



Date: June 13, 2014

To: HCHS/SOL Steering Committee, Publications Committee, and Statistical Committee

From: Psychosocial/Sociocultural Committee (Linda Gallo & Frank Penedo, Chairs);
Measurement/Psychometrics Working Group (Linda Gallo, Patricia Gonzalez, Maria Llabre, Frank Penedo, Scott Roesch, Vanessa Malcarne)

Re: Measurement Issues in HCHS/SOL Sociocultural/Psychosocial Scales

Introduction. When making inferences about a construct from scores obtained on any measurement instrument, the psychometric properties of the instrument are very important. For example, regression models assume that the predictors included are measured without error. Measurement error in a predictor is known to attenuate the influence of that predictor in a simple regression equation. Thus, the reliability of a measurement instrument is important. Validity is also important because it influences the meaning we can ascribe to a score on a measure. There are multiple aspects to validity, conceptually unified under construct validation (Messick, 1995). The structural aspect of validity refers to the correspondence between the scoring of the measurement instrument and the hypothesized constructs and involves examination at the level of the items. The extent of the evidence required for demonstrating validity depends on the types of inferences one wishes to make from the obtained scores, but at a minimum, the structure of the measure should be established. When using measurement instruments that have been translated into different languages, an added concern related to validity is the extent to which the measurement instrument has the same structure across languages. This is referred to as measurement invariance.

Measurement invariance refers to the ability of a measurement instrument to yield scores that can be interpreted in a similar fashion across different populations or groups, in our case different languages. There is an extensive literature describing how invariance can be established and confirmatory factor analysis (CFA) plays a prominent role in that literature. Using CFA one can assess several degrees or levels of invariance. Configural invariance exists when the number of factors and the items that load on each factor are the same across language. This is the most basic level of invariance. Metric invariance assumes configural invariance, but in addition, requires the relations between the factors and their items or indicators (factor loadings) to be the same across language. Metric invariance insures that relations between the construct and other external variables are comparable across language because the units would be the same. Scalar invariance refers to the case when the item intercepts when relating the factor to the items are the same across language. Scalar invariance assures that level or mean comparisons between languages are interpretable because the origin of the scale would be the same for both languages. With respect to metric and scalar invariance, often only partial invariance can be established. Partial invariance refers to situations where some but not all items are invariant. While there are other parameters that may be compared across different groups or populations, configuration, metric, and scaling are the basic invariance properties typically tested.

Psychometric studies. As a supplement to the Sociocultural Ancillary Study in HCHS/SOL, psychometric studies were conducted on several of the measures used to assess psychological or sociocultural constructs in the HCHS/SOL baseline and in the Sociocultural Ancillary Study. These studies assessed the internal consistency reliability of the measures, as is commonly done. Additionally, these studies also examined the structural validity of the measures using CFA, as well as the factorial invariance between the Spanish and English versions of the measures. This last examination was important because in many cases, data analysis is based on samples that are mixed with respect to the language in which participants responded to the items. These studies typically followed the following general procedures: 1. Items were assigned to scales according to the recommended scoring procedure. 2. Internal consistency was estimated for each subscale within language and for the total sample. 3. CFA was used to determine the fit of the model to the data, where the model was based on the recommended scoring procedure. 4. In the event of poor model fit, alternative models were tested. For example, if a scale is scored with a single total score, but a single factor model did not fit the data, a 2-factor solution was investigated. 5. Once a factorial structure was established, configural invariance was assessed by comparing the factor structure across languages. 6. Once configural invariance was established, metric invariance was tested by comparing the factor loadings across language. 7. In the event of lack of metric invariance, partial metric invariance was assessed. 8. Once metric invariance (full or partial) is established, scalar invariance was assessed by comparing the intercepts across language. 9. In the event of lack of scalar invariance, partial scalar invariance was assessed. Although there is an extensive literature on how to conduct invariance studies using CFA (see for example, Dimitrov, 2010), there is little if any guidance with respect to the minimum level of invariance that is deemed acceptable. In the HCHS/SOL psychometric studies we will follow the following criteria in establishing various levels of invariance. We are using the Root Mean Squared Error of Approximation (RMSEA) and Standardized Residuals (SRMR), both descriptive indices, as the gold standard chi-squared test might be highly sensitive given the sample size of the HCHS/SOL.

