



Published in final edited form as:

J Fam Psychol. 2009 October ; 23(5): 680–689. doi:10.1037/a0016142.

Cross-Language Measurement Equivalence of Parenting Measures for use with Mexican American Populations

Rajni L. Nair¹, Rebecca M.B. White², George P. Knight³, and Mark W. Roosa^{1,4}

¹Program for Prevention Research, Arizona State University, P. O. Box 876005, Tempe, AZ 85287-6005.

²School of Health Management and Policy, Arizona State University, P.O. Box 874506, Tempe, AZ 85287-4506.

³Department of Psychology, Arizona State University, P. O. Box 871104, Tempe, AZ 85287-1104.

⁴School of Social and Family Dynamics, Arizona State University, P. O. Box 873701, Tempe, AZ 85287-3701.

Abstract

Increasing diversity among families in the United States often necessitates the translation of common measures into various languages. However, even when great care is taken during translations, empirical evaluations of measurement equivalence are necessary. The current study demonstrates the analytic techniques researchers should use to evaluate the measurement equivalence of translated measures. To this end we investigated the cross-language measurement equivalence of several common parenting measures in a sample of 749 Mexican American families. The item invariance results indicated similarity of factor structures across language groups for each of the parenting measures for both mothers and children. Construct validity tests indicated similar slope relations between each of the four parenting measures and the outcomes across the two language groups for both mothers and children. Equivalence in intercepts, however, was only achieved for some outcomes. These findings indicate that the use of these measures in both within group and between group analyses based on correlation/covariance structure is defensible, but researchers are cautioned against interpretations of mean level differences across these language groups.

Keywords

Cross-Language Equivalence; Parenting; Mexican American; Spanish-speaking

Families in the United States are becoming increasingly diverse with regard to ethnicity and language use (U.S. Census Bureau, 2005). Family researchers interested in including minority families in their investigations or studying minority families exclusively, face several challenges, including the growing proportions of minority family members who speak a language other than English in the home. Thus, to study representative samples of minority families, researchers are often required to translate measures, previously used only with English-speaking European Americans, into different languages. Though recommendations regarding translation procedures abound (e.g., Erkut, Alarcon, Garcia-Coll, Tropp, & Vasquez Garcia, 1999), careful translation itself does not ensure that multiple language versions of an

Correspondence concerning this article should be addressed to Rajni L. Nair, Email: rajni.nair@asu.edu. Phone: (480) 727-6134; Fax (480) 965-5430.

This article may not exactly replicate the final versions published in the journal. It is not the copy of record.

instrument are measuring the same construct, in the same way, in different groups. Indeed, working with multilingual samples necessitates instruments that are similarly valid and reliable across language groups.

Though the value of careful translation cannot be underestimated, only empirical investigations of *measurement equivalence* are capable of ascertaining whether different language versions are similarly valid and reliable. Unless constructs are measured equivalently across language-diverse groups, findings from (a) data pooled across languages and (b) language-group comparisons may be misleading (Hui & Triandis, 1985; Knight&Hill, 1998). The analytic strategies employed to investigate measurement equivalence have been discussed in the methodological literature (Knight et al., 2002; Knight, Roosa, & Umaña-Taylor, 2009; Reise, Widaman, & Pugh, 1993; Widaman & Reise, 1997), but few substantive researchers have utilized such strategies to empirically examine the adequacy of translated measures (Millsap, 2007). The primary aim of this paper was to provide an example of the processes used to empirically evaluate the cross-language measurement equivalence of translated scales. Through an examination of the English-Spanish measurement equivalence of several measures of parenting, commonly employed in family research with Mexican-origin mothers and children (i.e., measures of warmth, consistent discipline, harsh parenting, and monitoring), we highlighted procedures that can be used to empirically evaluate any translations, not just translations from English to Spanish. A secondary aim of this paper was to provide evidence of the equivalence of Spanish language versions of the parenting measures and make recommendations for their use in future research with Mexican origin families.

English – Spanish Equivalence of Common Parenting Measures among Mexican Americans

Latinos are the largest and fastest growing ethnic minority group in the U.S. and Mexican Americans account for about 60% of the Latino population (U.S. Census Bureau, 2005). Given that nearly 80% of Mexican Americans speak a language other than English in the home and less than half of them speak English very well (U.S. Census Bureau, 2004), Spanish-language parenting measures are increasingly necessary. Though several studies have employed translated versions of parenting measures for use with Mexican-origin Latinos (Hill, Bush, & Roosa, 2003), none have empirically evaluated these translations for equivalence. For example, studies have reported that less acculturated, presumably more Spanish speaking Latinos scored lower on inconsistent discipline than English-speaking Latinos (Dumka, Roosa, & Jackson, 1997; Samaniego & Gonzales, 1999). However, absent evidence of measurement equivalence, researchers are unable to determine if differences across these groups reflect true group differences, or differences due to measurement artifact.

Cross-Language Measurement Equivalence

To determine that multiple language versions of measures are similarly valid and reliable across translated versions several aspects of measurement equivalence must be examined: item, functional, and scalar equivalence (Hui & Triandis, 1985). Item equivalence exists when individual scale items have the same meaning across translated versions. Functional equivalence exists when the construct being measured has similar antecedents, precursors, and consequents in the groups completing different language versions. Scalar equivalence exists when a scale score refers to the same degree, intensity, and magnitude of the construct in the groups completing the different language versions.

