Investigating the Structure and Measurement Invariance of the Multigroup Ethnic Identity Measure in a Multiethnic Sample of College Students

Stevie C. Y. Yap and M. Brent Donnellan
Michigan State University

Seth J. Schwartz
University of Miami

Su Yeong Kim
The University of Texas at Austin

Linda G. Castillo
Texas A&M University

Byron L. Zamboanga
Smith College

Robert S. Weisskirch
California State University, Monterey Bay

Richard M. Lee
University of Minnesota–Twin Cities

Irene J. K. Park
Indiana University

Susan Krauss Whitbourne
University of Massachusetts Amherst

Alexander T. Vazsonyi
University of Kentucky

In this article, we evaluate the factor structure of the Multigroup Ethnic Identity Measure (MEIM; Phinney, 1992) and test whether the MEIM exhibits measurement invariance across ethnic groups taken from a diverse sample of students from 30 different colleges and universities across the United States (N = 9,625). Initial analyses suggested that a bifactor model was an adequate representation of the structure of the MEIM. This model was then used in subsequent invariance tests. Results suggested that the MEIM displayed configural and metric invariance across 5 diverse ethnic groups (i.e., White, Black, Hispanic, East Asian, and South Asian). There were indications that the MEIM displayed a similar factor structure with roughly equivalent factor loadings across diverse ethnic groups. However, there was little evidence of scalar invariance across these groups, suggesting that mean-level comparisons of MEIM scores across ethnic groups should be interpreted with caution. The implications of these findings for the interpretation and use of this popular measure of ethnic identity are discussed.

Keywords: ethnic identity, measurement invariance, multigroup ethnic identity measure, MEIM, factor structure

Thoughts and feelings about one’s ethnic group membership and its importance to the self are referred to as ethnic identity (Phinney & Ong, 2007). Ethnic identity is considered to be an important individual characteristic, particularly for members of ethnic minority groups (e.g., Phinney & Alipuria, 1990). Ethnic identity is associated with aspects of positive adjustment including self-esteem (Kiang, Yip, Gonzales-Backen, Witkow, & Fuligni, 2006), subjective well-being (Yap, Settles, & Pratt-Hyatt, 2011), and reduced depressive symptoms (Settles, Navarrete, Pagano, Sidanius, & Abdou, 2010). One of the most
widely used measures of ethnic identity is the Multigroup Ethnic Identity Measure (MEIM; Phinney, 1992; Roberts et al., 1999). This self-report measure was designed to be used with members of different ethnic groups under the premise that it captures aspects of ethnic identity that are equally applicable across ethnic groups. However, relatively little research has tested this assumption using rigorous methods for evaluating measurement invariance. The goal of the present study was to test this assumption in a large and diverse sample of college students.

The Multigroup Ethnic Identity Measure (MEIM)

Some of the earliest measures of ethnic identity in the literature were constructed for use with specific ethnic groups (e.g., Suinn, Rickard-Figueroa, Lew, & Vigil, 1987). One consequence of using group-specific measures of ethnic identity is that variation in scale content across groups prevented researchers from making direct cross-group comparisons. In an attempt to address this concern, Phinney (1992) developed the MEIM as a general measure of ethnic identity that could be used with diverse ethnic groups. The MEIM was designed to assess general aspects of ethnic identity that were presumed to be relevant across all ethnic groups, which includes exploration of one’s ethnicity, a sense of belonging to one’s ethnic origin group, and involvement in the cultural practices of one’s ethnic group (see Table 1). The introduction of the MEIM coincided with an increase in the amount of research in ethnic identity and, as a consequence, a substantial expansion of the literature (Schwartz et al., 2014). The potential ability to measure ethnic identity in multiple groups with the MEIM is one reason why this measure plays an important role in ethnic identity research. The popularity of the MEIM makes a clear understanding of its structure and psychometric properties an important empirical issue with relevance to a number of literatures.

