NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Examining Equivalence of Concepts and Measures in Diverse Samples

Tracy W. Harachi,
Social Development Research Group, University of Washington, Box 354900, 4101 Fifteen Avenue Northeast, Seattle, WA, 98105, USA

Yoonsun Choi,
School of Social Service Administration, University of Chicago, USA

Robert D. Abbott,
College of Education, University of Washington, USA

Richard F. Catalano, and
Social Development Research Group, School of Social Work, University of Washington, USA

Siri L. Bliesner
Seattle, Washington, USA

Tracy W. Harachi: tharachi@u.washington.edu

Abstract

While there is growing awareness for the need to examine the etiology of problem behaviors across cultural, racial, socioeconomic, and gender groups, much research tends to assume that constructs are equivalent and that the measures developed within one group equally assess constructs across groups. The meaning of constructs, however, may differ across groups or, if similar in meaning, measures developed for a given construct in one particular group may not be assessing the same construct or may not be assessing the construct in the same manner in other groups. The aims of this paper were to demonstrate a process of testing several forms of equivalence including conceptual, functional, item, and scalar using different methods. Data were from the Cross-Cultural Families Project, a study examining factors that promote the healthy development and adjustment of children among immigrant Cambodian and Vietnamese families. The process described in this paper can be implemented in other prevention studies interested in diverse groups. Demonstrating equivalence of constructs and measures prior to group comparisons is necessary in order to lend support of our interpretation of issues such as ethnic group differences and similarities.

Keywords

Equivalence; Comparative research

The heterogeneity of the U.S. race and ethnic composition continues to increase. Recent census data suggest that the proportion of European Americans continues to decline, 75.1% in the 2000 census vs. 87.5% in the 1970 census (Hobbs & Stoops, 2002). Despite the increasing diversity especially among children, much of the developmental research has focused on majority populations (McLoyd, 1998). Thus, a dearth of empirical findings

Correspondence to: Tracy W. Harachi, `tharachi@u.washington.edu`.

guides the development of interventions specific to ethnic minorities to prevent problem behaviors among these youth. More attention is needed to determine whether existing theoretical frameworks of child development are generalizable across groups and whether there are shared or unique factors that predict problem behaviors with our growing heterogeneous population. If different groups have similar predictors, interventions may share common targets. If, however, predictors are different, interventions will need to tailor their targets depending on the focal groups (Catalano et al., 1993). This information is particularly relevant as legislation such as No Child Left Behind (U. S. Department of Education, 2002) encourages the adoption of empirically-based, effective prevention programs. Many of these programs have not been validated with diverse groups.

Much of the research involving ethnic minorities lacks a prerequisite examination of conceptual and measurement equivalence. The meaning of constructs may differ across groups. Even if constructs are similar in meaning, instruments developed for a given construct in one particular group may not be assessing the same construct or may not be assessing the construct in the same manner in other groups. Research involving different cultures or demographic groups cannot assume a universality of meaning across groups but must utilize strategies to ascertain whether constructs are first comparable and second whether instruments adequate and appropriate in one context remain so in another (Harkness, Mohler, & Van de Vijver, 2003). Gottfredson Koper (1997) and Rosay, Gottfredson, Armstrong, and Harmon (2000) provide an example of researchers who have examined measurement invariance across not only racial but also gender groups prior to examining relationships between risk and protective factors and adolescent problem behaviors.

The definitions and terms of various forms of equivalence differ by authors (see e.g., Hui & Triandis, 1985; Van de Vijver & Leung, 1997). For example, Hui and Triandis organize the concept of equivalence into conceptual, functional, item, and scalar. The first two forms relate to meaning while the later two forms refer to properties of instruments assessing the target construct. *Conceptual equivalence* is hence defined as a construct having the same meaning across groups. It is also the first requirement prior to conducting any comparison. *Functional equivalence* refers to constructs which share similar nomological networks across groups (i.e., groups with similar precursors, consequences, and correlates). Behaviors shown in different groups that achieve similar goals are said to be functionally equivalent (Berry & Dasen as cited in Hui & Triandis, 1985). *Item equivalence* assumes that the prior forms of equivalence have been met. It is a form of measurement equivalence in which empirical evidence demonstrates a construct has the same meaning across groups via a particular instrument. Lastly, *scalar equivalence* builds on the prior forms of equivalence and is attained when a construct is measured on the same metric. Hence, a particular score on an instrument represents the same degree, intensity, or magnitude of the construct across groups. Scalar equivalence is necessary, for example, in order for the same score on a diagnostic tool to reflect the same level of severity across groups. These four forms represent requisite building blocks necessary for comparative research.

A variety of procedures have been adopted to examine the different forms of equivalence. For example, focus groups may be used to investigate conceptual equivalence (Fuller, Edwards, Vorakitphokatorn, & Sermsri, 1993; Wu et al., 2002). A common strategy to investigate functional equivalence is to examine correlation patterns between the target construct and other factors. Knight, Tein, Shell, and Roosa (1992) used an r-to-z transformation procedure to examine the intercorrelations between a set of family socialization measures across European and Latin American families. Testing the moderating effects of group membership on a pattern of correlations is another option

(Knight et al., 1992), which can be done by using multi-group structural equation modeling or examining interaction terms within a regression approach.

