



Theses and Dissertations

2014-07-08

Reliability Generalization of the Multigroup Ethnic Identity Measure-Revised (MEIM-R)

Hayley Marie Herrington
Brigham Young University - Provo

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>



Part of the [Counseling Psychology Commons](#), and the [Special Education and Teaching Commons](#)

BYU ScholarsArchive Citation

Herrington, Hayley Marie, "Reliability Generalization of the Multigroup Ethnic Identity Measure-Revised (MEIM-R)" (2014). *Theses and Dissertations*. 4212.
<https://scholarsarchive.byu.edu/etd/4212>

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Reliability Generalization of the Multigroup Ethnic
Identity Measure–Revised (MEIM–R)

Hayley Marie Herrington

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of
Educational Specialist

Timothy Smith, Chair
Erika Feinauer
Derek Griner

Department of Counseling Psychology and Special Education
Brigham Young University

June 2014

Copyright © 2014 Hayley Marie Herrington

All Rights Reserved

ABSTRACT

Reliability Generalization of the Multigroup Ethnic Identity Measure–Revised (MEIM–R)

Hayley Marie Herrington

Department of Counseling Psychology and Special Education, BYU
Educational Specialist in School Psychology

The strength of ethnic identity has been linked to higher academic achievement, academic attainment, and overall psychological wellbeing. The Multigroup Ethnic Identity Measure (MEIM) was developed to measure ethnic identity development, and this measure has been used widely. The MEIM has since been updated to the Multigroup Ethnic Identity Measure–Revised (MEIM–R). The relatively new MEIM–R has not been widely studied. In order to solidify researchers' confidence in the measure, a reliability generalization analysis was performed and the data compared to the original MEIM. Random-effects weighted models evaluated internal consistency coefficients (Cronbach's alpha). The MEIM–R demonstrated adequate reliability coefficients (average $\alpha = .88$) and tended to produce more reliable data than the original MEIM (average $\alpha = .84$ across a sample of 75 studies). Reliability coefficients did not differ across participant age, gender, and race, but more reliability was generally obtained when participants had relatively higher levels of education. This study indicates that researchers and others may have confidence in using the MEIM–R, although future studies should specifically seek to confirm its validity.

Keywords: reliability, psychometrics, identity, ethnicity

TABLE OF CONTENTS

| | |
|---|-----|
| ABSTRACT..... | ii |
| TABLE OF CONTENTS..... | iii |
| LIST OF TABLES..... | iv |
| Introduction..... | 1 |
| Literature Review..... | 5 |
| Correlates of Ethnic Identity Relevant to the Work of School Psychologists..... | 7 |
| Academic achievement..... | 7 |
| Academic attainment..... | 9 |
| Psychological wellbeing..... | 10 |
| Measurement of Ethnic Identity..... | 10 |
| The Multigroup Ethnic Identity Measure (MEIM)..... | 11 |
| The Multigroup Ethnic Identity Measure–Revised (MEIM–R)..... | 12 |
| Measurement Reliability and Reliability Generalization..... | 13 |
| Method..... | 17 |
| Literature Search Procedure and Study Selection Criteria..... | 17 |
| Coding Procedures..... | 17 |
| Data Analysis..... | 18 |
| Comparison of Data from the Original MEIM..... | 19 |
| Results..... | 21 |
| Study Characteristics..... | 21 |
| Reliability Generalization Analyses..... | 23 |
| Analyses for Possible Publication Bias..... | 26 |
| Discussion..... | 27 |
| Limitations and Recommendations..... | 29 |
| Conclusion..... | 30 |
| References..... | 32 |
| APPENDIX A: Coding Sheet..... | 38 |

LIST OF TABLES

| | | |
|---------|---|----|
| Table 1 | <i>Descriptive Statistics For 46 Research Studies Reporting Reliability Coefficients For the MEIM–R.....</i> | 22 |
| Table 2 | <i>Random Effects Weighted Internal Consistency Coefficients (Cronbach’s α) for the MEIM–R and Original MEIM.....</i> | 24 |
| Table 3 | <i>Meta-regression of Participant Characteristics on Reliability Coefficients.....</i> | 25 |

Introduction

In the United States, ethnic diversity has been increasing. In 2008, the Non-White population represented 34% of the US population, compared to 20% in 1980. In 2025, it is expected that 42% of the US population will be represented by non-whites (National Center of Educational Statistics, 2009). Trends also show that children of color typically do not perform as well academically as their White peers. In 2007, 54% of Black 4th-graders, 50% of Hispanic 4th-graders, and 51% of American Indian/Alaska Native 4th-graders scored below the basic achievement level (National Center of Educational Statistics, 2009).

Over the past two decades psychologists and school psychologists have found that academic achievement and the psychological wellbeing of children and adolescences of color have been positively associated with strength of ethnic identity development. This link has been shown across different groups including race, age, and gender (Fuligni, Witkow, & Garcia, 2005; Martinez & Dukes, 1997; Umaña-Taylor, Vargas-Chanes, Garcia, & Gonzales-Backen, 2008; Yasui, Dorham, & Dishion, 2004). For example, given the growing number of ethnic minority students and the apparent disparity in academic achievement, ethnic identity is an important construct to measure since it is linked with academic achievement in children and adolescents of color.

Ethnic identity has been defined in several ways, but most scholars have settled on the conceptualization offered by Jean Phinney (2000):

A dynamic, multidimensional construct that refers to one's identity, or sense of self, in ethnic terms, that is, in terms of a subgroup within a larger context that claims a common ancestry and shares one or more of the following elements: culture, race, religion, language, kinship, or place of origin. Ethnic identity is not a fixed categorization, but

rather a fluid and dynamic understanding of self and group that changes with age, time, and context. (p. 254-255)

According to this definition, ethnic identity grows and develops within each individual. With its basis in understanding the self in relation to the group, ethnic identity has a valuable role in adolescent development (Erikson, 1994; Marcia, 1966). Adolescence is a time of identity exploration and ethnic identity is one area that is linked to strong psychological wellbeing in individuals of color.

Given the importance of ethnic identity, it is not a surprise that there are many instruments that measure ethnic identity (Bernal, Knight, Garza, & Ocampo, 1990; Resnicow, Soler, Braithwaite, Selassie, & Smith, 1999; Umaña-Taylor, Yazedjian, & Bámaca-Gómez, 2004). Issues of ethnic identity measurement have also proliferated throughout the past two decades.

By far, the most frequently used instrument is the Multigroup Ethnic Identity Measure (MEIM) (Phinney, 1992), one of the first measures designed to assess ethnic identity for multiple ethnic groups. Since then, the MEIM has been used to measure the ethnic identity formation and development in people across many ethnic groups, ages, and genders, becoming the dominant instrument in the field (Phinney & Ong, 2007). The original MEIM consisted of 14 items assessing three components of ethnic identity: attachment or belonging, achieved identity, and ethnic behaviors. In 1999, Roberts, Phinney, Masse, Chen, Roberts, and Romero performed an exploratory factor analysis finding the two items assessing ethnic behaviors on the MEIM could be removed. Over the next several years, researchers used various versions of the MEIM, removing items in many instances because of validity concerns (e.g., Roberts et al.) or to fit the designs of studies. Use of the MEIM became so widespread that it also became unsystematic

and to some degree overgeneralized. Given several concerns of validity that have arisen over time regarding the use of the original MEIM (and its variants), the original authors revised the MEIM in 2007 (Phinney & Ong, 2007). The resulting MEIM–R consists of six items representing a two-factor model. After being in use for over seven years, the psychometric properties of the MEIM–R require further analysis.