Phinney’s (1989, 1992) original theoretical model drew largely from an Eriksonian (1968) framework for identity development, and conceptualized the MEIM to have a two-dimensional underlying structure. These dimensions, often labeled exploration and commitment/affirmation (Phinney & Ong, 2007; Roberts et al., 1999), have been widely discussed as critical processes of identity development (e.g., Marcia, 1966, 1980; Schwartz, Donnellan, Rafter, Luyckx, & Zamboanga, 2013). Identity exploration is characterized by active efforts to discover and learn about one’s own ethnic group and to participate in the traditions and cultural practices of that group. Identity commitment is characterized by positive affirmation of one’s own ethnic group membership and a sense of belonging or attachment to that group. This two-dimensional structure for the MEIM has received empirical support using diverse samples from the United States (e.g., Pegg & Plybon, 2005; Spencer, Icard, Harachi, Catalano, & Oxford, 2000; Yancey, Anshenelson, & Driscoll, 2001) and other countries (e.g., Dandy, Durkin, McEvoy, Barber, & Houghton, 2008).

Although this two-factor model for the MEIM has empirical and conceptual support, alternative models for this instrument have also been reported. Some research has found a one-factor structure (e.g., Ponterotto, Gretchen, Utsey, Stracuzzi, & Saya, 2003; Reese, Vera, & Paikoff, 1998; Worrell, 2000; Worrell, Conyers, Mpofo, & Vandiver, 2006) or a three-factor structure (e.g., Gaines et al., 2010; Juang & Nguyen, 2010; Lee & Yoo, 2004). Thus, there is debate over the precise factor structure of the MEIM. Some researchers have suggested that discrepancies might be explained by the fact that various samples included different kinds of participants in terms of ethnic group membership, age, and other demographic differences. For instance, Ong, Fuller-Rowell, and Phinney (2010) suggested that discrepancies in the factor structure of the MEIM across studies could be due to differences in the contexts from which samples are drawn (e.g., United States vs. United Kingdom; Gaines et al., 2010). Indeed, many prior MEIM studies have included only a single ethnic group, and Ong et al. (2010) have theorized that members of different ethnic and demographic groups may conceptualize their ethnic identities in different ways, thereby producing some of the different factor structures found in the prior literature.

If members of different ethnic groups conceptualize their ethnic identities in distinct ways, then it might be unreasonable to expect the same set of items to measure this construct in a similar way across diverse groups. To be sure, administering the same set of items across groups does not guarantee that the same constructs are actually being assessed in each group. It may be that an item could be a stronger (or weaker) indicator of ethnic identity in one group as opposed to another. Consider the MEIM item that asks about involvement in organizations that include mostly members of one’s own ethnic group (i.e., Item 2; Roberts et al., 1999). In the United States, for example, an organization that consists of mostly Asian individuals would likely be construed as a culturally related organization, and affiliation with such an organization would likely reflect one’s level of ethnic identity. In contrast, some organizations in the United States consist of mostly White individuals (due to the majority status of this ethnic group), and membership in many of these groups likely has little or nothing to do with identifying with one’s ethnic background or cultural group. Thus, for White individuals, this item would likely be a much poorer indicator of ethnic identity.

Measurement Invariance and the MEIM

Ultimately, it is necessary to empirically evaluate the degree to which the MEIM measures the same construct in the same way across different ethnic groups. This is the process of establishing measurement invariance, and this process is vital to the use and interpretation of any measure that is used across diverse groups (see e.g., Chen, 2008; Schmitt & Kuljanin, 2008). The issue of measurement invariance draws on the critical distinction between constructs and measures (or indicators) of constructs. Constructs are unobservable theoretical variables, whereas the responses to measures are directly observed in a given study. Issues of measurement invariance broadly concern whether the psychometric connections between measures and constructs are the same across different groups.

There are various forms of measurement invariance (Chen, 2008; Schmitt & Kuljanin, 2008). The baseline and least psychologically demanding kind of invariance is referred to as configural invariance and involves testing whether scores on a measure are best represented by the same number of latent factors and general pattern of free and fixed factor loadings across different groups. This form of invariance is relevant to the debates as to whether the MEIM is best characterized by two factors as opposed to one or three in different ethnic groups. This level of invariance
is required before other more rigorous forms of invariance are considered.

**Metric invariance** (or weak invariance) involves testing whether the factor loadings for the items on their respective factor are equal across groups. A finding of metric invariance indicates that the strength of the association between each individual item and the corresponding latent factor is equal across groups. In essence, evidence of metric invariance suggests that a measure ranks individuals from different groups in the same way. This level of invariance is required for comparing the correlations between the MEIM and other variables across groups. Metric invariance is a stricter form of invariance than configural invariance; however, an even more rigorous form of invariance is required to conduct psychometrically meaningful comparisons of scale means across groups.