Methods to establish item and scalar equivalence vary extensively, and include the basic examination of computing reliability coefficients, such as Cronbach's alphas (Hui & Triandis, 1985), the examination of the internal structure of a measure and invariance using multiple group confirmatory factor analysis, or conducting an item response theory analysis (Van de Vijver & Leung, 1997; Vandenberg & Lance, 2000). Despite the availability of strategies, Steenkamp and Baumgartner (1998) note that measurement equivalence is not systematically examined, in part due to a bewildering array of strategies and a lack of agreement of guidelines. Vandenberg (2002) highlights the need for continued research to further strengthen the validity and applicability of various measurement equivalence procedures.

The aims of this paper are to demonstrate a process to examine conceptual, functional, item, and scalar equivalence using different methods. Data are from the Cross-Cultural Families Project, a study examining factors that promote the healthy development and adjustment of children among immigrant Cambodian and Vietnamese families. Family management is one of the most well researched proximal correlates of childhood problem behaviors (Griffin, Scheier, Botvin, Diaz, & Miller, 1999; Huebner & Howell, 2003; O'Neil, Parke, & McDowell, 2001), yet little research has been conducted to examine the equivalence of family management constructs or appropriateness of instruments to measure family management practices across groups. This study focused on two distinct questions: (1) Do Cambodian and Vietnamese U.S. immigrants hold general meanings of family management similar to western definitions of family management?; and (2) Is there empirical support of measurement equivalence of existing family management instruments among these Southeast Asian groups and a sample of non-immigrant European Americans? The process described in this paper can be utilized in other studies interested in diverse groups not only culturally but also demographically, for example, across gender groups. Such a process strengthens our confidence in the equivalence of constructs and instruments and hence provides support for interpretations of issues such as group differences and similarities.

Methods

The design of the pilot study of the Cross-Cultural Families Project followed the process Knight and colleagues (1992) used to examine the cross-cultural equivalence of family management measures between Hispanic and European American mothers and youth. Our study consisted of two phases. In Phase I, focus groups were conducted to explore the meaning of family management constructs among the two Southeast Asian groups. Subsequently a limited battery of family management measures were selected and translated for Phase II. In Phase II, data were collected from samples of European American, Vietnamese, and Cambodian mothers assessing aspects of family management to examine functional, item, and scalar equivalence of a select group of measures.

Phase I: conceptual equivalence

Focus groups were conducted within the local Vietnamese and Cambodian immigrant communities. A snowball sampling technique was used to identify parents through low-income housing organizations, refugee social service providers, and classes on parenting, citizenship and culture. Bilingual staff presented the project to groups of parents and invited interested parents to participate. A screening process was implemented to exclude non-immigrants and to ensure variation on education and length of time in the U.S. The principal investigator conducted the groups with the assistance of an interpreter in community locations convenient to the targeted parents. A total of 17 Cambodian (16 female and 1

male) parents and 18 Vietnamese (9 female and 9 male) parents participated. The average age for the Cambodian sample was 45 years and 51 years for the Vietnamese sample. The Cambodian participants migrated between 1980 and 1987 while the Vietnamese migrated between 1975 and 1995. The Cambodians participants had an average of 3.5 years of education and the Vietnamese had 8 years. Both groups were primarily Buddhist, though each had Protestants and the Vietnamese also had Catholic affiliation.

The focus group format included lead-in questions, probing statements, and sequencing more sensitive issues toward the end of the discussion. We developed questions to generate discussions on family management including: what parents want for their children, parents' roles, expectations and rules, monitoring and discipline, parent involvement, and changes in parenting made within the U.S. We did not ask participants to report about their own behavior, but to report on their perceptions and observations within their community. This strategy was chosen in order for participants to feel more comfortable sharing about the sensitive topic of family life and to reduce concerns that we would be asking about specific behaviors that might result in a need to report on child abuse.

The focus group data revealed that both Vietnamese and Cambodian parents reported using practices consistent with common, Western conceptions of family management. Both groups of parents had similar comments about expectations across developmental periods; both expressed that clear explanations about right and wrong, consistency in expectations, and routines and schedules are important for their children's success. Parents of both communities felt strongly that modeling expected behavior for children helped to teach them values and appropriate behaviors, and to gain respect and authority with their children.

Parents discussed the need to know a child's whereabouts and that this was particularly true, as the child grew older. For example, parents mentioned the use of relatives and friends as extensions of the family who help monitor their children's whereabouts. They felt it was important to know their child's friends, especially in the teen years. Parents stressed that, in order to foster positive behaviors in the teenage years, it was important to take time when children are young to teach them how to behave. During the teenage years, parents noted that establishing trust and developing a good relationship were very important factors that contributed to their ability to supervise and monitor their teenagers' behaviors.

When rules were broken or expectations not met, parents in both groups discussed the use of consequences and discipline which varied with the age of the child. For example, as children got older, parents from both groups mentioned the importance of talking, explaining, and teaching teenagers why a certain behavior was inappropriate. One subject that came up in both groups was the fact that their discipline practices had changed since their immigration to the U.S. Specifically, both groups expressed that parents were not able to use traditional means of corporal punishment now that they are in the U.S., and some expressed frustration that the inability to use traditional forms of corporal punishment undermined their ability to be effective parents.

Given the limitations of focus group methodology, including our sampling method and small number of focus group participants, the results were not used to make conclusive or generalizable statements about Cambodian or Vietnamese family management practices. Rather the intent of this qualitative investigation was to generate observations about parenting among families in these specific groups to consider whether the general meanings Vietnamese and Cambodian parents placed on family management were similar to those held by parents of Western cultures. The themes arising from the focus group discussions suggest some similarity in meanings regarding family management topics (e.g.

developmental expectations, rules, monitoring, and consequences) between the participating Vietnamese and Cambodian parents and commonly held western definitions.