An instrument is as useful as it is psychometrically sound. Reliability is an important aspect of measurement and a precursor to measuring validity. It is important to measure and report reliability because it may be an indicator of how much error there is in a test score. In its basic sense, reliability is the consistency of scores, meaning how well the items on a single test correlate together across multiple samples. When the intercorrelations are higher, the test can be said to be measuring something. In order to determine if the MEIM–R is a valid measure of ethnic identity, adequate reliability needs to be established first. Evaluating the reliability of the newly developed MEIM–R would allow researchers and practitioners to have greater confidence in its use.

In 1998, Vacha-Haase introduced a modern method for systematically evaluating data reliability, termed *reliability generalization*. Reliability generalization uses meta-analytic methods and is similar to validity generalization approaches that have been used since the 1970s (Vacha-Haase, 1998). Thus this method is the equivalent of a meta-analysis of reliability coefficients. It evaluates a broad set of data with a particular instrument to see under what conditions (populations, settings, etc.) the reliability coefficients generalize. The MEIM and subsequent MEIM–R is designed to measure ethnic identity across ethnicities, ages, and other variables. With so many variables that have the potential to affect reliability, it would be

valuable to understand which variables influence reliability. Reliability generalization does just that. It is a method perfectly suited to evaluate the newly developed MEIM-R.

This study evaluates the reliability generalization of data from the MEIM-R and compares it to data from the original version of the MEIM. It is hypothesized that the MEIM-R will have adequate reliability coefficients (averaging above 0.80) that generalize across participant composition (age, gender, and race) and across study settings (schools, clinics, etc.). This information will be useful to practitioners and researchers evaluating the construct of ethnic identity in our increasingly multicultural society.

Literature Review

Ethnic identity is often considered in light of the broader concept of identity development. Identity development refers to the individuals' construction of the self as a unique entity, which may or may not account for one's relationships with others (Erikson, 1994). Scholars have theorized identity development occurring in distinct stages throughout a person's life. These stage-models have become central to the study of identity development throughout the years.

One of the first scholars to develop a theory of identity was Sigmund Freud (1923/1961). In his theory of psychosexual development, Freud laid out five stages of development that lead to the development of the ego. Erik Erikson built on Freud's theory and explored identity development, particularly in adolescence, in his theory of psychosocial development. The adolescent conflict of *identity vs. role confusion* is according to Erikson, the initial formation of identity (1994). In this phase, adolescents reach identity commitment or stay in a period of identity confusion based on personal exploration. An individual who has achieved identity commitment is said to possess a strong sense of self and independence (Erikson, 1994).

James Marcia elaborated on Erikson's ideas and developed four possible outcomes in the process of identity achievement: *foreclosure*, *identity diffusion*, *moratorium*, and *identity achievement*. *Foreclosure* refers to an individual with apathy toward a sense of identity; they would not have explored nor committed to individual roles. An individual in *identity diffusion* would assume the roles and expectations of others for themselves. Though they have committed to certain roles, they have not explored their options. *Moratorium* refers to an individual in the process of exploring their sense of self. They may have had an identity crisis, which led them to question and explore possible options. A person who has achieved a sense of self is said to have

reached *identity achievement*. They have explored their options and made a commitment to their sense of identity. Like Erikson, Marcia states the processes of exploration and commitment determine an individual's outcome in identity development (Marcia, 1966).

Researchers like William Cross (1971) and Janet Helms (1990) expanded identity theory even further by incorporating race into identity development. Racial identity refers to “a sense of group or collective identity based on one's perception that he or she shares a common racial heritage with a particular racial group” (Helms, 1990, p. 3). An emphasis is placed on the perception of the individual. Racial identity theory is, in other words, an explanation of the varying degrees to which an individual identifies with his or her race.

Racial identity theory is based on group or social identity in that it examines an individual's perception of membership in a group. Like ethnic identity, racial identity is based on a social psychological perspective (Tajfel & Turner, 1986), as well as a developmental perspective. An important distinction should be made here between race and ethnicity. Ethnicity is defined as a “classification of individuals who share a unique social and cultural heritage (customs, language, religion, and so on) passed on from generation to generation” (Casas, 1984, p. 787). Therefore, ethnicity, and ethnic identity, is a broad concept that encompasses many groups. For example, the Encyclopedia of Psychology (Burlew, 2000) differentiates racial identity as “primarily to describe the idiosyncratic experiences unique to African Americans as a result of the rich African heritage as well as the unique legacy of slavery, oppression, and colonialism” (p. 259). In other words, ethnic identity is primarily used to explain broad ethnic groups, while racial identity is more specific in that it is used to explain African-American heritage. Thus, while there appears to be similarities between racial and ethnic identity, the purpose and research have differed.

Throughout the years, ethnic identity was defined in many ways before reaching a clear theoretical based definition. The variability in the definitions of ethnic identity leads to confusion. In 2007, the *Journal of Counseling Psychology* published an edition of the journal directly aimed at this problem. In that special issue, Phinney and Ong (2007) elaborated on the development of ethnic identity theory. Specifically, they described the empirical components of ethnic identity: self-categorization, commitment and attachment, exploration, behavioral involvement, evaluation, in-group attitudes, ethnic values and beliefs, importance or salience of group membership, and national identity. These components have been identified in the research, as opposed to exclusively theoretical based components, as relating and contributing to the overarching construct of ethnic identity. Clearly, ethnic identity is complex and multifaceted.

Correlates of Ethnic Identity Relevant to the Work of School Psychologists

Ethnic identity is related to many fields, including school psychology. The field of school psychology explores issues with students' mental health and academic success. Ethnic identity has been found to correlate with many of the same issues. There are three main influences of ethnic identity relevant to school psychologists: academic achievement, academic attainment, and psychological wellbeing.

Academic achievement. The relevance of ethnic identity to school psychology is seen in the link between ethnic identity and academic achievement. A few ideas have been proposed as to why this connection exists. One explanation is that students with a strong ethnic identity have rejected the negative stereotypes of their ethnicity, which may include negative attitudes about school. Their positive view of their ethnicity has translated into positive academic attitudes (Hogg, 2003). Another idea proposed by Phinney (1990) is that students with a strong ethnic identity give more meaning to their goals and aspirations, including academic goals. This in turn

leads to higher academic achievement. Regardless of the cause of the link between ethnic identity and academic achievement, the importance is clear: a strong sense of ethnic identity may foster academic success.

Academic achievement has been measured in different ways throughout the literature. Some indicators of academic achievement include grade point average (GPA), standardized test scores, and at times students' academic self-concept (Fan & Chen, 2001). Academic achievement of minority students has received much attention in research, particularly in light of the "achievement gap" between students of color and their white peers. The achievement gap refers to the disparity of academic performance between differing groups of students, in this case different ethnic groups. For example, while high school GPAs have increased across all ethnic groups between 1990 and 2009, White students have continued to outperform Black and Hispanic students (NCES, 2009). This disparity has led many researchers to study academic achievement across different ethnic groups.