Item equivalence is empirically assessed by exploring the factorial invariance of each language version of a measure, while functional and scalar equivalence are empirically assessed by exploring the construct validity equivalence of each language version of a measure. Factorial

invariance testing involves fitting a series of hierarchically-nested factor models simultaneously to each language group (Widaman & Reise, 1997). At each step in the series, additional constraints are added which further restrict distinctions between the measurement models in each language group. Only when model fit is deemed acceptable, are subsequent constraints applied. Under the theoretical assumption that items should operate similarly in both groups, the steps proceed as follows. The first step in the series, the configural invariance model, tests whether items form the same factor across language versions of a measure. Second, the weak factorial invariance model, tests whether factor loadings are the same in each language version of a measure. Next, testing proceeds to strong factorial invariance, where constraints on item intercepts across language versions are added. Strict factorial invariance is tested by adding an equality constraint on the unique error terms. Construct validity equivalence testing involves regressing the target scale on a theoretically related construct simultaneously in both language groups. Initially models are estimated freely in each language group. Under the theoretical assumption that the construct should demonstrate similar correlations with other variables among the language groups, the second step involves constraining the slopes in the model to be equal in both language groups. Pending equal slopes, an additional constraint is added requiring intercepts in the model to be equal in both language groups. Functionally equivalent measures demonstrate similar slopes across language groups and measures that are scalar equivalent across language versions of a measure demonstrate similar slopes and intercepts.

Convergent construct validity equivalence relies on selecting theoretically relevant constructs to which each measure being tested should be related; discriminant construct validity (which is not tested in the current study) relies on selecting theoretically relevant constructs to which the measure being tested should not be related. Based upon the extant literature, available study variables, and theoretical assumptions that the constructs should demonstrate similar associations among the language groups we generated the following convergent construct validity hypotheses: warmth would be positively related to relationship quality and negatively related to youth internalizing symptoms (e.g., Steinberg, 1990); consistent discipline would be negatively related to classroom task orientation, frustration tolerance, and shy anxious behavior (e.g., Baumrind, 1989; Maccoby & Martin, 1983); harsh parenting would be positively related to classroom acting out behavior and negatively related to grades (e.g., Kochanska, Padavich, & Koenig, 1996; Nix, Pinderhughes, Dodge, Bates, & McFadyen-Ketchum, 1999; Patterson, Reid, & Dishion, 1992); and monitoring would be positively related to classroom social competence and negatively to externalizing symptoms (e.g., Patterson, Reid, & Dishion, 1992). Given that the purpose of this manuscript was to highlight the procedures for evaluating the adequacy of the translation process by examining cross-language measurement equivalence, it was important that the convergent construct validity analyses were faithful to nature of the relations between these parenting measures and outcome measures as they are understood to be from the existing theoretical and empirical literature. Consequently, we chose to focus upon theoretically related convergent construct validity associations that were (a) likely to be the same among the language groups, (b) generally proximal, and (c) frequently highlighted in the theoretical and empirical literature.

Given that our research focused on the cross-language measurement equivalence of parenting measures often used in multi-ethnic samples, we chose to examine the equivalence of translated versions of these measures in a sample that is very representative of the larger Mexican American population. The types of language groups we have within our sample are representative of the types of language groups naturally occurring within the larger U.S. Mexican American population and as such would be the types of groups of interest to many family researchers. That is, the majority of Mexican Americans within the United States (and thus of interest to family researchers) are not monolingual English or Spanish speakers; rather they have varying degrees of comfort with both languages (Garcia, 2002). Thus in examinations

of differences across language groups it is critical to test cross language measurement equivalence in a sample with similar language qualities rather than focusing exclusively on cross language equivalence in mono-lingual speakers of either Spanish or English. The current study was uniquely situated to address this issue because it employed a highly representative sample of Mexican Americans adolescents and families (Roosa et. al., 2008). However, it is important to note that these naturally occurring language groups likely differ on other important variables that may impact measurement equivalence, such as socioeconomic status. We addressed the latter issue by controlling for socioeconomic differences in all construct validity analyses to provide more concise evidence of measurement equivalence across language groups.

Method

Data for this study come from a longitudinal study of the role of culture and context in the lives of Mexican American families in a southwestern metropolitan area (Roosa et. al, 2008). This sample was selected to represent the diversity of the Mexican American population on acculturation, social class, and the cultural/ecological niches in which they live. Participants were 5th grade students of Mexican heritage and their parents, selected from schools that served ethnically and linguistically diverse communities. Eligibility criteria included: (a) the mother was the child's biological mother, and self-identified as Mexican or Mexican American; (b) the child's biological father was of Mexican origin; (c) the child was not severely learning disabled; and (d) no step-father/boyfriend was living with the child. The current study includes the 749 families with complete data for the analyses. Approximately 70% of children and 26% of mothers were born in the United States. The average ages of children and mothers were 10.9 years and 35.9 years, respectively. About half the children were male (51%) and two-thirds of the families were two-parent. Average income was in the \$25,000 to \$30,000 range. Teachers provided data on children's school behavior and academic performance; because only 93% of children had data from teacher report, sample size drops to 692 for some validity analyses.

Procedures

Study procedures are detailed elsewhere (Roosa et. al, 2008) but are briefly reviewed here. Using both random and purposive sampling, the research team chose 47 public, religious, and charter schools from throughout the metropolitan area to represent the economic, cultural, and social diversity of the city. Recruitment materials that explained the research project in English and Spanish and asked parents who were interested in being in the study to provide contact information were sent home with all children in the 5th grade. Upon obtaining contact information, families whose ethnicity was indicated as Hispanic or who had Hispanic/Latino surnames were selected for screening. Over 85% of those who returned recruitment materials were eligible for screening (e.g., Hispanic) and 1,028 met eligibility criteria. In-home Computer Assisted Personal Interviews with mothers (required), fathers (optional), and children (required) from 750 families, 73% of those eligible, were conducted concurrently by trained professional interviewers. Interviewers read each survey question and possible response aloud in participants' preferred language to reduce problems due to variations in literacy levels. Families were paid \$45 per participating member and interviews lasted on average two and half hours. All procedures were reviewed and approved by the Institutional Review Board at the first author's university and conformed to APA ethical standards.

Measures

All measures were translated using translation/back translation procedures and had been used in both Spanish- and English-speaking Mexican American populations in prior work (e.g., Barrera et al., 2002; Gonzales, Dumka, Deardorff, Carter, & McCray, 2004; Hill et al., 2003). All translations were completed by bicultural and bilingual adults, with at least a bachelor's

degree, who were from or involved with the local Latino community and had a high degree of familiarity with the most common local Spanish dialect. All measures, except the teacher rating scale, had mother and child report.

