

What Are We Measuring? An Evaluation of the CES-D Across Race/Ethnicity and Immigrant Generation*

KRISTA M. PERREIRA, *University of North Carolina at Chapel Hill*
NATALIA DEEB-SOSSA, *University of North Carolina at Chapel Hill*
KATHLEEN MULLAN HARRIS, *University of North Carolina at Chapel Hill*
KENNETH BOLLEN, *University of North Carolina at Chapel Hill*

Abstract

The sociological study of the mental health of racial-ethnic minorities depends on the measurement quality of the instruments used to evaluate mental health. A commonly used instrument in research on mental health disparities, the Center for Epidemiologic Studies Depression Scale (CES-D), has not been thoroughly validated for use in the multiethnic and foreign-born populations currently living in the U.S. Using data from the National Longitudinal Study of Adolescent Health, this analysis provides the first multiethnic evaluation and psychometric analysis of the CES-D by acculturation level among youth ages 12–20. Correcting for the measurement problems contained in the CES-D improves the ability to detect differences in depression across ethnocultural groups, and to identify relationships between depression and other outcomes.

* We are grateful to Bernadette Gray-Little, participants in the Add Health Users Workshop and anonymous reviewers for comments and suggestions that contributed significantly to this paper. We are especially thankful for the Russell Sage Foundation for support of this research. This research uses data from Add Health, a project designed by J. Richard Udry, Peter S. Bearman and Kathleen Mullan Harris, and funded by grant P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is due to Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Please direct correspondence to Krista M. Perreira, Department of Public Policy, University of North Carolina at Chapel Hill, Abernethy Hall, CB#3435, Chapel Hill, NC 27599-3435. E-mail: perreira@email.unc.edu.

In 1998, President Bill Clinton's initiative to eliminate racial and ethnic disparities in health, including mental health, became a national priority (Office of Minority Health 2002; U.S. Department of Health and Human Services 1998). Census 2000 data, showing rapid growth in our foreign-born and non-white populations, make this initiative all the more imperative. Hispanic and Asian populations are the fastest-growing segments of American society (U.S. Census Bureau 2003). And by 2050, non-Hispanic whites will no longer be the majority population in the U.S. (U.S. Census Bureau 2000).

To evaluate health disparities in an increasingly multiethnic population, social and health scientists require measurement instruments that provide culturally equivalent measures of health outcomes. Without such instruments, cross-ethnic comparisons of physical and mental health are problematic (Finch et al. 2002; McGee et al. 1999; Roberts 1990).

Several health measurement instruments used in current U.S. community-based surveys were originally designed to measure health outcomes in a primarily European-American population with native English speaking abilities. These instruments have not been thoroughly validated for use in the multiethnic and foreign-born populations currently living in the U.S. (Flores 2002).

With respect to mental health outcomes, the Center for Epidemiological Studies Depression Scale (CES-D) is one of these key measurement instruments. Developed in 1976 for use in general adult population (aged 18 or older), the standard CES-D is a 20-item self-report scale that measures depressive symptoms (Radloff 1977). Table 1 lists the items used in the CES-D. Although it may be used to identify populations at risk of developing clinical depression or anxiety disorders, it is not intended as a clinical diagnostic tool. Since its introduction, the CES-D has been used to assess depression risks in several populations (e.g., adolescent, elderly, ethnic and clinical populations) for whom it was not originally designed. Some of the national surveys that have used short (i.e., 4 to 10 questions) and long forms (i.e., 20 questions) of the CES-D to measure the risk of depression include: the National Longitudinal Study of Adolescent Health (Add Health), the Health and Retirement Survey (HRS), U.S. National Alcohol Survey, the National Survey of Families and Households (NSFH), General Social Survey (GSS), the National Longitudinal Survey of Youth (NLSY), the National and Hispanic Health and Nutrition Examination Surveys (N-HANES, H-HANES), and The Commonwealth Fund Survey of the Health of Adolescent Girls.

We aim to evaluate the validity of the CES-D as a tool for making cross-cultural comparisons regarding mental health in an adolescent population ages 12–18. In this analysis, we focus on evaluating the measurement equivalence of the CES-D across four racial-ethnic groups (white, Asian, Hispanic and black). Within each racial-ethnic group, we also evaluate differences by immigrant generation (1st, 2nd and 3rd+). Through this systematic analysis of the CES-D we are able to identify a subset of five CES-D items that allow for comparisons in depression risk across the 12 ethnocultural groups in this sample.

**Table 1. Center for Epidemiological Studies-Depression (CES-D)
Scale Questions**

How often was each of the following things true during the past week?
Answers range from 0 (never or rarely) to 3 (most of the time or all of the time).

Negative Affect Items:

- (1) You felt **depressed**.
- (2) You thought your life had been a **failure**.
- (3) You felt **sad**.
- (4) You felt that you could not shake off the **blues**, even with help from your family and your friends.
- (5) You were **bothered** by things that usually don't bother you.
- (*) You felt **life** was not worth living.

Positive Affect Items:

- (6) You felt **hopeful** about the future.
- (7) You were **happy**.
- (8) You **enjoyed** life.
- (9) You felt that you were just as **good** as other people.

Somatic Complaint Items:

- (10) You didn't feel like eating, your **appetite** was poor.
- (11) You had trouble keeping your **mind** on what you were doing.
- (12) You felt that you were too **tired** to do things.
- (13) You felt **fearful**.
- (14) It was hard to get **started** doing things.
- (15) You felt **lonely**.
- (16) You **talked less** than usual.
- (17) I had **crying** spells.
- (18) My **sleep** was restless.

Interpersonal Relations Items:

- (19) People were **unfriendly** to you.
- (20) You felt that people **disliked** you.

* The item "life" is an addition to the CES-D first made by Garrison (1991). The items "crying" and "sleep" in numbers 17 and 18 were excluded from the CES-D contained in the Add Health survey.

To our knowledge, this is the first multiethnic evaluation of the CES-D for use with adolescents of various immigrant generations. Previous analyses have been limited to one or two ethnic groups including Japanese (Kikuzawa 2001), Chinese (Kuo 1976; Ying 1988), Koreans (Noh et al. 1998), American Indians (Chapleski 1997; Curyto et al. 1998;), Russians (Dershem et al. 1996), and Mexican Americans, Puerto Ricans and Cuban Americans (Guarnaccia et al. 1989; Morton et al. 1989). Few analyses have evaluated the psychometric quality of the CES-D across acculturation levels or immigrant generations within an ethnicity (Roberts, Roberts and Chen 1997).

By doing this analysis of the CES-D, we illustrate three methodological issues that should concern all social scientists developing scales for use in making cross-cultural comparisons. First, we demonstrate the importance of conducting multiple-group comparisons on the measurement structure of scales. Second, we illustrate the importance of creating scales from effect indicators (i.e., indicators that are determined by the latent variable or factor that they are presumed to be measuring). Third, we show how a failure to design cross-culturally equivalent measures and to model the measurement structure in analyses can lead to biased and inefficient estimates in statistical models.

In the first section of this article, we discuss the use of the CES-D in the sociology of mental health and illness and review the psychometric literature on the CES-D. We then review measurement theory and the requirements for a cross-culturally valid measure of the risk for depression. In the second section, we provide information on the data, measures and analytic techniques used in this analysis. The third section summarizes our results, and the fourth section concludes.

Background

THE SOCIOLOGY OF MENTAL HEALTH AND ILLNESS

In viewing emotional life as molded by social structures, Marx, Weber, Durkheim and Simmel began a long history of research on psychological distress (Smith-Lovin 1995). Since its development in 1977, the CES-D has become one of the most popular instruments for identifying psychological distress in community-based samples (Link 2002; Schwartz 2002; Vega and Rumbaut 1991). Sociologists have used it, in regards to socio-economic conditions, to identify risk and protective factors for depression as well as the short- and long-term consequences of mental illness for individuals and their families (Cockerham 1989; Mirowsky and Ross 1989).

Much of the research using the CES-D has been conducted on elderly populations, as depressive symptoms begin to increase in midlife and the elderly tend to have more risk factors that are associated with depressive symptoms (Akiyama and Antonucci 2002; Ranzijn and Luszcz 2000; Schieman et al. 2001). Risk factors for the elderly, for example, include impairment in physical functioning, disability, poor self-rated health, low education, lack of social support, retired status and widowhood or living alone (Hughes and Waite 2002; Hybels et al. 2001; Schieman et al. 2001). Among adults in general, risk factors for depression as measured by the CES-D include homelessness, poverty and material hardship, homosexuality, work among single mothers, transitions into a care-giving role, divorce and female gender (LaGory et al. 2001; Lewis et al.

2001; Marks et al. 2002; Petterson and Friel 2001; Sweeney and Horwitz 2001; Wilk 2001). Protective factors are represented commonly by the flipside of risk factors, with social support, network stability, volunteer work and marriage reducing symptoms of depression for adults (Bost et al. 2002; Simon 2002; Thoits and Hewitt 2001). Research that has focused on particular ethnic groups in the U.S. reports social support to be one of the most important protective factors and discrimination and acculturative stress important risk factors for depressive symptoms as measured by the CES-D (Finch et al. 2000; Vega, Kolody and Valle 1986).

The CES-D has also been used in numerous studies of risk and protective factors associated with adolescent or young adult emotional health, with a handful focusing on particular race or ethnic groups. Protective factors for adolescents that have been identified include the family context and in particular, parent-child relationships and parental support, while family risk factors include parental separation and divorce (Harker 2001; Videon 2002). Other family life stressors associated with depressive symptoms in young people include poverty and low socioeconomic status, single-parent household structure, and other negative life events (Gore et al. 1992; Harker 2001; Jackson and Finney 2002; Steinhausen and Metzke 2000). Involvement in certain risk behaviors has also been found to increase the risks of depression in youth, such as violence, especially in intimate adolescent relationships (Hagan and Foster 2001). Research on adolescents in specific ethnic groups finds that cultural orientation towards one's own ethnic group accompanied by low acculturation into U.S. society is associated with depressive symptoms among Asian American youth (Wong 2001).

Much less research has been done on the consequences of depressive symptoms as measured by the CES-D, and almost all this research examines adult populations. For example, depression has been found to be associated with greater alcohol consumption in the short-term, and the combination of alcohol use and depression increases the likelihood of disease and mortality in the long-run (Greenfield et al. 2002). Similarly, depressive symptoms lower individuals' self-reports of quality of life items, such as satisfaction with family life, financial security and recreation (Michalos et al. 2000). Among adolescents, one study found higher levels of emotional distress to be associated with increased sexual risk behaviors and use of alcohol and drugs (Murphy et al. 2001).

In sum, the sociological and broader social science literature has relied extensively on the CES-D as a reliable measure of depressive symptoms, both in the adult and adolescent populations and in multiple race and ethnic groups. In addition, various studies use the CES-D to make racial and ethnic comparisons in the protective and risk factors associated with depressive symptoms, as well as in the effects of depressive symptoms on outcomes. The starting point for such research, however, is establishing the different levels of depressive symptoms according to race and ethnicity.

In their review of the literature on the mental health of racial-ethnic minority groups in the United States, Vega and Rumbaut (1991) find few consistent patterns. Our own review, focusing on ethnic youth, also finds few consistent patterns. Comparisons between African Americans and their European American counterparts find a decreased likelihood of mental health disorders or depressive symptoms relative to European Americans, no difference, and an increased likelihood of mental health disorders and symptoms (Office of the Surgeon General 2001, Siegel 1998). Similarly, while some studies of Hispanic and Asian Americans find that they are more likely to report depressive symptoms than their European American peers (Portes and Rumbaut 2001; Roberts and Sobhan 1992; Tienda and Kleykamp 2000; Ying 1988), others find quite the opposite (Office of the Surgeon General 2001).