**Scalar invariance** (or strong invariance) involves testing whether item intercepts are equivalent across groups. Evidence of scalar invariance essentially indicates that individuals with the same underlying level of the construct in question have equivalent observed item scores (on average) regardless of group membership (see Schmitt & Kuljanin, 2008; Vandenberg & Lance, 2000). This level of invariance is necessary if one is interested in comparing scale means across groups (Chen, Sousa, & West, 2005). There are even more restrictive forms of invariance of which involving testing whether item-specific residuals can be constrained to the same value across groups (so called strict invariance). Such evidence indicates whether scores have equal precision across different groups. This level of invariance is often difficult to achieve in basic research and is probably unnecessary in applications that do not involve the use of cut scores or some form of selection or classification based on observed scores (Little, 2013).

Only a handful of studies have evaluated different levels of invariance of MEIM across different ethnic groups (e.g., Avery, Tonidandel, Thomas, Johnson, & Mack, 2007; Gaines et al., 2010; Roberts et al., 1999; Spencer et al., 2000; Yancey et al., 2001). Unfortunately, these studies have yielded inconsistent findings. For example, Roberts et al. (1999) examined the invariance of the MEIM across samples of White, African American, and Mexican American adolescents using confirmatory factor analysis. Using the two-factor exploration and commitment model, Roberts et al. found that the structure of the scale differed across ethnic groups (i.e., they did not find evidence of configural invariance). However, Spencer et al. (2000) found that a two-factor structure of the MEIM was similar across White, monoracial minority and multracial minority adolescents, although the sizes of the loadings differed across groups. This finding indicated that there was configural but not metric invariance. Last, Avery, Tonidandel, Thomas, Johnson, and Mack (2007) found evidence of configural and metric invariance across groups of White, African American, Hispanic, and Asian American adults, using Roberts et al.’s (1999) two-factor model.

Taken as a whole, previous studies are inconsistent with respect to the kind of invariance that the MEIM demonstrated across ethnic groups. This inconsistency raises critical questions about the applicability of the MEIM in research with diverse samples. For example, Roberts et al. (1999) interpreted the observed differences in the factor structure of the MEIM for White Americans compared with other ethnic groups as suggesting that ethnic identity may be less clearly defined for Whites due to their dominant status in American society. Roberts et al. regarded the lower MEIM scores for Whites as further support for this interpretation. However, from a psychometric perspective, it is inappropriate to compare mean levels if there are indications that scalar invariance does not hold (Chen et al., 2005). The safest interpretation in such cases is to conclude that the measure captures different constructs in the different groups. Indeed, recent articles have acknowledged that comparisons of mean MEIM scores across ethnic groups are inappropriate without establishing scalar invariance (e.g., Gaines et al., 2010; Ong et al., 2010). However, scalar invariance of MEIM scores across diverse groups has not been empirically examined for the MEIM in the published literature to date.

### The Present Study

Given the importance of the MEIM for the study of ethnic identity, the primary goal of the present study was to assess the invariance of the MEIM across a diverse sample of college-age students.
young adults in the United States. Given the conflicting evidence regarding the underlying structure of this measure, we evaluated the appropriateness of a one-factor, two-factor, three-factor, and a bifactor model of the MEIM to identify which of these approaches would serve as a reasonable baseline model to use in evaluating invariance of this measure across ethnic groups.

The one-factor, two-factor and three-factor models were evaluated because they have been identified and used in the past literature (e.g., Lee & Yoo, 2004; Roberts et al., 1999). We also evaluated a novel bifactor model for the MEIM. This model specified that each item loads on a general ethnic identity factor and one of two orthogonal substantive factors following Roberts et al.’s (1999) exploration/commitment model. The general factor in this model reflects the component of each indicator that is shared among all items—which includes both substantive variance (e.g., general ethnic identity) and shared method variance. A schematic diagram of the bifactor model is shown in Figure 1. This bifactor approach is advantageous because it distinguishes variance shared across all items from variance specific to the exploration and commitment dimensions. Further, this bifactor approach can be useful for evaluating the unique correlates of multifaceted constructs (Chen, Hayes, Carver, Laurenceau, & Zhang, 2012). This concern is relevant to the MEIM, given that a moderate to large correlation is typically found between the exploration and commitment scales in previous studies (e.g., Roberts et al., 1999). The bifactor model provides a tool for isolating the unique aspects of exploration and commitment and may therefore contribute to a further understanding of the correlates of ethnic identity in future studies.

