ORIGINAL ARTICLE

Measurement Invariance of English and Amharic Language Versions of the Multi-group Ethnic Identity Measure (MEIM) in Ethiopian Context

Dawit Demlie Messay[1]* & Tamirie Andualem Adal[2]

Abstract

This study aimed to examine measurement invariance (configural, metric, scalar and residual) of English and Amharic language versions of the Multi-group Ethnic Identity Measure (MEIM) on a sample of Ethiopian public university students (n = 286, for the English and n = 343 for the Amharic versions). After establishing a baseline model of the scale using CFA, we assessed the four levels of measurement invariance, applying more stringent and increasingly restrictive multi-group confirmatory factor analyses. We used chi-square differences and alternative fit indices to determine model fit. In the configural invariance testing of the arrangement across groups, the parameters were freely estimated in each group. For the metric, scalar, and residual, the parameters were tested with the imposition of the equality constraints. The partial invariance model finally yielded ∆CFI = .003, ∆RMSEA = -.002 and ∆SRMR = 0. The results, in general, showed that the national identity scale in the English and Amharic language versions showed measurement invariance. The tool in the Amharic version was reliable and valid to measure the national identity construct among public university students. We hope that this work will encourage researchers to further examine this topic, aiming at upholding togetherness and peaceful coexistence. Besides, the present study's result can be taken as an integral part of the knowledge domain of social identity, adding a methodological body of literature to the field.

Keywords: Multi-group Ethnic Identity Measure (MEIM), measurement invariance, language equivalence, confirmatory factor analysis, multi-group confirmatory analysis

Introduction

National identity, from the social psychological perspective, is one of the multiple forms of collective identity generated by the social psychological processes of individuals over time. Tajfel's studies have turned out to be most useful and employed for understanding the nature of human attachment/affiliation with a nation to see groups extendedly as a nation, as national identity is a specific type of social identity (Nigbur & Cirinnella, 2007;

1 *PhD candidate, School of Psychology, College of Education and Behavioral Studies, Addis Ababa University, Ethiopia*
Corresponding Author
Email: dawitd344@gmail.com or dawit.demlie@aau.edu.et

2 *Associate Professor of Psychology, School of Psychology, College of Education and Behavioral Studies, Addis Ababa University, Ethiopia*

DOI: https://dx.doi.org/10.4314/erjssh.v12i1.10

Smith et al., 2005).

Studies of how individuals relate to nations or how they identify themselves with ideological, institutional, and political structures and how nation-states are composed have been at the center of researchers' attention (Bellamy, 2008; Hutchinson, 2019; Storm, 2022; Woods & Debs, 2013). They emphasized on the importance of studying the concept because a stronger national identity has been found to be essential to a nation's prosperity, in general and to institutions' effectiveness. in particular. According to Espinosa *et al.* (2016), the greater the degree of identification with a nation is, the greater the acceptance, interpersonal confidence, sense of contribution to society, perception of better institutional functioning and comprehension of the system will be.

Hofstede (2001) posits that nations often converge on certain values while diverging on others. Hofstede's analysis indicates that the American cultural landscape, for example, is particularly marked by strong individualism, low uncertainty avoidance, and a preference for short-term results. In stark contrast, Ethiopia—where the Multi-group Ethnic Identity Measure (MEIM) is being validated—exhibits characteristics typical of collectivist societies, such as a high power distance, weak uncertainty avoidance, and a short-term orientation (Hofstede, 2001). Understanding these societal differences is crucial, as they play significant roles not only at the intrapersonal and interpersonal levels but also in shaping an individual's attachment to their nation. This suggests that due to the cultural differences among countries, the psychometric properties of measures assessing national identity, along with their factor structures, might differ significantly across contexts. Most of the countries that adapted and used the MEIM are based in Europe, serving as illustrative examples of this trend.

The Multi-group Ethnic Identity Measure (MEIM) was first developed by Phinney (1992) to measure the process of ethnic identity development. It has been used in several studies and has consistently shown good reliability, typically with alphas above .80 across a wide range of ethnic groups and ages (Phinney and Ong, 2007). For example, Roberts *et al.* (1999) found out that the measure comprised two factors: identity exploration (a process-oriented developmental and cognitive component) and commitment (an affective and attitudinal component).