Although some of these ethnic differences in mental health may be explained by differences in the gender, age and socio-economic composition of each ethnic group (Office of the Surgeon General 2001), Vega and Rumbaut (1991) emphasize the importance of controlling for acculturation levels within racial-ethnic groups. Studies of immigrant youth have demonstrated that higher levels of acculturation to the U.S. are positively related to psychiatric disorder (Burnam et al. 1987; Portes and Rumbaut 1996; Vega et al. 1998), depressive symptoms (Harker 2001; Harris 1999), and psychological distress (Kaplan and Marks 1990).

Psychometric Evaluations of the CES-D

Given the lack of consistent findings regarding differences in depression between ethnic youth, it is plausible that the scales used for detecting depression are measuring something different within groups being compared. Depression scores in various ethnocultural groups may be biased by response patterns that differ between immigrant and ethnic groups, not because a group actually has more/less disorders or symptoms, but because the group expresses psychopathology in a way not captured by measures developed from a Western conceptualization of depression (Vega and Rumbaut 1991). Previous studies suggest that Latinos and African Americans somatize their emotional/mental health into physical health constructs (Angel and Guarnaccia 1989; Angel and Worobe 1989; Finch 2002; Guarnaccia and Roberts 1990). Asians avoid expressing positive affect. They do not distinguish between emotional and physical well-being and identify a strong interpersonal component to depression (Kim 2002; Marsella et al. 1975; Noh 1998; Ying 1988). And some American Indian cultures have no concept of depression and instead focus on discouragement or demoralization (Roberts 1990).

In addition, recent research suggests that language significantly affects psychopathology in ethnic populations. Feelings reported in a native language may be expressed with more precision and emotion than those expressed in a second language (Cuellar and Roberts 1984; Roberts 1989). In part due to this

risk, the CES-D was originally not recommended for use in bilingual populations (Radloff 1977; Trieman 1975).

Lastly, unfamiliarity with questionnaires, social desirability, extreme responding and acquiescence can produce artificial cross-ethnic differences (Van de Vijver and Leung 2001). Researchers have found that Hispanics tend to show more extreme responses than non-Hispanic whites, and that this tendency decreases with time in the U.S. (Hui and Tiandis 1989; Marin et al. 1992; Ross and Mirkowsky 1984); younger and better educated respondents of a variety of ethnic backgrounds use more midpoint responses (Stening and Everett 1984); and respondents who are older, Mexican or of a lower socio-economic status are more likely to provide affirmative answers to survey questions (Mirowsky and Ross 1984; Ross and Mirowsky 1984).

Measurement Theory and the Use of the CES-D

Use of the CES-D to measure mental health outcomes for different racial, ethnic and immigrant generation groups (i.e., 1st generation, 2nd generation and 3rd+ generations) requires measurement equivalence (also called, invariance) in the CES-D across these groups. There are two main levels of measurement equivalence: equivalence in model *form* and equivalence in model *parameters* (for more detailed expositions of these issues, see Bollen 1989).

At the most basic level, a scale must have the same structural form within groups before meaningful group comparisons can be made. Comparisons of scale values across groups assume that scale indicators correspond to the same basic underlying concepts (i.e., factors or latent variables) in each group being compared. If these underlying concepts are equivalent across groups, then the factor structure is equivalent across groups. If the dimensionality of the scale (i.e. number of factors) differs across groups or if the patterns of fixed, free and constrained parameters differ between groups, then the assumption of equivalence in structural form is violated.

Figure 1 illustrates the four most commonly identified structural forms for the CES-D: a single-factor model, a four-factor model, and two different three-factor models with correlations between each factor. If the CES-D is structurally equivalent across groups, then we would expect to confirm the existence of the same form in each group. In other words, *one* of these diagrams should hold for all groups.

While structural equivalence is necessary for meaningful group comparisons, it is not sufficient. The next level of measurement equivalence requires that parameter estimates in the structural model be equivalent. These parameter estimates include factor loadings, factor co-variances and variances, factor means and intercepts, and error variances and co-variances. The equivalence of each of these sets of parameter estimates can be tested separately. It is up to the researcher

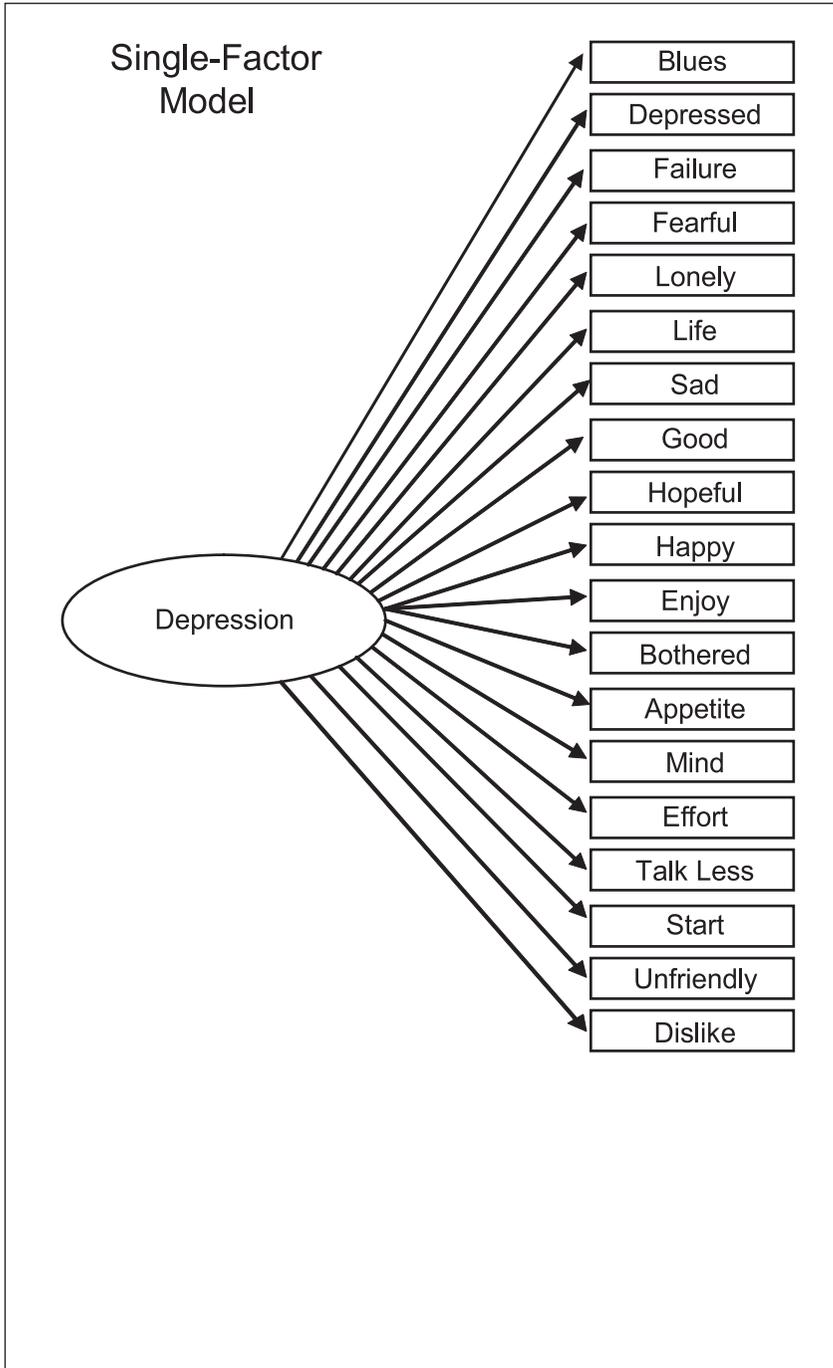
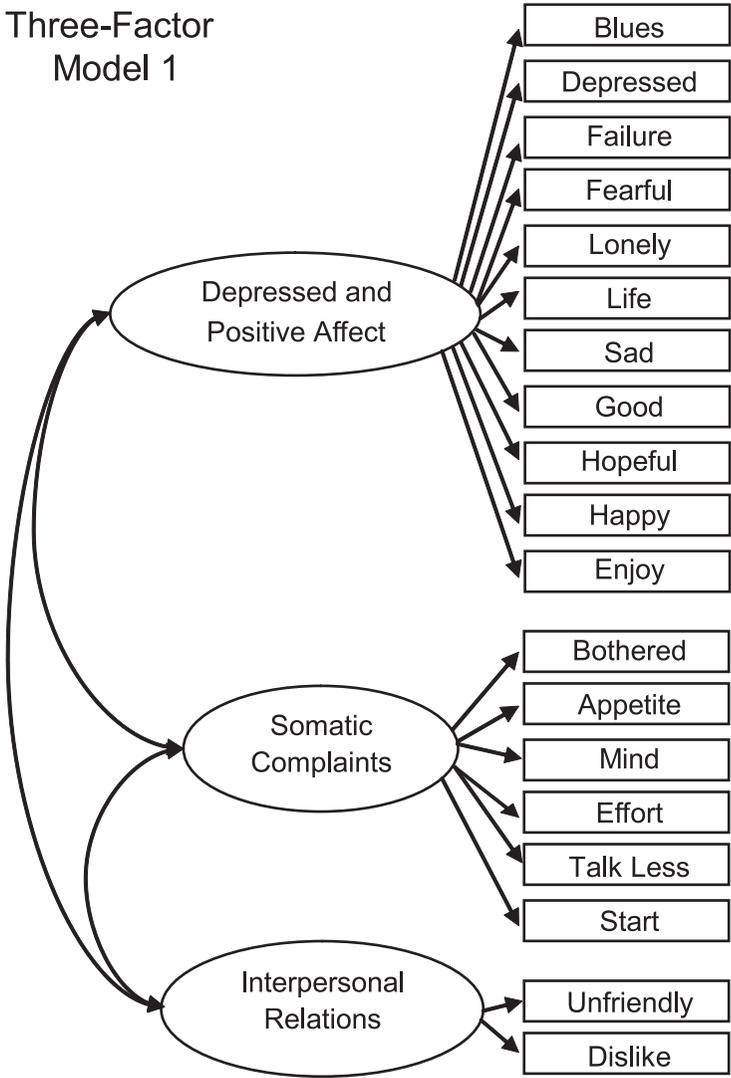
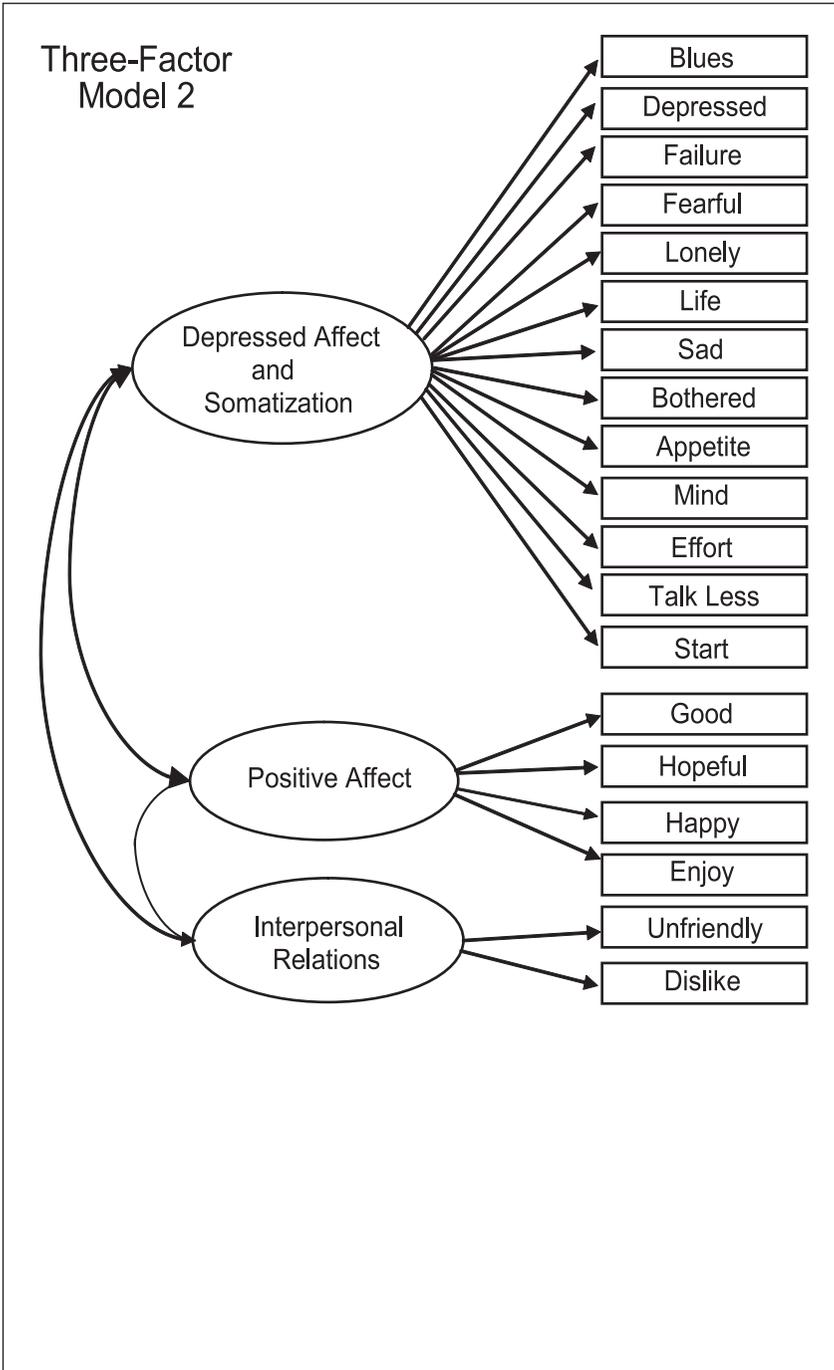
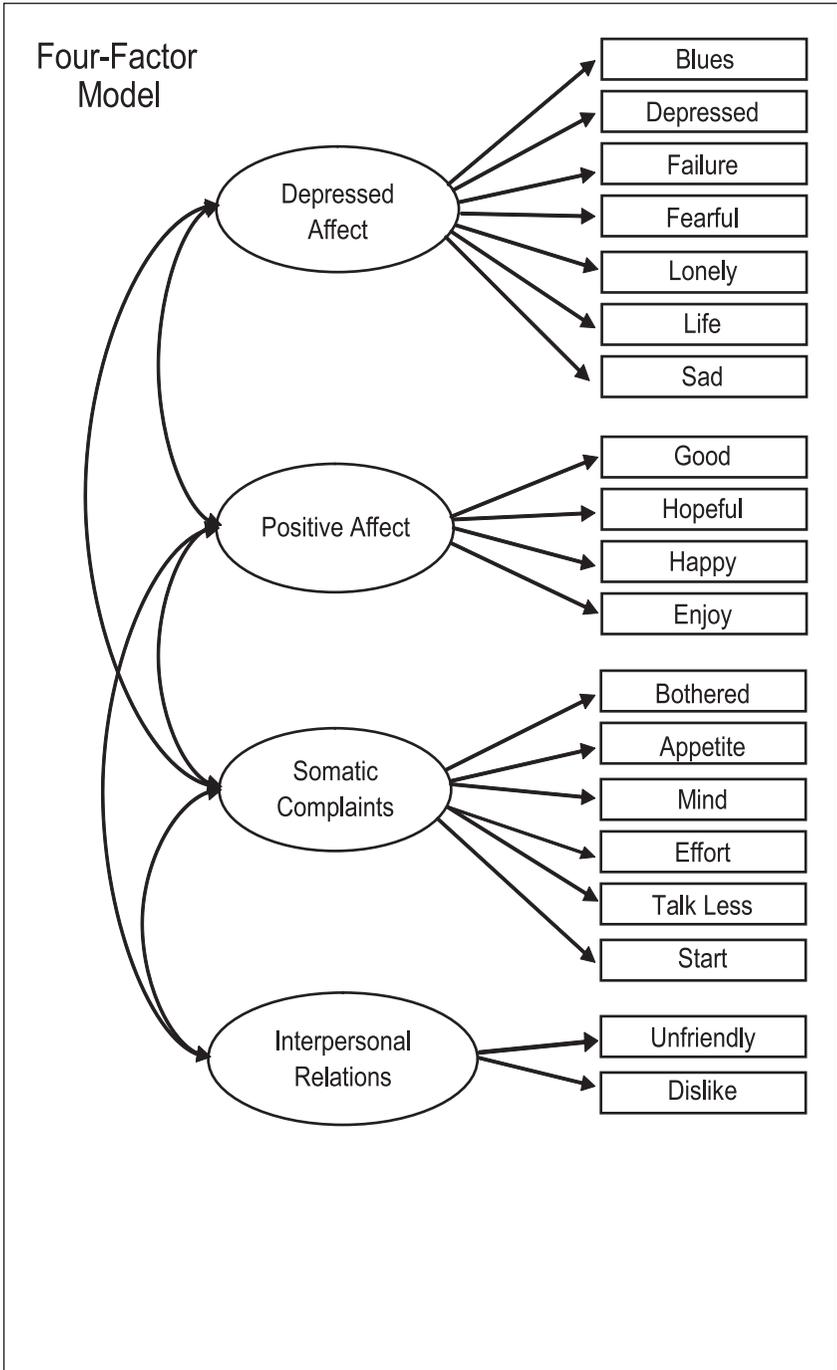


