Published in final edited form as:
J Couns Psychol. 2014 January ; 61(1): 154–161. doi:10.1037/a0034749.

The Multigroup Ethnic Identity Measure-Revised: Measurement invariance across racial and ethnic groups

Susan D. Brown, **Kirsten A. Unger Hu**, **Ashley A. Mevi**, **Monique M. Hedderson**, **Jun Shan**, **Charles P. Quesenberry**, and **Assiamira Ferrara**
Division of Research, Kaiser Permanente Northern California

Abstract

The Multigroup Ethnic Identity Measure-Revised (MEIM-R), a brief instrument assessing affiliation with one's ethnic group, is a promising advance in the ethnic identity literature. However, equivalency of its measurement properties across specific racial and ethnic groups should be confirmed before using it in diverse samples. We examined a) the psychometric properties of the MEIM-R including factor structure, measurement invariance, and internal consistency reliability, and b) levels of and differences in ethnic identity across multiple racial and ethnic groups and subgroups. Asian (n = 630), Black/African American (n = 58), Hispanic (n = 240), multiethnic (n = 160), and White (n = 375) women completed the MEIM-R as part of the "Gestational diabetes' Effect on Moms" diabetes prevention trial in the Kaiser Permanente Northern California health care setting (N = 1,463; M age 32.5 years, SD = 4.9). Multiple-groups confirmatory factor analyses provided provisional evidence of measurement invariance, i.e., an equal, correlated two-factor structure, equal factor loadings, and equal item intercepts across racial and ethnic groups. Latent factor means for the two MEIM-R subscales, exploration and commitment, differed across groups; effect sizes ranging from small to large generally supported the notion of ethnic identity as more salient among people of color. Pending replication, good psychometric properties in this large and diverse sample of women support the future use of the MEIM-R. Preliminary evidence of measurement invariance suggests that the MEIM-R could be used to measure and compare ethnic identity across multiple racial and ethnic groups.

Keywords

ethnic identity; measurement; psychometric; Multigroup Ethnic Identity Measure-Revised (MEIM-R)

Ethnic identity continues to be a highly relevant and vigorously debated construct in the field of counseling psychology (Cokley, 2007; Ponterotto & Mallinckrodt, 2007). Ethnic identity refers to the quality of an individual's affiliation with his or her ethnic group (Phinney & Ong, 2007). It is conceptualized as a multidimensional construct, the development of which involves a process of *exploring* the meaning of one's identity and a

Correspondence concerning this article should be addressed to Susan D. Brown, PhD, Division of Research, Kaiser Permanente Northern California, 2000 Broadway, Oakland, CA 94612. susan.d.brown@kp.org.

Portions of this research were presented in abstract form at the 2013 Society of Behavioral Medicine Annual Meeting.

sense of *commitment* or belonging to that identity (Marcia, 1980; Ong, Fuller-Rowell, & Phinney, 2010).

To understand individual variability in the value and personal salience of ethnic group membership, researchers and clinicians require a brief, easily administered measure of ethnic identity that is valid and reliable among people of various racial and ethnic backgrounds. In contrast to measures for a single group, Phinney (1992) designed the Multigroup Ethnic Identity Measure (MEIM) as a measure of the subjective sense of membership in any group. Numerous psychological and behavioral health studies have since relied on the MEIM, yet have yielded mixed results. For example, a meta-analysis of 184 studies, over 70% of which used the MEIM, associated ethnic identity with well-being among racial and ethnic minorities (e.g., greater self-esteem; Smith & Silva, 2011). Higher levels of ethnic identity have been associated with less alcohol use among Black students from immigrant families (Schwartz et al., 2011) and lower frequencies of unsafe sexual practices among European Americans (Espinosa-Hernández & Lefkowitz, 2009). Higher levels of ethnic identity have been associated with greater leisure-time physical activity, more healthful eating patterns, and fewer depressive symptoms (Siegel, Yancey, & McCarthy, 2000), and lower risk for disordered eating (Henrickson, Crowther, & Harrington, 2010; Rogers Wood & Petrie, 2010; Shuttlesworth & Zotter, 2011), among African American women. In contrast, some research using the MEIM associated ethnic identity with disordered eating among White women (Shuttlesworth & Zotter, 2011) and risky sexual behavior and substance use in some Hispanic samples (Schwartz, et al., 2011; Zamboanga, Schwartz, Jarvis, & Van Tyne, 2009); other studies of Hispanic adults found no significant associations (Cachelin, Phinney, Schug, & Striegel-Moore, 2006). Disparate patterns highlight the need to confirm that a measure of ethnic identity is psychometrically sound across racial and ethnic groups and that outcomes do not stem from differential measurement performance.

Following more than fifteen years of applied and psychometric research (e.g., Avery, Tonidandel, Thomas, Johnson, & Mack, 2007; Dandy, Durkin, McEvoy, Barber, & Houghton, 2008; Gaines Jr. et al., 2010; Roberts et al., 1999), Phinney and Ong (2007) recently revised the MEIM to improve its content and face validity and address disagreement about its factor structure. In the resulting Multigroup Ethnic Identity Measure-Revised (MEIM-R), the authors removed items that were redundant, conceptually divergent from ethnic identity, or poor indicators of the underlying latent variable they were designed to represent. The authors reduced the number of items by half and matched the number of items across two subscales, exploration and commitment, allowing the subscales to be weighted equally if using the measure as a whole.