The 12-item scale (MEIM, Roberts *et al.*, 1999) comprises five items in the *exploration* and seven items in the *commitment* dimensions, with a 5-point scale of degree of agreement. Because of its flexibility and openness to measure the different types of collective identities, it was frequently adapted to measure, for example, national identity (Maehler *et al.*, 2019; Schwartz *et al.*, 2014), gender and religious identities (Ashdown *et al.*, 2014), racial, religious and national identities (Gaines *et al.*, 2013). Translations of the measure into different languages are available as well (Habibi *et al.,* 2021; Maehler *et al.,* 2019; Mastrotheodoros *et al.,* 2012; Musso *et al.*, 2017; Schwartz *et al.*, 2014; Vedder, 2005).

To test measurement invariance, four basic steps are described by Widaman and Reise (1997). They are configural, weak factorial (metric), strong (Scalar), and strict (residual or invariant uniqueness). Diane and Marc (2016) also summarized the four measurement invariance steps considered as (1) *configural* (equivalence of model form); (2) *metric* (weak factorial or equivalence of factor loadings); (3) *scalar* (strong factorial or equivalence of item intercepts or thresholds); and (4) *residual* (strict or invariant uniqueness or equivalence of items' residuals or unique variances).

Configural invariance is a test of whether the constructs have the same pattern of free and fixed loadings across groups. If configural invariance is not met, it can be concluded that the assessment tool does not measure the same construct across groups. Metric invariance is considered satisfied when all item factor loadings are equal between groups. This means that each item contributes to the latent factor to an analogous degree across groups. To make a meaningful comparison of associations between groups, metric equivalence is necessary. Scalar invariance is satisfied when every item in each group exhibits the same point of origin (i.e., intercept) across the groups; this level of equivalence is necessary for comparing group means (Chen, 2008). When scalar invariance is met, all the mean differences in the shared variance of the items are captured by the mean differences in the latent factor. The final step in measurement invariance test is residual invariance, which means the error variances across groups are similar. According to Diane and Putnick (2016), residual invariance means that the sum of specific variance (variance of the item that is not shared with the factor) and error variance (measurement error) is similar across groups. They further noted that "…there could be more measurement error and less specific variance in one group than another, and residual invariance could still be supported if the totals of these two components were similar" (p. 6). Although testing residual invariance is important for full factorial invariance, according to Vandenberg and Lance (2000), it is not a prerequisite for testing mean differences because the residuals are not part of the latent factor.

Problem Statement

Following the restructuring of the ethnic federalism political system of Ethiopia, the trajectory of ethnic conflicts, in turn, has been shown to rise and has left the country's stability under serious concern. In the last three decades, conflicts attributed to negative ethnicity have resulted in the loss of many lives in different parts of Ethiopia (Fisseha, 2016; Zekarias, 2020). Most Ethiopian public universities in the past few years have continued to experience challenges concerning ethnic and religious conflicts, which have resulted in interruptions of academic activities, injuries and deaths (Adamu, 2020).

Despite the opinions heard about the importance of upholding Ethiopian national identity to avoid the disastrous result of 'narrow nationalism', adequate attention has not been paid to how to investigate the problem. Studies on Ethiopian national identity are hardly found. According to Windari (2021), national identity can moderate the impact of ethnic conflicts if members of one nation-state/country with high ethnic diversity have a strong sense of national identity.

Though the trends of the last decades have shown that Ethiopian public universities have been experiencing diversity-related conflicts among students, the universities have been shown passively working on studies that foster commonalities to help uphold peace, cooperation, collaboration and mutual co-existence. There are substantive calls for scholars to conduct investigations on Ethiopian national identity, more particularly at this time. From the extant literature review, however, one hardly finds reliable and valid measurement tools to examine the problem from the social-psychological perspective. Consequently, to have a reliable and valid measure that helps assess the problem remains essential. These are the ideas that fueled the present study.

