. As the Spanish version of the SDQ exhibited lower factor loadings, fit indices, and alpha levels, what instruments should be used when working with Spanish speaking students and parents in order to get an idea of how their behaviors may manifest in the classroom? Research has indicated that some scales remain psychometrically sound when translated into another language. As discussed earlier, The Social Anxiety Scale For Adolescents is a three-factor model scale that was confirmed and remained reliable and valid despite its translation. Compared to the SDQ, the SAS-A when translated still demonstrates good reliability and internal structural validity. However, it is not as similar to the SDQ as one would like because it can only be used in an adolescent population with self-raters. Another instrument that can be used by parents, which is quite similar to the SDQ in that it measures emotional problems as well as attention and social related concerns is the Child Behavior Checklist (Goodman & Scott, 1999). This is an instrument that is being widely used in schools today and has good psychometric properties in its Spanish translated version. A study by Gross, Fogg, Young, Ridge, Cowell, Richardson, and Silvan

(2006) examined the Child Behavior Checklist (CBCL) with parents who represented different races, ethnicities, incomes, and language backgrounds. The authors found that despite language, racial, and socio-economic differences, the model was considered a good fit when it was translated in Spanish. Furthermore, as mentioned earlier, studies have found that the SDQ and CBCL are comparable in many ways, in that they correlate highly, measure similar behaviors, and discriminate between low and high-risk populations (Goodman & Scott, 1999). Compared to the SDQ, the CBCL when translated still demonstrates similar internal structural validity. Therefore, it is plausible that using a more psychometrically sound instrument like the CBCL in Spanish, may be useful when working with Spanish-speaking parents.

With all this said, the SDQ can certainly still be used in a Spanish-speaking population. There are circumstances when a shortened form is more practical for use and this is one of the strengths of the SDQ. Further, if one were to utilize the SDQ in Spanish, results should be analyzed through item analysis, in order to determine areas of strengths and weaknesses. Since the scale scores are not considered adequately reliable, it is difficult to generalize that one possesses strengths or weaknesses in a certain domain. However, when evaluating the item level analysis, one can learn more about a child, and utilize the information to gain further information from the raters. In addition, since the Spanish measure is less reliable, the error bands around scores will be greater. Coupled with the fact that raters using the Spanish form rate persons lower than English raters, it might be helpful when assessing Spanish forms to lower the criterion required for those who are considered at risk. It is the hope that eventually, a more complete picture of the child can be obtained through using this as an initial screener.

**Limitations**

The generalizability of these findings is limited in several ways. The SDQ has forms for children up to age 16; however this study is limited in that only children 3 through 5 were rated. Perhaps, mean ratings would have differed if the sample represented a larger age range of students. Also, there was an unequal distribution of ethnicities represented in this sample, with Mexican-American children being the most highly represented. Furthermore, it is unknown how generalizable the results of this study would be in communities where the Mexican-American population is not as high. In addition, country of origin was utilized in this study in order to assess how that variable may affect mean ratings and reliability statistics. However, if a reliable and valid acculturation measure was used in this study, one might be better able to infer how one views their child based on how assimilated they are to society in the United States. A related issue is that there were only 24 people from the United States who completed the form in Spanish. Thus, it is very difficult to disentangle the impacts of language and parental country of origin in this data. Finally, only the United States and Mexico were assessed as far as how country of origin may impact one's ratings. However, there were other countries of origin represented in this study, which had even smaller number of participants.

**Implications for Future Research and Practice**

Future studies regarding the Strengths and Difficulties Questionnaire could center upon analyzing changes in mean ratings as children grow older. In this study, SDQ ratings were only taken at the point of entry into the California University (Irvine) Initiative for the Development of Attention and Readiness (CUIDAR) program. It would be helpful to examine how ratings may change over time as children develop. Similarly, it would be



interesting to interpret what similar ratings over time may indicate about the stability of problems or areas of strength that youth possess.

Future studies could also explore using a reliable and valid measure of acculturation in order to view how one identifies with their culture and thus how it affects their views and ratings of their children. In this study, country of origin was used as variable to assess whether where one originated from affected reliability statistics and mean ratings. A more direct and reliable measure of acculturation would make for more generalizable and reliable results.