Language use—Our language use measure was an 8 item measure adapted from Acculturation Rating Scale for Mexican Americans-II (ARSMA-II; Cuellar, Arnold, Maldonado, 1995). This use measure consists of two subscales: English language use and Spanish language use. Mothers and children were asked to endorse items indicating how often or how much they used Spanish or English across a variety of contexts (i.e. spoken language, music, television, and writing) with higher scores indicating higher levels of use of the language. Participants responded on a 5-point Likert scale from “almost never” to “almost always”. Reliabilities for the mother report of English use and Spanish use subscales were .90 and .86, respectively. Reliabilities for the child report of English use and Spanish use subscales were .71 and .74, respectively.

Warmth, Consistent Discipline, and Harsh Parenting—The original CRPBI was developed by Schaefer (1965) to assess children’s perceptions of parents’ behaviors and has since been adapted to assess parents’ perceptions as well (Barrera et al., 2002). This study used three 8 item subscales to assess parental warmth ($\alpha_{\text{mothers}}=.79$ $\alpha_{\text{children}}=.82$), consistent discipline ($\alpha_{\text{mothers}}=.81$ $\alpha_{\text{children}}=.82$), and harshness ($\alpha_{\text{mothers}}=.73$ $\alpha_{\text{children}}=.70$). Participants responded on a 5-point Likert scale from “almost never” to “almost always” to items like “Your mother clearly told you about the rules she expected you to follow.”

Parental Monitoring—This 10-item scale, adapted from Small and Kerns (1993) and Small and Luster (1994) assesses parent’s and children’s perceptions of the parent’s knowledge of children’s actions, whereabouts, and friends. An example item is “You knew what [Child name] was doing after school.” Responses ranged from 1 (almost never) to 5 (almost always). Reliability for this scale was .75 for both mothers and children.

Teacher report of classroom behavior—Teacher ratings of children’s school behaviors were obtained using the Teacher-Child Rating Scale (T-CRS; Hightower et al., 1986). First, teachers judge the severity of children’s problems: acting-out (5 items, $\alpha=.92$), and shy-anxious behavior (5 items, $\alpha=.83$), ranging from 1 (not a problem) to 5 (very serious problem). Next teachers rated children’s competencies with regards to frustration tolerance (5 items, $\alpha=.92$; e.g., copes well with failure) and task orientation (6 items, $\alpha=.95$; e.g., completes work) on a scale ranging from 1 (not at all) to 5 (very well). Third, teachers completed a five item social competence scale ($\alpha=.95$) (Stormshak, Bellanti, and Bierman, 1996), rating behaviors (e.g., listens carefully to others) on a scale ranging from 1 (almost never) to 6 (almost always). Finally, teachers ranked the child’s academic performance ranging from the top 1/5 to the bottom 1/5 of the class. We used this approach because of the diversity, and incompatibility, of grading systems used in participating schools.

Internalizing and Externalizing symptoms—Mothers and children reported on children’s internalizing/externalizing symptoms using the Diagnostic Interview Schedule for Children (DISC; Bravo, 2003; Shaffer et al., 1996). Symptom counts were calculated consistent with procedures used previously in the literature (Ge, Brody, Conger & Simons, 2006).

Relationship quality—The assessment of quality of the mother-child relationship utilized a single item adapted from the Matthews, Wickerama, and Conger (1996) marital relationship quality scale that provides a brief evaluation of each individual’s overall evaluation of the relationship: “please tell me what kind of relationship you have you’re your child/mother.” . Responses ranged from 1 (the worst) to 7 (the best).

Analysis Plan

All analyses were conducted using MPLUS statistical software (Muthen & Muthen, 2005) with the COMPLEX analysis option used to compute standard errors and tests of model fit while accounting for the non-independence of observations (i.e., families nested within neighborhoods). For all models fit was considered good (acceptable) if the Comparative Fit Index (CFI) is greater than or equal to 0.95 (acceptable if > 0.90), the Root Mean Square Error of Approximation (RMSEA) is less than or equal to 0.05 (acceptable if less than 0.08), and the Standardized Root Mean Square Residual (SRMR) is below 0.05 (acceptable if less than 0.08; Hu & Bentler, 1999; Kline, 2005). Additionally, we relied on the Akaike Information Criteria (AIC) to make nested model comparisons evaluating the impact of additional constraints on model fit. The utilization of the COMPLEX option precluded the use of chi square difference tests in this case (Muthen & Muthen, 2005). A more constrained model provides a better fit to the data when the AIC value decreases with the added constraint (Browne & Cudeck, 1993). Judgments regarding overall fit of the model were based on evaluating the evidence across all fit indices; that is, rather than exclusively utilizing any one fit statistic to evaluate each model we relied on the preponderance of evidence across multiple fit indices to inform conclusions regarding overall model fit.

Factorial Invariance—Factorial invariance was assessed using multi-group confirmatory factor analysis (CFA) to fit a series of hierarchically nested factor structures: configural invariance, weak factorial invariance, strong factorial invariance, and strict invariance. The sequence of nested CFA models included (a) a free factor model (b) a model requiring loadings to be equal across groups, (c) a model requiring loadings and intercepts to be equal, and (d) a model requiring loadings, intercepts and unique error terms to be equal. Testing proceeded until model fit was unacceptable or until strict factorial invariance was achieved.

Construct Validity Equivalence—In a series of hierarchically nested model testing, we specified models in which each parenting scale score was regressed on a convergent construct validity variable. For each convergent construct validity equivalence analysis, three models were tested. In the first model both the slope and the intercept were freely estimated in each language group. In the second model slopes were constrained to be equal across language groups. This model was used to determine if the freely estimated slopes were statistically different from each other. In the third model intercepts were constrained to be equal across language groups. This model was used to determine if the freely estimated intercepts were statistically different from each other. A measure was considered to be functionally equivalent if the slope constraint did not contribute to misfit. A measure was determined to be scalar equivalent if the added intercept constraint did not contribute to misfit.