To the best of the authors' knowledge, the MEIM has not been translated to Amharic and validated to measure Ethiopian national identity for the Amharic-speaking people, which comprise the majority segment of the overall population size. Consequently, we adapted, translated and validated the psychometric properties of the Amharic version measure (Demlie & Andualem, in press). In the current study, we examined the measurement equivalence or invariance between the English and Amharic language versions of the measure. By creating two groups, we examined the factor structure, the relationship between observed and underlying latent variables, and measurement invariance (configural, metric, scalar & residual), using a youth sample drawn from two Ethiopian public universities. We also conducted a thorough assessment of measurement reliability and construct validity.

Materials and Methods

Study Design

A cross-sectional study design was employed. We used a quantitative research method to meet its objectives. The data were obtained from young public university students in the undergraduate programs of the regular admission.

Participants and Sampling Procedure

The participants of the study were drawn from Dire Dawa University (DDU) and Wollo University (WU) of Ethiopia. They have to be also in their third and above years of their studies. To select these universities, factors related to resource utilization and practicality details, including accessibility, affordability, safety and security matters during trip and data collection were considered. There were no other methodological reasons than those mentioned in choosing the study areas. The assumption behind targeting year three and above students was it helped decrease ambiguities related to students' limited exposure and experience concerns, which are taken as the most important issues in understanding the social identity construct. We received an ethical clearance letter from the Committee of Ethics of the School of Psychology of Addis Ababa University.

To select the sample, Yamane's (1967) sample size formula and stratified proportional sample size formula were respectively used, taking the size of the population of the two strata (DDU and WU) as the general population. Thus, the sample size calculation was done b

$$n = \frac{N}{1 + N * (e)^2}$$

Where,

n is the sample size; N is the population size; e is the level of precision.

The assumption is that at a 95% confidence level, p = .05. The size of the population was 11,359. Thus,

$$n = \frac{11,359}{1 + 11,359*0.05^2}$$

$$386$$

In proportional stratified random sampling, it is recommended that the sample for each stratum should be proportional to the size of the subpopulation in each stratum. Accordingly, to have a proportional allocation of the sample of participants, the sample size for each stratum was computed by:

$$nk = \frac{Nk*n}{N}$$

Where,

nk is the sample size in each stratum; Nk is the size of the sub-population in stratum k; N is the total population; and n is the sample size for the population.

The two universities' population was ((DDU = 6,737) + (WU = 4,622) = 11,359). With a nonresponse rate of 5% (19 + 386 = 405), the sample taken was:

$$DDU = \quad n_{DU} = \frac{Nk*n}{N} = \frac{6,737*386}{11,359} = 228$$

$$WU = \quad n_{WU} = \frac{Nk*n}{N} = \frac{4,622*386}{11,359} = 158$$

Three hundred forty-three (343) students for the Amharic version and 286 students for the English version participated in the study. The participants chosen for the Amharic and English versions of the study are intentionally distinct, and this separation is crucial for the integrity of the data analysis. Maintaining data independence and ensuring that samples are distinct are essential to producing valid results. A violation of this principle can result in distorted statistical outcomes, including inflated standard errors that compromise the reliability of the tests conducted. For this reason, we employed two separate groups for the Amharic and English versions during the equivalence tests. By employing the lottery method of random sampling, thus, we drew the sample. We got their consent to collect the data. For the Amharic version, 85% of the respondents completed the questionnaire, , which showed an excellent rate of response. However, for the English version, one-fifth of the questionnaire was either not returned or properly filled out.

For the Amharic version, 237 participants (69%) were male, and 105 participants (31%) were female. One participant did not mention his or her sex. The mean age of the participants was 22.05 years (*SD* = 1.30). Twenty-six participants did not state their age. For the English version, the mean ages of 22 (*SD* = 1.304). In addition, 164 participants (57.3%) were male, and 76 participants (26.6%) female. Forty-six participants (16.1%) did not report their sex.

Measures

For the current measurement invariance assessment, we utilized both the adapted English and translated Amharic versions of the 12-item Multi-group Ethnic Identity Measure (MEIM). Prior to data collection, the validation process consisted of three key phases: adaptation, translation, and a comprehensive content validity assessment designed to ensure both reliability and validity from the outset.