While this recent work resulted in a psychometrically strengthened measure, some uncertainty remains as to whether the MEIM-R is best described as having a correlated two-factor structure. Phinney and Ong (2007) found that a correlated two-factor model and a higher-order model (in which the latent factors exploration and commitment load onto a single second-order factor) fit their data equally well. Yoon (2011) found that a correlated two-factor model fit well in a heterogeneous minority sample and was unable to test a higher-order model. As noted by Cokley (2007), the most rigorous approach to verifying factor structure requires confirmatory factor analyses (CFAs) rather than exploratory factor analyses, and testing models in subgroups rather than treating racially and ethnically heterogeneous samples in such a way as to mask group differences—problems that have plagued much prior research on the original MEIM.

Importantly, the few existing evaluations of the MEIM-R lacked thorough assessments of measurement invariance. Measurement invariance evaluation involves determining how well an instrument’s measurement model generalizes across subgroups of a population. This is a critical issue; evidence that the measurement properties of the MEIM-R are equivalent in different racial and ethnic groups is necessary to legitimately compare groups to one another, yet measurement invariance is frequently assumed and rarely tested (Ong, et al., 2010). If MEIM-R items do not comparably measure underlying constructs across diverse groups, the resulting test bias would undermine what the MEIM-R was explicitly designed to do. In the only study to date that examined this issue, Yoon (2011) found evidence of invariance between European Americans and a heterogeneous sample of racial and ethnic minorities. However, despite acknowledging the need to study specific racial and ethnic groups separately, the small sample precluded such analyses. The lack of formal testing to establish measurement invariance, and the lack of clarity regarding which factor structure provides the most parsimonious explanation of observed relationships between latent factors, limits application of the MEIM-R in diverse samples.

MEIM-R evaluations have further been limited by primarily student samples. Community samples could enhance generalizability across age and educational ranges. Evaluations among adults may also be valuable given the notion, rooted in theories of Erikson (1968) and Marcia (1980), of ethnic identity as a dynamic developmental process in which individuals strengthen their understanding of and commitment to their ethnicity over time (see Ong, et al., 2010).

We addressed gaps in the psychometric literature on the MEIM-R by a) evaluating its psychometric properties including factor structure, measurement invariance, and internal consistency across multiple specific racial and ethnic groups; and b) if we could establish measurement invariance, assessing levels of and differences in ethnic identity. Where possible given adequate subsample sizes, we also examined ethnic identity across subgroups.

Methods

We conducted this study within “Gestational diabetes’ Effect on Moms” or GEM, a large 12-month cluster-randomized trial testing the comparative effectiveness of type 2 diabetes prevention strategies among women with gestational diabetes mellitus (GDM). The trial was set in Kaiser Permanente Northern California (KPNC), a large integrated healthcare delivery system. KPNC membership represents the region’s ethnic and socioeconomic diversity well, except at the outermost extremes of income and education (Gordon, 2012). The GEM trial was approved by the KPNC institutional review board and registered at ClinicalTrials.gov (NCT01344278).

Procedure

GEM staff identified eligible women through the KPNC electronic medical record system (EMR). Women received letters followed by phone calls describing the trial and offering retail gift cards for participation in GEM surveys at baseline (pregnancy) and at 6 weeks, 6 months, and 12 months postpartum. We obtained data for the current study from the segment of the GEM baseline survey conducted between May 2011 and August 2012. The survey was offered in English for the first three months and both English and Spanish for the remaining 12 months.

Measures

Ethnic Identity—We assessed ethnic identity using the MEIM-R (Phinney & Ong), administered by mail or online. The scale is preceded by an open-ended question to identify one's ethnic group (self-categorization). Six close-ended items then assess *exploration* of (items 1, 4, and 5) and *commitment* to (items 2, 3, and 6) one's ethnic identity on a 5-point scale from (1) strongly disagree to (5) strongly agree. Sample items include "I have often talked to other people in order to learn more about my ethnic group" and "I feel a strong attachment towards my own ethnic group." Scores for the three-item subscales and the overall scale are calculated by averaging item values. Prior research in primarily college student samples has indicated good reliability, with internal consistency (Cronbach's α) ranging from .76 to .91 for the two subscales and .81 to .89 for the overall scale (Phinney & Ong, 2007; Yoon, 2011).

Demographics—We assessed self-reported race and ethnicity (using a 16-category checklist including multiple Asian and Hispanic subgroups), education (six categories), income (nine categories), employment status (seven categories), marital status (six categories), country of birth, and number of years in the U.S. with single-item questions administered by computer-assisted telephone interview. We assessed age and preferred language through the EMR.

Phinney & Ong noted that "Measurement of ethnic identity must begin with verifying that the individuals being studied in fact self-identify as members of a particular group….For this purpose, it does not matter whether the label is an ethnic group or a racial group (regardless of how these terms are defined and whether they are broad or narrow in scope)" (2007, p. 272). They indicated that researchers may use appropriately inclusive checklists and/or open-ended questions to obtain self-categorizations. We reported self-categorization into what are commonly referred to as racial groups (e.g., Asian) and ethnic groups (e.g., Hispanic) at the broadest level using the former method; we also reported granular ethnicity data where groups were large enough to warrant separate analyses. Informed by recent recommendations from the Institute of Medicine (2009), we took this approach acknowledging that ethnicity and race refer to complex, socially constructed ideas (Markus & Moya, 2010); that their meanings remain contested, and can overlap and inform one another (Markus & Moya, 2010); and that they are often used interchangeably (Quintana, 2007), despite having divergent historical origins (Snipp, 2010; Trimble, 2007). For a discussion, see the recent special section of the *Journal of Counseling Psychology* (Ponterotto & Mallinckrodt, 2007) and Markus and Moya (2010).