Figure 1. Four Alternative Factor Patterns for the 19-item CES-D

Three-Factor Model 1







to decide which parameter estimates to test for equality across groups and in what hierarchical order. Ultimately, the similarity in parameter values, like the similarity in structural form, is a matter of degree.

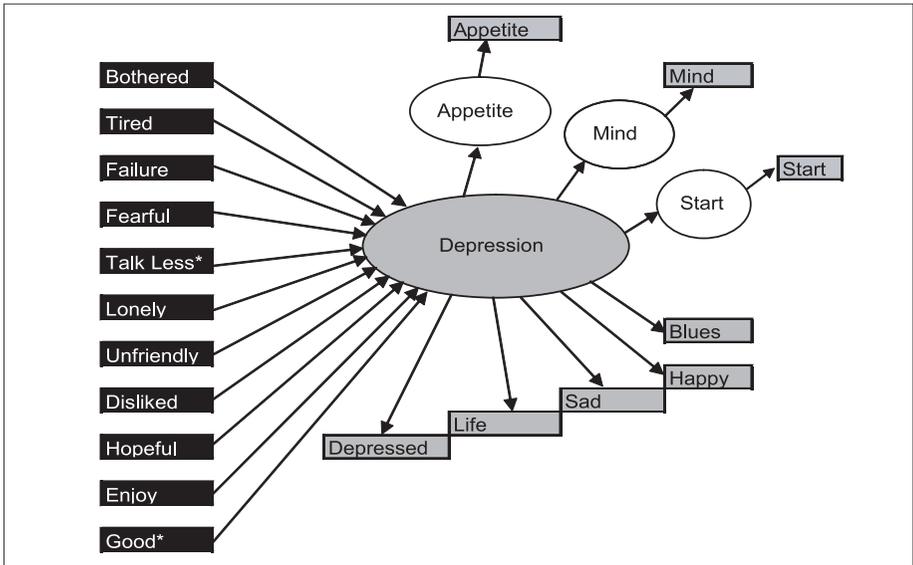
When both the assumptions of structural and parameter equivalence are violated, the scale in question may be measuring different constructs (e.g., discrimination or self-esteem vs. depression) in each group; it is inappropriate to compare group means on the scale, and common statistical procedures such as regressions are no longer meaningful.

When scales are found to lack measurement equivalence across groups, researchers have two options. First, they can disaggregate the construct being measured into more discrete and equivalent sets of underlying concepts. Second, they can include a measurement model in their estimations. By explicitly modeling the measurement structure and allowing it to vary between groups, researchers can obtain unbiased and more efficient estimates of the primary relationships of interest. In this analysis, we use both approaches.

To help us disaggregate the CES-D and identify a cross-cultural equivalent measure of the risk for depression in each group, we focus on the fact that the 19-item CES-D mixes effect indicators, cause indicators, and outcomes (Bollen and Lennox 1991). Techniques for the evaluation of multi-item scales (e.g., reliability estimates) rest on the assumption that all scale items are effect indicators. Effect indicators are *determined by* the latent variable or factor that they are presumed to be measuring and should be positively correlated. Causal indicators, in contrast, are indicators that *determine* the latent variable and may be either positively or negatively correlated with each other. Outcomes are not indicators at all but are consequences of the latent variable.

Figure 2 shows a hypothetical model of the 19-item CES-D with five effect indicators (depressed, life, happy, sad and blues), 11 cause indicators, and three outcomes (appetite, focus and start). As discussed by Bollen and Lennox (1991), "I felt depressed" and "I felt sad" are clearly effect indicators because we expect change in the latent depression variable to lead to a change in the response to these items. However, "I felt lonely" could be a cause indicator because loneliness may cause depression rather than vice versa. Similarly, feeling that people "disliked" you or were "unfriendly" could lead to depression. Appetite, focus and start are hypothesized to be outcomes because persons with depression often exhibit poor appetite, a reduced ability to concentrate or focus on what they are doing, and difficulty starting new projects. However, these outcomes may be associated with various other personality traits and/or physical and mental health problems. They are not exclusively indicators of depression.

To illustrate the consequence of correctly modeling the measurement structure of the CES-D, we evaluate probit estimates of the effects of depression on suicidal ideation using the traditional CES-D scale; an all-effect indicator version of the scale; and a structural equation model that incorporates the measurement model. Results from this analysis are presented in the final section of this paper.



Note: Asterisked items were removed from the Tetrad Analysis due to memory limitations in available computers.

Figure 2. Hypothetical Model of Effect Indicators, Cause Indicators, and Outcomes in the and Outcomes in the CES-D scale

Through these probit analyses, we show that treating variables as effect indicators when they are actually causal indicators leads to model specification error, which can bias parameter estimates and lead to incorrect assessments of the relationships between variables (Bollen and Ting 2000). We can best eliminate this bias by focusing only on effect indicators in the measurement of the CES-D and incorporating the measurement model into our estimation.

Methods

DATA

The National Longitudinal Study of Adolescent Health (Add Health) is a nationally representative study of adolescents in grades 7 through 12 in the United States 1995. The study was designed to explain the causes of adolescent health and health behavior, focusing on the multiple contexts in which young people live. Add Health used a multistage, stratified, school-based, cluster sampling design. Included in the sample were students from 80 high schools (both public and

private) and a corresponding feeder junior high or middle school. While some minority ethnic groups were sampled in proportion to their size within the U.S. population, smaller ethnic groups, including Chinese, Puerto Rican and Cuban youth, were over sampled. This aspect of the Add Health makes possible the analyses we conduct by generation within racial-ethnic group (Harris 1999).