Subsequently, we selected a sub-set of existing measures of family management for review and translation. Two panels of cultural reviewers were convened to help determine the best set of items. The first panel was composed of six bilingual Cambodian immigrants with some training or knowledge in research and the second panel consisted of five similarly skilled Vietnamese community members. Each panel member received a list of items and was asked to evaluate the translatability of the item into their native language and to rate the cultural relevance of each item. Both groups provided feedback regarding problematic wording and translation difficulties. The review panels provided another opportunity to examine the meaning of the target family management constructs. For example, panel members had a lively discussion regarding "time-out" as a strategy to discipline children. Panel members felt that "time-out" was not a practice commonly used in their countries of origins, however, felt it fit within their definition of discipline and was appropriate to include on a survey to be administered among families from their culture living in the U.S. In general, most items appeared to have the potential to translate and maintain con struct validity according to the cultural experts.

The selected items were subsequently translated into Khmer and Vietnamese utilizing a method of close translation (Harkness, Van De Vijver, & Mohler, 2003). A back translation process was conducted by another individual. In a review process, inconsistencies in the initial translation and back translation were reconciled. The translated measurement package was pretested with three to four parents from both communities using cognitive interviewing. Cognitive interviewing asks respondents to tell the interviewers what they thought about as they formulated their response to a particular item (Forsyth, Lessler, & Hubbard, 1992). This is different from traditional survey pretesting, and cognitive interviewing has been shown to provide data that can be used to improve survey quality through the identification of problems and through suggestions for terminology and time frames (Jobe & Mingay, 1990). In addition to identifying problematic terminology, the process allowed for testing of alternative wording and confirming face validity. Revisions based on the pretest data were made and a final review of comparability across translations was conducted.

Phase II: functional, item, and scalar equivalence

The study recruited samples of Vietnamese and Cambodian immigrant mothers and European American non-immigrant mothers in order to examine the remaining forms of equivalence. Locator information was obtained from an urban school district in the Pacific Northwest, and a random sample of Vietnamese, Cambodian, and European American families with children in grades 2 thru 4 was contacted. Project interviewers were the same ethnicity as the respondent. It was expected that a number of the Vietnamese and Cambodian parents might be unfamiliar with the concept of research and the process of consent gathering (Yu, 1985). Our procedures for training interviewers ensured that they were adept at establishing rapport with parents and addressing potential concerns. Additionally, interviewers did not interview families with whom they had an existing personal or professional relationship to address issues of confidentiality. All interviews were conducted in the native language of the respondent in the homes of respondents. In order to examine issues of quality assurance, interviewers audiotaped the interviews. These taped reviews suggested a high degree of consistency and adherence to study protocols by the interviewer team.

Participant characteristics—A high consent rate was achieved with 89% of individuals contacted agreeing to participate (n = 153 Vietnamese, 149 Cambodian, and 150 European American mothers). The average age of the respondents was 40 years for both the European American and Vietnamese mothers and 39 years for Cambodian mothers. There were significant group differences on income status with 17% of the European American, 88% of the Cambodians, and 90% of the Vietnamese households reporting receiving food stamps or eligibility for the federally funded free school lunch program. Ninety-six percent of European American mothers said they had completed at least high school in contrast to 34% of the Vietnamese mothers and 8% of the Cambodian mothers. The length of time in the U.S., at the time of interview, ranged from less than 1 year to 24 years, averaging 8 years for Vietnamese mothers and 13 years for Cambodian mothers. All of the Cambodian participants and 92% of the Vietnamese participants were first-generation immigrants or refugees.

Measures—Results from two parenting measures are examined in this paper: the cohesion items of the Family Adaptability and Cohesion Evaluation Scale (FACES III) (Olson, Portner, & Bell, 1982); and the warmth and involvement items from the Parenting Practices Questionnaire (PPQ) (Robinson, Man-dleco, Olsen, & Hart, 1995). The cohesion subscale has a reported alpha reliability of .77 (Olson, 1986) and has shown strong predictive and discriminant validity applied across a broad range of family groups (Sawin & Harrigan, 1995). The warmth and involvement subscale comprises one of three factors of authoritative parenting whose overall alpha reliability was reported at .91 (Robinson, Mandleco, Olsen, & Hart, 1995). Some items in both scales were dropped during the survey development process due primarily to difficulties with translation. Eight of 10 original items from the FACES III cohesion subscale were used and six of 11 items from the PPQ parental warmth and involvement subscale were kept. Wu and colleagues reported using seven of 11 PPQ items in their study of Chinese and American mothers (Wu et al., 2002). Additionally, response options were revised to standardize them in the survey. Response options for both measures ranged from "Never or almost never," "Once in a while," "Frequently," and "Always or almost always." Standardized scores of the measured survey items were first computed and then used to create indicators for subsequent analyses. The means for the cohesion subscale for the European American, Vietnamese, and Cambodian participants were 4.16 (SD = 0.42, α = .78), 4.27 (SD = 0.53, α = .81), and 3.83 (SD = 0.80, α = .88), respectively; while the means for warmth and involvement among the European American, Vietnamese, and Cambodian participants were 4.50 (SD = 0.43, α = .82), 4.16 (SD = 0.64, α = .82), and 3.90 (SD = 0.82, α = .82), respectively.