Another future area of research would involve replicating exploration of mean parent ratings of the Spanish and English form of the SDQ, with parents born from the United States and Mexico, in order to analyze whether cultural or language differences impact parent ratings of their children's behavior on the SDQ. In the present study, there were too few participants who were born in the United States but spoke Spanish as a primary language. If studies were to replicate this one with equal number of participants, we could obtain a deeper understanding and either confirm or disconfirm how language and country of origin affects SDQ ratings. This design could be expanded to other countries with English and Spanish-speaking parents.

Finally, the factor structure of the SDQ should be evaluated in all of the languages into which the measure has been translated. Doing so would indicate whether the translation of the SDQ items into different languages has resulted in lower internal structural validity.

For practitioners who want to get a better picture of a child's behavior, it is essential that they choose an instrument that is reliable and valid, and in the language with which respondents feel the most comfortable. It is of the utmost importance to have an instrument to measure behavioral strengths and weaknesses and how these characteristics will affect children's progress within the classroom. When working with a parent of any student, it is important to understand their values and perspectives when they are being asked to rate their child's behavior. Beyond using a reliable and valid instrument, one must follow up with parents in order to ensure understanding of their background and how they feel about their participation in the schooling process.

In addition, since the reliability at the scale level is low, regardless of the language of the form, practitioners should consider using additional measures when certain scales indicate high levels of difficulty. For example, it may be helpful to utilize the SDQ as an initial screener to indicate where areas of difficulty may lie for a child, but to then use a more reliable instrument to ensure this indeed is a problem for the client or student.

## **Conclusion**

As part of their entrance into the California University (Irvine) Initiative for the Development of Attention and Readiness (CUIDAR) program, parent raters assessed their preschool age children's behaviors using the Strengths and Difficulties Questionnaire. Data was collected over a four-year period, from 2004-2008. In this study, the psychometric properties of the instrument were assessed in order to explore mean rating differences between the English and Spanish versions of the SDQ, along with coefficient alpha indicators of reliability at the scale and composite level, and factor structure differences. Results indicated that mean ratings of the individual scales and the TD scales

were very similar across both forms of the SDQ. Mothers who were born in the United States and completed the form in English rated their children the highest among all subgroups in the study. Reliability coefficients indicated alphas were higher for the English form compared to the Spanish form at the scale and composite level, although there were no statistically significant differences.. On the TD scale, there was good reliability when the form was completed in English. Finally, the Five First Order Factor (5F) Model was the best-fit and most valid representation of all 25 items of the SDQ, despite the language of the form. When comparing Spanish and English factor models of the SDQ, it is apparent that English models attain better fit. Thus, it is important for practitioners to utilize caution when advising to use the SDQ in a Spanish-speaking population. There are other similar assessment measures, which have adequate psychometric properties despite their language translation.

## References

- Achenbach, T.M. (1991). *Manual for the Child Behavior Checklist and 1991 Profile*. Burlington, VT: University of VT, Department of Psychiatry.
- American Educational Research Association. (1999). *Standards for Educational and Psychological Testing*. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. Washington, DC: American Educational Research Association.
- Bentler, P.M (1990). Comparative fit indices in structural models. *Psychological Bulletin*, 107, 238-246.
- Buckley, J.A., Ryser, G., Reid, R., & Epstein, M.H. (2006). Confirmatory factor analysis of the Behavioral and Emotional Rating Scale-2 (BERS-2) Parent and Youth Rating Scales. *Journal of Child and Family Studies*, 15, 27-37.
- Campbell, T.C., Gillaspay, J.A., & Thompson, B. (1995, January). The factor structure of the Bem Sex-Role Inventory (BSRI): A confirmatory factor analysis. Paper presented at the annual meeting of the Southwest Educational research Association, Dallas. (ERIC Document Reproduction Service No. ED 380 491)
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78, 98- 104.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*, 38, 581-586.
- Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire (SDQ). *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 1337-1345.