The instrument was subjected to scrutiny by a panel of thirteen experts, encompassing fields such as psychology, linguistics, and social anthropology, all of whom hold academic positions across various universities. Among these, four linguistic experts executed both forward and backward translations of the instrument, while eight psychology experts and one social anthropology expert conducted the content validity evaluation. All panel members were at least bilingual, facilitating nuanced assessments.

Each expert rated the items using a three-tier scale—“Essential,” “Useful but not Essential,” and “Nonessential”—which quantified the relevance of each item in measuring the intended construct. With a scoring system of essential = 3, useful but not essential = 2, and nonessential = 1, items categorized as “essential” were retained without modification. Items categorized as “useful but not essential” underwent revisions based on expert feedback, with the exceptional circumstance that no items were deemed “nonessential.” The feedback from the experts was substantial. It leads to significant modifications throughout the tool’s items. Items classified as “useful but not essential” that received minor revisions were further appraised for morphological accuracy by one of the language experts. To better align with the target population, the original MEIM’s ethnic-item formulations were substituted with terminology reflecting “Ethiopian national identity”.

Both the adapted English and Amharic versions functioned as self-report instruments, employing a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The experts also reached a consensus to introduce three additional items to the *exploration* dimension, customizing the original five items. The flexibility to add or remove items was supported by the premise that the measure "can be complemented by additional measures" (Phinney and Ong, 2007, p. 28). This adjustment resulted in an initial pool of 15 items.

To quantify expert consensus on content validity, we applied Lawshe's (1975) content validity ratio (CVR) formula. The formula is defined as follows:

$$CVR = \frac{ne - \left(\frac{N}{2}\right)}{\frac{N}{2}}$$

where *ne* represents the number of panel members who deemed the item "essential," and *N* denotes the total number of panel members (Lawshe, 1975). According to Lawshe's guidelines, items are expected to be assessed on a three-point scale (1 to 3). For retention, an item must achieve a CVR of at least .75 when evaluated by eight experts (Lawshe, 1975, p. 568). Consequently, our analysis yielded CVR scores ranging from .75 to 1 for each item, with an overall content validity index (CVI) of .91.

The finalized instruments in both English and Amharic versions underwent a final review by a supervisor, thereby preparing them for data collection and subsequent statistical analysis. In the course of the study, two items did not sufficiently measure the *exploration* dimension. Consequently, the Cronbach's alpha reliability coefficients for the Amharic versions of *exploration* and *commitment* were determined to be .74 and .88, respectively. After the removal of the poorly functioning items, the reliability coefficient for *exploration* increased to .88. For the English version, the reliability coefficients stood at .69 for *exploration* and .86 for *commitment*; post-removal of the ineffective items, the coefficient for exploration adjusted to .83.

Statistical Analysis

We conducted five series of analyses. First, we carried out data screening and assumptions checking for the univariate and multivariate levels of statistical analyses in each group of the dataset; the Amharic and the English group. A few outlying cases were found, albeit they were not excluded from the multivariate analyses because keeping them in the data did not result in errant data outcomes. Hence, normality, the correlational strength of the items, linearity and significant outliers were checked. We employed a multiple imputation method to address some of the missing values. Second, by employing descriptive statistics, scores of the key variables and demographic information were analyzed and presented descriptively via percentages and means. Third, correlational analysis was employed to obtain an overview of poor items and subsequently determine the degree of relationship between items and factors. We used Pearson correlation coefficients for this purpose. Fourth, we assessed the relationships between the observed factors and latent variables via confirmatory factor analysis (CFA). We examined the factor structure in each group.

By employing a test of confirmatory factor analysis, the observed indicators (items) for each variable with its metric value were identified. Before the MI test, two CFAs were conducted for the Amharic and English versions. Given that the latent structure was multifactorial (i.e., made up of two factors), the pattern of factors and factor loadings was confirmed by the CFA. Item□factor relationships were also defined. Fifth, we conducted multi-group CFA to test the four levels of MI. For each CFA model, factor variances were set to one. We used IBM SPSS version 25 and AMOS 23 to analyze these statistics, employing maximum likelihood estimation.