1. Configural invariance will be declared when the model with the same number of factors, each with the same items in both English and Spanish fits the data according to descriptive criteria for approximate fit (i.e., $RMSEA \leq .08$ and $SRMR \leq .08$; Hu & Bentler).
2. Metric invariance (invariance in an item loading) will be declared when RMSEA and SRMR differ by .015 or less between a model with item loading(s) constrained equal and a model with item loading(s) estimated freely.
3. Invariance in an item intercept will be declared when RMSEA and SRMR differ by .015 or less between a model with item intercept(s) constrained equal and a model with item intercept(s) estimated freely.
4. Partial metric invariance will be deemed acceptable when over 60% of the items are invariant.
5. Partial scalar invariance will be deemed acceptable when over 60% of the items are invariant.

Minimal standards recommended

1. We are recommending that all instruments used in analyses meet minimum psychometric standards. These include configural invariance across English and Spanish versions, and internal consistency estimates of the resulting subscales of .7 or higher, unless the subscale consists of 6 items or fewer, or if a binary response format is used. In this case lower estimates (in the .5 to .6 range) may be considered.
2. When configural invariance cannot be established, we do not recommend combining English and Spanish responses into one analysis.

References

- Dimitrov, D.M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development, 43*, 121-149.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.

Results from completed analyses

The table below indicates the results of studies evaluating the minimal standards for internal consistency and configural invariance across language.

Instrument	Internal consistency (Cronbach's alpha)	Language Invariance	Recommendation
Analyses performed by/field center			
ISEL-12 (perceived social support) Erin Merz/San Diego	Total Score (Yes) (E=.86, S=.80) Subscales (No)	Total Score (Yes)	Use total score Do not use subscales
CES-D 10* (depression) Patty Gonzalez/San Diego	Total Score (Yes) (E=.82, S=.82)	Total Score (Yes)	Use total score
Familism (6-item version)* Scott Roesch/San Diego	No (E=.57, S=.46)	No	Do not use
Familism (original scale) Scott Roesch/San Diego	Total Score (No) (E=.74, S=.79) Subscales (Yes) Obligations 6 items (E=.68, S=.68) Support 3 items (E=.74, S=.59) Referents 5 items (E=.59, S=.68)	Total Score (No) Subscales (No)	Use subscale scores only Note. For additional information please refer to the Familism memo. We recommend investigators note the relatively low alphas as a limitation in their manuscripts.
Perceived Stress	Total (Yes) (E=.86, S=.84)	Yes	Use total score

Instrument Analyses performed by/field center	Internal consistency (Cronbach's alpha)	Language Invariance	Recommendation
Scale Scott Roesch/San Diego			
Simpatia Scott Roesch/San Diego	Total (Yes) (E=.75, S=.76)	Yes	Use total score
Brief PEDQ-CV (discrimination) Scott Roesch/San Diego	Total Score (Yes) (E=.91, S=.87) Subscales (Yes) Exclusion/Rejection (E=.78, S=.76) Stigmatization/Evaluation (E=.82, S=.69) Work/School Discrimination (E=.79, S=.71) Threat/Aggression (E=.82, S=.72)	Yes	Use total score and/or subscale scores as desired given research question Scoring should be as follows: Total score (dcea1- dcea17) Exclusion subscale (dcea8, dcea11-dcea13) Stigma subscale (dcea5, dcea10, dcea15-17) Work/School Discrimination (dcea1, dcea2, dcea9, dcea14) Threat/Aggression (dcea3, dcea4, dcea6, dcea7) (Scoring revised from original) (Note, syntax/scoring in master SCAS file is