When both mother and child reports were available for a given construct validity variable, within-reporter models were computed (e.g., mother report of child internalizing symptoms on mother report of consistent discipline). This decision is consistent with prior work showing low levels of agreement between parent and child reports on parenting behaviors and child mental health and recommendations to conduct separate analyses for each reporter (Tein, Roosa, & Michaels, 1994). In addition, we controlled for socioeconomic status by entering mother report of family income on a first step in all convergent construct validity analyses.

Results

The language groups were based on the language in which the respondents felt more comfortable (i.e., the language each respondent chose) to complete the interview. There were 618 English-speaking children, 131 Spanish-speaking children, 226 English-speaking mothers, and 523 Spanish-speaking mothers. To explore language use patterns of our sample we

examined the frequency of their use of Spanish and English. Mothers and children who completed the measures in English had a mean English use scores of 4.46 and 4.43 (on a 5-point scale) and mean Spanish usage scores of 2.86 and 2.73, respectively. Children who completed measures in Spanish had very similar means in their English and Spanish language use scores, 3.73 and 3.71, respectively. Mothers who completed measures in Spanish had a mean Spanish usage score of 4.60 and mean English usage score of 2.30. This suggests that the children who completed the measures in Spanish may be more bilingual, showing adeptness in both languages, rather than being monolingual Spanish speakers. This pattern of language use is representative of the larger Mexican American population (Roosa et. al, 2002) in which the majority of individuals display heterogeneity in their use of English and Spanish rather than being exclusively monolingual (Garcia, 2002). English-speaking mothers, and mothers of English-speaking children, reported statistically higher levels of education and income compared to Spanish-speaking mothers and mothers of Spanish-speaking children. Most Spanish-speaking participants (72% of the children and 97.1% of the mothers) were born in Mexico.

Factorial Invariance of Parenting Measures

Child report of consistent discipline—This scale demonstrated configural and weak invariance across language groups when unique item variances were permitted to correlate among the three items that used the phrase “you broke a rule” and between the two items that used the phrase “thought carefully” (Table 1). Though theoretical explanations regarding correlations among unique factor variances were not evident, the problems occurred in both language groups, suggesting that translation was not the underlying issue. Strong invariance testing resulted in a small increase in the AIC, however the remaining fit indices indicated good to adequate fit. Thus a strict invariance model was attempted, but holding the item level unique variances invariant across groups contributed to significant misfit.

Mother report of consistent discipline—Similar to findings with children, this scale demonstrated configural invariance across language groups when unique item variances were permitted to correlate in both groups among the same sets of items described above (see Table 1). Building on the configural model, weak invariance was achieved, but strong invariance resulted in poorer fit indices and an increased AIC.

Child report of warmth—This scale demonstrated configural and weak invariance across language groups (see Table 1). Strong invariance testing resulted in a small increase in the AIC, however the remaining fit indices indicated good to adequate fit (RMSEA =.04, CFI=.96, SRMR=.07), thus a strict invariance model was attempted. The strict invariance model showed misfit as evidenced by increasing AIC values combined with less than adequate fit indices.

Mother report of warmth—This scale demonstrated configural and weak invariance across language groups (see Table 1). Strong invariance testing resulted in an increase in the AIC which is indicative of some misfit. In addition, the fit indices fell from the good range to the adequate or poor range. A strong invariance model therefore did not provide good fit to the data. Model testing did not proceed further.

Child report of harsh parenting—This scale demonstrated configural invariance across language groups (see Table 1). Weak invariance testing indicated some problems with fit (the AIC increased), however the remaining fit indices indicated adequate to good fit and thus a strong invariance model was tested. The intercept constraints added during testing of strong invariance contributed to misfit of the model as evidenced by the large increase in the AIC and

less than adequate fit indices. A strong invariance model therefore did not provide good fit to the data and model testing did not proceed further.

Mother report of harsh parenting—This scale demonstrated configural invariance when unique item variances between two items were allowed to correlate (both items employ the phrase “around the house”) (see Table 1). Although theoretical reasons for correlations among unique factors variances are not evident, common phrasing may be the cause. In addition, the problem occurred across both language groups, indicating that the issue was not due to translation. Building on the configural model, weak invariance was achieved, but strong invariance resulted in a large increase in the AIC value, suggesting that the more restrictive model did not provide better representation of the data. Further, all fit indices fell from the adequate and good range into the poor range during testing of the strong model. Modification indices revealed that four of the eight item intercepts were not invariant across the language groups. The strict model was not tested due to failure at the strong invariance level.

Child report of monitoring—This scale demonstrated configural invariance when unique variances were free to correlate in both groups between two items using the phrase “your friends” and between two items using the phrase “when you went out” (see Table 1). Using this final configural model, weak and strong invariance were achieved according to the AIC and other fit indices. When a strict invariance model was tested, the AIC increased but other fit indices were adequate to good.

Mother report of monitoring—Similar to the child monitoring scale results, this scale demonstrated configural and weak invariance when unique variances were free to correlate in both groups between the same sets of items that employed the word “child’s friends” and “when child went out” (see Table 1). Strong invariance testing revealed adequate fit according to the RMSEA, CFI, and SRMR. However, the AIC value increased, suggesting some misfit. Consequently, we also tested a strict model but fit indices clearly moved into the poor range.

Convergent Construct Validity Equivalence Analyses of Parenting Measures

To control for socioeconomic status mother report of family income was entered as a first step in all regressions. Across language groups mothers’ reports of family income were significantly positively related to the following outcomes: teacher report of task orientation, teacher report of frustration tolerance, teacher report of academic rank, and teacher report of social competence. In addition, across language groups mothers’ reports of family income were significantly negatively related to teachers’ reports of shy anxious behavior and teachers’ reports of acting out behavior.

Child report of consistent discipline—The consistent discipline subscale, as reported by children, demonstrated invariance in slopes across all three outcomes: shy anxious behavior, task orientation, and frustration tolerance (Table 2). Though slopes were invariant, contrary to our hypothesis, the association between consistent discipline and the three outcomes only achieved statistical significance in one case: task orientation for Spanish-speakers. This subscale also demonstrated invariance in intercepts for all outcomes.