We used different indices to evaluate model goodness and determine model fitness. The relative chi-square test (x^2/df), incremental fit index (IFI), Tucker–Lewis index (TLI), Bentler's comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR) were employed. Each measurement model's quality was assessed via well-recognized standards: RMSEA ≤ .08, SRMR ≤ .10, and CFI ≥ .90 for adequate model fit and values of RMSEA ≤ .05, SRMR ≤ .08, and CFI ≥ .95 for excellent model fit (Browne & Cudeck, 1993; Collier, 2020). In addition, a relative chi-square test (x^2/df) < 3 was used as an indicator of good model fit. The Akaike information criterion (AIK; Akaike, 1987) was used to compare the quality of competing baseline models; a smaller number indicates a more ideal model. As measures of invariance, we applied other alternative fit indices (e.g., ΔRMSEA, ΔSRMR: Meade *et al.*, 2008; CFI: Cheung & Rensvold, 2002). According to these authors, a change in RMSEA (i.e., ΔRMSEA), with a change ≤ -.015, indicates invariance between MI levels (Meade *et al.*, 2008). Change in CFI (i.e., ΔCFI), where invariance is encountered on the basis of a change in CFI ≤ -.015 (Cheung & Rensvold, 2002). That is, if there was no decrease in ΔCFI and ΔRMSEA between increasingly restrictive MI models of greater than .01 and .015, respectively, the assumption of MI was accepted at each level. We used these statistics.

Results

Descriptive

During the preliminary analysis phase, a number of cases were found to have missed a few items on the scale, with each instance being individually below 5%. These cases were addressed using statistical methods, specifically through multiple imputations, during the multivariate analyses. However, two cases were identified as extreme outliers, missing seven items (46.7%) from the overall Amharic version of the national identity scale and four items (26.6%) from the English version. As a result, these two cases were excluded from the dataset.

Additionally, the assumptions of normality, linearity, and the absence of significant outliers were thoroughly examined. These assumptions pertain to various aspects of score distribution and the relationships between variables. The data in the present study satisfied the criteria for normality and linearity (see Table 1). Moreover, factor analysis, like other multivariate analyses, can be sensitive to outliers. Thus, we inspected the histogram and the boxplot to check for outliers. Seven cases in the Amharic version and five cases in the English version were found to be outliers. However, they were maintained in the analyses because they were less likely to affect the MI test. The descriptive statistics (Table 1) show that these cases could not have a significant effect on the multivariate analysis. The inference we can draw from this marginal difference between the *mean* (50.38 [52.77]) and the *5% trimmed mean* (50.83 [53.56]) is that the values were not very much different from the remaining distributions. Therefore, they were retained in the dataset of the study. However, two cases were found to be extreme outliers in the English version and were discarded from the dataset.

Since CFA is grounded in correlation, it is assumed that the relationships among the variables are linear. In the absence of clear evidence indicating a curvilinear relationship, authors such as Pallant (2020) suggest that it is generally safe to proceed with the analysis, as long as an adequate sample size and appropriate ratio of cases to variables are achieved. In this case, the specific assumption was not violated.

Table 1.

Results of the Descriptive Statistics of Participants for both Language Versions: Amharic [English]

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness	Kurtosis
Age	317 [264]	20 [20]	26 [25]	22.05 [22]	1.30 [1.304]	.445 [.370]	-.192 [-.50]
Exploration	343 [286]	7 [8]	30 [30]	21.30 [23.51]	4.63 [4.13]	-.661 [-1.12]	.354 [1.23]
Commitment	343 [286]	8 [9]	35 [35]	29.09 [29.26]	5.65 [5.41]	-1.292 [-1.21]	1.091 [.97]
Overall National Identity	343 [286]	15 [24]	65 [64]	50.38 [52.77]	8.49 [8.18]	-.927 [-1.04]	.987 [.57]

Note. 5% Trimmed Mean of age 22 [22], exploration = 21.25 [23.99], commitment = 31.59 [29.79], overall national identity = 50.83 [53.56].