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Participants

GEM trial inclusion criteria included being 18 years of age or older and having a pregnancy complicated by GDM diagnosed with a standard 100-gram, 3-hour oral glucose tolerance test (American College of Obstetricians and Gynecologists, 2011). Of 2,305 women eligible to complete the MEIM-R as part of the GEM baseline survey, 67.7% (n = 1,560) responded. Using EMR data, responders did not differ from non-responders (n = 745) in age, p = .63; but responders were somewhat less likely than non-responders to be Hispanic (20.9% vs. 28.2%) and more likely to prefer English (89.6% vs. 82.0%), $p < .001$. However, once the survey was offered in Spanish, responders (n = 1,015) no longer differed from non-responders (n = 452) by ethnicity, p = .19; the modest difference in preferred language was further attenuated (89.3% vs. 84.3%), p = .03. Among all 1,560 responders, small proportions skipped one or more MEIM-R items (0.8%, n = 12), completed the MEIM-R in Spanish (n = 59), or identified their ethnicity as Native Hawaiian/Other Pacific Islander, Native American/Alaska Native, “other,” or unknown (1.7%; n = 26) and were excluded from analyses due to small cell sizes.

The final analytic sample (N = 1,463) was highly diverse. Three quarters identified as racial or ethnic minorities including Asian (43.1%, n = 630), Black/African American (3.9%, n = 58), Hispanic (16.4%, n = 240), multiethnic (10.9%, n = 160), and White participants (25.6%, n = 375). Reflecting local demographics, most Asian participants identified as Chinese (26.7%, n = 168), Filipina (26.2%, n =165), or South Asian (e.g., Asian Indian, Pakistani, or Afghan; 22.9%, n = 144); most Hispanic participants identified as Mexican/ Mexican American (77.9%, n = 187). Most participants preferred English (92.9%, n = 1,360) versus Spanish (2.5%, n = 37) or other languages (4.5%, n = 66). About half were born outside the U.S. (47.4%; n = 693) yet had resided in the U.S. for a substantial number of years (M = 15.7, SD = 10.0; range 0–42 years). Immigration status varied across groups with 80.6% of Asian (n = 508), 15.5% of Black/African American (n = 9), 41.7% of Hispanic (n = 100), 12.5% of multiethnic (n = 20), and 14.9% of White participants (n = 56) born outside the U.S. Participants were primarily adults (M age 32.5 years, SD = 4.9; range = 18–50 years) who were married or living with a partner (90.9%; n = 1,330). Over half (53.3%; n = 779) had a college or professional degree, 33.8% (n = 495) had completed a two-year degree or some college, and 12.8% (n = 187) had a high school diploma or less education. Most were employed full-time (52.1%; n = 763) or part-time (10.3%; n = 151); 24.2% (n = 354) were not employed, 10.3% (n = 151) were on maternity leave, and 2.9% (n = 43) were students. Annual household income ranged from < $50,000 (26.2%; n = 383) to $50,000-$99,999 (36.0%, n = 527) to > $100,000 (32.7%; n = 479). (Categories that sum to < 100% reflect missing values.) Similar proportions completed the MEIM-R by mail (50.7%; n = 741) as online (49.4%; n = 722) with no difference by race or ethnicity, p = .16.

Statistical Analyses

We used Mplus 7.11 for CFAs and multiple-groups CFAs, and SAS® 9.3 for all other analyses. The data were not normally distributed, as Wilks-Shapiro tests were significant for all MEIM-R items, $p < .001$. We therefore used robust maximum likelihood estimation and report the Satorra-Bentler scaled chi-square (χ^2_{SB}) for all CFAs (Satorra & Bentler, 1999).

Factor Structure—We determined the best-fitting factor structure in our sample using CFAs. Following prior work (Phinney & Ong, 2007; Yoon, 2011), we evaluated the fit of three competing theoretical models: a one-factor structure in which all six MEIM-R items loaded onto a single factor; an uncorrelated two-factor structure in which items 1, 4, and 5 loaded exclusively onto the factor “exploration” and items 2, 3, and 6 loaded exclusively onto the factor “commitment”; and the same two-factor structure in which factors were allowed to correlate.

We relied on multiple goodness-of-fit indices to evaluate model fit in CFAs. Recognizing that our large sample size would likely lead a chi-square test to reject models based only on minor deficiencies (Brown, 2006), we examined three goodness-of-fit indices in tandem with one another: the standardized root mean square residual (SRMR), an absolute fit index reflecting the average discrepancy between observed correlations and correlations predicted by the model; the root mean square error of approximation (RMSEA), a fit index that incorporates a penalty for having additional parameters, thereby reflecting model parsimony (Steiger & Lind, 1980); and the Bentler comparative fit index (CFI), in which model fit is tested against a nested “independence” model positing no relationships among variables (Bentler, 1990). We also reported the Akaike Information Criterion (AIC), a comparative fit index that does not require nesting. In general, a well-fitting model will have SRMR values close to or less than 0.08, RMSEA values close to or less than 0.06, and CFI values close to or greater than 0.95 (Hu & Bentler, 1999); comparatively better fitting models will have lower AIC values.