Mother report of consistent discipline—The consistent discipline subscale, reported by mothers, demonstrated invariance in slopes across all three outcomes: shy anxious behavior, task orientation, and frustration tolerance (Table 2). Although slopes were invariant, consistent discipline was not significantly associated with any of the three outcomes. This subscale also demonstrated invariance in intercepts across all three outcomes.

Child report of warmth—The warmth subscale, reported by children, demonstrated invariance in slopes across both outcomes, relationship quality, and internalizing symptoms (Table 2). Further, the association between warmth and (a) relationship quality was positive and significant, and (b) internalizing symptoms was negative, and significant but only for English speaking Mexican American children. This subscale also demonstrated invariance in intercepts for both relationship quality and internalizing symptoms.

Mother report of warmth—The warmth subscale, reported by mothers, demonstrated invariance in slopes across both outcomes, relationship quality and internalizing symptoms (Table 2). The association between warmth and (a) relationship quality was positive and significant, and (b) internalizing symptoms was negative, but not statistically significant. This subscale also demonstrated invariance in intercepts for both outcomes.

Child report of harsh parenting—The harsh parenting subscale demonstrated invariance in slopes across both outcomes: acting out behavior, and academic rank (Table 2). In these analyses two significant slope coefficients emerged. The relation between harsh parenting and (a) acting out behavior was significant but only for the English speaking group, and (b) academic rank was significant but only for the English speaking children. However, these coefficients were in unanticipated directions and contrary to hypotheses; in these regression models *higher* harsh parenting was associated with *lower* acting out behavior and *higher* academic ranking. This subscale also demonstrated invariance in intercepts for one out of the two outcomes, namely academic rank. There was lack of invariance in intercepts for acting out behavior.

Mother report of harsh parenting—The harsh parenting subscale, reported by mothers, demonstrated invariance in slopes across both outcomes: acting out behavior and academic rank (Table 2). Two significant slope coefficients emerged: the relation of harsh parenting to (a) acting out behaviors was significant in both groups and (b) academic rank was significant and positive but only for English speakers. However, similar to the results using child report, both coefficients were in unanticipated directions and therefore contrary to our hypothesis. This subscale demonstrated intercept invariance for both outcomes.

Child report of monitoring—The monitoring subscale, as reported by children, demonstrated invariance in slopes across both outcomes: social competence, and externalizing symptoms (Table 2). The relation between monitoring and (a) social competence was positive and significant but only for English speakers, and (b) externalizing symptoms was negative and significant and negative for both groups. This subscale demonstrated intercept invariance for social competence but not for externalizing symptoms.

Mother report of monitoring—The monitoring subscale, as reported by mothers, demonstrated invariance in slopes across both outcomes: social competence, and externalizing symptoms (Table 2). However, none of the slope coefficients were significant. This subscale demonstrated intercept invariance for both outcomes.

Summary—The results of the item invariance analyses and the convergent construct validity equivalence analyses are summarized in Table 3. All mother measures achieved at least weak invariance across groups while three out of the four child measures achieved strong invariance (child report of harsh parenting only achieved weak invariance). Tests of similarity of validity coefficients for the mother and child report variables across groups, however, showed a complex pattern of results. For analyses involving mother report of items, three of four parenting scales achieved both slope and intercept invariance across language groups. The warmth subscale showed slope and intercept invariance in one out of two tested associations.

Analyses involving child report showed that these scales achieved slope invariance across groups, but were only able to achieve intercept invariance across all outcomes for two of the four parenting scales.

Discussion

The increasing demand for research focused on expanding ethnic and linguistic minority populations in the U.S. generates the need to ensure that measures demonstrate measurement equivalence across heterogeneous populations including subsets of ethnic populations that may speak different languages. A demonstration of measurement equivalence provides evidence that measured constructs represent similar entities across language groups, providing confidence that observed differences between or similarities across groups reflect differences/similarities in parenting processes without the confound of measurement artifact. Hence, the present study focused on demonstrating the process of empirically evaluating the cross-language measurement equivalence of translated scales through examining the equivalence of common parenting measures in a highly representative sample of Mexican American mothers and children who varied in their preference for using Spanish and English in the research context.

The groups in our sample are representative of the types of language groups naturally occurring within the larger Mexican American population and as such would be the types of groups of interest to many family researchers. That is, the majority of Mexican Americans within the United States are not monolingual English or Spanish speakers; rather they have varying degrees of comfort with both languages (Garcia, 2002). These language groups also vary on several related demographic characteristics. For example, families in which mother or child completed measures in Spanish reported lower levels of income and education, and were more likely to have an immediate family member born in Mexico, than families in which mother and/or child completed measures in English. These observed distributions of socioeconomic status and nativity (or generational status) across language groups are exactly what one would expect for a relatively representative sample of Mexican American families. Hence, any evidence of factor bias in the measures across these language groups could be the result of translation failures, or differences associated with those demographic characteristics on which these language groups differ in this broader Mexican American population. Fortunately, the factorial invariance analyses provided considerable support for cross-language invariance of factor structure, supporting the equivalence of the translation in groups that differ in language use and the many demographic characteristics confounded with language preference. Further, the construct convergent validity analyses controlled for family income, substantially reducing the possibility that the very few observed failures of construct validity equivalence could be a function of SES or nativity differences across language groups.

Item invariance results indicated similarity of factor structures across language groups for each of parenting measures for mothers and children. Item level invariance in intercepts/means could not be achieved for mother reports of any parenting variable except monitoring. However, with one exception (harsh parenting), strong invariance was achieved for all child reports. This pattern suggests that across all subscales, utilization of models based on the covariance/correlation structure of the items is justified in both English and Spanish mother and child groups with the present translations. Conversely, analyses focused on mean level differences across these groups may be more problematic, particularly for mothers, given the inability of all but one of the mother report scales to demonstrate strong invariance.