Instrument Analyses performed by/field center	Internal consistency (Cronbach's alpha)	Language Invariance	Recommendation
			incorrect)
Marianismo (Female Gender Roles) Patty Gonzalez, Alicia Nunez/San Diego	Subscales (Yes) Family Pillar (E=.78, S=.79) Virtuous & Chaste (E=.79, S=.81) Subordinate to Others (E=.76, S=.80) Silencing to Maintain Harmony (E=.82, S=.82) Spiritual Pillar (E=.79, S=.79)	Subscales (Yes)	Use 5 subscale scores
Life Engagement Test (LET) Carrie Brintz/Miami	Total Score (Yes) (E=.80, S=.72) Positive items (E=.63,S=.68) Negative items (E=.74,S=.69)	Total score (yes with correlated residuals) Two factors (Yes)	Use total score
Rosenberg Self-Esteem Carrie Brintz/Miami	Total Score (Yes) (E =.87, S = .79) Positive items (E= .80), S = .77) Negative items (E = .82, S=.73)	Total Score (Yes with correlated residuals) Two factors (Yes)	Use total score
Optimism (LOT) Carrie	Total Score (Yes) (E =.97, S=.89) Optimism (E = .52,	Total Score (yes with correlated residuals)	Use two factors Optimism (IPEA11,

Instrument Analyses performed by/field center	Internal consistency (Cronbach's alpha)	Language Invariance	Recommendation
Brintz/Miami	S=.54) Pessimism (E=.59, S=.54)	Two factors (Yes)	IPEA14, IPEA 19) Pessimism (IPEA13, IPEA16, IPEA18) (New subscale scores need to be created; data dictionary only contains total score)
Spielberger Trait Anxiety Inventory* Patty Gonzalez/ San Diego	Total Score (Yes) (E=.92, S=.94)	Total score (Yes)	Use total score
State Trait Anger Expression Inventory-2 (STAXI-2) Patty Gonzalez/San Diego	Total Score (Yes) (E=.87, S=.85) Subscales (Yes) Angry Temperament (E=.93, S=.95) Angry Reaction (E=.95, S=.94)	Total Score (Yes)	Use total score and/or subscale scores as desired given research question. Note. For purposes of calculating a total score all 10 items should be included. In accordance with the STAXI-2 scoring manual, when planning to include the 2 subscales in analyses, only use 8 items: Angry Temperament (emea21, emea22, emea23, emea26) Angry Reaction (emea24, emea25, emea28, emea30).

Instrument Analyses performed by/field center	Internal consistency (Cronbach's alpha)	Language Invariance	Recommendation
Fatalism (8 items; binary response format) Scott Roesch/San Diego	Total (E = .61, S = .59)	Total Score (Yes)	Use total score
Cook Medley Cynicism Scale (Hostility, 13 items; binary response format) Scott Roesch/San Diego	Total (E = .82, S = .77)	Total Score (Yes)	Use total score
Machismo (Male Gender Roles; derived 8-item scale) Patty Gonzalez, Alicia Nunez/San Diego	Total score (No) Subscales (Yes, based on revised scoring) Traditional Machismo (E = .58, S = .59) Caballerismo (E = .69, S = .73)	Subscales (Yes, based on revised scoring)	Use 2 subscales: Traditional machismo (GNEA1 GNEA3 GNEA4 GNEA5 GNEA10) Caballerismo (GNEA7 GNEA8 GNEA9) Do not include items GNEA2 and GNEA6 (Scoring revised from original)
Family Cohesion Carrie	(E = .75, S = .75)	Total score (Yes)	Use total subscale score – remove item 46 from scoring due to very low