Analysis—First, to test functional equivalence, correlations between the two parenting subscales were conducted across the three groups. Second, using the EQS program (Bentler, 1993) multiple-group confirmatory factor analyses (CFA) were conducted in which we compared two nested models (i.e., unconstrained and constrained) to examine item equivalence (Byrne, 1994). The unconstrained model allows all parameters to be estimated freely for each group to establish the adequacy of factor loadings, model fit, and the pattern of intercorrelations among the latent factors for each group. This also determines whether the factor structure is invariant across groups, showing a consistent pattern of free and fixed factor loadings imposed on each construct including the direction and strength of factor loadings (Vandenberg & Lance, 2000), also called configural invariance in some studies. All factor loadings were allowed to vary freely while the factor variances were constrained to 1.00. The constrained model was then run in which equality constraints were placed on the factor loadings and covariances to examine whether there were statistical differences in the magnitudes of parameters across groups. Having no significant differences in parameters is termed metric invariance or equivalence. The European American group was selected as the

reference group and comparisons were made between the reference group and other two ethnic groups.

The goodness of fit of the measurement was evaluated by three statistics: model chi-square (χ^2), the Comparative Fit Index (CFI) (Bentler, 1990), and Root Mean Square Error of Approximation (RMSEA) (Hu & Bentler, 1998; Kline, 1998; Steiger & Lind, 1980). CFI values of greater than .90 indicate a good fit (Bentler, 1990). In addition, Knight and colleagues (1992) have suggested that values between .80 and .89 indicate an adequate fit, between .60 and .79 a poor fit, and less than .60 a very poor fit. RMSEA values of less than .05 are considered evidence of a good fit, between .05 and .08 a fair fit, between .08 and .10 a mediocre fit, and greater than .10 a poor fit (MacCallum, Browne, & Sug-awara, 1996). The statistical significance of the estimated parameters was examined with z-statistics at a .05 level of significance. The difference between the unconstrained and constrained model was examined by the change in χ^2 relative to the change in degrees of freedom (Byrne, 1994). If the change in χ^2 is significant, the measurement is considered to be nonequivalent across the groups. Each factor and indicator also needs to be further examined to understand where the differences exist. The Lagrange Multiplier (LM) test on the constrained models indicates which equality constraints contribute most to degradation in model fit (Bentler, 1990) and, hence, is used to provide information about potentially nonequivalent parameters.

Lastly to examine scalar equivalence, we examined invariance in the latent mean structures of each subscale using a multi-group structured means model with European American being the reference group. These steps follow the sequence recommended by Vandenberg (2000, 2002).

Functional equivalence—Group differences in correlations among the two parenting scales were examined. The correlation between the two measures was significant in all three groups (European American $r = .44$, $p < .05$; Vietnamese $r = .40$, $p < .01$; Cambodian $r = .44$, $p < .01$) suggesting functional equivalence across groups.

Item equivalence: cohesion subscale—The two-factor measurement model was estimated in both the unconstrained and constrained models. The unconstrained model fit the data well with $\chi^2(57) = 156.38$, a CFI of .92 and RMSEA of .06. All factor loadings were significant and in the hypothesized direction in all three groups. However, the constrained model showed adequate but marginal fit indices with $\chi^2(75) = 274.14$, a CFI of .83 and RMSEA of .08. The χ^2 difference between the constrained and unconstrained model was significant, $\Delta\chi^2(18) = 117.77$, $p < .05$, indicating that the measurement models for the three groups were nonequivalent. The LM tests, based on the unstandardized loadings, showed that the factor loadings of all indicators, except V8, were statistically different between the European American and the Cambodian groups. However, the differences were in magnitude not in direction. There was no statistically significant factor loading difference between the European American and Vietnamese groups (see Fig. 1). Loading differences between the European Americans and the Cambodian groups ranged from .058 to .482. Most of the differences were less than .10. In general, factor loadings were stronger in the Vietnamese group with the exception of "We like to do things with just the family in our household" (V2) and "Family members consult other family members on their decisions" (V8).

Item equivalence: warmth and involvement subscale—A one-factor measurement model was estimated. The unconstrained model suggested adequate but marginal fit with $\chi^2(27) = 140.97$, a CFI of .87 and RMSEA of .10. All factor loadings were significant and in the hypothesized direction in all three groups. The constrained model showed worse fit

indices with $\chi^2(39)= 186.95$, a CFI of .83 and a slightly better RMSEA of .09. The χ^2 difference between the constrained and unconstrained model was significant, $\Delta\chi^2(12) = 45.97$, $p < .05$, indicating that the measurement models for the three groups were nonequivalent. Statistically significant differences were found for all of factor loadings between the European American and Cambodian mothers with the LM test (see Fig. 2). Differences were again in magnitude not in direction. There was no statistically significant difference in factor loadings between the European American and Vietnamese mothers. Differences between the European Americans and Cambodians (range from .008 to .137) were again bigger than between the European Americans and the Vietnamese (range from .011 to .054), although both had more restricted ranges of differences than the cohesion measure.

Scalar equivalence—In the latent mean structures models, factor intercepts were interpretable only in a relative sense; more specifically, factor intercepts of the reference group were fixed to zero and the intercepts of the two Southeast Asian groups were compared in relation to those of the reference group (European American). For the cohesion subscale, the F1 unstandardized factor intercept for the Vietnamese group was .191 (SE.056, $t = 3.401$), and for the Cambodian group it was $-.159$ (SE.080, $t= -1.989$), hence, the latent means were significantly different with the Vietnamese group having a higher mean and the Cambodian group having a lower mean in relation to European American group. The F2 unstandardized factor intercept for the Vietnamese was $-.024$ (SE.069, $t= -0.348$) and $-.420$ (SE.097, $t= -4.320$) for Cambodians. Again the latent means were significantly different between the reference group and the Cambodian group, but not with the Vietnamese group.