Given the research of the past few decades, it is clear that strength of ethnic identity is positively correlated with academic achievement across different ethnic groups (Martinez & Dukes, 1997; Umaña-Taylor et al., 2008; Yasui et al., 2004). Fuligni et al. (2005) were interested in whether the ethnic label students subscribed to mattered in academic achievement. They compared students identifying with national origin labels (i.e., Mexican, Chinese) versus students who identified with hybrid labels (i.e., Mexican-American, Chinese-American). They found that the label a student identifies with matters less than the strength of the identification. Indeed, the more the students identified with their ethnic background, regardless of which label they chose, the more likely they were to succeed academically. Altschul, Oyserman, and Bybee (2006) also noted this connection. They cite that while GPAs decrease across a student's

educational career, ethnic identity is still associated with better academic achievement, suggesting ethnic identity may buffer against even more of a decrease in grades (p. 1165). In brief, the research establishes the link between ethnic identity and academic achievement across many settings.

Academic attainment. Just as there is a link between ethnic identity and academic achievement, there is also a link with academic attainment. Academic attainment is defined as high school graduation status, college attendance, and highest level of education completed. One of the reasons for the increased rate of academic attainment for students with a strong sense of ethnic identity may be that these students also have a higher level of academic achievement. However, Chavous and colleagues (2003) found that for African-Americans, ethnic identity was more important for college enrollment than GPA. This may be because ethnic identity keeps students engaged and determined in school (Fulgini et al., 2005). Altschul et al. (2006) suggested that having a stronger sense of one's ethnic identity leads to a fortification against stereotype threat. Despite the reason behind the link between ethnic identity and academic attainment, the connection should be explored.

Since ethnic identity is so multifaceted, some scholars argue a strong sense of ethnic identity leads to positive outcomes only when specific beliefs are held. For example, it has been argued by Fordham and Ogbu (1986) that because of societal barriers, African-American youths may develop oppositional identities in regards to schooling. This belief purports the idea that ethnic identity in African-American youths leads to poorer outcomes in academics. However, ethnic identity is not the same as internalizing stereotypes (Steele & Aronson, 1995). In fact, it is possible to have a strong sense of ethnic identity and the belief that society values or devalues the group's ethnicity. It has been shown that as long as students have a strong sense of ethnic

identity, regardless of their perception of society, they have higher academic attainment than those without a strong sense of ethnic identity (Chavous et al., 2003). Furthermore, the research has primarily shown ethnic identity leads to better academic outcomes, thus the importance of ethnic identity development should not be overlooked.

Psychological wellbeing. It is generally agreed upon in the literature that ethnic identity is positively correlated with psychological wellbeing. In a recent meta-analysis (Smith & Sylva, 2011), results indicated a relationship between ethnic identity and psychological wellbeing. The link was strongest with adolescents and young adults. While no definite causation has been found, a few ideas have been proposed. Outten, Schmitt, Garcia, and Branscombe (2009) state ethnic identity may provide a sense of belonging to a larger group, promoting psychological wellbeing. Ethnic identity may also provide youth a sense of purpose and security and can foster psychological resilience and validation (Yip & Fulgni, 2002). Another idea is that ethnic identity buffers against the stress of discrimination (Mossakowski, 2003). While we don't know the cause of the link between ethnic identity and wellbeing, ethnic identity is positively linked to self-esteem, coping, sense of mastery, and optimism and negatively linked to depression and loneliness (Roberts et al., 1999). The fact that the meta-analysis identified over 150 research studies on the topic of ethnic identity and well-being would at the very least seem to suggest substantial interest in the topic. With so many research studies being conducted, the importance of investigating the psychometric properties of the most common measure of ethnic identity becomes obvious.

Measurement of Ethnic Identity

Given the many theories of ethnic identity, multiple measures have been developed and used in ethnic identity research. Some measures have been created for use with specific groups

such as African Americans (Smith & Brookins, 1997), Asians (Barry, 2002), and Mexican Americans (Der-Karabetian & Ruiz, 1997). One of the disadvantages of having multiple measures for each ethnicity is that the results of these measures cannot be compared. Seeing this difficulty, Phinney (1992) developed the MEIM with the belief that ethnic identity could be measured as a single construct across all ethnic groups. In 2004, Umaña-Taylor and colleagues also developed a universal ethnic identity measure, the Ethnic Identity Scale (EIS), based on three domains of ethnic identity: exploration, resolution, and affirmation (2004). However, given that the MEIM still remains the most widely used measure of ethnic identity our efforts were focused on analyzing this measure.

The Multigroup Ethnic Identity Measure (MEIM). The MEIM was designed to assess three aspects of ethnic identity: affirmation and sense of belonging, ethnic behaviors or practices, and ethnic identity achievement. The first two aspects, affirmation and sense of belonging and ethnic behaviors or practices, were derived from other measures assessing ethnic identity. The third component, identity achievement, had less of a research base from which it was derived. Identity achievement describes how developed a person's ethnic identity is, and its development was heavily influenced by various identity development theories. The MEIM measures identity achievement as a continuous variable with complete achievement suggesting a secure sense of self. The original scale consisted of 14 items representing the three aspects of ethnic identity: 5 items representing the Affirmation/Belonging subscale, 7 items for the Ethnic Identity Achievement subscale, and 2 items for the Ethnic Behavior subscale.

Phinney (1992) calculated the psychometric properties of the original MEIM. The reliability coefficients (Cronbach's alpha) for the entire 14-item scale was .81 for high school students and .90 for college students, indicating strong reliability for the original MEIM data

when used with those populations. Cronbach's alpha was calculated for each subscale as well, with results indicating acceptable reliability for each subscale. Factor analysis identified one factor for the MEIM suggesting a unified construct for ethnic identity as measured by the MEIM.

However, Roberts et al. (1999) suggested the MEIM consisted of two factors: exploration and commitment. After an exploratory factor analysis, two negatively worded items were found to not fit and were thus removed, resulting in a 12-item scale. The resulting scale consisted of five items representing exploration and seven items representing commitment. Several exploratory factor analyses followed (Spencer, Icard, Harachi, Catalano, & Oxford, 2000; Yancey, Aneshelsel, & Driscoll, 2003) and confirmed the two-factor model. Inconsistent factor loadings may indicate an issue with the validity of the measure. Thus, the conflicting factor loadings between the original factor analysis and the recent exploratory factor analyses led Phinney to consider a two-factor model of the MEIM and ultimately to consider revising the measure.

The Multigroup Ethnic Identity Measure–Revised (MEIM–R). Phinney and Ong (2007) reviewed the literature on the MEIM and made revisions to the measure in order to account for the conflicting findings in the factor analyses. The researchers conducted focus groups and interviews to assess face validity and content validity. As a result, the two ethnic behavior items were removed with the recognition that ethnic behaviors are a distinct construct from ethnic identity. The authors also revised the wording in some items to increase face validity. The most considerable change comes from the creation of two conceptually based subscales representing the two-factor model that was found in previous factor analyses. Each subscale, exploration and commitment, was created with an equal number of five items in order

to be equally weighted in analyses. This revised version of the MEIM was then tested to assess the psychometric properties.