A summary of the descriptive statistics for the national identity scale results in both language versions is provided in Table 1. For the exploration dimension, data were collected from 343 [286] respondents, with scores ranging from 7 [8] to 30 [30] (M = 21.30 [23.51], SD = 4.63 [4.13]). In a similar vein, the commitment dimension data were gathered from 343 [286] participants, with values spanning from 8 [9] to 35 [35] (M = 29.09 [29.26], SD = 5.65 [5.41]). Overall, the data for the national identity scale came from

343 [286] respondents, with scores ranging from 15 [24] to 65 [64] (M = 50.38 [52.77], SD = 8.49 [8.18]).

Table 1 also illustrates the data symmetry, presenting the skewness and kurtosis values. According to Curran *et al.* (1996), the normality thresholds for skewness and kurtosis are 2.0 and 7.0, respectively, when evaluating multivariate normality in analyses such as factor analysis and MANOVA.

Confirmatory Factor Analysis Result

Prior to conducting the MI test, two separate Confirmatory Factor Analyses (CFAs) were performed for the Amharic and English versions of the assessment. During these analyses, we identified the observed indicators (items) associated with each latent variable, along with their respective metric values. This process ensured that we thoroughly understood the relationships and characteristics of the items within each language version, setting a solid foundation for the subsequent testing of measurement invariance (See Figure 1 & 2).

Figure 1.

Measurement Model of the Two-factor National Identity Measure for the Amharic Version

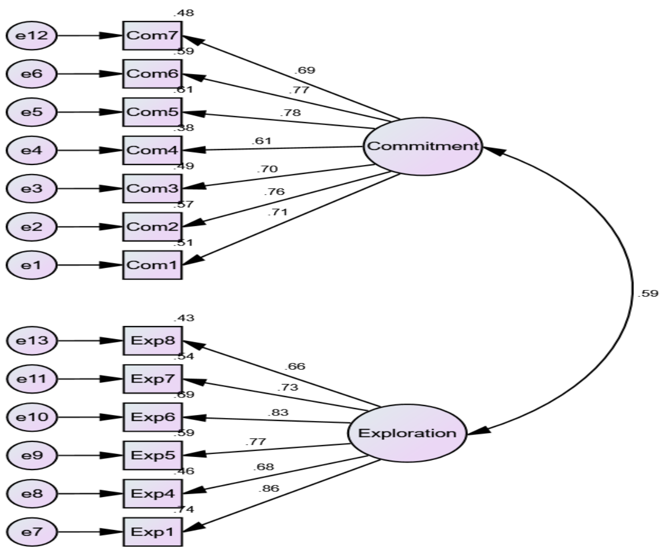


Figure 2.

Measurement Model of the Two-factor National Identity Measure for the English Version

Considering that the latent structure consisted of multiple factors—specifically two—the patterns of these factors and their loadings were validated through CFA. The relationships between items and factors were also established. To illustrate the predictive power and the degree to which variations were accounted for by each item, R-squared values are provided (see Table 2). All values presented in square brackets in this paper correspond to the English version.

Table 2.

Standardized Regression Weights of the Measure: Amharic [English]

Item	Factor Loading		
	Commitment	*Exploration*	R^2
7. I have a lot of pride in my Ethiopian nationality	.781 [.749]		.610 [.561]
3. I am happy that I am an Ethiopian	.778 [.712]		.605 [.506]
11. I have a strong attachment to my Ethiopian nationality	.757 [.756]		.573 [.571]
2. I have a clear understanding of my Ethiopian nationality	.711 [.674]		.505 [.454]
4. I have a strong sense of belonging to my Ethiopian nationality	.697 [.657]		.486 [.431]
12. I feel good about my Ethiopian cultural, historical, traditional, etc., backgrounds	.687 [.666]		.471 [.443]
5. I understand fairly well what my Ethiopian nationality group membership means to me	.642 [.625]		.412 [.390]
1. I have spent time trying to find out more about my Ethiopian nationality, such as its history, traditions, customs, etc.		.875 [.586]	.766 [.343]
9. I have spent time trying to know the historical facts and cultural practices of other subgroups under Ethiopian nationality membership		.807 [.601]	.651 [.362]
8. I have explored seeking to know more about cultural issues of my Ethiopian nationality, such as special food, music, dressing, etc.		.736 [.718]	.541 [.515]
10. I have explored more about the historical facts of symbols of my Ethiopian nationality, such as the Ethiopian national flag		.730 [.749]	.471 [.561]
6. To learn more about my Ethiopian nationality background, I have often talked to others about it		.682 [.673]	.465 [.452]
13. I have spent time trying to learn about the festivals of *Irreecha, Fichee-Chambalaalla, Ashenda*, and so on		.644 [.674]	.415 [.454]