Given the results of the CFA, further examination of scalar equivalence for the warmth and involvement sub-scale were not warranted; however, for purposes of illustration, the model was examined. The unstandardized factor intercept for the Vietnamese was $-.382$ (SE.074, $t= -5.164$) and $-.718$ ($SE .094$, $t = -7.660$) for the Cambodians, suggesting both latent means were significantly different between the reference group and the other two groups.

Discussion

Preventive interventions are based on interrupting causal processes or buffering against risk factors in the developmental etiology of problem behaviors. Thus, the similarity of the developmental etiology across groups must first be established (Catalano et al., 1993). A prerequisite but frequently overlooked step to determine etiology similarity is the examination of equivalence among relevant constructs within the developmental model and their indicators across groups. This study illustrated a process for examining conceptual and measurement equivalence that can be adopted by other prevention researchers in an effort to enhance our confidence of results examining different populations. This is particularly relevant for universal prevention strategies that seek to prevent development of problem behaviors across broad populations.

Research often overlooks the initial step of examining construct equivalence across groups. In the case of Gottfredson and Koper (1997), the analyses began with an investigation of measure equivalence and assumed that the constructs of delinquency and drug use held similar meanings across racial and gender groups. In their particular example, it is possible that these constructs at a conceptual level were similar given that groups in their study were derived from similar socio-environmental contexts. Such assumptions of conceptual equivalence may be less appropriate when comparisons are being made across less similar contexts, for example, the concept of self-efficacy across Chinese versus American students (Vandenberg, 2002) or the concept of life satisfaction across adolescent and elderly individuals (Pons, Balaguer, & Garcia Merita, 2000). This study began the process of

examining equivalence by conducting focus groups to elicit information about how family management was defined by immigrant Vietnamese and Cambodian parents. While focus groups are a cost-effective and flexible methodology to gather data, they are not without limitations. In particular these data are not readily generalizable given their limited sample sizes and often non-probabilistic sampling. Thus evidence derived from focus groups or other qualitative methodologies would be further strengthened by reporting from larger, probabilistic samples. Herein lies a dilemma since it makes little sense to develop a large-scale study which likely requires measurement equivalence when the initial step of conceptual equivalence has not been supported. This study relied on observations from focus groups to offer some indication that dimensions of family management were conceptualized similarly across the target groups. A subsequent study could be developed to examine in greater depth a particular family management construct. Detailed qualitative data could be gathered to more exhaustively define the universe of the target construct, e.g. monitoring within a particular developmental range. This level of detail can provide information to determine the range of indicators required to adequately cover the universe of the construct. Additionally, one may want to recruit group members to more specifically represent particular segments of the target population. For example, participants in our focus groups were chosen to represent a range in terms of length of time in the U.S. and education level which would be representative of the likely range in our local community. However, to examine conceptual equivalence with greater specificity, it would be useful to limit participation based on particular attributes, for example, conduct separate groups for recent immigrants and for those with longer U.S. residence.

The next phase of the study requires use of a common instrument to examine functional and the two aspects of measurement equivalence, item and scalar. Results from Phase 2 suggest that mothers of the three sample groups, European American and Vietnamese and Cambodian immigrants, have both similarities and differences in equivalence. Functional equivalence is demonstrated by similar precursors, consequences, and correlates. The results of the correlation analysis demonstrated that the subscales were similarly related to one another across groups, providing evidence of functional equivalence. Additional data and analyses which investigated the relationships between other precursors and consequences would have provided further support for this form of equivalence had it been available. It should also be noted that functional equivalence is not dependent on the use of the same instrument for comparison. One can utilize a culture or group specific set of indicators and establish functional equivalence of the construct. For example, buying apples in one culture may be functionally equivalent to buying mangos in another, so one measure may refer to one type of fruit and a second type in the comparison measure.

Turning to the remaining two forms of measurement equivalence, the use of alpha reliability coefficients has been given as a method to investigate item equivalence (Hui & Triandis, 1985) though more sophisticated and accurate procedures are now readily available (Vandenberg & Lance, 2000). Studies which rely solely on alpha reliability as a demonstration of psychometric properties would conclude that the two measures examined in this study were acceptable across groups, and would perhaps draw erroneous conclusions about group similarities or differences. Our findings underscore the need to utilize alternative measurement invariant procedures such as those described in Phase 2. The CFA suggest an adequate fit for the cohesion subscale and a marginal fit for the warmth and involvement subscale. There were significant differences in factor loadings across groups though the differences were in magnitude rather than direction, suggesting configural invariance but not metric invariance.

Positions vary among researchers on the extent of acceptable invariance when establishing equivalence. Horn and McArdle (1992) suggest that configural invariance may be sufficient

to compare factors across groups. Configural invariance is indicated when the magnitude of salient loadings are not identical, but the overall pattern of non-zero and zero loadings and factor structure is consistent across groups. Horn and McArdle argue that the criteria of evidence for metric invariance are too stringent in light of sample-specific artifacts that influence item variances and covariances. Byrne, Shavelson, and Muthen (1989) suggest partial invariance as more appropriate, in which at least one of the freely estimated loadings is equivalent across groups in addition to one factor loading fixed to 1.00 for identification purpose.