In order to examine the reliability, a new study was carried out (Phinney & Ong, 2007) with 192 ethnically diverse college students. The 10-item measure yielded Cronbach's alpha of .83 for the exploration subscale and .89 for the commitment subscale, indicating good internal consistency (Cronbach, 1951). Both an exploratory, and then confirmatory, factor analysis indicated a two-factor model, supporting Phinney and Ong's 2007 conceptualization of ethnic identity in this measure. Through item-total correlations, two items in both exploration and commitment subscales were dropped from the test. The remaining 6-item measure (3 items for each subscale) makes up the current Multigroup Ethnic Identity Measure-Revised (MEIM-R). Phinney and Ong investigated the properties of this current measure, finding a Cronbach's alpha of .81, with .76 for the exploration subscale and .78 for the commitment subscale. The MEIM-R has now been in use for over seven years across different ethnic groups.

Measurement Reliability and Reliability Generalization

Reliability is usually referred to as the consistency of test scores. Classical Test Theory assumes that every test score has a true score and an observed score. The difference between the true score and the observed score is measurement error. Reliable test scores are said to have minimal measurement error, and the data is thought to be closer to true scores (Gall, Gall, & Borg, 2007). Conversely, unreliable results are comparable to completely random results, and can be said to measure nothing. The importance of reliability lies in its description to measure something rather than nothing (Vacha-Haase & Thompson, 2011).

While reliability demonstrates the measurement of something rather than nothing, what that something is remains unknown. Determining what a test is measuring is a study of validity.

Reliability is a necessary first step to determining validity as it determines there is something being measured. The original MEIM was revised because of concerns of validity, and this study is the first step to determine the effectiveness of that revision.

An important, and sometimes overlooked, aspect of reliability is that it is a property of the test's data, and not of the instrument itself. In other words, the results of a test can be reliable or unreliable, but the test itself remains the same. The same test can generate reliable scores in one instance and unreliable scores in another depending on the sample. Because reliability is a property of data, different features of studies (age, gender, ethnicity of participants) can affect variations in reliability (Vacha-Haase, 1998). For this reason, it is important to examine reliability for each sample of data and not just for an instrument.

There are different approaches to estimating reliability. These methods include inter-rater, split-half, and test-retest reliability. Inter-rater reliability measures the reliability between multiple raters, and is not applicable for administering the MEIM or MEIM-R. Split-half reliability takes the scores from half the test and compares it with the other half to determine consistency. Given the brevity of the MEIM and MEIM-R, split-half reliability would not be an appropriate measurement of reliability. Test-retest reliability compares scores from two separate administrations, and measures consistency over time. This type of reliability is thus better suited for constructs that are stable over time. As previously noted, ethnic identity is a fluid construct that changes within the individual over time. Because of this, test-retest reliability is not best suited for measuring reliability in the MEIM-R. In order to analyze the variables that produce the most reliable results in the MEIM-R, another measurement of reliability will be analyzed.

The most widely used measure of reliability is internal consistency. Internal consistency analyzes the consistency of individual test items by looking at how well they correlate with each

other. For example, if an individual answered that they enjoyed going to school and on another item said they felt happy in class, this would indicate strong internal consistency and those scores would be said to be reliable. Consistency of responses suggests that items can be combined because they are measuring the same thing.

The most widely used measure of internal consistency is Cronbach's alpha coefficient. Like other reliability coefficients, scores range between .00 and 1.00, with a score of 1.00 indicating perfect reliability and .00 indicating no reliability. Cronbach's alpha is the average of all the possible split-half coefficients for a test (Cronbach, 1951). It is similar to the Kuder-Richardson formula 20 (Kuder & Richardson, 1937), with the difference being that Cronbach's alpha is not limited to dichotomous variables. Since its creation in 1951, Cronbach's alpha has been reported more than any other statistic as an indicator of reliability and has over 6,500 citations in the literature (Sijtsma, 2009).

Acknowledging the variability of reliability coefficients across different datasets, researchers have explored ways to systematically analyze the variables that contribute to reliability. Vacha-Haase (1998) recognized that reliability scores changed across studies and set out to create a meta-analytic method for reliability, termed Reliability Generalization (RG). RG developed out of Schmidt and Hunter's (1977) validity generalization, a meta-analytic approach to validity. In RG, instead of effect sizes, reliability coefficients are analyzed. RG seeks to characterize "(a) the typical reliability of scores for a given test across studies, (b) the amount of variability in reliability coefficients for given measures, and (c) the sources of variability in reliability coefficients across studies" (Vacha-Haase, 1998, p. 6). This method is not widely used, with the first 12 years of RG garnering 47 studies, but it remains a tool to inform the use of popular instruments, such as the MEIM and the current MEIM-R.

Because the MEIM–R is a relatively new measure, an examination of its use across studies will be helpful for researchers. Using the RG method, researchers would be able to access the source of variation in reliability scores and the amount of variability in the MEIM–R’s reliability scores. With that information researchers will be better able to design studies measuring ethnic identity. Through a RG study, researchers will also be aware of the sources of the variability so that the design of studies can be made more appropriate to the study’s purposes.

Using RG methods, it is hypothesized that the MEIM–R data will have adequate reliability coefficients (averaging above 0.80) comparable or higher than those obtained for the original MEIM. It is also hypothesized that the reliability coefficients will be similar across participant composition (age, gender, and race) and across study settings (schools, clinics, etc.). With our increasingly multicultural society, it is important to understand factors that may influence multicultural students’ academics. For the researchers and practitioners working with students, having confidence in a common measure of ethnic identity can be very useful.

Method

Like a meta-analysis, a reliability generalization analyzes past studies. This section describes that process. First, how the studies were selected for analysis is discussed. Second, how those studies were coded for analysis is discussed. Then, a description of the data analysis is provided, and following that a brief description of how the original MEIM was analyzed is provided.

Literature Search Procedure and Study Selection Criteria

Articles that cited the original MEIM-R article by Phinney and Ong (2007) were found by using Google Scholar and EBSCO Publishing (PsycINFO and ERIC databases). Articles not written in English and research not conducted in the United States or Canada were excluded. By definition, only quantitative research would administer the MEIM-R, but all quantitative research designs were eligible for inclusion. In order to avoid the potential confound of repeated measures that would characterize longitudinal studies, data extraction was limited to first time administration within studies. For sake of consistency, only manuscripts reporting Cronbach's alpha coefficient were coded.

A literature search found 391 articles via Google Scholar that cited the article by Phinney and Ong (2007) that contained the description of the MEIM-R. Of those articles, 120 contained quantitative data collected in North America. Of those 120 articles, 46 reported Cronbach's alpha reliability coefficients for the MEIM-R with the data collected in that study. Manuscript searches were completed in October 2013.