Instrument Analyses performed by/field center	Internal consistency (Cronbach's alpha)	Language Invariance	Recommendation
Brintz/Miami			loading New scale consists of: FES38 FES40r FES42 FES44 FES48 FES50r FES52 FES54 (Scoring revised from original)
Family Conflict Carrie Brintz/Miami	(E = .72, S = .54)	Total Score (Yes)	Use total subscale score – remove item 41 from scoring due to very low loading New scale consists of: FES39 FES43 FES45r FES47 FES49 FES51r FES53 FES55r (Scoring revised from original)
Hispanic Stress Inventory (2 factors or 5 factors) Carrie Brintz/Miami	<u>Two Subscales</u> Intrafamilial (8 items) E = .71 , S = .74 Extrafamilial (9 items) E = .72, S = .85 <u>Five subscales</u> Occupational (5 items) E = .69 , S = .78 Parental (4 items) E = .63, S = .63	Total subscale scores (Yes) Total subscale scores (Yes)	Use the total scores of two subscales or the total scores of 5 subscales. Do not use the immigration subscale for English speakers due to low internal consistency. <i>Note:</i> The marital subscale, which is composed of only 1 item (item 2), fits very well in the Parental subscale (psychometrically and

Instrument Analyses performed by/field center	Internal consistency (Cronbach's alpha)	Language Invariance	Recommendation
	Immigration (4 items) E = .33 , S = .77 Familial (3 items) E = .74, S = .75 Marital (1 item) Cannot calculate		conceptually), with comparable model fit and configural language invariance. Could use a 4-factor solution with item 2 in Parental.
DUREL Two Factors: Intrinsic Religiosity subscale (3 items) Religious Behavior Subscale (2 items) Carrie Brintz/Miami	Intrinsic religiosity total score – 3 items (Yes) (E = .86, S = .88) Religious behavior total score – 2 items (E = .68, S = .60)	Subscales (Yes)	Use total score for intrinsic religiosity subscale. Commonly, the attendance and private activity items are examined as separate 1 item subscales, but these analyses suggest they can also be used to calculate one 2-item subscale.
FACIT (spirituality) Original Scoring – due to copyright agreement, use this scoring until further notice Carrie Brintz/Miami	Total score and subscales (yes) Total score 23 items (E = .92, S = .92) Meaning/Peace 8 items (E=.82, S=.82) Faith 4 items (E=.87, S=.89)	Total score and subscales (Yes)	Use total score (23 items) or 2 subscales (8 items and 4 items)
Neighborhood Social Cohesion Gigi Lopez/San Diego	Total score (5 items) (E=.70, S=.69)	Total score (Yes)	Use total score

Instrument Analyses performed by/field center	Internal consistency (Cronbach's alpha)	Language Invariance	Recommendation
Neighborhood Problems Gigi Lopez/San Diego	Total Score (7 items) (E=.79, S =.78)	Total score (Yes)	Total score (yes)
SASH (Acculturation) Patty Gonzalez, Alicia Nunez/San Diego	Total Score (No) Subscales (Yes) Language (6 items) (E=.80, S =.85) Socialization (4 items) (E= .65, S = .71)	Subscales (Yes)	Use 2 subscales (10 items) SCEA6 fits well in the SASH Language Use subscale (SCEA1-SCEA6) SASH Ethnic Social Relations subscale (SCEA7-SCEA10) Note. Results support the use of the 2 subscales. Results do not support the use of a total score.
Scale of Ethnic Experience Vanessa Malcarne, Erin Merz/San Diego			Use of measure (scoring recommendations) on hold until psychometric analyses are completed.

*Analyses conducted in HCHS/SOL baseline cohort (all other analyses conducted in Sociocultural Cohort)

Note: Recommendations shown in bold indicate that we are suggesting revisions from the original scoring.

List of psychometric/measurement manuscripts – completed

ISEL/perceived social support [Merz et al. (2013). Psychological Assessment) – San Diego]

List of psychometric/measurement manuscripts - in progress

CES-D 10/depression (Gonzalez et al. – San Diego)

Speilberger Trait Anxiety (Gonzalez et al. – San Diego)

Familism (Campos et al. – San Diego)

Marianismo (Castillo et al. – San Diego)

Scale of Ethnic Experience (Malcarne et al. – San Diego)

FACIT/spirituality (Brintz et al. - Miami)

SASH/acclulturation (Navas-Nacher et al. - Chicago)

Measurement analyses still to be conducted

Scale of Ethnic Experience/ethnic identity (to be conducted at San Diego/Vanessa Malcarne)