In sum, the present study helps advance the literature on the assessment of ethnic identity in a number of ways. First, we tested a range of models for the MEIM including a novel application of a bifactor model. Second, and perhaps most important, we examined the invariance of the MEIM across a greater number of ethnic groups than has been considered in most previous measurement invariance studies using the MEIM. In particular, this study includes White, Black, Hispanic, East Asian, South Asian, and Middle Eastern participants. Third, we used one of the largest samples in the literature to test the invariance of the MEIM (over 9,700 participants). The MEIM continues to be the most widely used measure of ethnic identity in this area of research. Thus, researchers need to know more about the psychometric properties of the MEIM.

Method

Participants

The present sample consisted of 9,756 college students (72.8% women) with a mean age of 20.30 years (SD = 3.37). Participants in this sample self-identified as belonging to one of the following ethnic groups: White (n = 5,998), Black (n = 832), East Asian (n = 1,023), Hispanic (n = 1,465), South Asian (n = 307), and Middle Eastern (n = 131). Participants were drawn from a large, multisite data set collected at 30 colleges and universities from diverse geographic regions across the United States. There was substantial variation in the type of institutions sampled, ranging from large public universities to small private colleges. Participants were recruited using various methods, including institutional subject pools, printed flyers, and e-mail invitations. Study participants were given partial course credit or provided entry into a prize drawing. At each site, participants were directed to a single study website and completed all questionnaires online. All data were collected between September 2008 and October 2009.

The Multi-Group Ethnic Identity Measure

Each participant responded to the 12-item version of the MEIM (Roberts et al., 1999). Participants responded to each item using a 5-point Likert-type scale ranging from 1 (strongly disagree) to 5 (strongly agree).

Analytic Approach

To identify the most appropriate baseline structure to use in our tests of invariance, we used confirmatory factor analysis (CFA) in Mplus Version 5.0 (Muthén & Muthén, 2007). We then used multigroup CFA to evaluate the measurement invariance of MEIM scores across the six ethnic groups evaluated in this study. This
multigroup CFA procedure for testing invariance (see Cheung & Rensvold, 2002; Schmitt & Kuljanin, 2008; Vandenberg & Lance, 2000) consists of a series of analyses that evaluate increasingly restrictive levels of measurement invariance. The first step evaluates evidence for configural invariance. Testing configural invariance in CFA involves the specification of a single baseline factor model across the various groups. The requirement is that the number of factors and the particular items that load on to each factor are similar across groups. This model serves as a starting point for subsequent tests of more rigorous forms of invariance.

The next step tests for metric invariance, which involves comparing the fit of the configural invariance model to the fit of a nested model in which the factor loadings are constrained equal across groups. The absence of a meaningful decrement in fit between the configural invariance model and the more constrained metric invariance model would suggest metric invariance. Scalar invariance is evaluated by adding constraints on item intercepts and comparing model fit to the metric invariance model in which only the factor loadings are constrained to be equal across groups. Finding that the metric and scalar invariance models do not differ meaningfully in fit is taken as evidence in favor of scalar invariance.

Although the steps for evaluating measurement invariance in CFA are relatively clearly defined in the literature (e.g., Cheung & Rensvold, 2002), the appropriate criteria that should be used to evaluate decrement in fit among the nested models are less well established. Given the well-known dependence of the chi-square statistic on sample size, several researchers have questioned the appropriateness of the chi-square statistic as an indicator of model fit or as an indicator of change in fit between nested models (e.g., Cheung & Rensvold, 2002). Thus, it is often argued that other fit indices besides the chi-square test should be reported and considered to evaluate practical measurement invariance. This issue of practical invariance is typically investigated by evaluating differences in alternative indices of model fit (e.g., the comparative fit index or the root-mean-square error of approximation) when testing particular sets of parameter constraints (Cheung & Rensvold, 2002; Meade, Johnson, & Braddy, 2008).