Coding Procedures

Information reported in the obtained studies was coded to enable statistical analyses. The categorical variables coded were (a) publication status (published or unpublished), (b) other

variables measured in the same administration as the MEIM–R, (c) immigrant status of study participants, and (d) clinical or adjudicated status of the participants. Coded variables with continuous level data were (a) mean age and years of education of the participants, (b) number of participants in the sample, (c) race/ethnicity of participants, (d) percentage of female participants, (e) the number of other measures administered at the same time as the MEIM–R, and (f) the internal consistency reliability coefficient reported for scores on each scale/subscale used in the study. A copy of the coding sheet is presented in Appendix A.

To achieve sufficient inter-rater reliability, two trained raters coded articles independently and recorded each article. Discrepancies were resolved by the two raters discussing and reaching a consensus based on review of the manuscript. When a consensus could not be reached, the coders consulted with an expert in the field to resolve the issue.

Data Analysis

For this study, random-effects weighted models with alpha coefficients were used. Fisher's z transformation was performed on the data prior to analyses, and then following analyses the data were transformed back to the metric of a reliability coefficient for purposes of interpretation.

The main analysis ascertained the averaged reliability coefficient across all studies coded. However, variability was expected across studies. To determine the homogeneity of alpha coefficients, both Q and I^2 statistics were calculated. Both a scatterplot of alpha coefficients by number of participants and a fail-safe N were calculated as indicators of possible publication bias (unreliable data are most often characteristic of small- N studies, which are also less likely to be published).

To evaluate the possibility that study and participant characteristics moderate the reliability of alpha coefficients, two sets of subsequent analyses were conducted. Random effects weighted correlations were calculated between the continuous level data coded and the coefficient alpha. For example, it is possible that the average participant age could be associated with the reliability coefficients obtained within studies (perhaps with data obtained from younger participants being less reliable on average). Random effects weighted analyses of variance (ANOVAs) were calculated for the categorical level data coded and the alpha coefficients. For example, it was possible that data from student populations could be less reliable than data from general community members. These analyses will thus inform researchers who intend to use the MEIM–R about possible study and participant characteristics that they should consider when designing their research.

Comparison of Data from the Original MEIM

The resulting data from the MEIM–R was also compared to the reliability coefficients from the original MEIM using data from 75 research manuscripts included in a meta-analysis of the association between ethnic identity and well-being (Smith & Silva, 2011). The authors of that meta-analysis had identified published and unpublished studies on that topic by searching multiple electronic databases including PsycINFO and using the phrase *ethnic identity* crossed with descriptors of psychological well-being such as *mental health*, *self-esteem*, *happiness*, and *depression*. Inclusion criteria consisted of the study being written in English and providing quantitative data regarding well-being and participants' identification of African American, Asian American, Hispanic/Latino(a) American, Native American Indian, and/or Pacific Islander American. Two separate pairs of coders extracted reliability coefficients on the MEIM coded the articles with discrepancies resolved by further examination of the paper. Across pairs, inter-rater

agreement was acceptably high, with an average intraclass correlation coefficient of .87 for extraction of the reliability coefficient. For further information on the methods, see Smith and Silva (2011).

Results

The purpose of this study is to determine the variables that contribute to reliability variance, if any, in the MEIM–R. Another purpose was to compare the average reliability of the MEIM–R to the MEIM. In order to fulfill these aims, we analyzed the study characteristics and performed the reliability generalization. An analysis for possible publication bias was also performed.

Study Characteristics

A total of 46 studies were located that contained Cronbach's alpha coefficients on the MEIM–R or one of its two subscales. Reliability coefficients on the total MEIM–R were obtained across 41 unique samples within 37 studies; two studies reported total MEIM–R coefficients separately for two unique samples of participants, and one study reported coefficients separately for three unique samples of participants. Reliability coefficients from 26 unique samples on the three-item Exploration subscale were reported across 21 studies, and reliability coefficients from 27 unique samples on the three-item Commitment subscale were reported across 22 studies.

The studies located were both unpublished dissertations (46%) and published in peer-reviewed journals (54%). The average sample size across all studies was 264, with the greatest number of participants being 1,884 and the fewest number being 28. All of the participants were obtained from normal community samples or were students enrolled in schools, elementary through university. Immigrant status was reported in only 46% of the studies. Out of the 41 reported studies, 18 used samples of immigrants to the United States, one study used a sample of non-immigrants, and 22 were unreported. Across studies, participants were 65% female, 27% White/European American, 22% Hispanic/Latinos(as), 21% Asian American, 18%

Black/African American, 0.5% Native American, 0.06% Pacific Islander, and 11.6% reported as “other” ethnic minorities. The average age of all participants was 26.3 years (SD = 9.9), and the average level of education was 13.9 years (SD = 2.8). The mean level of ethnic identity measured on a 1-5 scale was 3.58 (SD = .46). These data are presented in Table 1.

Table 1

Descriptive Statistics for 46 Research Studies Reporting Reliability Coefficients for the MEIM–R

| Variable | M | Minimum | Maximum |
|---------------------------------|--------|---------|---------|
| Sample Size | 264.20 | 28 | 1884 |
| Age of Participants | 26.25 | 13 | 52 |
| Years of Education | 13.87 | 7 | 16 |
| Percent Female | 64.89 | 0 | 100 |
| Percent White/Caucasian | 26.82 | 0 | 100 |
| Percent Black/African American | 17.87 | 0 | 100 |
| Percent Hispanic/Latino(a) | 22.20 | 0 | 100 |
| Percent Asian American | 20.88 | 0 | 100 |
| Percent Native American Indian | 0.51 | 0 | 8 |
| Percent Pacific Islander | 0.06 | 0 | 2 |
| Percent “Other” Ethnic Minority | 11.60 | 0 | 100 |

Every study measured variables in addition to ethnic identity, with 63% of studies measuring more than one additional variable and with an average of four other variables measured. The most common variables measured were related to academic achievement and psychological well-being.

Reliability Generalization Analyses

The first analysis was specific to the coefficients reported for the full scale MEIM–R. The overall random effects weighted average reliability coefficient for the MEIM–R was .88, a very respectable value for a six-item measure considering the number of studies involved. The internal consistency coefficients of the Exploration and Commitment subscales were .84 and .86, respectively. The overall MEIM–R was statistically significantly more reliable than both of the subscale coefficients (see 95% confidence intervals in Table 2). Furthermore, the average reliability coefficient of the MEIM–R was statistically significantly higher than that of the original MEIM ($\alpha = .84, p < .05$), evaluated using coefficients obtained from 75 studies obtained from a previous meta-analytic review (Smith & Silva, 2011).

Subsequent analyses were conducted to ascertain whether differences in reliability coefficients were associated with participant characteristics. No differences were found across coefficients reported across different ethnic groups. Also, the percentage of participants from a specific ethnic group was not correlated with reliability coefficients within studies. Thus the MEIM–R yielded equivalent reliability coefficients across participant race/ethnicity. Similarly, no differences in reliability coefficients were found when considering (a) the number of other variables measured within the studies, (b) whether the participants were students or community members, and (c) whether the participants included immigrants to North America or not.