Convergent construct validity results provided similar evidence. In general, after controlling for socioeconomic status, results indicated similar slope relations between each of the four subscales and the outcomes across language groups. As would be expected based on item

invariance results, intercept equivalence was achieved only for some of the convergent construct validity analyses. While the complex pattern of intercept invariance across construct validity analyses could be suggestive of non-equivalence of these measures across language groups, it is equally plausible that non-equivalence in particular outcomes contributed to this finding. Without further testing it is not possible to discern whether lack of invariance in intercepts in the relation between, for example, warmth and mom report of internalizing, is due to problems of translation equivalence in the warmth measure, in the internalizing measure, or both. In summary, given the preponderance of evidence, the parenting measures examined here demonstrated adequate evidence of convergent construct validity equivalence across language groups. A noted limitation of these construct validity analyses is that an assessment of discriminant validity was not included. Instead we focused on demonstrating the processes of evaluating construct validity equivalence by focusing on convergent properties because these associations are well-documented in the extant literature. Focusing on them facilitated our aim to clearly highlight the process of empirically evaluating translation equivalence. Notably, however, the basic procedures are the same whether assessing discriminant or convergent validity and a complete assessment of construct validity equivalence would evaluate both convergent and discriminant associations.

Interestingly, while harsh parenting was related to acting out behavior and grades equivalently across language groups, the relation was opposite what had been hypothesized (i.e., negative for acting out behavior; positive for grades). Deater-Deckard and colleagues (2005) have suggested that racial group differences in associations between harsh parenting and child outcomes may result from differences in (a) the way that parent's use various parenting strategies, (b) the meaning children attach to different parenting strategies, and (c) the strategies' effect on adjustment. The unexpected associations identified in this study may reflect similar differences in the (a) manner, (b) meaning, and (c) effects of parenting among Mexican Americans. More substantive research is needed to address this theoretical lacuna among Mexican Americans specifically, and Latinos more generally. Specifically, it is important to determine if differences in the manner, meaning, and effects of parenting that have previously been identified between African Americans and European Americans also exist among Mexican Americans.

The harsh parenting findings highlight an important way in which theory should and will influence the process of empirically evaluating the cross-language measurement equivalence of translated scales. To demonstrate the analytical processes we evaluated the English-Spanish measurement equivalence of parenting measures that were, based on theoretical and empirical work among Mexican Americans (which has largely focused on between group differences, not within group variability), expected to operate similarly with identified convergent validity constructs among the two language groups. However, in some cases, family researchers may hypothesize that various constructs are expected to operate differently across language groups. In these cases, measurement equivalence is evidenced when the measures perform as one would expect based on theory. Indeed, if the culturally informed theory indicates that the construct is differentially related to some other construct across the language groups, or if the theory suggests some subtle differences in the item functioning and factor structure across language groups, then measurement equivalence would be indicated by somewhat different but theoretically consistent interrelations with other constructs or among the items. Hence, in some cases partial factorial invariance or partial construct validity equivalence may be indicative of measurement equivalence as long as the specific failures correspond precisely with culturally informed theory (Knight et. al., 2002). In the highlighted example, had warmth been theoretically expected to operate differently in the two language groups, then findings consistent with that theory in the construct validity analyses (e.g., different slopes in the association between warmth and an outcome) would have been viewed as evidence of measurement equivalence. Consequently, culturally informed theory should and will inform

the ways in which researchers interpret findings from empirical assessments of cross-language measurement equivalence.

In light of the empirical findings, we conclude that analyses based on the correlation or covariance structure using these translated measures are warranted in both within and between group research designs sampling English speaking and Spanish speaking Mexican American children and mothers. However, researchers are cautioned about interpreting any between language group analyses based on mean structure and mean level comparisons in these parenting measures for Mexican Americans. It is unclear whether mean level differences between these language groups are reflective of differences in the underlying parenting dimensions or whether they are due to measurement artifact associated with the translation process. However, the impact of socioeconomic status and cultural orientations represent dimensions on which these groups differ that offer additional sources of within group diversity with the potential to lead to measurement bias. Future work should examine the potential influence that these dimensions have on the within-group equivalence of important family measures.

This study investigated the cross-language measurement equivalence of several common parenting measures in an ethnic homogenous sample of Mexican American families and is designed to provide a model for researchers to identify language problems that may not be detected even within a very rigorous translation process. Despite the fact that this study utilized common guidelines for translation, some of the measures failed to achieve the highest level of item invariance or measurement equivalence. Though more work is needed to determine if these invariance and equivalence failures are associated with the translation, or with other within and between group differences, this study highlights the statistical procedures that can be used to explore empirically the results of various translation approaches.

Acknowledgments

We gratefully acknowledge the families for their participation in the project. Work on this project was supported, in part, by NIMH grants P30-MH39246, R01-MH68920, and T32-MH018387, and by the Cowden Fellowship Program of the School of Social and Family Dynamics at Arizona State University.

References

- Baumrind, D. Rearing competent children. In: Damon, W., editor. *Child development today and tomorrow*. San Francisco: Jossey-Bass; 1989. p. 349-378.
- Barrera MJ, Prelow HM, Dumka LE, Gonzales NA, Knight GP, Michaels ML, Roosa MW, Tein J. Pathways from family economic conditions to adolescents' distress: Supportive parenting stressors outside the family, and deviant peers. *Journal of Community Psychology* 2002;30:135-152.
- Bravo, M. Instrument development. In: Bernal, G.; Trimble, J.; Burlew, AK.; Leong, FTL., editors. *Handbook of racial & ethnic minority psychology*. Thousand Oaks, CA: Sage Publications; 2003. p. 220-236.
- Browne, MW.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA.; Long, JS., editors. *Testing structural equation models*. Newbury Park: CA: Sage; 1993. p. 136-162.
- Cuellar I, Arnold B, Maldonado R. Acculturation Rating Scale for Mexican Americans-II: A Revision of the original ARSMA Scale. *Hispanic Journal of Behavioral Sciences* 1995;17:275-304.
- Deater-Deckard, K.; Dodge, KA.; Sorbring, E. Cultural differences in the effects of physical punishment. In: Rutter, M.; Tienda, M., editors. *Ethnicity and causal mechanisms*. New York: Cambridge University Press; 2005. p. 204-226.
- Dumka LE, Roosa MW, Jackson KM. Risk, conflict, mothers' parenting, and children's adjustment in low-income, Mexican immigrant, and Mexican American families. *Journal of Marriage and the Family* 1997;59:309-323.