There is, however, a lack of consensus regarding the amount of relative change in model fit statistics that would demonstrate a practically meaningful decrement in fit between more restrictive invariance models. Cheung and Rensvold (2002) suggested that it is useful to report change in three fit indices: comparative fit index (CFI), McDonald’s noncentrality index (Mc), and gamma hat. These authors suggested that a change in CFI (i.e., ΔCFI) greater than .01, a change in Mc (i.e., ΔMc) greater than .02, and a change in gamma hat greater than .001 among nested invariance models would indicate a meaningful decrement in fit between invariance models. In more recent research, Meade et al. (2008) recommended stricter criteria. For instance, they suggested that a ΔCFI greater than .002 between invariance models would be indicative of meaningful decrement in model fit.²

It is clear that there is variation regarding the appropriate criteria in evaluating whether measurement invariance exists (see also Chen, 2008). Because our study is one of the first comprehensive examinations into the invariance of the MEIM across different ethnic groups, we adopted a more liberal perspective in testing measurement invariance. Thus, we followed the recommendations of Cheung and Rensvold (2002) and evaluated ΔCFI and ΔMc to evaluate different forms measurement invariance of the MEIM.³ In addition to the CFI and Mc, we also reported the root-mean-square error of approximation (RMSEA), standardized root-mean-square residual (SRMR), Akaike information criterion (AIC), and Bayesian information criterion (BIC). Following the recommendations of Hu and Bentler (1999; but see Marsh, Hau, & Wen, 2004, for a discussion of the limitations of “golden rules” when interpreting misfit in CFA contexts), we adopted the following approximate cutoff values for evaluating “good fit” for these fit indices: CFI > .95; Mc > .90; RMSEA < .06; SRMR < .08.

### Results

#### Identifying the Baseline Model

Correlations among items and descriptive statistics for each item are presented in Table 1. Preliminary analyses indicated that 9.9% of the sample participants did not provide responses to at least one of the MEIM items. Among participants with missing data, participants were missing responses for a mean of 5.62 items (SD = 5.21; mdn = 1.00; min = 1; max = 12). All subsequent models were estimated using full information maximum likelihood. Moreover, descriptive analyses indicated that most MEIM items were moderately skewed and kurtotic,⁴ and thus all subsequent analyses reported here used robust parameter estimation procedures (i.e., maximum-likelihood robust estimator, or MLR, in Mplus).

Given debates over the structure of the MEIM, we used CFA to test four potential MEIM factor structures in the overall combined sample (N = 9,756). In the one-factor model, all items were specified as indicators of a single latent ethnic identity factor. In the two-factor model, items were specified to reflect either exploration (Items 1, 2, 4, 8, and 10) or commitment (Items 3, 5, 6, 7, 9, 11, and 12) following the specifications of Roberts et al. (1999). In the three-factor model, items were specified as indicators of three correlated factors—cognitive clarity (Items 3, 6, and 7), affective pride (Items 5, 9, 11, and 12), and behavioral engagement (Items 1, 2, 4, 8, 10, and 11) following Lee and Yoo (2004). Last, the bifactor model was used to isolate the Roberts et al. (1999) exploration and commitment items from a general identity factor as depicted in Figure 1.

Goodness-of-fit indicators for each baseline structural model are shown in Table 2. These data indicate that there is less support for a one-factor model compared with the other models evaluated here. Although Roberts et al.’s (1999) two-factor model is the most consistent with the prevalent theories in this area of research and is the most widely used model for this measure in past literature, this model had the worst fit relative to the bifactor and three-factor models (e.g., the AIC and BIC values were the largest of the three models). Model fit was relatively better for the three-factor model and the bifactor model, suggesting that there is empirical support

---

² Meade et al. (2008) suggested using condition-specific ΔMc criteria, rather than a general cutoff value. However, their suggested critical values are stricter than those suggested by Cheung and Rensvold. For instance, the cut-off value for a two factor, 12-item measure (i.e., Roberts et al.’s, 1999 model of the MEIM) would be .0066.

³ Gamma hat was not reported because Meade et al. (2008) suggest that it is redundant with CFI.

⁴ Additional information regarding the shape of item distributions is available upon request from the first author.
Invariance of the Bifactor Model

The complexity of the bifactor model led to initial convergence problems when specified as a multigroup CFA. However, once we respecified our model as a standardized model (i.e., where the latent variable variances were set to 1 and all factor loadings were freely estimated) and discarded the smallest ethnic group (Middle Eastern, \(n = 131\)), the multigroup bifactor model was able to converge.