Reliability coefficients did significantly vary ($p < .05$) as a function of four variables: participants' average age, education level, gender (operationalized as percentage of females), and level of self-reported ethnic identity (mean MEIM–R total). Specifically, higher averaged reliability coefficients were obtained when (a) the samples were adults (the older the participants, the more reliable the data); (b) the samples were college educated (the greater number of years of

education, the more reliable the data); (c) the samples contained relatively more female participants (the more females in the sample, the more reliable the data); and (d) the samples reported relatively high levels of ethnic identity (the higher the average MEIM–R score, the more reliable the data).

Table 2

Random Effects Weighted Internal Consistency Coefficients (Chronbach's α) for the MEIM–R and Original MEIM

| Scale/Subscale | Mean ES | 95% CI | k |
|----------------|---------|-----------|-----|
| MEIM–R | .88 | .87 - .89 | 38 |
| Exploration | .84 | .81 - .86 | 27 |
| Commitment | .86 | .84 - .88 | 29 |
| Original MEIM | .84 | .83 - .85 | 75 |

These four variables could interact with one another (e.g., years of education is obviously related to participant age), so it was essential to evaluate the relative contribution of each variable in meta-regression models. Because only 32 studies reported MEIM–R total scores, a single model with all four variables would have restricted statistical power, so it was optimal to conduct two regression models: one without MEIM–R scores and the other with that data included, after removing variables that did not reach statistical significance in the first model.

Both random effects weighted models reached statistical significance ($p > .0001$). The first model that included three predictor variables (see Table 3) explained 39.7% of the variance in reliability coefficients across 40 studies. The only one of the three predictor variables to remain statistically significant was participant education (standardized beta = .50, $p = .006$).

Thus the univariate associations observed with the two variables of participant age and percentage of female participants were better accounted for by participant level of education.

The second model added participant level of ethnic identity (MEIM–R total scores) to participant level of education but excluded participant age and gender because those variables did not reach statistical significance in the first model. The combination of participant education and level of ethnic identity explained 38.5% of the variance in reliability coefficients across 32 studies. Only participant level of education remained statistically significant (standardized beta = .45, $p = .004$), although the association with level of ethnic identity contributed to the model and likely did not reach statistical significance due to the low power associated with only 32 cases (standardized beta = .273, $p = .08$).

Table 3

Meta-Regression of Participant Characteristics on Reliability Coefficients

| Variable | <i>B</i> | <i>SE</i> | <i>p</i> | β | R^2 | <i>k</i> |
|--------------------|----------|-----------|----------|---------|-------|----------|
| Model 1 | | | | | .397 | 40 |
| Constant | 0.780 | .110 | <.0001 | .00 | | |
| Level of Education | 0.029 | .011 | .006 | .50 | | |
| Age | 0.002 | .003 | .60 | .08 | | |
| Female | 0.002 | .002 | .30 | .15 | | |
| Model 2 | | | | | .385 | 32 |
| Constant | 0.537 | .221 | <.0001 | .00 | | |
| Ethnic Identity | 0.121 | .068 | .08 | .27 | | |
| Education | 0.027 | .009 | .004 | .45 | | |

Overall, reliability coefficients differed as a function of participant education. Higher reliability coefficients can be expected when the MEIM–R is administered to populations with greater levels of education. Nevertheless, the average reliability coefficients obtained on the MEIM–R with populations having lower levels of education remained above acceptable levels. No coefficient across 41 distinct samples was less than .71.

Analyses for Possible Publication Bias

To evaluate the possibility of publication bias influencing the findings just presented, we first compared the coefficients from published studies to those reported in unpublished studies. The results were very similar, and the small difference did not approach statistical significance ($p > .10$). The results of a “trim and fill” analysis (Duval & Tweedie, 2000) did not identify any possibly missing values in the data distribution. Moreover, Orwin’s fail-safe N was calculated to be 220 using a conservative estimate of Chronbach’s $\alpha = .70$ for “missing” studies. That is, over 220 studies with coefficients of .70 would have to be found in the literature for the present results to be reduced to an overall average of .75. Given that the lowest reliability coefficient observed was .71, it seems highly unlikely that 220 unpublished studies with coefficients of .70 would exist. Thus, publication bias was not a threat to the overall results.

Discussion

The purpose of this study was to assess the reliability of the newly developed MEIM–R. It was hypothesized that the MEIM–R would prove to be reliable across participant characteristics and across research settings. Random-effects weighted models with alpha coefficients were used to ascertain reliability generalization. The overall average reliability coefficient was very respectable at .88, with similar reliability coefficients for the two MEIM–R subscales as well.

It was also hypothesized the MEIM–R would have greater reliability than the original MEIM. The MEIM–R had been designed based on extensive analyses of MEIM data, and it was intended to improve the psychometrics of the original. The results from this study confirm that the MEIM–R resulted in data characterized by significantly greater internal consistency reliability than the original MEIM. Even with only six items, the MEIM–R appears to be an improvement upon the original MEIM both in terms of brevity of administration and reliability coefficients.

Because of the widespread use of the MEIM and now the MEIM–R, it was important to ascertain whether the instrument can yield reliable data across multiple settings and populations. The MEIM–R is designed to measure ethnic identity across different ethnicities, ages, and settings. Since it is possible that those variables could influence the scores obtained on the measure, it was important to understand those possible differences. The analyses indicated that the MEIM–R is equally reliable across several variables, including ethnicity. Data from participants of different ethnic groups yielded about the same averaged reliability coefficients. Similarly, student status, immigrant status, and whether the study was published or unpublished

did not seem to influence the findings. Thus the MEIM–R can be used with confidence across different ethnicities, ages, and settings.

Reliability coefficients did tend to differ as a function of participant gender, participant age, participant level of ethnic identity, and participant level of education. Because it is possible these variables may account for one another (e.g, the older a person is the more education they may have) we analyzed the variables simultaneously in order to determine which variables were significant. When these variables were considered simultaneously, only participant education remained statistically significant. That is, the participant’s education level had a positive relationship with the internal consistency of the measure in many studies. Therefore, the higher level of participants’ education, the more reliable the data obtained within studies. This may be due to the fact that individuals with higher education may have had made more decisions in regards to ethnic identity and respond differently than those without those opportunities. Regardless of the reasons, participant education level has shown to be a source of variability in the MEIM–R.

Establishing the reliability of the MEIM–R is important as a first step to assess the validity of the measure. Reliability is a necessary but insufficient for establishing validity. Reliability can be described as the consistency of scores, that is, how well the items on a single test correlate together. When scores tend to correlate together and are said to be reliable, as is the case in this study, it can be said that the test is measuring something. Determining the validity of the measured construct is beyond the scope of this analysis. This study determined that the MEIM–R indeed produces reliable data. Further studies are needed to assess validity.

As previously noted, reliability is a property of data, not of the measure itself. The MEIM–R has shown to produce reliable data across many settings and with many variables, with

level of education producing the most variability. Since ethnic identity has shown to be an important predictor of academic achievement, academic attainment, and overall psychological wellbeing, an accurate measure of ethnic identity is an important aspect of research. This study supports the MEIM–R as an important measure in ethnic identity research.