Add Health involves three waves of data collection and several data collection components. For the majority of this analysis, we use data from the in-home interviews conducted during the first wave of data collection. (For details on the Add Health design, see Bearman, Jones, and Udry 1997.) Interviews were completed with 20,745 youths at Wave I and 14,738 youths at Wave II. Due to missing data on sampling weights we excluded 1,782 youths from the analysis. An additional 59 youths were excluded because data on gender, age, ethnicity or immigrant generation were missing; or their primary ethnic identity was American Indian/Native American ($N = 331$). Comparisons across immigrant generations were not appropriate for the American Indian/Native American subsample. Finally, after an analysis of non-response patterns to the 19-item CES-D measure contained in the Add Health survey, we excluded 113 observations for which data were missing on at least one of the CES-D items.¹ The final analytic sample included 18,460 youths ages 12–21.²

MEASURES

The primary variables of interest in this analysis are the 19-items of the CES-D measured in Add Health. Each item has a possible value ranging from 0 to 3 (0 = *rarely or none of the time*, 1 = *some of the time*, 2 = *a lot of the time*, and 3 = *most or all of the time*). As shown in Table 1, the items factor along four dimensions: negative affect, positive affect, somatic complaints and interpersonal relations (Radloff 1977). While some researchers have recommended scoring each dimension of the CES-D separately, most sum across all the items (positively worded items are reverse scored) to create a composite score ranging from 0 to 60 (57 in our case). Higher scores indicate more prevalent depressive symptoms and a correspondingly greater risk of depression. Scores of 24 in females and 22 in males are thought to represent clinical (DSM-defined) cases of depression in adolescents (Garrison 1991; Roberts et al. 1991). In clinical samples, these scores maximize the sensitivity (i.e., true positive rate) and specificity (i.e., true negative rate) of the CES-D instrument. (For a good overview of the properties of the CES-D and other widely-used symptom checklists, see Link and Dohrenwend 1980.)

The CES-D instrument used in Add Health was missing two items: *I had crying spells* and *my sleep was restless*.³ Two items were worded differently than in the original CES-D: *I felt that everything I did was an effort* and *I could not get going*. One item was an addition to the original CES-D: *I felt that life was not*

**Table 2. Sample Size by Race-Ethnicity and Immigrant Generation
(a percentage of total sample)^a**

Race-Ethnicity	Immigrant Generation			Total
	3rd+	2nd	1st	
White	9,332	527	65	9,924 (54%)
Black	3,731	205	77	4,013 (22%)
Asian	174	569	579	1,322 (7%)
Hispanic	1,013	1,418	770	3,201 (17%)
Total	14,250 (77%)	2,719 (15%)	1,491 (8%)	18,460

^a unweighted

worth living. The two changes in wording were in keeping with the development of the CES-D for children (Faulstich et al. 1986; Weissman et al. 1980). The additional item (*I felt that life was not worth living*) was first added to the CES-D by Garrison and colleagues (1991) as an important component of depression in adolescents.

For this analysis, we evaluate the measurement of the CES-D across 12 ethnocultural groups constructed by cross-classifying each adolescent's primary race-ethnicity identification with his/her immigrant generation. Race-ethnicity was defined using the respondent's self-reported ethnic identity (i.e., Hispanic/Latino) and race (i.e., white, black/African American, Asian/Pacific Islander). The final variable included 4 ethnicity-race groups: (1) Hispanic of any race, (2) non-Hispanic Asian, (3) non-Hispanic black, and (4) non-Hispanic white. For immigrants and the children of immigrants, we validated their self-reported ethnic background with country of birth information (Harris 1999). Table 2 presents the distribution on this variable. Of the analytic sample, 54% were white, 22% black, 17% Hispanic, and 7% Asian.

Immigrant generation is defined by a three-category variable signifying that the adolescent is foreign-born to foreign-born parents (1st generation), U.S.-born to foreign-born parents (2nd generation) and U.S.-born adolescents to U.S.-born parents (3rd+ generations or native) (Harris 1999). Adolescents in native-born families (3rd+ generations) may have grandparents or great grandparents who were immigrants, but because the immigration experience is much farther removed from the social context of their childhood and adolescent development, this category is considered the native population and the fundamental comparison group for immigrant children and the children of immigrants. The majority (77%) of the sample was native born (Table 2); and 2nd and 1st first generation immigrants comprised 15% and 8% of the sample, respectively. Reflecting general demographic trends in the U.S., there were few blacks ($N = 77$) and whites

($N=65$) and relatively larger numbers of Hispanics ($N = 770$) and Asians ($N = 579$) within the 1st generation.

For use in our probit models, we also included measures of age, gender (Females = 1) and suicidal ideation. To measure suicidal ideation we combined data from Waves I (1995) and II (1996) of Add Health. Respondents were considered at risk for suicide in 1996 if they indicated that they had “seriously thought about committing suicide” in the 12 months preceding either their Wave I or Wave II interview. Due to missing data on suicidal ideation in either Wave I or II, our sample size for analyses using this variable was reduced to 18,397 adolescents. Within this sample, 18% of the adolescents were at risk for suicide in either 1995 or 1996; 49% were female, and the average age was 16.

STATISTICAL MODELS

We conduct confirmatory factor analyses (CFA) and multiple group analyses (MGA) using a maximum likelihood structural equation program called Mplus Version 2.13 (Muthen and Muthen 2003). Both CFA and MGA analyses require the estimation of measurement models that use structural equations to describe the relationship between latent and observed variables while allowing for measurement error.⁴ In these models, all variables are measured as deviations from their means and, by assumption, the measurement errors are uncorrelated. To adjust for the complex sample design of the Add Health data, we use weights in all analyses and adjust the estimates for clustering at the school-level (Chantala and Tabor 1999).⁵

As with any SEM package, Mplus provides several measures of fit that can be used to evaluate the model. These include a χ^2 value, the Comparative Fit Index (CFI), and the Root Mean Square Error (RMSEA). (For details on these fit statistics, see Bollen 1989:256–289.) We report all three of these statistics. The null hypothesis that the model fits is not rejected if the χ^2 -value is less than expected given the degrees of freedom in the model, the CFI is approximately one, and the RMSEA is approximately zero. At a given level of accuracy, a more parsimonious model is preferable. Unlike the chi-square, the CFI and RMSEA are sensitive to parsimony.

ANALYTIC TECHNIQUES

We begin our analysis by evaluating the dimensionality of the 19-item CES-D through a series of nested confirmatory factor models estimated among 3rd+ generation whites. Following Riddle (2002), four different factor structures are hypothesized (Figure 1).⁶ First, we estimate a single-factor model that represents the use of a composite CES-D score, and that assumes depression can be measured as a global, uni-dimensional construct. Second, we estimate the four-factor model

first identified by Radloff (1977) and repeatedly replicated by others using both exploratory and confirmatory factor analyses (Golding and Aneshensel 1989; Hertzog et al. 1990; Roberts et al. 1989, 1990; Sheehan et al. 1995). Next we estimate two three-factor models that have been identified as superior for some populations (Riddle, Blais and Hess 2002). The first two-factor model treats the negative and positive affect dimensions as opposite poles of a single affect continuum. In the second two-factor solution, the negative affect dimension is combined with the somatic complaints dimension. This dimensionality has been supported by research in ethnic and non-Western communities (Guarnaccia et al. 1989; Iwata and Roberts 1996; Manson et al. 1990; Somervell et al. 1993; Ying 1988).

To make statistical comparisons between a hierarchical pair of estimated CFAs, we set the metric for the latent variables by setting their variances equal to one. We then estimate an unrestricted CFA in which the correlations between latent variables are freely estimated. Finally, we estimate a second CFA in which we constrain the correlations between the latent variables to one. We can reject the null hypotheses that the latent variables are perfectly correlated if the change in χ^2 between the restricted and unrestricted models is significantly large. If we reject the null hypothesis, then we can conclude that the factors measure distinct dimensions of depression and must be treated separately. By making a series of these nested comparisons, we can identify the best fitting factor model for the native (3rd+ generations) white population.⁷

After identifying the best fitting structural form for the measurement model, including its dimensionality in the native white population, we use MGA to evaluate the invariance of this dimensionality across our 12 ethnocultural groups. If we reject the hypothesis of structural invariance, this suggests that, at some probability level, each ethnocultural group manifests depression differently. Moreover, the CES-D is probably measuring something different within the groups being compared and mean values on the CES-D should not be used to make cross-group comparisons regarding the risk of depression.

If we reject the hypothesis of structural invariance, we use regression analysis to explore difference between groups in factor loadings, variance and covariance patterns, and factor means and intercepts. Using the parameter estimates from the MGA for the native white population as the dependent variables, we run three sets of regressions on the parameter estimates for each of the 11 other ethnocultural groups.⁸ The first regression set evaluates the equivalence of the factor loadings and co-variances between native whites and the other subpopulations. The second regression set evaluates the equivalence of the variances and residual variances between the native whites and the other subpopulations. The third regression set evaluates the equivalence of the means and intercepts between the groups. Thus, a total of 33 regressions are estimated and the number of observations in each regression is equal to the number of parameter estimates being compared. For

each regression, we can conclude that the estimated parameters are equivalent if both the intercept in the regression equation is zero and the slope in the regression equation is one.

As discussed earlier, factor analyses assume that all items in a scale are effect indicators. Therefore, one reason for a failure to find structural invariance across groups may be that a scale mixes effect, outcome and cause indicators. We hypothesize that only five of the CES-D items (depression, life, start, happy and blues) are effect indicators and that the remaining items in the scale are either outcomes of depression or cause indicators (Bollen and Lennox 1991). Using the Confirmatory Tetrad Analysis (CTA) designed by Bollen and Ting (2000), we test this hypothesis against the alternative hypothesis, assumed in the CFA, that all items are effect indicators.⁹ If, based on the Confirmatory Tetrad Analysis, we find that the new five-item CES-D is an improvement over the 19-item CES-D, we then repeat the previous analyses and evaluate the structural invariance of the five-item CES-D.

If we do not reject the hypothesis that the structural forms are invariant across groups, we can move down the measurement hierarchy and formally test for equivalence in the parameter estimates of our measurement model. Our formal test for equivalence in the parameters include the following: (1) a test for the equivalence of factor loadings, (2) a test for the equivalence of factor loadings and means, (3) a test for the equivalence of factor loadings, means and intercepts, (4) a test for the equivalence of factor loadings, means, intercepts and variance-covariances, and (5) a test for the equivalence of factor loadings, means, intercepts, variance-covariances and error variances-covariances.¹⁰ We conduct these equivalence tests using MGA and comparing the difference in fit statistics (χ^2 , CFI, and RMSEA) of the more constrained model with equal parameters to the unconstrained model. For several of our analyses, differences in χ^2 -values are significant but the CFI and RMSEA fit statistics suggest that the more constrained model fits well. Because of our large sample size, it is not unusual for the χ^2 -difference to be large despite the fact that the constrained model fits well. This is similar to the problem in regression analyses of large samples where researchers find that coefficients are statistically different from zero although these differences are very small and not particularly meaningful. For this reason, it is more prudent to base our findings on the CFI and RMSEAs.

Lastly, to demonstrate the importance of correcting for measurement inequivalence and measurement error when using the CES-D, we estimate three probit regressions of suicidal ideation on depression. The first model is a standard probit model of suicidal ideation on the 19-item CES-D measure. The second model is a standard probit model of suicidal ideation on the five-item CES-D measure. The third probit model is estimated through structural equations and includes a measurement model of the five-item CES-D. To ensure the comparability of the coefficients on depression in these probit models, the standard deviations of the depression variable are adjusted to equal each other.

Results

EXPLORATORY ANALYSIS

Although we question its appropriateness for this scale¹¹, with Cronbach's alphas ranging from .85 to .89, the 19-item CES-D appears to be highly reliable within each of our 12 ethnocultural groups. Moreover, these reliability estimates are consistent with Cronbach's alpha measured in other samples (Doerfler et al. 1988; Roberts et al. 1990; Schoenbach et al. 1982; Tolor and Murphy 1985). Similarly, mean CES-D scores and case rates (i.e., the percentage of the sample identified with DSM-defined depression) are comparable to the means and rates found in other studies of adolescent youth (Radloff 1991; Roberts et al. 1990; Roberts, Roberts and Chen 1997; Vega and Rumbaut 1991). The mean CES-D score in our sample is 11, and 7.7% of the sample is at risk of clinical depression. A comparison of means across our 12 ethnocultural groups suggests that 3rd+ generation blacks; 1st generation Asians; and 1st, 2nd, and 3rd+ generation Hispanics are at significantly ($p < .05$) greater risk for depression than their white counterparts.¹² Clinical cases of depression were lowest (4–5%) in 1st and 2nd generation blacks and highest (12%) in 2nd generation Hispanics. However, as discussed above, the Cronbach's alpha, means, and case statistics are not appropriately measured unless we find that the CES-D meets the criteria for invariance in structural form and parameters estimated for the measurement model.