The explained variance of the observed variables, commonly denoted as R^2, showed differences between the Amharic and English versions, yet both remained within acceptable ranges, indicating their comparable functionality. For the Amharic version, the variance for exploration items exhibited considerable variability, ranging from a low of 41.5% for item 13 to a high of 76.6% for item 1. This indicates that responses to these items differed significantly in terms of how effectively they captured the underlying constructs being measured. In contrast, the English version displayed a more modest range of explained variance, with exploration item 9 registering 36.2% and commitment item 11 achieving a maximum of 57.1%. Overall, it is evident that the Amharic version outperformed the English version, accounting for a greater proportion of variance across the assessed items.

Table 3.

Model Fit Indices for the Adapted MEIM (Amharic, n = 343; English, n = 286)

Model	x^2	*df*	x^2/df	CFI	IFI	TLI	RMSEA	SRMR
A. 2-Factor Amharic Version	140.250	61	2.299	.966	.966	.956	.062	.048
B. 2-Factor English Version	140.942	64	2.202	.945	.945	.933	.065	.051

Note. x^2 = chi-square; *df* = degree of freedom; CFI = comparative fit index; IFI = incremental fit index; TLI = Tucker-Lewis index; RMSEA = root-mean-square error of approximation; SRMR = standardized root-mean-square residual

Table 3 presents a comprehensive overview of the model fitness indices for both the Amharic and English versions of the measurement tool. For both versions, the statistical analyses produced favorable model fit indices, indicating robust support for the underlying constructs. Specifically, the Amharic version of the model yielded x^2 = 140.250; $p < .05$, CFI = .966; IFI = .966; TLI = .956; RMSEA = .062; and SRMR =.048. In comparison, the English version produced a slightly different set of indices, with x^2 = 140.942; $p < .05$, CFI = .945; IFI = .945; TLI = .933; RMSEA = .055; and SRMR = .051.

Furthermore, we conducted a composite reliability (CR) analysis to evaluate the overall reliability of the measure in addition to assessing internal consistency reliability through Cronbach's alpha. The CR values for the exploration and commitment constructs of the Amharic version were found to be .93 and .92, respectively. In contrast, the English version of the same constructs yielded CR values of .70 and .76.

To determine the extent to which all items within each factor measured the intended underlying construct, we examined convergent validity. One key criterion for establishing convergent validity is the Average Variance Extracted (AVE). According to Hair *et al.* (2019), an AVE greater than .50 indicates the presence of convergent validity. In this study, the AVE values for the exploration and commitment constructs in the Amharic version were .70 and .64, respectively, signifying convergent validity. Similarly, the AVE for the English version of these constructs was .67 and .69, further supporting convergent validity.

Table 4.

Reliability and Validity of the Measurement Scales: Amharic and English

Constructs	Amharic Version			English Version			No. of Items
	CR	*AVE*	*√AVE*	*CR*	*AVE*	*√AVE*	
Exploration	.93	.70	.83	.70	.67	.81	6
Commitment	.92	.64	.80	.76	.69	.83	7

Note: CR = composite reliability; AVE = average variance extracted; √AVE = square root of the AVE.

We also examined discriminant validity to determine the degree to which the variables correlated more strongly with their corresponding factors than with other factors.

Discriminant validity can be assessed using various methods, including the square root of AVE. In this study, the square root of the AVE for the latent variables in the Amharic version was .83 and .80, both of which exceeded the correlation between the variables (*r* = .59), thus confirming the discriminant validity of the measurement model. Likewise, the square root of the AVE for these variables in the English version was .81 and .83, which was greater than (*r* = .51) also indicating discriminant validity. Table 4 presents the reliability and validity of the measure.