Limitations and Recommendations

As with meta-analyses, a reliability generalization is only as good as the studies it analyzes, and there are limitations to this particular study that should be addressed. First of all, 46% of the studies coded were unpublished doctoral dissertations. Studies that are unpublished are often not of the same caliber since they are not as rigorously reviewed as published articles. Because those articles may not be as high quality, it is possible that the results from those studies, and therefore this study, may not be as reliable. Therefore, it is possible that the results from this study may be even underestimating the reliability of the MEIM–R. Nevertheless, there were no differences in the average reliability coefficients obtained from published vs. unpublished studies, so any underlying differences may either be small or have evaded detection in this instance.

Another limitation is that three of the studies coded only reported subscale reliability for the MEIM–R, as opposed to reliability for the entire measure. When this occurred, we were unable to add that study to the overall analysis (although they were used when analyzing the subscales). There were potentially valuable data that were not utilized. It is therefore recommended that in the future researchers report reliability coefficients for the entire measure, as well as subscales.

One of the limitations to this study was the relatively small number of studies that were analyzed. Since 2007, only 41 studies were found that used the full MEIM–R. Many researchers

were still using the original MEIM, and others were only administering part of the measure. Given its brevity of only six items, researchers are advised to use the entire MEIM–R rather than only one three-item subscale. Researchers were also still using the original 12-item MEIM instead of the revised measure. Because the MEIM–R yielded higher average reliability coefficients than the original MEIM, researchers are encouraged to replace using the MEIM with the MEIM–R.

Another consideration concerns sample size. Only one study with a large number of participants (N=1,884) was located. Several of the studies located had fewer than 100 participants, yet studies with relatively few participants are more likely to be impacted by sampling error. In general, the greater number of participants, the more likely the results can be generalized to other populations.

A final limitation of the literature was that most of the participants were recruited using convenience samples of normal community members. It would be worthwhile to evaluate the reliability of the MEIM–R with different samples, such as clinical populations and at-risk groups. As it now stands, the reliability of the MEIM–R has been established for normal community members only.

Conclusion

Given the importance of ethnic identity to academic success and overall psychological well-being, it is valuable to have a measure that yields reliable data. The MEIM was developed to assess ethnic identity with participants from all ethnic backgrounds, and it became the most widely used measures of ethnic identity. As researchers used the MEIM, the construct validity of the measure was questioned. After some modifications, a new measure, the MEIM–R was developed with a two-factor model. Now in use for over six years, the MEIM–R was in need of

evaluation. In this reliability generalization study, the MEIM–R was shown to yield data with acceptable internal consistency reliability. Studies with participants having relatively higher levels of education yielded data with the strongest internal consistency, but the data from the MEIM–R was shown to be reliable across all populations evaluated to this point. Researchers and others may have confidence in using the MEIM–R, although future studies should specifically seek to confirm its validity.

References

- Altschul, I., Oyserman, D., & Bybee, D. (2006). Racial-ethnic identity in mid-adolescence: Content and change as predictors of academic achievement. *Child Development, 77*(5), 1155–1169. doi: 10.1111/j.1467-8624.2006.00926.x
- Barry, D. T. (2002). An ethnic identity scale for East Asian immigrants. *Journal of Immigrant Health, 4*(2), 87–94. doi:10.1023/A:1014598509380
- Bernal, M. E., Knight, G. P., Garza, C. A., & Ocampo, K. A. (1990). The development of ethnic identity in Mexican-American children. *Hispanic Journal of Behavioral Sciences, 12*(1), 3–24. doi: 10.1177/07399863900121001
- Burlew, A. (2000). Ethnic and racial identity: Racial identity. In A. E. Kazdin (Ed.), *Encyclopedia of Psychology, Vol. 3* (pp. 259–263). Washington, DC: American Psychological Association. doi: 10.1037/10518-123
- Casas, M. (1984). Policy, training and research in counseling psychology: The racial/ethnic minority perspective. In S. Brown & R. Lend (Eds.), *Handbook of counseling psychology* (pp. 785– 831). New York, NY: Wiley.
- Chavous, T. M., Bernat, D., Schmeelk-Cone, K., Caldwell, C. H., Kohn-Wood, L., & Zimmerman, M. A. (2003). Racial identity and academic attainment among African American adolescents. *Child Development, 74*(4), 1076–1090. doi: 10.1111/1467-8624.00593
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 97–334. doi: 10.1007/BF02310555
- Cross, W. (1971). The Negro-to-black conversion experience: Toward a psychology of liberation. *Black World, 20*, 13–26.

- Der-Karabetian, A., & Ruiz, Y. (1997). Affective bicultural and global-human identity scales for Mexican-American adolescents. *Psychological Reports, 80*(3, Pt 1), 1027–1039. doi: 10.2466/pr0.1997.80.3.1027
- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*, 455–463.
- Erikson, E. H. (1994). *Identity: Youth and crisis* [Google Book version]. Retrieved from <http://books.google.com>
- Fan, X., & Chen, M. (2001). Parental involvement and students' academic achievement: A meta-analysis. *Educational Psychology Review, 13*(1), 1–22. doi:10.1023/A:1009048817385
- Fordham, S., & Ogbu, J. (1986). Black students' school success: Coping with the burden of acting White. *Urban Review, 18*, 176–206.
- Freud, S. (1961). *The ego and the id*. In J. Strachey (Ed. and Trans.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 19, pp. 3–66). London, England: Hogarth Press. (Original work published 1923)
- Fulgini, A. J., Witkow, M., & Garcia, C. (2005). Ethnic identity and the academic adjustment of adolescents from Mexican, Chinese, and European backgrounds. *Developmental Psychology, 41*(5), 799–811. doi: 10.1037/0012-1649.41.5.799
- Gall, M. D., Gall, J. P., & Borg, W. P. (2007). *Education research: An introduction*. Boston, MA: Pearson Education.
- Helms, J. E. (1990). *Black and White racial identity: Theory, research, and practice*. Westport, CT: Greenwood Press.
- Hogg, M. (2003). Social identity. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 462–479). New York, NY: Guilford Press.

- Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, 2(3), 151–160. doi: 10.1007/BF02288391
- Marcia, J. E. (1966). Development and validation of ego-identity status. *Journal of Personality and Social Psychology*, 3(5), 551–558. doi: 10.1037/h0023281
- Martinez, R. O., & Dukes, R. L. (1997). The effects of ethnic identity, ethnicity, and gender on adolescent well-being. *Journal of Youth and Adolescence*, 26(5), 503–516. doi: 10.1023/A:1024525821078
- Mossakowski, K. N. (2003). Coping with perceived discrimination: Does ethnic identity protect mental health? *Journal of Health and Social Behavior*, 44(3), 318–331. doi: 10.2307/1519782
- National Center of Education Statistics. (2009). *IPED report: Nova Southeastern University* [Data file]. Retrieved from <http://www.nces.org/datasets/>
- Outten, H., Schmitt, M. T., Garcia, D. M., & Branscombe, N. R. (2009). Coping options: Missing links between minority group identification and psychological well-being. *Applied Psychology: An International Review*, 58(1), 146–170. doi: 10.1111/j.1464-0597.2008.00386.x
- Phinney, J. S. (1990). Ethnic identity in adolescents and adults: Review of research. *Psychological Bulletin*, 108(3), 499–514. doi: 10.1037/0033-2909.108.3.499
- Phinney, J. S. (1992). The multigroup ethnic identity measure: A new scale for use with diverse groups. *Journal of Adolescent Research*, 7(2), 156-176. doi: 10.1177/074355489272003
- Phinney, J. S. (2000). Ethnic and racial identity: Ethnic identity. In A. E. Kazdin (Ed.), *Encyclopedia of Psychology, Vol. 3* (pp. 254–259). Washington, DC: American Psychological Association. doi: 10.1037/10518-122