Aside from the SRMR, all model fit indices would be conventionally interpreted as relatively poor in the configural model (see Table 5). All subsequent invariance tests used the bifactor model as a baseline and fit indices for each invariance model (i.e., configural, metric, scalar) are shown in Table 5.

Table 2
Fit Indices Across Models

<table>
<thead>
<tr>
<th>Model</th>
<th>(\chi^2) (df)</th>
<th>CFI</th>
<th>RMSEA</th>
<th>SRMR</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-factor</td>
<td>6739.968 (54)</td>
<td>.850</td>
<td>.114</td>
<td>.066</td>
<td>297472.947</td>
<td>297730.733</td>
</tr>
<tr>
<td>Two-factor</td>
<td>5093.527 (53)</td>
<td>.887</td>
<td>.100</td>
<td>.055</td>
<td>295006.118</td>
<td>295271.065</td>
</tr>
<tr>
<td>Three-factor</td>
<td>3710.181 (50)</td>
<td>.918</td>
<td>.088</td>
<td>.047</td>
<td>292904.057</td>
<td>293190.486</td>
</tr>
<tr>
<td>Bifactor</td>
<td>3623.303 (42)</td>
<td>.919</td>
<td>.095</td>
<td>.035</td>
<td>292114.584</td>
<td>292458.299</td>
</tr>
</tbody>
</table>

Note. All chi-square values were statistically significant. CFI = comparative fit index; RMSEA = root-mean-square error of approximation; SRMR = standardized root-mean-square residual; AIC = Akaike information criterion; BIC = Bayesian information criterion.

For these latter two models in this large and diverse sample. However, given the similarity of their fit indices, it is not clear empirically which of these models better characterizes the dimensional structure of the MEIM.\(^5\)

Ultimately, we favored the bifactor model to test for invariance across groups. This choice was based on several considerations. First, the bifactor model had superior model fit compared with the one- and two-factor models. Second, although the bifactor and three-factor models fit similarly, the bifactor model had the conceptual advantage of retaining a focus on the two dimensions of exploration and commitment given their theoretical importance in the literature (e.g., Roberts et al.’s, 1999, two-factor exploration/commitment model). Third, the bifactor model has utility as a method of isolating exploration and commitment from a general factor for use in future studies designed to evaluate the correlates of ethnic identity. In other words, the bifactor model provides researchers with a tool for studying the correlates of general ethnic identity separate from the unique aspects of exploration and commitment. Factor loadings for the overall CFA model and by ethnic group are shown in Tables 3 and 4. All subsequent invariance tests used the bifactor model as a baseline and fit indices for each invariance model (i.e., configural, metric, scalar) are shown in Table 5.\(^6\)

\(^5\) It is worth noting that we also replicated this pattern of results regarding the relative fit of these four models in another large diverse multisite sample of college students (\(N = 1,612\)).

\(^6\) CFIs were computed using an appropriate null model for multigroup invariance analysis (in which intercepts were constrained to equality across groups). The default, standard null model used to compute CFI in Mplus (which allows intercepts to be freely estimated across groups) is not nested within the scalar invariance model (Widaman & Thompson, 2003). We also tested invariance across ethnic groups using the two-factor and three-factor models. These analyses were consistent with the results reported here for the bifactor model and indicated that there was evidence of configural and metric invariance, but not scalar invariance. Our evaluations of partial scalar invariance for the two- and three-factor models were also consistent with our results for the bifactor model. Further details regarding these analyses are available from the first author.
Table 3
Bifactor Confirmatory Factor Analysis Loadings