Factor Structure Among 3rd+ Generation Whites

The first panel of Table 3 shows the fit indices for the four hypothesized measurement models (i.e., the single-factor model, the three-factor model combining the somatic and negative affect dimensions, the three-factor model combining the negative and positive affect dimensions, and the four-factor model). The CFI value of .91 for the four-factor model is just above the conventional cutoff value of .90 (Bentler and Bonet 1980) and the estimated correlations between the four latent variables ranged between .52 and .84. While the two three-factor models present an improvement over the single-factor model, the chi-square difference test indicates that the four factor model provides a statistically better fit than either the combined negative-positive affect three factor model ($\Delta\chi_{df=3} = 2,420$) or the combined somatic complaints-negative affect model ($\Delta\chi_{df=1} = 319$). Thus, we use the four-factor CFA model as the primary bases for our multiple group comparisons. Complete results for the four-factor model are reported in the online Technical Appendix on the *Social Forces* Web site (<http://socialforces.unc.edu>).

Table 3. Comparisons of Competing Models

Model	N	Chi-sq	df	CFI	RMSEA
CFAs in 3rd+ Generation Whites					
Single-Factor Model	9,332	7,287	152	.804	.071
Three-Factor Model	9,332	5,438	149	.855	.062
Three-Factor Somatization Model	9,332	3,181	149	.917	.047
Four-Factor Model	9,332	2,616	146	.932	.043
MGAs in 12 Ethnocultural Groups					
Four-Factor Model (19-item CES-D)	18,460	9,256	1,752	.908	.053
Single-Factor Model (5-item CES-D)	18,460	192	60	.992	.038
Confirmatory Tetrad Analysis in 3rd+ Generation Whites					
All-Effect Model	9,332	1,907	119	—	—
Cause-Effect-Outcome Model	9,332	1,643	83	—	—
Cause-Effect Model	9,332	1,267	54	—	—
MGA Equivalence Tests for 5-Item CES-D in 12 Ethnocultural Groups					
(1) Factor Loadings (FL)	18,460	349	104	.985	.039
(2) FL + Means (M)	18,460	400	115	.982	.040
(3) FL + M + Intercepts (I)	18,460	605	159	.972	.043
(4) FL + M + I + Variances (V)	18,460	629	170	.971	.042
(5) FL + M + I + V + Error Variances	18,460	1,498	225	.920	.061

Note: The power of the chi-square difference test is larger in big samples than in small ones.

Given the large size of our sample, it is helpful to examine other measures of fit as well. The CFI and RMSEA statistics are less sensitive to sample size and take account of the degrees of freedom in the model. Mplus version 2.13, which was used for this analysis, did not allow estimation to simultaneously account for design effects and the categorical nature of the CES-D. Therefore, all models were estimated with and without clustering and weighting. Our results were robust to the estimation technique.

MULTIPLE GROUP AND REGRESSION ANALYSIS OF THE 19-ITEM CES-D

The second panel in Table 3 shows the fit statistics for the multiple group comparison of the 19-item CES-D. These results reveal that the four-factor measurement model does not fit the data well for the 11 other ethnocultural groups in our sample. We can reject the hypothesis that the structural form of the CES-D is equivalent within each of our ethnocultural groups. Thus, for one or more of our ethnocultural groups, one of the alternative measurement models pictured in Figure 1 may be more appropriate.

As described above, we explore the sources of variation in the measurement models between 3rd+ generation whites and each of the 11 other ethnocultural groups using regression analyses. These results (not shown) indicate that the structural invariance between groups may be driven not only by differences in the dimensionality of the CES-D within each group but also by differences between groups in the loadings, variance-covariance patterns, factor means and factor intercepts. The differences in parameter estimates from the MGA are greatest for Hispanics of all generations, 2nd and 3rd+ generation blacks, and 3rd+ generation Asians. In contrast, the measurement model identified for the 3rd+ generation whites appears to fit 1st and 2nd generation whites well.¹³

CONFIRMATORY TETRAD ANALYSIS

We hypothesize that the reason why the four-factor solution model does not fit the non-white groups well is because the current CES-D mixes effect indicators of depression, causal indicators of depression and outcomes of depression. Because classic measurement theory and factor analyses presume that all indicators are effect indicators, the inclusion of causal and outcome indicators in the CES-D leads to model specification error, which can bias parameter estimates and lead to incorrect assessments of the relationships between variables (Bollen and Lennox 1991). Using Confirmatory Tetrad Analysis (Bollen and Ting 2000) with the native white subsample¹⁴, we tested whether all items in the CES-D were effect indicators (all-effect model); whether only the items “blues,” “depressed,” “happy,” “sad” and “life” are effect indicators and the other items are cause indicators (cause-effect model); and whether the items “appetite,” “focus” and “start” are better treated as outcomes (cause-effect-outcome model).^{15,16}

Because these models are nested within each other, identifying the best fitting model simply requires calculating three chi-square differences. The null hypotheses in these tests are that the cause-effect-outcome model is a better fit than the all-effect model; the cause-effect model is a better fit than the all-effect model; and the cause-effect model is a better fit than the cause-effect-outcome model. As reported in the third panel in Table 3, results from the Confirmatory Tetrad Analysis indicate that the cause-effect model is the best fitting model. Because the inclusion of cause indicators in the CES-D violates classic test theory, we can improve the measurement quality of the CES-D by eliminating these indicators. This leaves us with a new five-item CES-D.

MULTIPLE GROUP ANALYSIS OF THE FIVE-ITEM CES-D

By conducting a new MGA of the five-item CES-D, we evaluated the extent to which our shortened CES-D improved the structural invariance among our 12 ethnocultural groups. As before, the measurement model fitted for the 3rd+ generation whites is set as the comparison group. Complete results for this

model are provided in the online Technical Appendix.¹⁷ All the fit indices from the MGA clearly support the hypotheses that the five-item CES-D is structurally invariant across the groups (see Table 3, panel 3).¹⁸

Since we do not reject the hypothesis that the structural form of the five-item CES-D is invariant across the groups, we move down the measurement hierarchy to evaluate the equivalence of parameters. The last panel of Table 3 provides the results of these equivalence tests. We fail to detect substantively large differences in the parameter estimates (i.e., loadings, covariances, variances, means and intercepts) for all indicators considered together.¹⁹ These results support the conclusion that all parameters estimates are equivalent across the groups. In contrast to the 19-item CES-D, the five-item index of depressive symptoms can be appropriately used to make inter-group comparisons.²⁰

ADDITIONAL PSYCHOMETRIC PROPERTIES OF THE FIVE-ITEM CES-D

Since each of the items in our shortened CES-D scale is an effect indicator, we can appropriately evaluate the internal consistency of the scale by looking at inter-item correlations and standard measures of reliability such as Cronbach's alpha. The Cronbach's alpha for the shortened CES-D is .78 in the full sample. The alpha is somewhat lower, for non-white, first-generation immigrant youth (.74), and blacks (.73).²¹

We can also consider the appropriateness of using the scale items to create a linear composite as a substitute for the latent variable in regressions or other similar analyses (Bollen and Lennox 1991). One common linear composite is an equally weighted sum of the scores on each item. For our short CES-D measure, this would produce a score ranging from 0–15 for each respondent. For our sample, the mean on this score is 2.5. It varies from 2.0 among 1st generation whites to 3.1 among 2nd generation Hispanics. Comparisons of means across groups indicate that black 3rd+ generation youth and Hispanic 2nd and 3rd+ generation youth are at greater risk of depression than their white counterparts. Thus, the pattern of mean differences in the CES-D scores using the five-item scales is similar to the pattern identified using the 19-item scale. While the same general pattern for differences in means across ethnicity and generation was identified using the 19-item and five-item scales, this does not diminish the importance of using a structurally invariant measure of the risk of depression or any other scale. The lack of structural invariance can have substantial effects on mean differences. Moreover, as can be seen in the next section, the measurement structure of a scale can influence the relationships between other independent and dependent variables in multivariate analyses.

Using a composite score poses no problem as long as depression is being evaluated as the dependent variable. However, using a composite score as an independent variable may introduce bias into any estimation unless the squared

correlation between the composite score and the latent variable depression is high (a correlation of at least .90 is recommended by Bollen 1980).

To calculate the squared correlation of the index and its latent variable, we used the formula for reliability given by Bollen (1980):²²

$$\text{reliability} = \frac{(\sum \text{factor} - \text{loading}_i)^2 \text{Var}(\text{latent} - \text{variable})}{(\sum \text{factor} - \text{loading}_i)^2 \text{Var}(\text{latent} - \text{variable}) + \sum \text{Var}(\text{factor} - \text{error}_i)}$$

The reliability of our five-item CES-D is .81. This means that 19% of the variance in the five-item CES-D is attributable to measurement error. The reliability is not high enough so that we can safely say that a researcher can ignore the measurement error. Thus, whenever possible, researchers should follow our example below and incorporate a measurement model into their estimations.

ILLUSTRATIONS OF THE EFFECTS OF BIAS IN THE CES-D

In Table 4, we present the results of our three probit estimations of suicidal ideation. The estimated relationship between suicidal ideation and depression change substantially between each model. The 19-item composite has the second highest coefficient of the three approaches. This is not surprising given our support for the mixed cause-effect model where some of these measures were determinants of depression while others were effects of it. As such the indicators are associated with depression. Adding indicators together results in an index that will correlate with the latent depression variable even if the indicators are not all effects of the construct. There are two main problems with this strategy: (1) even an index of 19 items is not a replacement for the latent variable, and (2) our earlier tests showed that the relation of these 19 indicators to the depression construct is not the same across groups. The former problem means that measurement error is present in the depression measure and will bias the estimates. The latter means that the composite scores are not directly comparable across groups, and this too can distort the estimates.

The estimated relationship between depression and suicidal ideation is almost 20% lower when we use the five-item CES-D in place of the 19-item CES-D. The estimated relationship is over 10% higher when the five-item CES-D is used and we adjust for the measurement error in the scale. These results demonstrate a sizable upward bias in the estimated relationship between depression and suicidal ideation due to the structural inequivalence of the scale across ethnocultural groups. In other words, the full CES-D predicts suicidal ideation better than the five-item form, but it achieves this because it taps constructs other than depression.

This bias not only affects the results on the depression variable but it also affects independent variables included in the model. For example, under the

correctly specified model, we find that the blacks are approximately 50% less likely to report suicidal ideation, whereas Asians are 100% more likely to report suicidal ideation. The effects of being in the 1st generation are reduced 14%; the effects of being female are increased over 50%; and the effects of age are reduced to near zero. These are sizable and substantive changes. Because of the correlation between other independent variables and measurement error in the depression index, the use of standard probit estimation techniques also results in considerable inefficiency in the estimates of other independent variables. The standard errors of the estimates in the SEM model are reduced considerably and significant relationships are more easily detected. Clearly, the structural and measurement problems with the original CES-D have important consequences for the conclusions we make regarding the effects of depression risk and other correlated variables on adolescent well-being.

Discussion

As the diversity of the U.S. population increases, researchers have sought to identify and explain mental health disparities between minority youth and the majority population. Based on this research, mental health providers and policymakers are designing treatments and interventions to eliminate mental health disparities.

The legitimacy of these research findings and the programs and policies developed from them hinges, in part, on the validity of the instruments used to measure mental health outcomes. To allow cross-ethnic or cross-cultural comparisons, survey instruments designed to measure mental health must be evaluated for their psychometric or measurement equivalence within each ethnocultural group.