- Erkut S, Alarcon O, Garcia-Coll C, Tropp LA, Vasquez Garcia HA. The dual focus approach to creating bilingual measures. *Journal of Cross-Cultural Psychology* 1999;30:206–218.
- Garcia EE. Bilingualism and schooling in the United States. *International Journal of the Sociology of Language* 2002;1–92.
- Ge X, Brody GH, Conger RD, Simons RL. Pubertal maturation and African American children's internalizing and externalizing symptoms. *Journal of Youth and Adolescence* 2006;35(4):528–537.
- Gonzales NA, Dumka LE, Deardorff J, Carter SJ, McCray A. Preventing poor mental health and school dropout of Mexican American adolescents following the transition to junior high school. *Journal of Adolescent Research* 2004;19:113–131.
- Hightower AD, Work WC, Cowen EL, Lotyczewski BS, Spinell AP, Guare JC, et al. The teacher-child rating scale: A brief objective measure of elementary children's school problem behaviors and competencies. *School Psychology Review* 1986;15(3):393–409.
- Hill NE, Bush KR, Roosa MW. Parenting and family socialization strategies and children's mental health: Low-income Mexican-American and Euro-American mothers and children. *Child Development* 2003;74:189–204. [PubMed: 12625445]
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* 1999;6:1–55.
- Hui CH, Triandis HC. Measurement in cross-cultural psychology. *Journal of Cross Cultural Psychology* 1985;16:131–152.
- Kline Rex, B. *Principles and Practice of Structural Equation Modeling*. Vol. 2nd ed.. New York: Guilford Press; 2005.
- Knight, GP.; Hill, NE. Measurement equivalence in research involving minority adolescents. In: McLoyd, V.; Steinberg, L., editors. *Studying minority adolescents: Conceptual, methodological and theoretical issues*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; 1998. p. 183-210.
- Knight, GP.; Roosa, MW.; Umaña-Taylor, AJ. *Methodological challenges and best practices*. Washington, DC: American Psychological Association; 2009. *Studying ethnic minority and/or economically disadvantaged populations*.
- Knight, GP.; Tein, J.; Prost, JH.; Gonzales, NA. Measurement equivalence and research on Latino children and families: The importance of culturally informed theory. In: Contreras, JM.; Kerns, KA.; Neal-Barnett, AM., editors. *Latino children and families in the United States: Current research and future directions*. Westport, CT: Greenwood Publishing Group; 2002.
- Kochanska G, Padavich EL, Koenig AL. Children's narratives about hypothetical moral dilemmas and objective measures of their conscience: Mutual relations and socialization antecedents. *Child Development* 1996;67:1420–1436. [PubMed: 8890492]
- Maccoby, E.; Martin, J. Socialization in the context of the family: Parent-child interaction. In: He&erington, EM.; Mussen, PH., editors. *Handbook of child psychology Vol. 4. Socialization, personality, and social development*. New York: Wiley; 1983. p. 1-101.
- Marin, G. Issues in the measurement of acculturation among Hispanics. In: Geisinger, KF., editor. *Psychological testing of Hispanics*. Washington D.C.: American Psychological Association; 1992. p. 235-251.
- Matthews LS, Wickerama KAS, Conger RD. Predicting marital instability from spouse and observer reports of marital interaction. *Journal of Marriage and the Family* 1996;58:641–655.
- Millsap RE. Structural equation modeling made difficult. *Personality and Individual Differences* 2007;42:875–881.
- Muthén, LK.; Muthén, BO. *Mplus user's guide*. Vol. Fourth Edition. Los Angeles, CA: Muthén & Muthén; 2005.
- Nix RL, Pinderhughes EE, Dodge KA, Bates JE, Pettit GS, McFadyen-Ketchum SA. The relation between mothers' hostile attribution tendencies and children's externalizing behavior problems: The mediating role of mothers' harsh discipline practices. *Child Development* 1999;70:896–909. [PubMed: 10446725]
- Patterson, GR.; Reid, JB.; Dishion, TJ. *A social learning approach for antisocial boys*. Eugene, OR: Castalia; 1992.

- Reise SP, Widaman KF, Pugh RH. Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin* 1995;114:552–566. [PubMed: 8272470]
- Roosa MW, Liu FF, Torres M, Gonzales NA, Knight GP, Saenz D. Sampling and recruitment in studies of cultural influences on adjustment: a case study with Mexican Americans. *Journal of Family Psychology* 2008;22:293–302. [PubMed: 18410216]
- Samaniego RY, Gonzales NA. Multiple mediators of the effects of acculturation status on delinquency for Mexican American adolescents. *American Journal of Community Psychology* 1999;27:189–210. [PubMed: 10425699]
- Small SA, Kerns D. Unwanted sexual activity among peers during early and middle adolescence: Incidence and risk factors. *Journal of Marriage and the Family* 1993;55:941–952.
- Small SA, Luster T. Adolescent sexual activity: An ecological, risk-factor approach. *Journal of Marriage and the Family* 1994;56:181–192.
- Schaefer ES. Children's reports of parental behavior: An inventory. *Child Development* 1965;36:413–424. [PubMed: 14300862]
- Shaffer D, Fisher P, Dulcan MK, Davies M, Piacentini J, Schwab-Stone ME, et al. The NIMH Diagnostic Interview Schedule for Children Version 2.3 (DISC-2.3): Description, acceptability, prevalence rates, and performance in the MECA Study. *Methods for the Epidemiology of Child and Adolescent Mental Disorders Study. Journal of the American Academy Child Adolescent Psychiatry* 1996;35(7):865–877.
- Steinberg, L. Interdependency in the family: Autonomy, conflict, and harmony. In: Feldman, S.; Elliot, G., editors. *At the threshold: The developing adolescent*. Cambridge, MA: Harvard University Press; 1990. p. 255-276.
- Stormshak EA, Bellanti CJ, Bierman KL. The quality of sibling relationships and the development of social competence and behavioral control in aggressive children. *Developmental Psychology* 1996;32(1):79–89.
- Tein J, Roosa MW, Michaels ML. Agreement between parent and child reports on parental behaviors. *Journal of Marriage and Family* 1994;56:341–355.
- Tienda M, Neidert LJ. Language, education, and socioeconomic achievement of Hispanic origin men. *Social Science Quarterly* 1984;65:519–536.
- U.S Census Bureau. *We the People: Hispanics in the United States*. Census 2000 Special Reports. Washington, D.C.: U.S. Department of Commerce, Economics and Statistics Administration; 2004.
- U.S Census Bureau. *Annual Demographic Survey: March Supplement*. 2005 [Retrieved December 3]. from <http://pubdb3.census.gov/macro/032005/pov/new01100.html>
- Widaman, KF.; Reise, SP. Exploring the measurement invariance of psychological instruments: Applicants in the substance use domain. In: Bryant, KJ.; Windle, M.; West, SG., editors. *The science of prevention: Methodological advances from alcohol and substance abuse research*. Washington, DC: American Psychological Association; 1997. p. 281-324.