Measurement Invariance—We evaluated measurement invariance using several sets of multiple-groups CFAs and a four-step process (Brown, 2006). First, we determined the best fitting factor structure separately in each group. Second, we simultaneously tested for equal factor structure (i.e., “configural invariance”) by specifying the same factor structure for all groups but allowing loadings, intercepts, and residual variances to differ. Third, if equal factor structure is established, then it is possible to test for equal factor loadings (i.e., “weak factorial” or “metric” invariance) by constraining loadings to equality and evaluating any decrement in model fit. Factor loading invariance indicates that the magnitudes of relationships between items and latent factors are equivalent across groups. Fourth, if equal factor loadings are established, then it is possible to test for equal item intercepts (i.e., “strong factorial” or “scalar” invariance) by constraining intercepts to be equal and evaluating any decrement in model fit. Equal item intercepts indicate that when the factor score is 0 the expected item scores are the same across groups. Only when equal item intercepts are established can absolute item scores be compared across groups. Signals that the constraints placed on each model had reduced model fit included CFI decrements ≥ 0.01 (Cheung & Rensvold, 2002) and higher AIC values. While recognizing the limitations of chi-square tests noted above, we also reported χ^2_{SB} difference tests ($\Delta\chi^2_{SB}$; Satorra & Bentler, 1999) in which significant values may have suggested reduced model fit.

Internal Consistency—We assessed internal consistency reliability with Cronbach’s α.

Ethnic Identity Across Groups—If measurement invariance holds, then groups may be compared. Taking a CFA approach, we tested for equality of latent factor means by constraining means to be equal to those of a reference group and evaluating any decrement in model fit.

Results

Preliminary Analyses

We examined psychometric properties of the MEIM-R by modality to determine the equivalency of mail and online forms. CFA results and descriptive statistics are available online as supplementary tables. Briefly, the correlated two-factor structure demonstrated the best fit in each modality group according to key goodness-of-fit indices. Multiple-groups CFAs suggested measurement invariance, i.e., equal factor structure, factor loadings, and item intercepts. Internal consistency in each modality was good; all subscale and overall scale values for Cronbach's α were near or above .80. We therefore combined data from the two modalities in all analyses.

Factor Structure

CFA goodness-of-fit indices for models in the entire sample revealed meaningful differences, although all χ^2_{SB} values were significant (Table 1, Models 1–3). A one-factor model showed adequate fit according to the SRMR but poor fit according to the RMSEA and CFI. An uncorrelated two-factor model showed overall poor fit. A correlated two-factor model fit best in the entire sample (Figure 1), showing good fit according to the SRMR and CFI, adequate fit according to the RMSEA, and the best, i.e., lowest AIC value relative to other models.

Measurement Invariance

We addressed measurement invariance across Asian, Black/African American, Hispanic, multiethnic, and White racial and ethnic groups. The correlated two-factor structure exhibited the best fit in each group, demonstrated by good fit based on the SRMR and CFI, adequate fit based on the RMSEA, and the best, i.e., lowest AIC values (Table 1, Models 4a–4e). A model simultaneously fitting the correlated two-factor structure in the five groups exhibited good fit based on the SRMR and CFI and adequate fit based on the RMSEA (Model 5). This and models for each racial and ethnic group suggested equal factor structure. A model with factor loadings held equal across groups (Model 6) resulted in little decrease in model fit, as demonstrated by a CFI decrement of only 0.005 and a positive reduction in AIC of 3.41; these indicators provided initial evidence of equal factor loadings, tempered by a significant χ^2_{SB} difference test, $\Delta\chi^2_{SB}(16) = 26.94$, $p = .04$. Next, we held item intercepts equivalent (Model 7) resulting in a further CFI decrement of only 0.006 and a minor increase in AIC of 3.16, but a significant χ^2_{SB} difference test, $\Delta\chi^2_{SB}(16) = 35.52$, $p = .003$. These results provided some preliminary evidence of equal item intercepts and measurement invariance across the five racial and ethnic groups.

At a granular level, we examined measurement invariance across three Asian subgroups with substantial sample sizes ($n > 50$): Chinese, Filipina, and South Asian. The correlated two-factor structure demonstrated the best fit of three competing models in each subgroup, generally fitting well or adequately according to the selected goodness-of-fit indices (Table

1, Models 11a–11c). These models and one simultaneously fitting the correlated two-factor structure suggested equal factor structure (Model 12). Goodness-of-fit indices for models with factor loadings (Model 13) and item intercepts (Model 14) held equal across groups signaled little decrease in fit given CFI decrements of 0.009 and 0.01, and minor AIC differences of 0.30 and 1.27, respectively; however, a significant χ^2_{SB} difference test for Model 14, $\Delta\chi^2_{SB}$ (8) = 15.68, p = .04 suggested only provisional evidence of measurement invariance across Asian subgroups.

Internal Consistency

The MEIM-R showed adequate internal consistency for the exploration and commitment subscales and the overall scale. All values for Cronbach's α were near or above .70 (Table 2).