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Item 3</td>
<td>0.17 0.71</td>
<td>0.20 0.85</td>
<td>0.12 0.71</td>
<td>−0.22 0.83</td>
<td>0.10 0.79</td>
<td>−0.08 0.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 5</td>
<td>0.67 0.50</td>
<td>0.65 0.45</td>
<td>0.64 0.55</td>
<td>0.06 0.72</td>
<td>0.53 0.58</td>
<td>0.49 0.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 6</td>
<td>0.47 0.70</td>
<td>0.60 0.78</td>
<td>0.35 0.81</td>
<td>−0.18 0.99</td>
<td>0.35 0.81</td>
<td>0.28 0.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 7</td>
<td>0.38 0.72</td>
<td>0.47 0.82</td>
<td>0.29 0.80</td>
<td>−0.38 0.95</td>
<td>0.18 0.80</td>
<td>0.07 0.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 9</td>
<td>0.39 0.71</td>
<td>0.45 0.75</td>
<td>0.29 0.70</td>
<td>0.23 0.89</td>
<td>0.51 0.65</td>
<td>0.45 0.76</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 11</td>
<td>0.22 0.84</td>
<td>0.31 0.98</td>
<td>0.09 0.90</td>
<td>0.10 1.03</td>
<td>0.31 0.87</td>
<td>0.32 0.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 12</td>
<td>0.43 0.66</td>
<td>0.45 0.61</td>
<td>0.28 0.67</td>
<td>0.26 0.77</td>
<td>0.52 0.58</td>
<td>0.61 0.75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 1</td>
<td>0.42 0.53</td>
<td>0.47 0.64</td>
<td>0.56 0.59</td>
<td>0.64 0.58</td>
<td>0.48 0.65</td>
<td>0.57 0.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 2</td>
<td>0.14 0.45</td>
<td>0.06 0.64</td>
<td>0.48 0.61</td>
<td>0.63 0.52</td>
<td>0.25 0.76</td>
<td>0.06 0.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 4</td>
<td>0.41 0.53</td>
<td>0.49 0.70</td>
<td>0.55 0.49</td>
<td>0.66 0.52</td>
<td>0.24 0.66</td>
<td>0.22 0.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 8</td>
<td>0.44 0.66</td>
<td>0.66 0.82</td>
<td>0.51 0.70</td>
<td>0.51 0.78</td>
<td>0.37 0.83</td>
<td>0.42 0.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Item 10</td>
<td>0.03 0.77</td>
<td>0.05 0.97</td>
<td>0.12 0.78</td>
<td>0.17 0.87</td>
<td>−0.06 0.84</td>
<td>0.14 0.93</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Standardized coefficients reported. All coefficients were significant, other than those in boldface. E/C column indicates loading onto substantive factor labeled exploration (E) or commitment (C). Middle Eastern participants were removed in the invariance testing stage due to the small sample size of this group, leading to problems with model convergence in the multigroup confirmatory factor analysis.

Discussion

The present study evaluated the structure and invariance of the MEIM in a diverse set of college students. In general, there seems to be evidence for configural and metric invariance of the 12-item version of the MEIM (Roberts et al., 1999) across five different ethnic groups of college students (i.e., White, Black, Hispanic, East Asian, South Asian) when the MEIM is modeled to have a bifactor structure. We found little evidence of overall scalar invariance and were not able to isolate a subset of scalar invariant items. If these results generalize to other samples, it suggests that the MEIM is useful for studying the correlates of ethnic identity in diverse groups, whereas it may not be suited for making mean-level comparisons across diverse groups.

However, our initial analyses indicate considerable ambiguity regarding the factor structure of the MEIM. More work is needed to clarify the dimensionality of this measure. For example, we ultimately retained and tested the bifactor model, but our analyses also showed that the three-factor model had a very similar fit. Although the two-factor model of exploration and commitment is consistent with previous theoretical work in the identity literature, the three-factor model (i.e., cognitive clarity, affective pride, and behavioral engagement; Lee & Yoo, 2004) is consistent with other theoretical models of ethnic identity (e.g., Umaña-Taylor, Yazedjian, & Bámaca-Gómez, 2004). However, the number of indicators per latent factor in the three-factor model is fairly low. To be sure, the MEIM is a relatively short measure, and it is possible that this small pool of items is being spread too thin when one attempts to specify a three (or even two) dimensional structure to the measure. Further work could be done to construct a pool of items that captures the broader theoretical concerns related to ethnic identity given evolutions in the current literature. Moreover, future work is needed to establish whether the dimensions captured by the MEIM have varying nomological nets—that is, to determine whether constructs like exploration and commitment are differentially associated with important outcomes. Indeed, one of the advantages of the bifactor model is that it isolates exploration and commitment from a general ethnic identity dimension. This distinction makes it possible to establish the unique correlates of these two latent dimensions (see Chen et al., 2012, for a discussion of the utility of the bifactor model for studying the correlates of multifaceted constructs). This point underscores the idea that future studies with the MEIM should be performed with structural equation modeling (SEM) approaches to account for a possible general dimension given the correlations between the latent factors in two- and three-factor models.