Table 1
 Cross-Language Factorial Invariance Results for each Parenting Measure and each Reporter.

	<i>df</i>	Child				Mother			
		AIC	RMSEA	CFI	SRMR	<i>df</i>	AIC	RMSEA	SRMR
<i>Consistent Discipline</i>									
Configural Invariance	34	16433.87	0.05	0.97	0.04	34	15482.74	0.06	0.04
Weak Invariance	41	16425.25	0.04	0.97	0.04	41	15473.13	0.05	0.05
Strong Invariance	48	16432.47	0.05	0.96	0.05	48	15495.88	0.06	0.05
Strict Invariance	56	16496.76	0.07	0.91	0.09				
<i>Warmth</i>									
Configural Invariance	40	13635.46	0.05	0.97	0.04	40	12253.74	0.04	0.04
Weak Invariance	47	13632.43	0.04	0.97	0.06	47	12251.68	0.04	0.07
Strong Invariance	54	13633.65	0.04	0.96	0.07	54	12297.84	0.06	0.10
Strict Invariance	62	13682.91	0.05	0.93	0.18				
<i>Harsh Parenting</i>									
Configural Invariance	40	18743.90	0.03	0.98	0.03	38	16687.09	0.05	0.04
Weak Invariance	47	18753.03	0.04	0.96	0.06	45	16680.64	0.05	0.05
Strong Invariance	54	18803.01	0.06	0.89	0.07	52	16881.55	0.11	0.14
<i>Monitoring</i>									
Configural Invariance	36	18073.03	0.05	0.96	0.04	36	16964.81	0.04	0.04
Weak Invariance	43	18067.32	0.04	0.96	0.04	43	16957.40	0.04	0.06
Strong Invariance	50	18603.54	0.04	0.96	0.05	50	16966.72	0.04	0.06
Strict Invariance	58	18067.30	0.04	0.96	0.08	58	17174.78	0.08	0.30

Table 2
 Cross-Language Convergent Construct Validity Equivalence Results for each Parenting Measure and each Reporter.

	Child		Mother	
	Slope ^a English (Spanish)	Intercept ^b English (Spanish)	Slope ^a English (Spanish)	Intercept ^b English (Spanish)
<i>Consistent Discipline</i>				
Shy-Anxious Behavior	0.04 (0.09)	1.54 (1.36)	-0.01(-0.03)	1.68 (1.63)
Task Orientation	-0.10 (-0.24*)	3.80 (4.50)	0.09 (0.63)	3.04 (3.31)
Frustration Tolerance	-0.05 (-0.08)	4.06(4.43)	0.05 (-0.04)	3.66 (4.30)
<i>Warmth</i>				
Relationship Quality	0.85* (0.79*)	2.70 (2.97)	0.78* (0.69*)	2.64 (3.04)
Internalizing Symptoms	-1.90* (-3.22)	31.84 (33.67)	-1.86(-0.31)	19.91(12.24)
<i>Harsh Parenting</i>				
Acting Out Behavior	-0.12* (-0.12)	2.30 (1.88)	-0.31* (-0.16*)	2.76 (2.19)
Grades (Class Rank)	0.30* (0.17)	1.81(1.81)	0.29** (0.04)	1.87 (2.34)
<i>Monitoring</i>				
Social Competence	0.10* (0.08)	4.10 (4.28)	0.08 (0.27)	4.23 (3.52)
Externalizing Symptoms	-1.59** (-1.15*)	13.77 (8.54)	-0.39(-0.26)	6.88(5.61)

* $p < .05$

** $p < .01$

^a All slope coefficients are partial coefficients after controlling for income.

^b All intercepts were significantly different from zero.

Note. **Bold** indicates that English-speaking and Spanish-speaking coefficients are significantly different.

Table 3

Summary of all Equivalence Analyses

	Child				Mother			
	Consistent Discipline	Warmth	Harsh Parenting	Monitoring	Consistent Discipline	Warmth	Harsh Parenting	Monitoring
Factorial Invariance								
Configural	Y	Y	Y	Y	Y	Y	Y	Y
Weak	Y	Y	Y	Y	Y	Y	Y	Y
Strong	Y	Y	N	Y	N	N	N	Y
Strict	N	N	N	N	N	N	N	N
Convergent Construct Validity Equivalence								
Slope Invariant *	All	All	All	All	All	All	All	All
Intercept Invariant **	All	All	1 of 2	1 of 2	All	1 of 2	All	All

Y=yes; N=no.

* = number of models out of all models run in which slope was invariant.

** = number of models out of all models run in which intercept was invariant.