In this analysis, we have shown that the most commonly used survey instrument to measure mental health, the CES-D, is not psychometrically equivalent across either race-ethnicity or immigrant generation. As a result, research using the CES-D may lead to erroneous conclusions about mental health disparities in the U.S. adolescent population. We have designed an abbreviated, five-item version of the CES-D that can be used for cross-cultural comparisons among adolescents living in the U.S.

The five items included in our abbreviated CES-D (i.e., blues, depressed, happy, sad and life) are all effect indicators that loaded strongly on either the positive affect or negative affect dimensions of the full CES-D. Thus, the abbreviated CES-D has been stripped of causal indicators, outcomes of depression and items related to somatic complaints and interpersonal relationships. Previous research has shown that somatic complaint items and items regarding interpersonal relationships are more likely to be biased by health status, gender, age and race than other items in the CES-D (Cole 2000; Office of the Surgeon General 2002;

Table 4. Probit Models for Effects of Depressive Symptoms on Suicidal Ideation

Variables	<u>Standard Estimation</u>		<u>Standard Estimation</u>		<u>SEM Estimation</u>	
	Betas	(s.e.)	Betas	(s.e.)	Betas	(s.e.)
CES-D 19-Item ^a	.540	(.019)**	—	—	—	—
CES-D 5-Item ^a	—	—	.438	(.017)**	.608	(.008)**
White	—	—	—	—	—	—
Black	-.302	(.052)**	-.241	(.047)**	-.172	(.026)**
Asian	.064	(.084)	.095	(.080)	.131	(.056)*
Hispanic	-.058	(.058)	-.012	(.058)	.050	(.031)
1st Generation	-.409	(.073)**	-.357	(.074)**	-.350	(.052)**
2nd Generation	-.097	(.061)	-.060	(.060)	-.051	(.031)
3rd+ Generation	—	—	—	—	—	—
Female	.204	(.034)**	.182	(.034)**	.316	(.016)**
Age	-.047	(.009)**	-.029	(.008)**	-.006	(.004)
Constant	-1.000	(.147)**	-1.345	(.136)**	—	—
Chi-sq (df = 21)	—	—	—	—	2,798	—
F (8, 121)	140	—	130	—	—	—
Observations	18,397	—	18,397	—	18,397	—

^a To allow for comparisons across models, the 19-item and five-item CES-Ds have been transformed so that their standard deviations equal .896, the standard deviation of the latent variable for depression in the SEM estimation.

* p -value < .05, ** p -value < .01

Vega and Rumbaut 1991). Moreover, these items are more likely to be related to other constructs (e.g., loneliness, discrimination and physical illness) and have less discriminate validity. On a more technical level, the use of causal indicators and outcome indicators in the CES-D violates the assumptions of classical measurement theory. As a result, standard psychometric statistics such as Cronbach's alpha and factor analysis are not valid for the full CES-D.

Previous research on the effects of reducing the length of the CES-D scale shows that, on average, five-item versions of the CES-D are nearly as sensitive and specific as the full 20-item scale (Shrout and Yager 1989). The selection of items for these abbreviated scales is typically based on high correlations with the total CES-D score, discriminant function analyses or factors analyses that focus on items with the highest loadings in each factor (Andresen et al. 1994; Shrout and Yager 1989). These selection methods, however, presume that all items are effect indicators and that the scale is psychometrically equivalent across groups. Although derived through different means, our five-item CES-D is quite similar to other abbreviated versions of the CES-D. The items "depressed," "sad" and "happy" are commonly included in most abbreviated versions of the CES-D (Andresen et

al. 1994; Kohout et al. 1993; Shrout and Yager 1989).²³ The item “life” is not often included and may sometimes be unavailable to researchers using adult versions of the CES-D scale. Fortunately, this poses no problem since the “Life” item can be eliminated from our abbreviated CES-D without undermining the invariance and reliability of the shortened scale.

While our proposed five-item CES-D improves the cross-cultural comparability of the CES-D, researchers studying the causes and consequences of depression within a particular ethnocultural group may prefer using more culturally-grounded instruments for depression. This is especially true when researchers aim to diagnose clinical depression in minority populations. Through our analysis we identify five items that measure aspects of depression common to our ethnocultural groups, and that can be included for comparative purposes in population-based surveys. However, culturally bounded measures of depression may be more appropriate when the researchers do not intend to make cross-cultural comparisons.

Finally, we do not claim that the five-item CES-D developed and evaluated here is the best cross-cultural measure of depression risk. Other researchers may extend this analysis by evaluating alternative measurement models using other items that are effect indicators and that allow for measurement equivalence across ethnocultural groups. Others may also extend this analysis by evaluating the equivalence of the scale within smaller ethnocultural groups and by gender within these groups. We have based this analysis on the pan-ethnic categories most commonly used by social scientists and health researchers. An important next step is to establish the equivalence of the CES-D within these pan-ethnic categories.

Although focused on the CES-D, this research has called attention to the overall importance of evaluating the structural equivalence of multiple-item scales before meaningful group comparisons can be made. The common techniques used to evaluate the reliability and empirical validity of multiple-item scales are not sufficient to establish the appropriateness of scale. As the diversity of the U.S. population grows and as researchers increasingly adapt measurement instruments for use in multiethnic populations, evaluations of the cross-cultural comparability of these instruments must become a priority.

Notes

1. The rate of item non-response did not differ across items. However, persons who did not respond to at least one item were more likely to be non-white and a member of the first-generation. In fact, logit analyses indicated that first generation immigrants were more likely than U.S.-born, 3rd+ generation immigrants to have missing data on all CES-D items; and 2nd generation immigrants were *less* likely to have missing data on three CES-D items — sad, good and hopeful. Controlling for immigrant generation, only three significant race-ethnicity differences in item non-response were found. Hispanics were more likely not to respond to the CES-D items failure, unfriendly and disliked. There were no significant

differences in item non-response by gender or age. Because item non-response was so minor, these differences do not bias our analyses.

2. The vast majority of adolescents in Add Health are ages 13–19. Adolescents younger than 13 or older than 19 either entered school early or were held back during high school.

3. In studies of item-response biases, the “crying” item, the item “talked less,” and two interpersonal items, “people unfriendly” and “people disliked me,” have been found to be gender-, age- and race-biased (Cole et al. 2000; Stommel et. al. 1993).

4. Using LISREL notation, the measurement model for observed independent and dependent variables can be written respectively as the following:

$$x = v_x + \Lambda_x \xi + \delta \tag{1}$$

$$y = v_y + \Lambda_y \eta + \epsilon \tag{2}$$

In equations (1) and (2), v (upsilon) is the intercept, Λ (lambda) is a matrix of factor loadings describing the expected change in the observed vector of x - and y -variables for a unit change in the exogenous latent variable vector ξ (x_i) and the endogenous latent variable vector η (η). The measurement errors are given by the vectors δ (δ) and ϵ (ϵ). The covariances among these errors constitute Θ_δ and Θ_ϵ , respectively. The factor means are equal to K (κ) such that $E(\xi) = K$. The structural part of the model is given by the following:

$$\eta = \alpha + B\eta + \Gamma\xi + \zeta \tag{3}$$

Regarding the above equation, α is the intercept, Γ is the expected effects of the endogenous latent variable and exogenous latent variables, B describes the relationship among endogenous latent variables, and ζ is a vector of measurement errors. The covariance among exogenous latent variables is given by Φ . In these models, $E(\zeta) = E(\epsilon) = E(\delta) = 0$, $E(\zeta\xi') = E(\epsilon\eta') = E(\delta\xi') = 0$, and $E(\epsilon_i\epsilon_j) = 0$ for $i \neq j$

5. Ideally, we would like to account for the fact that each CES-D item is a categorical (not continuous) variable. We used Mplus version 2.13. It and the other SEM packages that were available at the time of analysis were not able to simultaneously account for both the design effects and categorical nature of our observed variables. The use of clustering and weighting in SEM is an area of intense methodological research (Satorra and Muthen 1995). For this reason, we tested the sensitivity of our results to different estimation approaches. First, we estimated models that took into account the categorical nature of the CES-D items in lieu of the design effects. Our results were not altered. Second, we estimated all models with and without adjustments for clustering and weighting. Again, our results were not altered by differences in the estimation. For the interested reader, precise information about estimation with clustering and weights in Mplus can be found in Asparouhov (2004).

6. Additional two-factor and three-factor models were evaluated. In the two-factor models, one factor was presumed to be one of the following: negative affect, positive affect, somatic complaints or interpersonal relations. All the other items were presumed to load on a single second factor. In the other three-factor models, the somatic complaints and interpersonal relations factors were combined; the somatic complaints and positive affect factors were combined; the negative affect and interpersonal relations factors were combined; and the positive affect and interpersonal relations factors were combined. Thus, we actually compared all possible single-factor, two-factor, three-factor and four-factor solution models by estimating a total of 12 CFAs.

7. To estimate these models, we must also set the variance of the latent variables to 1. This sets the metric of the latent variables.

8. Three separate regression sets are required because of the differences in the scales of each set of parameter estimates.

9. Our focus on these five items as effect indicators was driven by previous empirical research and discussions with psychologists. This is a quasi-empirical approach designed to reduce the number of alternative hypotheses regarding the structure of the scale to be tested. A fully empirical approach would require testing every possible combination of effect, outcome and cause indicators possible using the nineteen items in the CES-D scale. There are several thousand possibilities. Thus, a fully empirical approach would be infeasible.

10. Using LISREL notation, these hypotheses can be written as the following: (1) $\Lambda^1 = \Lambda^2 = \dots = \Lambda^{12}$, (2) $K^1 = K^2 = \dots = K^{12}$, (3) $\nu_x^1 = \nu_x^2 = \dots = \nu_x^{12}$, (4) $\Phi^1 = \Phi^2 = \dots = \Phi^{12}$, and (5) $\Theta_\delta^1 = \Theta_\delta^2 = \dots = \Theta_\delta^{12}$.

11. We have reported the Cronbach's alpha because it is so commonly used as a measure of reliability for the CES-D and other scales. However, Cronbach's alphas are only appropriate if a scale is uni-dimensional with no correlated errors between items, equal factor loadings across items and equal error variances.

12. A bonferroni adjustment for multiple group comparisons was made.

13. Detailed results from the Multiple Group Analysis and the regression analyses are available in an online Technical Appendix (see the *Social Forces* Web site: <http://socialforces.unc.edu>).

14. Bollen (1991) derived a test statistic, T, that tests multiple vanishing tetrads simultaneously: $T = Nr \sum_{ii}^{-1} t$ where N is the sample size, t is a vector of the independent sample tetrad differences, and \sum_{ii}^{-1} is the inverse of the covariance matrix of the limiting distribution of t as N goes to infinity. The T statistic asymptotically approximates a chi-square statistic with degrees of freedom equal to the number of vanishing tetrads considered in the test. Therefore, differences between nested models can be evaluated as the change in chi-squares between the restricted and unrestricted model.

15. These types of analysis require a significant amount of computer memory, due to limitations in the memory available for this analysis, only 17 of the 19 CES-D indicators in Add Health were included in the analysis. The two excluded indicators were the following: (1) You felt that you were just as good as other people, and (2) You talked less than usual. The item "talk less" has been found to be a gender-biased item in previous analyses (Stommel et al. 1993). And the item "good" is thought to measure self-esteem. Both items also had the lowest factor loadings among the native white adolescent population.

16. This is a quasi-empirical approach that relies on first hypothesizing a model of effect, cause and outcome indicators. Several thousand models are possible. For example, future researchers could test whether a different 10-item measure might be an improvement over both the 19-item and our five-item measure. However, our discussions with a psychologist and our review of the empirical literature suggest that the five items identified are the only true effect indicators in the CES-D scale.

17. Complete results for each of the ethnocultural groups are included in the online Technical Appendix, located on the *Social Forces* Web site: <http://socialforces.unc.edu>.

18. The χ^2 statistic is particularly large due to our large sample size. For this reason, we also present the CFI and RMSEA.