Ethnic Identity Across Groups

Observed scores appear in Table 2. Given suggestions of measurement invariance, we evaluated equality of latent factor means for exploration (E) and commitment (C). Across racial and ethnic groups, White participants served as the reference group given that lower scores would facilitate the interpretation of relative scores. Starting with Model 7 (Table 1), we held factor means equal across groups (Models 8 and 9); these models fit more poorly, e.g., resulting in CFI decrements of 0.038 and 0.044, and AIC increases of 121.98 and 144.03, respectively. We examined which groups differed using Model 7, in which latent means were set at zero for White participants and freed for minority groups. Here, predicted latent means for minority groups and absolute effect sizes (Cohen's d) represented differences relative to White participants. Multiethnic participants had similar predicted means (E = .11, d_E = .21; C = −0.07, d_C = .11), Hispanic participants moderately higher means (E = .30, d_E = .48; C = .42, d_C = .57), and Asian (E = .52, d_E = .52; C = .64, d_C = .53) and Black/African American participants (E = .55, d_E = 1.86; C = .72, d_C = 2.03) the highest means. A "summary" model equivalent to Model 7 but where latent means were set at zero and held equal for White and multiethnic, freed for Hispanic, and held equal for Asian and Black/African American participants resulted in a CFI decrement of 0.002 and minor AIC increase of 0.751, yet significant χ^2_{SB} (Model 10). Similar to Model 7, the results implied that White and multiethnic participants appeared to have similar and the lowest means; Hispanic participants had moderately higher means (E = .27; C = .45) than White (d_E = .45; d_C = .64) and multiethnic participants (d_E = .36; d_C = 52); and Asian and Black/African American participants had similar and the highest means (E = .49; C = .67). Effect sizes were moderate comparing Asian to White (d_E = .56; d_C = .64) and multiethnic participants (d_E = .49; d_C = .57), large comparing Black/African American to White (d_E = 4.01; d_C = 4.59) and multiethnic participants (d_E = 2.84; d_C = 3.26), and small comparing Hispanic to Asian (d_E = .21; d_C = 17) and Black/African American participants (d_E = .26; d_C = .21).

In similar analyses among Asian subgroups, Chinese participants served as the reference group given relatively low MEIM-R scores. Starting with Model 14 (Table 1), we held factor means equal across groups (Models 15 and 16); these models fit more poorly, e.g., resulting in CFI decrements of 0.024 and 0.021, and AIC increases of 20.84 and 16.99, respectively. Model 14, in which latent means were set at zero for Chinese and freed for

Filipina and South Asian participants, suggested that South Asian had similar predicted means to Chinese participants (E = .03, d_E = .06; C = 0.03, d_C = .04), while Filipina participants had moderately higher means (E = .25, d_E = .46; C = .35, d_C = .44). A "summary" model in which latent means were set at zero and held equal for Chinese and South Asian and freed for Filipina participants fit fairly well, e.g., resulting in a CFI decrement of 0.002 and positive AIC reduction of 3.77, but a significant χ^2_{SB} (Model 17). Similar to Model 14, results implied that Chinese and South Asian participants had similar and lower means while Filipina participants had moderately higher means (E = .24; C = .34) than Chinese (d_E = .50; d_C = .51) and South Asian participants (d_E = .48; d_C = .49).

Discussion

Researchers and clinicians require a brief, easily administered measure of ethnic identity that is valid and reliable in diverse groups. To our knowledge, this study is the first to examine the psychometric properties of the MEIM-R, including measurement invariance, across multiple separate racial and ethnic groups in a community-based, primarily non-student sample. Like Phinney and Ong (2007) and Yoon (2011), we found that a correlated two-factor model in which ethnic identity was comprised of two distinct but related factors, exploration and commitment, fit the data fairly well. Key multiple-groups CFA goodness-of-fit indices indicated provisional evidence of measurement invariance, i.e., equal two-factor structure, factor loadings, and item intercepts across Asian, Black/African American, Hispanic, multiethnic, and White racial and ethnic groups. This extends limited prior research and suggests that the MEIM-R could be used to measure ethnic identity across multiple broadly assessed racial and ethnic groups. Additional granular analyses suggested that the MEIM-R measurement model could be considered equivalent among Chinese, Filipina, and South Asian subgroups. However, we observed inconsistencies between goodness-of-fit indices and chi-square tests which may have arisen from the sensitivity of chi-square to overall sample size and unbalanced sample sizes across groups being compared (Brown, 2006). Thus, these findings remain to be replicated and expanded.

With provisional evidence of measurement invariance suggesting that observed group differences in ethnic identity might be valid and meaningful, we examined levels across racial and ethnic groups. White and multiethnic women appeared to have similar and the lowest levels, while Hispanic women had moderately higher levels and Asian and Black/African American women had similar and substantially higher levels. Effect sizes ranged from small (e.g., comparing Hispanic with Asian participants) to large (e.g., comparing Black/African American with White participants). Similar to results reported by Yoon (2011), the general pattern of group differences was consistent with the notion that ethnic identity tends to be more salient and a more important component of self-concept among people of color than among people of White ethnic backgrounds (Phinney, 1996), particularly in the U.S. where minority group membership has been connected with common experiences of discrimination and oppression (Cokley, 2007). Of interest, multiethnic participants reported lower levels of ethnic identity than Asian, Black/African American, and Hispanic participants. Given the growth of the multiethnic population (e.g., Lopez, 2001), examining people who identify with specific racial and ethnic combinations would advance the emergent literature on the social and cultural experiences of multiethnic

Americans (for a discussion see Rockquemore, Brunsma, & Delgado, 2009). Also of interest, Chinese and South Asian participants appeared to have approximately equal and lower levels of ethnic identity than Filipina participants, differences which reached the medium effect size threshold. Investigation among additional Asian subgroups is needed.

Strengths of the present study include the large overall sample size and high diversity, which permitted separate analyses in multiple groups. This study also addressed a primarily non-student sample, extending research among adults and enhancing the measure's generalizability. However, limitations of our data assessment tool and restricted subsample sizes meant we were unable to address numerous granular subgroups. For example, we were unable to separate the diverse South Asian category into more ethnically uniform subgroups. The relatively low prevalence of GDM among Black/African American women (Ferrara, Kahn, Quesenberry, Riley, & Hedderson, 2004) resulted in a small subsample size. Similarly, we were limited in our ability to examine heterogeneity by immigration status. Finally, given the sample of women with GDM, it is unknown whether the results are applicable to women of other ages, at varying stages in the life course, or without a significant medical condition.