- Phinney, J. S., & Ong, A. (2007). Conceptualization and measurement of ethnic identity: Current status and future directions. *Journal of Counseling Psychology, 54*, 271–328. doi: 10.1037/0022-0167.54.3.271
- Resnicow, K., Soler, R. E., Braithwaite, R. L., Selassie, M., & Smith, M. (1999). Development of a racial and ethnic identity scale for African American adolescents: The survey of Black life. *Journal of Black Psychology, 25*(2), 171–188. doi: 10.1177/0095798499025002003
- Roberts, R. E., Phinney, J. S., Masse, L. C., Chen, Y. R., Roberts, C. R., & Romero, A. (1999). The structure of ethnic identity of young adolescents from diverse ethnocultural groups. *Journal of Early Adolescence, 19*, 301–322. doi: 10.1177/0272431699019003001
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107–120. doi: 10.1007/s11336-008-9101-0
- Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62*, 529–540. doi: 10.1037//0021-9010.62.5.529
- Smith, E., & Brookins, C. C. (1997). Toward the development of an ethnic identity measure for African American youth. *Journal of Black Psychology, 23*(4), 358–377. doi: 10.1177/00957984970234004
- Smith, T. B., & Silva, L. (2011). Ethnic identity and personal well-being of people of color: A meta-analysis. *Journal of Counseling Psychology, 58*(1), 42–60. doi: 10.1037/a0021528
- Spencer, M., Icard, L., Harachi, T., Catalano, R., & Oxford, M. (2000). Ethnic identity among monoracial and multiracial early adolescents. *Journal of Early Adolescence, 20*, 365–387.

- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797–811.
- Tajfel, H., & Turner, J. (1986). The social identity theory of intergroup behavior. In S. Worchel & W. Austin (Eds.), *Psychology of intergroup relations* (pp. 7–24). Chicago, IL: Nelson-Hall.
- Umaña-Taylor, A. J., Vargas-Chanes, D., Garcia, C. D., & Gonzales-Backen, M. (2008). A longitudinal examination of Latino adolescents' ethnic identity, coping with discrimination, and self-esteem. *The Journal of Early Adolescence*, *28*(1), 16–50. doi: 10.1177/0272431607308666
- Umaña-Taylor, A. J., Yazedjian, A., & Bámaca-Gómez, M. (2004). Developing the ethnic identity scale using Eriksonian and social identity perspectives. *Identity: An International Journal of Theory and Research*, *4*, 9–38. doi: 10.1207/S1532706XID0401_2
- Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement*, *58*(1), 6–20. doi: 10.1177/0013164498058001002
- Vacha-Haase, T., & Thompson, B. (2011). Score reliability: A retrospective look back at 12 years of reliability generalization studies. *Measurement and Evaluation in Counseling and Development*, *44*(3), 159–168. doi: 10.1177/0748175611409845
- Yancey, A., Aneshensel, C., & Driscoll, A. (2003). The assessment of ethnic identity in a diverse urban youth population. *Journal of Black Psychology*, *27*, 190–208.
- Yasui, M., Dorham, C., & Dishion, T. J. (2004). Ethnic identity and psychological adjustment: A validity analysis for European American and African American adolescents. *Journal of Adolescent Research*, *19*(6), 807–825. doi: 10.1177/0743558403260098

Yip, T., & Fulgni, A. J. (2002). Daily variation in ethnic identity, ethnic behaviors, and psychological well-being among American adolescents of Chinese descent. *Child Development, 73*(5), 1557–1572. doi: 10.1111/1467-8624.00490

Appendix A

Coding Sheet

- A. Short Description
- a. For authors with only one article included in our study: Last name of main author and the last two digits of year (ex. Belavich 98)
 - b. For authors that have more than one article included in our study that have also been published in the same year: Last Name + last two digits of year + Journal abbreviation in CAPS (ex. Koenig 99 AJG)
 - c. For authors with multiple studies in one article: Last name + last two digits of year + lower case letter (in sequential order) (ex. Levin 99 a, Levin 99 b....etc)
- B. APA citation: Author, A.B., & Colleague, J.F. (year). Title of the article. *Journal Title*, Vol, pages (you can COPY and PASTE from the internet)
- C. Year of publication
- D. Publication Outlet
- 0 = Unpublished (Dissertation/Thesis, conference presentation, etc.)
 - 1 = Published (Journal Article or Book Chapter)
- E. Coders: First names
- Note: participant information pertains to
those who completed the MEIM-R
- F. Participant Immigrant Status
- 0 = Not specified
 - 1 = Not immigrant/US Born
 - 2 = Immigrant to the US
- G. Participant Clinical or Adjudicated Status
- 0 = normal community members
 - 1 = students
 - 2 = at-risk groups (prisoners/delinquents, people with physical illness, ...)
 - 3 = clinical populations (in treatment for mental health issues)
- H. Mean Age of participants (or median of given range, IF a narrow range of < 15 years)
- I. Educational Level = *mean years of education*
- FOR EXAMPLE: 8th graders = 8, 9th graders = 9, etc.
- 11= 11th grade average or community sample or low SES without indication
 - 12= high school graduates
 - 13= college freshmen
 - 14= college sophomores
 - 15= college juniors

16= college seniors and college graduates

- J. Number of participants (N) completing the MEIM–R
- K. Percent female participants
- L. Percent White/European American (for participants included in *this* row, same for all below)
- M. Percent African American
- N. Percent Hispanic/Latin American
- O. Percent Asian American
- P. Percent Native American
- Q. Percent Pacific Islander/Hawaiian
- R. Percent “other” American (race not specified or not included in one of the above)
- S. Number of other measures administered in the same study
- T. Other Variables Also Measured
 0 = No other variables (just ethnic identity)
 1 = Psychological wellbeing
 2 = Psychological stressors (measures of depression, anxiety, etc.)
 3 = Variables related to prejudice and racial discrimination
 4 = Academic achievement
 5 = Other
 9 = More than one of the above
- U. Level of ethnic identity
 Enter the average item score on the MEIM–R in item units (must be on a 4-point scale, ranging from 1-4).
 For example, if they administered only 9 items, divide the mean for the total group by 9. Values entered must be between 1 and 4. If the scaling was modified by including more responses (e.g., 5-point scale), divide the mean item score by the number of points in the scale and then multiply by 4 (to make the scaling equivalent to a 4-point scale). If the scale started at 0, rather than 1, then simply add 1 to the mean.
 (Note the number of items used in the next column.)
- V. Number of items included in this version of the MIEM-R administered (blank if not reported)
- W. Reliability coefficient (alpha) of the MEIM–R
 (If more than one value is reported, average the values, weighted by N)
- X. Comments and clarifications regarding coding for this article (insert only as needed)