19. As discussed in a previous methodology section, our findings are based on the CFI and RMSEA. Because of the large N , differences in chi-squares are statistically significant without being analytically meaningful.
20. Detailed results from these Multiple Group Analyses are available in the online Technical Appendix.
21. Measures of reliability generally increase with the addition of items to a scale and decrease as they are removed. This is true even when those items are relatively poor indicators of the latent variable and indicators that are subject to measurement variance across ethnocultural groups. The mean of inter-item correlations is less sensitive to the number of indicators in a scale. For each of the ethnocultural groups in our study, the mean of the inter-item correlations for the five-item CES-D is higher than the mean of the inter-item correlations for the 19-item CES-D.
22. This measure of reliability can be calculated using the loadings and variances of the loadings provided in any standard output for factor analyses. This measure of reliability is more general and less restrictive in its assumptions than the Cronbach's alpha. The Cronbach's alpha assumes that factor loadings for each item in the scale equal 1 and that the variance of the factors is equal to one another.
23. The item "enjoy" is also commonly included in abbreviated CES-D scales and may be an effect indicator. However, including it in the abbreviated scale would result in a two-dimensional scale with correlated errors between the four negatively and two positively worded items. By including only one of the positively worded effect indicators in the model, we avoid this complication without any significant loss in the reliability of the abbreviated index.

References

- Akiyama, Hiroko, and Toni Antonucci. 2002. "Gender Differences in Depressive Symptoms: Insights from a Life Span Perspective on Life Stages and Social Networks." *Advances in Medical Sociology* 8:343–58.
- Andresen, E.M., J.A. Malmgren and W.B. Carter. 1994. "Screening for Depression in Well Older Adults: Evaluation of a Short Form of the CES-D." *American Journal of Preventive Medicine* 10(2):77–84.
- Angel, Ronald, and Peter J. Guarnaccia. 1989. "Mind, Body and Culture: Somatization Among Hispanics." *Social Science and Medicine* 28(12):1229–38.
- Asparouhov, Tihomir. 2004. Weighting for Unequal Probability of Selection in Latent Variable Modeling. *Mplus Web Notes*. <http://www.statmodel.com/mplus/examples/webnote.html>.
- Bearman, P.S., J. Jones and J.R. Udry. 1997. The National Longitudinal Study of Adolescent Health: Research Design. Carolina Population Center. <http://www.cpc.unc.edu/projects/addhealth/design.htm>.
- Bentler, P.M., and D.G. Bonett. 1980. "Significance tests and goodness of fit in the analysis of covariance structures." *Psychological Bulletin* 88:588–606.
- Bollen, Ken A. 1980. "Issues in the Comparative Measurement of Political Democracy." *American Sociological Review* 45(Jun):370–90.

- . 1989. *Structural Equations with Latent Variables*. John Wiley & Sons.
- Bollen, Ken A., and Kwok-fai Ting. 2000. "A Tetrad Test for Causal Indicators." *Psychological Methods* 5:3–22.
- Bollen, Ken A., and R. Lennox. 1991. "Conventional Wisdom on Measurement: A Structural Equation Perspective." *Psychological Bulletin* 110:305–14.
- Bost, Kelly K., Martha J. Cox, Margaret R. Burchinal and Chris Payne. 2002. "Structural and Supportive Changes in Couples' Family and Friendship Networks across the Transition to Parenthood." *Journal of Marriage and the Family* 64(2):517–31.
- Burnam, M.A., R.L. Hough, M. Karno, J.I. Escobar and C. Telles. 1987. "Acculturation and lifetime prevalence of psychiatric disorders among Mexican Americans in Los Angeles." *Journal of Health and Social Behavior* 28:89–102.
- Carolina Population Center. 1994. National Longitudinal Study of Adolescent Health. Computer file. Carolina Population Center.
- Chantala, Kim and Joyce Tabor. 1999. "Strategies to Perform a Design-based Analysis Using the Add Health Data." *Add Health Paper Series*. Carolina Population Center.
- Chapleski, Elizabeth E., James K. Lamphere, Richard Kaczynski, Peter A. Lichtenberg and Jeffrey W. Dwyer. 1997. "Structure of a Depression Measure among American Indian Elders: Confirmatory Factor Analysis of the CES D Scale." *Research on Aging* 19(4):462–85.
- Cockerham, William. 1989. *Medical Sociology, 4th Edition*. Prentice Hall, Inc.
- Cole, Stephen, Ichiro Kawachi, Susan Maller, Lisa Berkman. 2000. "Test of Item-Response Bias in the CES-D Scale: Experience from the New Haven EPESE Study." *Journal of Clinical Epidemiology* 53:285–89.
- Cuellar, I., and Robert.E. Roberts. 1984. "Psychological disorders among Chicanos." Pp 133–61 in *Chicano Psychiatry, 2nd Ed.*, edited by J.L. Martinez Jr. and R.H. Mendoza. Academic.
- Curyto, Kim J., Elizabeth E. Chapleski, Peter A. Lichtenberg, Elise Hodges, Richard Kaczynski and Sobeck. 1998. "Joanne Prevalence and Prediction of Depression in American Indian Elderly." *Clinical Gerontologist* 8(3):19–37.
- Dershem, Larry D., Valeri V. Patsiorkovski and David J. O'Brien. 1996. "The Use of the CES-D for Measuring Symptoms of Depression in Three Rural Russian Villages." *Social Indicators Research* 39(1):89–108.
- Doerfler, Leonard. 1988. "Depression in Children and Adolescents: A Comparative Analysis of the Utility and Construct Validity of Two Assessment Measures." *Journal of Consulting and Clinical Psychology* 56(5):769–72.
- Faulstich, M.E., M.P. Carey, L. Ruggiero, P. Enyart and F. Gresham. 1986. "Assessment of depression in childhood and adolescence: An evaluation of the Center for Epidemiological Studies Depression Scale for Children (CES-DC)." *American Journal of Psychiatry* 143(8):1024–27.
- Finch, Brian Karl, Robert Hummer and William A. Vega. 2002. "Validity of Self-rated Health among Latino(a)s." *American Journal of Epidemiology* 155(8):755–59.
- Finch, Brian Karl, Bohdan Kolody and William A. Vega. 2000. "Perceived Discrimination and Depression among Mexican Origin Adults in California." *Journal of Health and Social Behavior* 41(3):295–313.

- Flores, Glenn, Elena Fuentes-Afflick, Oxiris Barbot, Olivia Carter-Pokras, Luz Claudio, Marielena Lara, Jennie A. McLaurin, et al. 2002. "The Health of Latino Children: Urgent Priorities, Unanswered Questions, and a Research Agenda." *Journal of the American Medical Association* 288(1):82–90.
- Garrison, C.Z., C.L. Addy, K.L. Jackson, R.E. McKeown and J.L. Waller. 1991. "The CES-D As a Screen for Depression and Other Psychiatric Disorders in Adolescents." *Journal of the American Academy of Child and Adolescent Psychiatry* 30:636–41.
- Garrison, C., P. Lewinsohn, F. Marsteller, J. Langhinrichsen and I. Lann. 1991. "The assessment of suicidal behavior in adolescents." *Suicide and Life-Threatening Behavior* 21:217–30.
- Greenfield, Tom K., Jurgen Rehm and John D. Rogers. 2002. "Effects of Depression and Social Integration on the Relationship between Alcohol Consumption and All Cause Mortality." *Addiction* 97(1):29–38.
- Golding, J.M. and C.S. Aneshensel. 1989. "Factor Structure of the Center for Epidemiologic Studies Depression Scale among Mexican Americans and Non-Hispanic Whites." *Psychological Assessment: Journal of Consulting and Clinical Psychology* 1:163–68.
- Gore, Susan, Robert H. Aseltine Jr. and Mary Ellen Colton. 1992. "Social Structure, Life Stress and Depressive Symptoms in a High School-Aged Population." *Journal of Health and Social Behavior* 33(2):97-113.
- Guarnaccia, Peter J., Ronald Angel and Jacqueline Lowe Worobey. 1989. "The Factor Structure of the CES-D in the Hispanic Health and Nutrition Examination Survey: The Influences of Ethnicity, Gender and Language." *Social Science and Medicine* 29(1):85–94.
- Hagan, John and Holly Foster. 2001. "Youth Violence and the End of Adolescence." *American Sociological Review* 66(6):874–99.
- Harker, Kathryn. 2001. "Immigrant Generation, Assimilation, and Adolescent Psychological Well-Being." *Social Forces* 79(3):969–1004.
- Harris, Kathleen M. 1999. "The Health Status and Risk Behavior of Adolescents in Immigrant Families." Pp. 286–347 in *Children of Immigrants: Health, Adjustment and Public Assistance*, edited by Donald J. Hernandez, National Academy of Sciences Press.
- Hertzog, C., J.V. Van Alstine, P.D. Usala, D.F. Hulstsch and R. Dixon. 1990. "Measurement Properties of the Center for Epidemiological Studies Depression Scale (CES-D) in older populations." *Psychological Assessment: Journal of Consulting and Clinical Psychology* 2:64–72.
- Hughes, Mary Elizabeth, and Linda J Waite. 2002. "Health in Household Context: Living Arrangements and Health in Late Middle Age." *Journal of Health and Social Behavior* 43(1):1–21.
- Hui, C.H., and H.C. Triandis. 1989. "Effects of culture and response format on extreme response style." *Journal of Cross-Cultural Psychology* 20:296–309.
- Hybels, Celia F., Dan G. Blazer, and Carol F. Pieper. 2001. "Toward a Threshold for Subthreshold Depression: An Analysis of Correlates of Depression by Severity of Symptoms Using Data from an Elderly Community Sample." *Gerontologist* 41(3):357–65.
- Iwata, N., and Robert E. Roberts. 1996. "Age differences among Japanese on the Center for Epidemiologic Studies Depression Scale: An ethnocultural perspective on somatization." *Social Science and Medicine* 43:967–74.
- Jackson, Pamela Braboy and Montinique Finney. 2002. "Negative Life Events and Psychological Distress among Young Adults." *Social Psychology Quarterly* 65(2):186–201.

- Kaplan, M.A., and G. Marks. 1990. "Adverse effects of acculturation: psychological distress among Mexican American young adults." *Social Science and Medicine* 31:1313–19.
- Kim, M.T. 2002. "Measuring Depression in Korean Americans: Development of the Kim Depression Scale for Korean Americans." *Journal of Transcultural Nursing* 13(2):109–17.
- Kikuzawa, Saeko. 2001. "The Structure of Self Reported Distress in the United States and Japan: NFR and NSFH." *Japanese Journal of Family Sociology* 12(2):247–59.
- Kohout, Frank J., Lisa F. Berkman, Denis A. Evans and Joan Cornoni Huntley. 1993. "Two Shorter Forms of the CES D Depression Symptoms Index." *Journal of Aging and Health* 5(2):179–93.
- Kuo, Wen. 1976. "Theories of Migration and Mental Health: An Empirical Testing on Chinese-Americans." *Social Science and Medicine* 10:297–306.
- LaGory, Mark, Kevin Fitzpatrick and Ferris Ritchey. 2001. "Life Chances and Choices: Assessing Quality of Life among the Homeless." *The Sociological Quarterly* 42(4):633–51.
- Lewis, Robin J., Valerian J. Derlega, Andrea Berndt, Lynn M. Morris and Suzanna Rose. 2001. "An Empirical Analysis of Stressors for Gay Men and Lesbians." *Journal of Homosexuality*. 42(1):63–88.
- Link, Bruce. 2002. "The Challenge of the Dependent Variable." *Journal of Health and Social Behavior* 43(Jun):247–53.
- Link, B., and B.P. Dohrenwend. 1980. "Formulation of hypotheses about the true prevalence of demoralization." Pp. 114–32 in *Mental Illness in the United States: Epidemiological Estimates*, edited by B.P. Dohrenwend. Praeger.
- Manson, S.M., L.M. Ackerson, R.W. Dick, A.E. Baron et al. 1990. "Depressive Symptoms among American Indian Adolescents: Psychometric Characteristics of the Center for Epidemiologic Studies Depression Scale (CES-D)." *Psychological Assessment* 2(3):231–37.
- Marin, G., R.J. Gamba and B.V. Marin. 1992. "Extreme response style and acquiescence among Hispanics: The role of acculturation and education." *Journal of Cross-Cultural Psychology* 23:498–509.
- Marks, Nadine F, James David Lambert, and Heejeong Choi. 2002. "Transitions to Caregiving, Gender, and Psychological Well-Being: A Prospective U.S. National Study." *Journal of Marriage and the Family*. 64(3):657–67.
- Marsella, A.J., K.O. Sanborn, V. Kameoka, L. Shizuru and J. Brennan. 1975. "Cross-Validation of Self-Report Measures of Depression Among Normal Populations of Japanese, Chinese, and Caucasian Ancestry." *Journal of Clinical Psychology* 31(2):281–87.
- McGee, D.L., Y. Liao and G. Cao. 1999. "Self-reported health status and mortality in a multiethnic U.S. Cohort." *American Journal of Epidemiology* 149:41–46.
- Michalos, Alex C., Bruno D. Zumbo and Anita Hubley. 2000. "Health and the Quality of Life." *Social Indicators Research* 51(3):245–86.
- Mirowsky, John, and Catherine E. Ross. 1984. "Mexican culture and its emotional contradictions." *Journal of Health and Social Behavior* 25:2–13.
- . 1989. *Social Causes of Psychological Distress*. Aldine de Gruyter.