Future research should address men, older adults, and heterogeneity within racial and ethnic groups; examine equivalence across immigration statuses and levels of acculturation; evaluate non-English language versions; confirm findings in other regions, where attitudes about diversity and ethnic identity may differ; and examine concurrent validity with measures such as the Scale of Ethnic Experience (Malcarne, Chavira, Fernandez, & Liu, 2006) in specific groups. The brief MEIM-R could be easily incorporated into longitudinal and multi-occasion sampling research to investigate intra-individual variations in ethnic identity over the lifespan and across social contexts, two domains identified as priorities for ongoing research (Cokley, 2007; Ong, et al., 2010; Smith & Silva, 2011). Finally, future studies could use the MEIM-R to clarify relations between ethnic identity and health outcomes such as depression and lifestyle behaviors.

In summary, this study demonstrated good psychometric properties of the MEIM-R and provisional evidence of measurement invariance across specific racial and ethnic groups in a large sample of adult pregnant women. Given its psychometric and practical advantages, researchers and clinicians seeking to understand the salience of ethnicity may find the MEIM-R a useful tool. Pending replication, these findings support its use in diverse samples and suggest that results may be meaningful, rather than artifacts of differential measurement performance.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by grant R01 HS019367 to Assiamira Ferrara from the Agency for Healthcare Research and Quality. Assiamira Ferrara was also supported by grant P30 DK092924 from the National Institutes of Health. We thank the GEM trial participants and research staff, as well as Andrew Karter and Michaela Kiernan for their comments on earlier drafts of this manuscript.

References

American College of Obstetricians and Gynecologists. Committee opinion no. 504: Screening and diagnosis of gestational diabetes mellitus. Obstetrics and Gynecology. 2011; 118:751–753. [PubMed: 21860317]

Avery DR, Tonidandel S, Thomas KM, Johnson CD, Mack DA. Assessing the Multigroup Ethnic Identity Measure for measurement equivalence across racial and ethnic groups. Educational and Psychological Measurement. 2007; 67:877–888.

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

Brown, TA. Confirmatory factor analysis for applied research. New York: Guilford; 2006.

Cachelin FM, Phinney JS, Schug RA, Striegel-Moore RM. Acculturation and eating disorders in a Mexican American community sample. Psychology of Women Quarterly. 2006; 30:340–347.

Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. Structural Equation Modeling. 2002; 9:233–255.

Cokley K. Critical issues in the measurement of ethnic and racial identity: A referendum on the state of the field. Journal of Counseling Psychology. 2007; 54:224–234.

Dandy J, Durkin K, McEvoy P, Barber BL, Houghton S. Psychometric properties of multigroup ethnic identity measure (MEIM) scores with Australian adolescents from diverse ethnocultural groups. Journal of Adolescence. 2008; 31:323–335. [PubMed: 17645938]

Erikson, E. Identity: Youth and crisis. New York: Norton; 1968.

Espinosa-Hernández G, Lefkowitz ES. Sexual behaviors and attitudes and ethnic identity during college. Journal of Sex Research. 2009; 46:471–482. [PubMed: 19288336]

Ferrara A, Kahn HS, Quesenberry CP, Riley C, Hedderson MM. An increase in the incidence of gestational diabetes mellitus: Northern California, 1991–2000. Obstet Gynecol. 2004; 103:526–533. [PubMed: 14990417]

Gaines SO Jr. Bunce D, Robertson T, Wright B, Goossens Y, Heer D, Minhas S. Evaluating the psychometric properties of the Multigroup Ethnic Identity Measure (MEIM) within the United Kingdom. Identity: An International Journal of Theory and Research. 2010; 10:1–19.

Gordon, NP. Similarity of the adult Kaiser Permanente membership in Northern California to the insured and general population in Northern California. Oakland, CA: Kaiser Permanente Division of Research; 2012.

Henrickson HC, Crowther JH, Harrington EF. Ethnic identity and maladaptive eating: Expectancies about eating and thinness in African American women. Cultural Diversity and Ethnic Minority Psychology. 2010; 16:87–93. [PubMed: 20099968]

Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling. 1999; 6:1–55.

Institute of Medicine Subcommittee on Standardized Collection of Race/Ethnicity Data for Healthcare Quality Improvement. Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement. Washington, DC: National Academies Press; 2009.

Lopez, A. Race and Ethnicity in California: Demographics Report Series. Stanford: Center for Comparative Studies in Race and Ethnicity; 2001. The population of two or more races in California (Report No. 4).

Malcarne VL, Chavira DA, Fernandez S, Liu P. The Scale of Ethnic Experience: Development and psychometric properties. Journal of Personality Assessment. 2006; 86:150–161. [PubMed: 16599789]

Marcia, J. Identity in adolescence. In: Adelson, J., editor. Handbook of adolescent psychology. New York: Wiley; 1980. p. 159-187.

Markus, HR.; Moya, PML., editors. Doing race. New York: W. W. Norton; 2010.

Ong AD, Fuller-Rowell TE, Phinney JS. Measurement of ethnic identity: Recurrent and emergent issues. Identity: An International Journal of Theory and Research. 2010; 10:39–49.

Phinney JS. The Multigroup Ethnic Identity Measure: A new scale for use with diverse groups. Journal of Adolescent Research. 1992; 7:156–176.

Phinney JS. When we talk about American ethnic groups, what do we mean? American Psychologist. 1996; 51:918–927.