One strategy to address invariance in factor loadings is to omit significantly different items across groups. Gottfredson and Koper (1997) point out, however, that eliminating items may have costs in terms of reducing the breadth of measurement of the underlying construct and thereby risking the potential loss in measurement validity for all groups. Further, they argue that eliminating the very measures most likely to capture the variation of interest seems imprudent. In the case of our results, one may argue that the cohesion model fit adequately across groups and the support for configural invariance is sufficient to demonstrate item equivalence for this subscale. On the other hand, items assessing warmth and involvement were not equivalent across these three groups based on the marginal fit. Further, five out of the 11 items of the original subscale were eliminated during the translation process which may have reduced coverage of the construct's domain. A test of all items of both subscales would have been an ideal comparison of the original construction of the instruments.

Our analyses suggest that correlational analyses which examine the similarities and differences of impact of the cohesion subscale on positive and negative developmental outcomes would be appropriate. The examination of equivalence in this study suggest that there would be a good deal of confidence that group similarities and differences found in subsequent analyses involving cohesion would likely not be artifacts of measurement.

Given the evidence for configural invariance for the cohesion subscale, it was appropriate to proceed to the next step of examining scalar equivalence. The results, however, suggest that the same score on the cohesion subscale did not represent the same degree, intensity, or magnitude of the construct across groups. While this scale is not a diagnostic tool (e.g., where a specified score indicates a state or problem), our finding suggests that analyses involving a mean score would be inappropriate. Many studies that utilize diagnostic measures frequently omit the steps outlined in this study. Further, studies that utilize normed measures should be cautious in assuming that particular criterion scores are applicable across groups without first examining scalar equivalence.

Knight and Hill (1998) state that heterogeneity across groups may complicate the evaluation of the cross-ethnic equivalence of measures. In the case of our sample, the European American group had greater heterogeneity on variables such as income and education and may have a greater likelihood of in-group heterogeneity on other dimensions in contrast to the other two groups (McLoyd & Steinberg, 1998). Cauce, Coronado, and Watson (1998) commented that ethnic group differences often decrease and or disappear once socioeconomic status has been statistically or experimentally controlled. When differences were found, they were typically between the European American and Cambodian mothers. Hence, limitations of the present study include the restricted range of the two Southeast Asian samples on these variables which prevents disaggregating the effects of ethnicity and social class.

Lastly, though this study focused on the fundamental and mechanical issues of equivalence, one cannot forget the need for a theoretical framework to hypothesize possible group differences. For example, would one expect a similar demonstration of cohesion across the

three groups? One might predict a higher level of family cohesion among the two Asian sub-groups on the basis of greater interdependent emphasis among collectivist cultures (Triandis, 1993). However, given the level of higher rates of single parent households and psychiatric disorders related to war trauma among Cambodian Americans (Cahn & Stansell, 2005; Marshall, Schell, Elliott, Berthold, & Chun 2005), one might hypothesize that family cohesion has been impacted. Our results found that Vietnamese mothers expressed the highest means for Factor 1 of the cohesion subscale while they appeared to be similar to European mothers on Factor 2. On both subscales, Cambodian mothers reported the lowest means. Subsequent research may investigate whether household composition or disorder related to trauma are in fact related to these mean differences. Thus, similar attention that was given by this study to the mechanics of equivalence must be given in subsequent research to interpret possible group etiological differences on the basis of theory and substantive understanding of the target groups.

Acknowledgments

References

Bentler PM. Comparative fix indexes in structural models. Psychological Bulletin. 1990; 107:238–246. [PubMed: 2320703]

Bentler, PM. EQS: Structural equations program manual. Los Angeles, CA: BMDP Statistical Software; 1993.

Byrne, BM. Structural equation modeling with EQS and EQS-Windows: Basic concepts, applications, and programming. Thousand Oaks, CA: Sage Publications; 1994.

Byrne BM, Shavelson RJ, Muthen B. Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. Psychological Bulletin. 1989; 105:456–466.

From Refugee to Deportee: How U.S. Immigration Law Failed the Cambodian Community. In: Cahn, D., editor; Barrett, KH.; George, WH., editors. Race, culture, psychology, and law. Thousand Oaks, CA, US: Sage Publications, Inc; 2005. 2005. p. 237-254.p. xvii-478.

Catalano RF, Hawkins JD, Krenz C, Gillmore M, Morrison D, Wells E, et al. Using research to guide culturally appropriate drug abuse prevention. Journal of Consulting and Clinical Psychology. 1993; 61:804–811. [PubMed: 8245277]

Cauce, AM.; Coronado, N.; Watson, J. Conceptual, methodological, and statistical issues in culturally competent research. In: Hernandez, M.; Isaacs, R., editors. Promoting cultural competence in children's mental health services. Baltimore, MD: Paul H. Brookes; 1998. p. 305-329.

Forsyth, BH.; Lessler, JT.; Hubbard, ML. Cognitive evaluation of the questionnaire. In: Turner, CE.; Lessler, JT.; Gfroerer, JC., editors. Survey measurement of drug use: Methodological studies. Rockville, MD: National Institute on Drug Abuse; 1992. DHHS Publication No. ADM 92–1929, pp. 13–52

Fuller, TD.; Edwards, JN.; Vorakitphokatorn, S.; Sermsri, S. Using focus groups to adapt survey instruments to new populations: Experience from a developing country. In: Morgan, DL., editor. Successful focus groups: Advancing the state of the art. Vol. 156. Thousand Oaks, CA: Sage Publications; 1993. p. 89-104.

Gottfredson DC, Koper CS. Race and sex differences in the measurement of risk for drug use. Journal of Quantitative Criminology. 1997; 13:325–347.