- Morton, Deborah J., Susan A. Schoenrock, E. Percil Stanford, K. Michael Peddecord and Craig A. Molgaard. 1989. "Use of the CES D among a Community Sample of Older Mexican Americans." *Journal of Cross Cultural Gerontology* 4(4):289–306.
- Murphy, Debra A., Stephen J. Durako, Anna Barbara Moscicki, Sten H. Vermund, Yong Ma, Donald F. Schwarz and Larry R. Muenz. 2001. "No Change in Health Risk Behaviors over Time among HIV Infected Adolescents in Care: Role of Psychological Distress." *Journal of Adolescent Health* 29(supplement 3):57–63.
- Muthen, Bengt, and Linda Muthen. 2003. Mplus Version 2.13. Computer file. Muthen and Muthen.
- Noh, Samuel, Violet Kaspar and Xinyin Chen. 1998. "Measuring Depression in Korean Immigrants: Assessing Validity of the Translated Korean Version of CES D Scale" *Cross Cultural Research* 32(4):358–77.
- Office of Minority Health. 2002. *Statement of Purpose*. <http://www.omhrc.gov/OMH/sidebar/aboutOMH.htm>.
- Office of the Surgeon General. 2001. *Mental Health: Culture, Race, and Ethnicity, A Supplement to Mental Health: A Report of the Surgeon General*. <http://www.surgeongeneral.gov/library/mentalhealth/cre>.
- Petterson, Stephen M., and Lisa V. Friel. 2001. "Psychological Distress, Hopelessness and Welfare" *Women and Health* 32(12):79–99.
- Portes, Alejandro, and Ruben.G. Rumbaut. 1996. *Immigrant America: A Portrait, 2nd Edition*. University Of California Press.
- . 2001. *Ethnicities: Children of Immigrants in America*. University of California Press.
- Radloff, Lenore. S. 1977. "The CES-D Scale: A Self-Report Depression Scale for Research in the General Population." *Applied Psychological Measurement* 1(3):385–401.
- . 1991. The Use of the Center for Epidemiologic Studies Depression Scale in Adolescents and Young Adults. *Journal of Youth Adolescence* 20(2):149–66.
- Ranzijn, Rob, and Mary Luszcz. 2000. "Measurement of Subjective Quality of Life of Elders." *International Journal of Aging and Human Development* 50(4):263-78.
- Riddle, A.S., R.B. Marc, and U. Hess. 2002. "A Multi-Group Investigation of the CES-D's Measurement Structure across Adolescents, Young Adults, and Middle Aged Adults." *Scientific Series*, CIRANO, Québec. <http://www.cirano.qc.ca/pdf/publication/2002s-36.pdf>.
- Roberts, Robert. E., Sally. W. Vernon, and Howard M. Rhoades. 1989. "Effects of Language and Ethnic Status on Reliability and Validity of the Center for Epidemiologic Studies-Depression Scale With Psychiatric Patients." *Journal of Nervous and Mental Disease* 177(10):581–92.
- Roberts, Robert. E., Judy A. Andrews, Oeter M. Lewinsohn and Hyman Hops. 1990. "Assessment of Depression in Adolescents Using the Center for Epidemiologic Studies Depression Scale." *Psychological Assessment* 2(2):122–28.
- Roberts, Robert. E., P.M. Lewinsohn and J.R. Seeley. 1991. "Screening for adolescent depression: a comparison of depression scales." *Journal of American Academy of Child and Adolescent Psychiatry* 30:58–66.

- Roberts, Robert. E., and M. Sobhan. 1992. "Symptoms of depression in adolescence: A comparison of Anglo, African, and Hispanic Americans." *Journal of Youth and Adolescence* 21:639–51.
- Roberts, Robert. E., Catherine Roberts and Y. Richard Chen. 1997. "Ethnocultural differences in prevalence of adolescent depression." *American Journal of Community Psychology* 25:95–110.
- Ross, Catherine E., and John Mirowsky. 1984. "Socially-desirable response and acquiescence in a cross-cultural survey of mental health." *Journal of Health and Social Behavior* 25:189–97.
- Satorra, A., and B. Muthen. 1995. "Complex Sample Data in Structural Equation Modeling," *Sociological Methodology* 25:267–316.
- Schieman, Scott, and Heather A. Turner. 2001. "'When Feeling Other People's Pain Hurts': The Influence of Psychosocial Resources on the Association between Self-Reported Empathy and Depressive Symptoms." *Social Psychology Quarterly* 64(4):376–89.
- Schoenbach, V.J., B.H. Kaplan, T.C. Grimson and E.H. Wagner. 1982. "Use of Symptom Scale to Study the Prevalence of a Depressive Syndrome in Young Adolescents." *American Journal of Epidemiology* 116:791–800.
- Schwartz, Sharon. 2002. "Outcomes for the Sociology of Mental Health: Are We Meeting Our Goals?" *Journal of Health and Social Behavior* 43:223–35.
- Sheehan, T.J., J. Fifield, S. Reisine and H. Tennen. 1995. "The measurement structure of the Center for Epidemiologic Studies Depression Scale." *Journal of Personality Assessment* 64:507–21.
- Shrout, Patrick, and Thomas J. Yager. 1989. "Reliability and Validity of Screening Scales: Effect of Reducing Scale Length." *Journal of Clinical Epidemiology* 42(1):69–78.
- Siegel, J.M., C.S. Aneshensel, B. Taub, D.P. Cantwell and A.K. Driscoll. 1998. "Adolescent depressed mood in a multiethnic sample." *Journal of Youth and Adolescence* 27:413–27.
- Simon, Robin W. 2002. "Revisiting the Relationships among Gender, Marital Status, and Mental Health." *American Journal of Sociology* 107(4):1065–96.
- Simoni, Jane M., and M.T. Ng. 2000. "Trauma, Coping, and Depression among Women with HIV/AIDS in New York City." *AIDS Care* 12(5):567–80.
- Somervell, Phillip. D., Janette Beals, David Kinzie, James Boehnlein et al. 1992-93. "Use of the CES-D in an American Indian Village." *Culture, Medicine and Psychiatry* 16(4):503–17.
- Steinhausen, Hans-Christoph, and Christa Winkler Metzke. 2000. "Adolescent Self-Rated Depressive Symptoms in a Swiss Epidemiological Study." *Journal of Youth and Adolescence* 29(4):427–40.
- Stening, B.W., and J.E. Everett. 1984. "Response Styles in a Cross-Cultural Managerial Study." *Journal of Social Psychology* 122:151–56.
- Stommel, M., B.A. Given, C.W. Given, H.A. Kalaian, R. Schulz and R. McCorkle. 1993. "Gender Bias in the Measurement Properties of the Center for Epidemiologic Studies Depression Scale (CES-D)." *Psychiatry Research* 49:239-50.
- Sweeney, Megan M., and Allan V. Horwitz. 2001. "Infidelity, Initiation, and the Emotional Climate of Divorce: Are There Implications for Mental Health?" *Journal of Health and Social Behavior* 42(3):295–309.

- Thoits, Peggy A., and Lyndi N. Hewitt. 2001. "Volunteer Work and Well-Being" *Journal of Health and Social Behavior* 42(2):115–31.
- Tienda, M., and M. Kleykamp. 2000. *Physical and mental health status of Hispanic adolescent girls: A comparative perspective*. Manuscript in preparation. Princeton University.
- Tolor, A., and V.M. Murphy. 1985. "Stress and Depression in High School Students." *Psychological Reports* 57:535–41.
- Trieman, B. 1975. "Depressive mood among middle class urban ethnic groups." Technical Report, Contract HSM 42-73-238, National Institute of Mental Health.
- U.S. Census Bureau, Population Projections Program, Population Division. 2000. *Projections of the Resident Population by Race, Hispanic Origin, and Nativity: Middle Series, 2050 to 2070*. <http://www.census.gov/population/projections/nation/summary/np-t5-g.txt>.
- U.S. Census Bureau. 2003. "Census Bureau Releases Population Estimates by Age, Sex, Race and Hispanic Origin." *United States Department of Commerce News*. <http://www.census.gov/Press-Release/www/2003/cb03-16.html>.
- U.S. Department of Health and Human Services. 1998. "President Clinton Announces New Racial And Ethnic Health Disparities Initiative." *White House Fact Sheet*. <http://www.hhs.gov/news/press/1998pres/980221.html>.
- Van de Vijver, Fons J.R., and Kwol Leung. 2001. "Personality in Cultural Context: Methodological Issues." *Journal of Personality* 69(6):1007–31.
- Vega, William A., Bohdan Kolody, S. Aguilar-Gaxiola, Ethel Alderete, Ralph Catalano and Jorge Caraveo-Anduaga. 1998. "Lifetime Prevalence of DSM-III-R Psychiatric Disorders Among Urban and Rural Mexican Americans in California." *Archives of General Psychiatry* 55(9):771–78.
- Vega, William A., and Ruben Rumbaut. 1991. "Ethnic Minorities and Mental Health." *Annual Review of Sociology* 17:351–83.
- Vega, William A., Bohdan Kolody and Juan Ramon Valle. 1986. "The Relationship of Marital Status, Confidant Support, and Depression among Mexican Immigrant Women." *Journal of Marriage and the Family*. 48(3):597–605.
- Videon, Tami M. 2002. "The Effects of Parent-Adolescent Relationships and Parental Separation on Adolescent Well-Being." *Journal of Marriage and the Family* 64(2):489–503.
- Weissman, N.M., H. Orvaschel and N. Padian. 1980. "Children's symptoms and social functioning self-reported scales: Comparison of mothers' and children's reports." *Journal of Nervous Mental Disorder* 168(12):736–40.
- Wilk, Piotr. 2001. "Women's Employment Transitions and Changes in Psychological Distress." *Canadian Studies in Population* 28(2):513–33.
- Wong, Sandra L. 2001. "Depression Level in Inner City Asian American Adolescents: The Contributions of Cultural Orientation and Interpersonal Relationships." *Journal of Human Behavior in the Social Environment* 3(34):49–64.
- Ying, Yu-Wen. 1988. "Depressive Symptomatology Among Chinese-Americans As Measured by the CES-D." *Journal of Clinical Psychiatry* 44(5):739–46.

Copyright of Social Forces is the property of University of North Carolina Press. The copyright in an individual article may be maintained by the author in certain cases. Content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.