Phinney JS, Ong AD. Conceptualization and measurement of ethnic identity: Current status and future directions. Journal of Counseling Psychology. 2007; 54:271–281.

Ponterotto JG, Mallinckrodt B. Introduction to the special section on racial and ethnic identity in counseling psychology: Conceptual and methodological challenges and proposed solutions. Journal of Counseling Psychology. 2007; 54:219–223.

Quintana SM. Racial and ethnic identity: Developmental perspectives and research. Journal of Counseling Psychology. 2007; 54:259–270.

Roberts RE, Phinney JS, Masse LC, Chen YR, Roberts CR, Romero A. The structure of ethnic identity of young adolescents from diverse ethnocultural groups. The Journal of Early Adolescence. 1999; 19:301–322.

Rockquemore KA, Brunsma DL, Delgado DJ. Racing to theory or retheorizing race? Understanding the struggle to build a multiracial identity theory. Journal of Social Issues. 2009; 65:13–34.

Rogers Wood NA, Petrie TA. Body dissatisfaction, ethnic identity, and disordered eating among African American women. Journal of Counseling Psychology. 2010; 57:141–153. [PubMed: 21133566]

Satorra, A.; Bentler, P. Department of Statistics Papers. University of California, Los Angeles: Department of Statistics, UCLA; 1999. A scaled difference chi-square test statistic for moment structure analysis.

Schwartz SJ, Weisskirch RS, Zamboanga BL, Castillo LG, Ham LS, Huynh Q-L, Cano MA. Dimensions of acculturation: Associations with health risk behaviors among college students from immigrant families. Journal of Counseling Psychology. 2011; 58:27–41. [PubMed: 21171742]

Shuttlesworth ME, Zotter D. Disordered eating in African American and Caucasian women: The role of ethnic identity. Journal of Black Studies. 2011; 42:906–922. [PubMed: 22073427]

Siegel JM, Yancey AK, McCarthy WJ. Overweight and depressive symptoms among African-American women. Preventive Medicine: An International Journal Devoted to Practice and Theory. 2000; 31:232–240.

Smith TB, Silva L. Ethnic identity and personal well-being of people of color: A meta-analysis. Journal of Counseling Psychology. 2011; 58:42–60. [PubMed: 21171745]

Snipp, CM. Defining race and ethnicity: The Constitution, the Supreme Court, and the Census. In: Markus, HR.; Moya, PML., editors. Doing Race. New York: W. W. Norton; 2010. p. 105-122.

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

Trimble JE. Prolegomena for the connotation of construct use in the measurement of ethnic and racial identity. Journal of Counseling Psychology. 2007; 54:247–258.

Yoon E. Measuring ethnic identity in the Ethnic Identity Scale and the Multigroup Ethnic Identity Measure-Revised. Cultural Diversity and Ethnic Minority Psychology. 2011; 17:144–155. [PubMed: 21604838]

Zamboanga BL, Schwartz SJ, Jarvis LH, Van Tyne K. Acculturation and substance use among Hispanic early adolescents: Investigating the mediating roles of acculturative stress and self-esteem. The Journal of Primary Prevention. 2009; 30:315–333. [PubMed: 19408121]

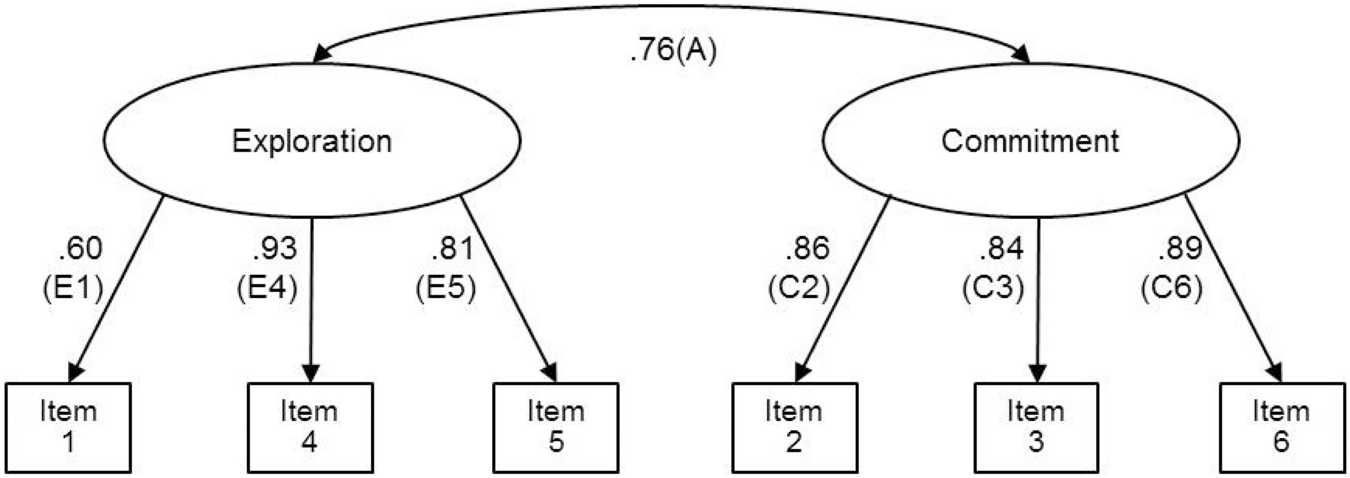

Figure 1. Confirmatory factor analysis model for the MEIM-R within the GEM trial analytic sample

Completely standardized parameter estimates for a correlated two-factor model of ethnic identity in the entire sample, *N* = 1,463. Path labels in parentheses correspond to estimates presented in Table 2. MEIM-R item descriptions are available in Phinney & Ong (2007).