Griffin KW, Scheier LM, Botvin GJ, Diaz T, Miller N. Interpersonal aggression in urban minority youth: Mediators of perceived neighbhorhood, peer, and parental influences. Journal of Community Psychology. 1999; 27:281–298.

Harkness, JA.; Mohler, PPh; Van de Vijver, FJR. Comparative research. In: Harkness, JA.; Van de Vijver, FJR.; Mohler, PPh, editors. Cross-cultural survey methods. Hoboken, NJ: John Wiley & Sons; 2003. p. 3-16.

Harkness, JA.; Van De Vijver, FJR.; Mohler, PPH. Cross-cultural survey methods. Hoboken, NJ: John Wiley & Sons; 2003.

Hobbs, E.; Stoops, N. Demographic trends in the 20th century: Census 2000 special reports. Washington, DC: U.S. Government Printing Office; 2002. Series CENSR-4

Horn JL, McArdle JJ. A practical and theoretical guide to measurement invariance in aging research. Experimental Aging Research. 1992; 18:117–144. [PubMed: 1459160]

Hu L, Bentler PM. Fitindices in covariance structure modeling: Sensitivity to underparameterized model misspecification. Psychological Methods. 1998; 3:424–453.

Huebner A, Howell LW. Examining the relationship between adolescent sexual risk-taking and perceptions of monitoring, communication, and parenting styles. Journal of Adolescent Health. 2003; 33:71–78. [PubMed: 12890597]

Hui CH, Triandis HC. Measurement in cross-cultural psychology—a review and comparison of strategies. Journal of Cross-Cultural Psychology. 1985; 16:131–152.

Jobe JB, Mingay DJ. Cognitive laboratory approach to designing questionnaires for surveys of the elderly. Public Health Reports. 1990; 105:518–524. [PubMed: 2120731]

Kline, RB. Principles and practice of structural equation modeling. NY: Guilford Press; 1998.

Knight, GP.; Hill, NE. Measurement equivalence in research involving minority adolescents. In: McLoyd, VC.; Steinberg, L., editors. Studying minority adolescents: Conceptual, methodological, and theoretical issues. Mahwah, NJ: Erlbaum; 1998. p. 183-210.

Knight GP, Tein JY, Shell R, Roosa M. The cross-ethnic equivalence of parenting and family interaction measures among Hispanic and Anglo-American families. Child Development. 1992; 63:1392–1403. [PubMed: 1446558]

MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. Psychological Methods. 1996; 1:130–149.

Marshall GN, Schell TL, Elliott MN, Berthold SM, Chun CA. Mental health of Cambodian refugees 2 decades after resettlement in the United States. JAMA: Journal of the American Medical Association. 2005; 294:571–579.

McLoyd, VC.; Steinberg, L., editors. Studying minority adolescents. Mahwah, NJ: Lawrence Erlbaum Associates; 1998.

McLoyd, VC. Changing demographics in the American population: Implications for research on minority children and adolescents. In: McLoyd, VC.; Steinberg, L., editors. Studying minority adolescents: Conceptual, methodological, and theoretical issues. Mahwah, NJ: Erlbaum; 1998. p. 3-28.

Olson, DH.; Portner, J.; Bell, R. Family adaptability and cohesion evaluation scales. St. Paul: University of Minnesota Press; 1982.

O'Neil R, Parke RD, McDoweU DJ. Objective and subjective features of children's neighborhoods: Relations to parental regulatory strategies and children's social competence. Applied Developmental Psychology. 2001; 22:135–155.

Pons DA, FL, Balaguer I, Garcia Merita ML. Satisfaction with Life Scale: Analysis of factorial invariance for adolescents and elderly persons. Perceptual and Motor Skills. 2000; 91:62–68. [PubMed: 11011872]

Robinson CC, Mandleco BF, Olsen S, Hart CH. Authoritative, authoritarian, and permissive parenting practices: Development of a new measure. Psychological Reports. 1995; 77:819–830.

Rosay AB, Gottfredson DC, Armstrong TA, Harmon MA. Invariance of measures of prevention program effectiveness: A replication. Journal of Quantitative Criminology. 2000; 16:341–367.

Steenkamp JBEM, Baumgartner H. Assessing measurement invariance in cross-national consumer research. Journal of Consumer Research. 1998; 25:78–90.

Steiger, JH.; Lind, J. Statistically based tests for the number of common factors. Paper presented at the annual meeting of the Psychometric Society; Iowa City, IA. 1980.

Triandis HC. Collectivism and individualism as cultural syndromes. Cross-Cultural Psychology. 1993; 27:155–180.

U. S. Department of Education. Inside No Child Left Behind. Title IV–21st Century Schools, Part A– Safe and Drug-Free Schools and Communities. 2002. Retrieved the World Wide Web: http://www.ed.gov/legislation/ESEA02/pg51.html

Van de Vijver, FJR.; Leung, K. Methods and data analysis for cross-cultural research. Newbury Park, CA: Sage Publications; 1997.

Vandenberg RJ. Toward a further understanding of and improvement in measurement invariance methods and procedures. Organizational Research Methods. 2002; 5:139–158.

Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. Organizational Research Methods. 2000; 3:4–69.

Wu P, Robinson CC, Yang C, Hart CH, Olsen SF, Porter CL, Jin S, Wo J, Wu X. Similarities and differences in mothers' parenting of preschoolers in China and the United States. International Journal of Behavioral Development. 2002; 26:481–491.