Indeed, it appears that there may be some value to modeling the MEIM with a bifactor structure in future research. This model fits better than the conventional two-factor model, yet retains
Roberts et al.’s (1999) conceptual distinction between exploration and commitment. The current results also suggest that the MEIM provides a similar underlying structure and similar factor loadings across diverse college students within the United States when modeled with this bifactor approach. Thus, we believe the MEIM has at least some utility as a multigroup measure. The key issue is that it may not be strictly appropriate to make mean-level comparisons across diverse groups using the MEIM given the absence of evidence for scalar invariance. Researchers may need to be restricted to considering the correlates of the MEIM across different groups.

Although these results suggest that future research may benefit from using the bifactor model, further work should be done to clarify what the latent factors within this model represent and their implications on the other constructs and outcomes. For instance, the degree to which the general factor reflects shared item response effects is unclear. The bifactor model also estimates the two latent substantive factors as orthogonal constructs, and it is unclear the extent to which these factors, which represent the unique aspects of exploration and commitment, reflect the notions of exploration and commitment that have been discussed in past research on ethnic identity (e.g., Roberts et al., 1999). Future empirical research is needed to clarify the nature of exploration and commitment as specified by the bifactor model considered here. However, these two dimensions are often treated as orthogonal theoretical constructs in the general identity literature (e.g., the 2 × 2 matrix used to generate different identity statuses as specified by Marcia, 1966, 1980), and thus, the bifactor specification may provide a cleaner way to test hypotheses about aspects of identity and identity development than alternative models.

Limitations and Conclusions

There are a number of methodological limitations to the current study that warrant discussion. The most notable is the use of a convenience sample of college student participants. It is possible that the psychological processes associated with the development and maintenance of ethnic identity in college students may be different from those found in other demographic groups and periods of the life span. Furthermore, feelings of belongingness, attachment, and pride to one’s ethnic group (ethnic identity commitment) may be more salient to non-White college students attending predominately White institutions than the general population of late adolescents (Castillo et al., 2006). However, identity issues are especially salient during the college years (Schwartz et al., 2013), and thus college student samples may be useful for testing associations between ethnic identity and psychosocial outcomes.

Another potential limitation of this study is that the sample was predominately female (72.8%). Thus, it is possible that the results garnered in the present analyses may not generalize to samples with higher proportions of male participants. Past research provides hints that ethnic-identity-related processes differ between men and women (e.g., Yap et al., 2011), and this is an important direction for future study. Accordingly, future studies should seek to generate more balanced samples in terms of the ratio of women and men.

We should also note that recent research by Phinney and Ong (2007) has produced a revised, six-item version of the MEIM. However, rigorous examinations of this revised measure’s equivalence across diverse ethnic groups have not yet been conducted. Four items of the revised scale are drawn directly from Roberts et al.’s (1999) 12-item version used in this study, but this recent revision includes two items with updated wording. Thus, we could not evaluate the revised scale with the present data. Nonetheless, the procedures used here could be used to test whether the factor structure of this measure is invariant across different groups. Whether this revised version of the MEIM displays a consistent dimensional structure among various samples remains an unanswered empirical question.

In conclusion, we have found evidence of some forms of invariance for the MEIM (i.e., configural, metric) among five diverse ethnic groups examined in our study. This finding means that the MEIM captures the construct of ethnic identity relatively similarly in these groups and suggests that the MEIM is a valuable tool for investigating and comparing the correlates of ethnic identity for members of ethnic minority college students in the United States. At the same time, the results of this study indicate that researchers should continue to evaluate whether there are differences in how various ethnic groups of college students interpret and respond to the items on the MEIM. Comparisons of MEIM scores across various groups should be performed with caution (and at the latent level of analysis) until more work has been done to identify and understand the potential differences in item intercepts among diverse groups. Likewise, the bifactor approach specified here might prove useful for clarifying the unique correlates of exploration and commitment.

References

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.


Received May 31, 2013
Revision received January 24, 2014
Accepted January 27, 2014