- Goodman, R., Ford, T., Simmons H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *British Journal of Psychiatry*, 177, 534-539.
- Goodman R, Scott S (1999). Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: Is small beautiful? *Journal of Abnormal Child Psychology*, 27, 17-24.
- Gross, D., Young, M., Fogg, L., Ridge, A., Cowell, J., Richardson, R., & Sivan, A. (2006). The equivalence of the child behavior checklist/11/2-5 across parent race/ethnicity, income level, and language. *Psychological Assessment*, 18, 313-323.
- Guilford, J. P. (1946). New standards for test evaluation. *Educational and Psychological Measurement*, 6, 427-439.
- Hair, J., Black, W., Babin, B., Anderson, R. (2010). *Multivariate Data Analysis*. Upper Saddle River, NJ: Prentice Hall.
- Hawes DJ, Dadds MR (2004). Australian data and psychometric properties of the Strengths and Difficulties Questionnaire. *Australian and New Zealand Journal of Psychiatry*, 38, 644-651.
- Hopkins, W. G. (2002). A scale of magnitudes for effect statistics. *A New View of Statistics*. Retrieved October 9, 2011, from <http://sportsci.org/resource/stats/effectmag.html>
- Felt, L. & Seonghoon, K. (2006). Testing the difference between two alpha coefficients with small samples of subjects and raters. *Educational and Psychological Measurement*, 66, 589-600.
- Jimerson, S. R., Sharkey, J. D., Nyborg, V. M., & Furlong, M. J. (2004). Strength-based

assessment and school psychology: A summary and synthesis. *California School Psychologist*, 9, 9-20.

Joreskog, K.G. & Sorbom, D. (1986). Lisrel VI. 4th Ed. Mooresville, Indiana: Scientific Software, Inc.

Kettler, R.J., Elliott, S.N., Beddow, P.A., Compton, E., McGrath, D., Kaase, K., Bruen, C., Ford, L., & Hinton, K. (2010). What does an alternate assessment measure? A multitrait-multimethod analysis. *Exceptional Children*, 76(4), 457-474.

LaGreca, A.M. & Lopez, N. (1998). Social anxiety among adolescents: Linkage with peer relations and friendships. *Journal of Abnormal Child Psychology*, 26, 83-94.

Lakes, K.D., Kettler, R.J., Schmidt, J., Haynes, M., Feeney-Kettler, K.A., Kamptner, L., et al. (2009). The CUIDAR early intervention parent-training program for preschoolers at risk for behavioral disorders: An innovative practice for reducing disparities in access to service. *Journal of Early Intervention*, 31, 167-178.

Muris, P., Meesters, C., & Van den Berg, F. (2003). The Strengths and Difficulties Questionnaire (SDQ): Further evidence for its reliability and validity in a community sample of Dutch children and adolescents. *European Child and Adolescent Psychiatry*, 12, 1-8.

Olivares, J., Ruiz, J., Hidalgo, M., Garcia-Lopez, L., Rosa, A., & Piqueras, J. (2005). Social anxiety scale for adolescents (SAS-A): psychometric properties in a Spanish-speaking population. *International Journal of Clinical and Health Psychology*, 5, 85-97.

- Pedrotti, J. T., & Edwards, L. M. (2010). The intersection of positive psychology and multiculturalism in counseling. In J. G. Ponterotto, J.M. Casas, L. A. Suzuki, & C. M. Alexander (Eds.), *Handbook of multicultural counseling* (3rd ed., pp. 165–174). Thousand Oaks, CA: Sage
- Sharkey, J. D., You, S., Morrison, G. M., & Griffiths, A. J. (2009). Behavioral and Emotional Rating Scale-2 Parent Report: Exploring a Spanish Version with At-Risk Students. *Behavioral Disorders, 35*(1), 53-65.
- Smedje, H., Broman, J.E., Hetta, J., Von Knorring, A.L. (1999). Psychometric properties of a Swedish version of the "Strengths and Difficulties Questionnaire". *European Child and Adolescent Psychiatry, 8*,63-70.
- Smokowski, P.R., Reynolds, A.J., & Bezruczko, N. (1999). Resiliency and protective factors in adolescence: An autobiographical perspective from disadvantaged youth. *Journal of School Psychology, 37*(4), 425-448.
- Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement, 56*, 197-208.
- Tinkler, B. (2002). *A review of literature on Hispanic/Latino/a parent involvement in K-12 education*. Retrieved October 28, 2011, from <http://www.eric.ed.gov/PDFS/ED469134.pdf>
- Youth in Mind Organization. (2012). *SDQ Info*. Retrieved November 20, 2011, from <http://www.sdqinfo.com/>