Yu, ESH. Studying Vietnamese refugees: Methodological lessons in transcultural research. In: Owan, TC., editor. Southeast Asian mental health: Treatment, prevention, services, training and research. Washington, DC: U.S. Government Printing Office; 1985. DHHS Publication No. ADM 85–1399, pp. 517-541
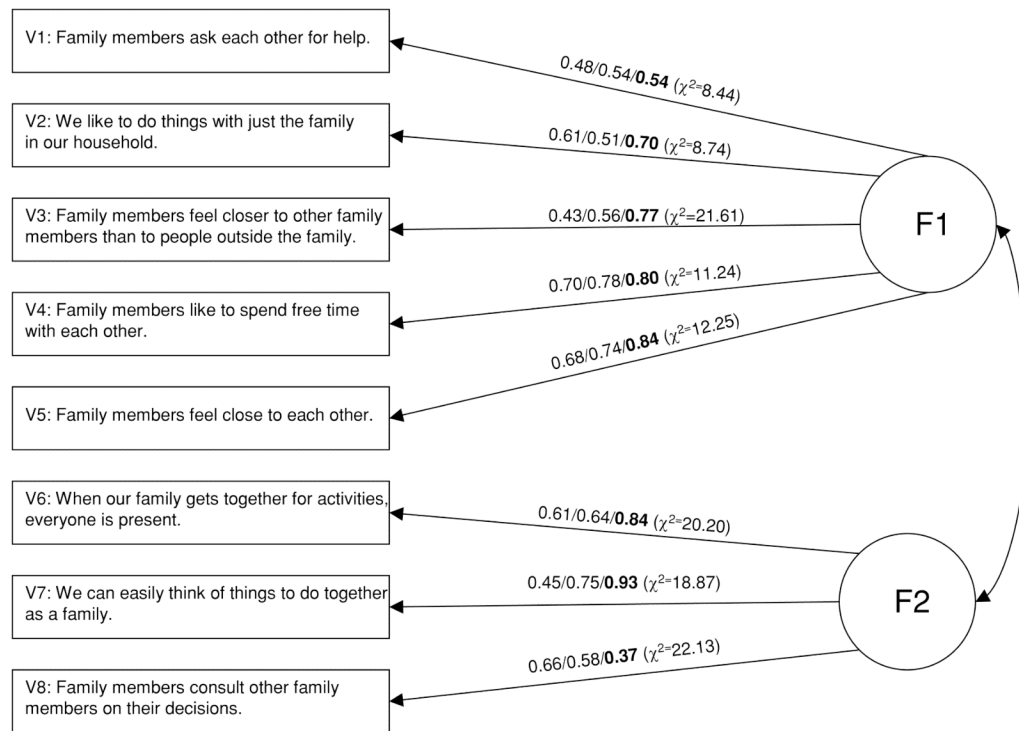
Fig. 1.

Cohesion Scale: Standardized Factor Loadings of Measurement Model. Note: Standardized factor loadings are shown for European American, Vietnamese immigrant and Cambodian immigrant mothers, respectively. All factor loadings were statistically significant ($p>0.05$). Constraints were placed on all factor loadings. Significant differences ($p>0.05$) between the reference group (European Americans) and the comparison group are indicated by **bold** and χ^2 difference from LM test
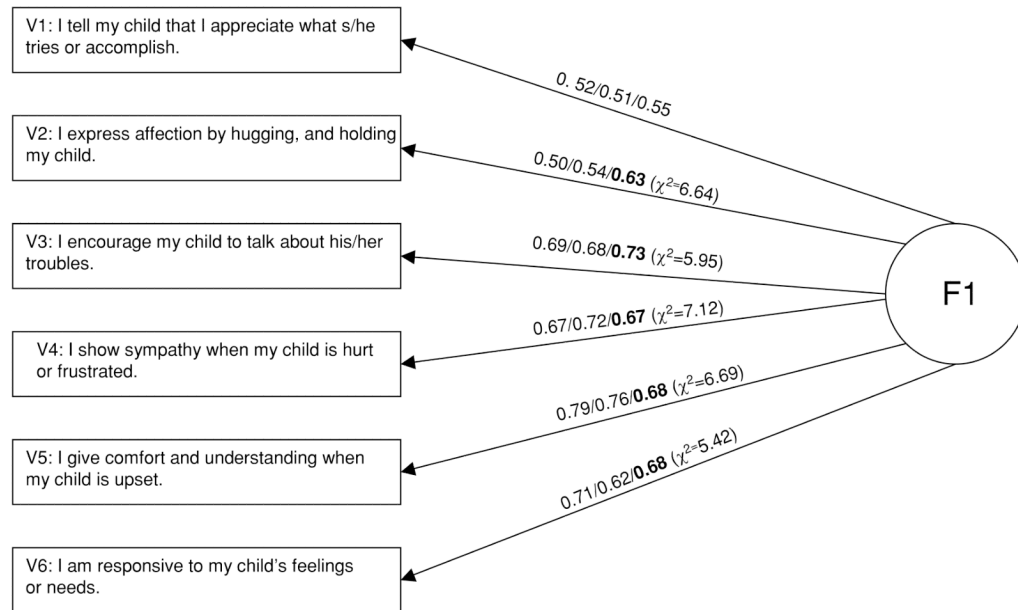
Fig. 2.
Warmth & Involvement Scale: Standardized Factor Loadings of Measurement Model. Note: Standardized factor loadings are shown for European American, Vietnamese immigrant and Cambodian immigrant mothers, respectively. All factor loadings were statistically significant (p>0.05). Constraints were placed on all factor loadings. Significant differences (*p*>0.05) between the reference group (European Americans) and the comparison group are indicated by **bold** and χ^2 difference from LM test