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 1

CFA goodness-of-fit indices for the MEIM-R in the GEM trial analytic sample

Model	χ^2_{SB} *(df)*	SRMR	RMSEA [90% CI]	CFI	AIC
Models in the entire sample (N = 1,463)					
1. One-factor	508.55 (9)***	.08	.20 [.18–.21]	.82	21110.28
2. Uncorrelated two-factor	672.15 (9)***	.29	.22 [.21–.24]	.76	21260.16
3. Correlated two-factor	68.99 (8)***	.03	.07 [.06–.09]	.98	20391.31
Models across racial and ethnic groups					
4. Fit of correlated two-factor model					
4a. Asian (*n* = 630)	32.60 (8)***	.03	.07 [.05–.10]	.98	8255.73
4b. Black/African American (*n* =58)	6.34 (8)	.04	.00 [.00–.13]	1.0	846.53
4c. Hispanic (*n* = 240)	21.57 (8)**	.04	.08 [.04–.13]	.97	3535.77
4d. Multiethnic (*n* = 160)	16.14 (8)*	.03	.08 [.02–.14]	.98	2248.46
4e. White (*n* = 375)	15.44 (8)	.03	.05 [.00–.09]	.99	5280.44
5. Equal factor structure	94.29 (40)***	.03	.07 [.05–.09]	.98	20166.93
6. Equal factor loadings	123.07 (56)***	.06	.06 [.05–.08]	.98	20163.52
7. Equal item intercepts	158.52 (72)***	.06	.06 [.05–.08]	.97	20166.68
8. Equal factor means: Exploration	268.45 (76)***	.12	.09 [.08–.11]	.93	20288.66
9. Equal factor means: Commitment	285.58 (76)***	.13	.10 [.09–.11]	.93	20310.71
10. Equal factor means: Summary	166.44 (76)***	.07	.06 [.05–.08]	.97	20167.43
Models across Asian subgroups					
11. Fit of correlated two-factor model					
11a. Chinese (*n* = 168)	26.59 (8)***	.05	.12 [.07–.17]	.95	2239.39
11b. Filipina (*n* = 165)	15.95 (8)*	.05	.08 [.01–.13]	.96	2116.57
11c. South Asian (*n* = 144)	6.74 (8)	.02	.00 [.00–.09]	1.00	2066.74
12. Equal factor structure	47.40 (24)**	.04	.08 [.05–.11]	.97	6422.70
13. Equal factor loadings	62.78 (32)***	.08	.08 [.05–.11]	.96	6423.40

Model	χ^2_{SB} (*df*)	SRMR	RMSEA [90% CI]	CFI	AIC
14. Equal item intercepts	78.46 (40)***	.07	.08 [.05–.10]	.95	6422.67
15. Equal factor means: Exploration	100.31 (42)***	.12	.09 [.07–.12]	.93	6443.51
16. Equal factor means: Commitment	97.84 (42)***	.12	.09 [.07–.12]	.93	6439.66
17. Equal factor means: Summary	79.20 (42)***	.07	.08 [.05–.10]	.95	6418.90

Note. χ^2_{SB} = Satorra–Bentler scaled chi-square; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; CI = confidence interval; CFI = Bentler comparative fit index; AIC = Akaike Information Criterion.

* $p < .05$,

** $p < .01$,

*** $p < .001$.

Table 2

CFA parameter estimates, means, standard deviations, and Cronbach's alphas for the MEIM-R in the GEM trial analytic sample

		CFA completely standardized parameter estimates[a]							Exploration subscale		Commitment subscale		Overall scale	
			Exploration			Commitment								
Group	*n*	*A*	*E1*	*E4*	*E5*	*C2*	*C3*	*C6*	*M (SD)*	α	*M (SD)*	α	*M (SD)*	α
Entire sample	1,463	.76	.60	.93	.81	.86	.84	.89	3.13 (0.95)	.82	3.68 (0.91)	.90	3.41 (0.83)	.88
Asian	630	.71	.47	.92	.77	.88	.85	.84	3.37 (0.83)	.76	3.94 (0.79)	.89	3.65 (0.71)	.85
Chinese	168	.70	.28	.98	.72	.95	.86	.85	3.21 (0.78)	.69	3.81 (0.89)	.91	3.51 (0.72)	.83
Filipina	165	.66	.40	.96	.62	.85	.84	.79	3.55 (0.78)	.71	4.16 (0.67)	.86	3.85 (0.61)	.79
South Asian	144	.70	.59	.88	.86	.78	.79	.83	3.27 (0.92)	.81	3.84 (0.81)	.84	3.56 (0.77)	.86
Black/African American	58	.56	.59	.84	.95	.81	.78	.93	3.36 (0.97)	.80	4.04 (0.85)	.87	3.70 (0.78)	.83
Hispanic	240	.78	.67	.89	.81	.77	.79	.88	3.10 (0.98)	.84	3.73 (0.89)	.86	3.41 (0.85)	.88
Multiethnic	160	.80	.72	.97	.79	.85	.83	.90	2.98 (1.04)	.87	3.27 (0.92)	.89	3.13 (0.90)	.91
White	375	.72	.65	.91	.82	.88	.81	.90	2.78 (0.95)	.83	3.33 (0.93)	.90	3.06 (0.84)	.88

Note. Means and standard deviations represent observed scores.

[a]Labels correspond to paths in Figure 1.