Overall, these results indicate a strong fit for both language versions according to the established criteria for model evaluation, underscoring the reliability and validity of the measurement tool across different cultural contexts.

Measurement Invariance

Table 5.

The Goodness of Fit Indices and Variations of Statistical Values of the MI for Amharic and English language Versions of the Adapted MEIM

Model	x^2	df	CFI	RMSEA, 90% CI	SRMR	Δx^2	Δdf	ΔCFI	ΔRMSEA	ΔSRMR
Configural	252.157	122	.965	.041 (.034, .048)	.047	-	-	-	-	-
Metric	259.537	133	.966	.039 (.032, .046)	.053	7.38	11	.001	-.002	.006
Scalar	276.841	144	.964	.038 (.032, .045)	.046	17.304	11	-.002	-.001	-.007
Residual	389.423	157	.937	.049 (.043, .055)	.053	112.582	13	-.027	.011	.007
Partial invariance	278.060	155	.967	.036 (.029, .042)	.046	1.219	11	.003	-.002	0

Note. x^2 = chi-square; *df* = degree of freedom; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; CI = confidence interval; SRMR = standardized root-mean-square residual; ΔCFI = change in CFI (all values not significant); ΔRMSEA = change in RMSEA (all values not significant); ΔSRMR = standardized root-mean-square residual.

Table 5 shows the full results of the examination of measurement invariance. In the *configural invariance* testing of the arrangement across groups, the parameters were freely estimated in each group, resulting in a good general model. All of the paths were statistically significant in each group. The values of the CFI, RMSEA and SRMR were found to be more than adequate, yielding .965, .041 and .047, respectively. In the past, significant changes in the chi-square for two nested models were used as a criterion to assess MI (Byrne *et al.*, 1989; Reise *et al.*, 1993). Since chi-square is unduly sensitive to small and nonsignificant deviations from a "precise" model, researchers have shifted their emphasis from chi-square to alternative fit indices in the search for absolute model fit (Cheung & Rensvold, 2002; Meade, *et al.*, 2008). Consequently, we used alternative fit indices more frequently to determine model fit in this study.

By constraining the factor loadings to make them equal across groups, we next generated a model to examine the *metric invariance*. One approach is to carry out a chi-square difference test, which tests whether the model represents a significantly worse fit to the data than the previous model does (i.e., configural invariance; $x^2_{difference} = x^2_{metric} - x^2_{configural} =$

259.537 – 252.157 = 7.38). We performed the same computation for the degree of freedom; $df_{difference} = df_{metric} - df_{configural} =$ 133 – 122 = 11. This is rooted in the assumption that the model with the equality constraints (i.e., metric invariance) was nested within the configural model so that we can test whether there is a statistically significant reduction in fit as a result of adding the quality constraints. Assuming $df = 11$, $\alpha = .05$, the chi-square test p value is .76. This demonstrated that there was no statistically significant difference in the chi-square values, confirming that the metric invariance model's assumption holds.

Another way to test MI is to inspect the differences in values of indices such as the CFI, RMSEA and SRMR between models. A model with a higher CFI is assumed to locate the data more accurately than a model with a lower CFI. When equality constraints are imposed, it is usually expected that the model with the imposed equality constraints fits at least somewhat worse than the model without those constraints. A frequently applied criterion for a nested model is ΔCFI ≤ -.01 to determine whether the use of equality constraints leads to a significantly worse fit than the earlier model (without those constraints) (Cheung & Rensvold, 2002). Other scholars have suggested the use of other alternative fit indices for determination (e.g., ΔRMSEA, ΔSRMR; Meade et al., 2008). Chen (2008) recommended a criterion of a −.01 change in CFI, paired with changes in RMSEA of .015 and SRMR of .030 (for metric invariance) or .015 (for scalar or residual invariance).

For the present metric invariance test, ΔCFI, ΔRMSEA and ΔSRMR were found to be .001, -.002 and .006, respectively. The values indicated that the improvement in model fit was not significant with the imposition of the equality constraints. Therefore, the results of the fit indices change with the results of the chi-square test, which shows that the application of the equality constraints results in a nonsignificant drop-in fit. Consequently, we concluded that metric invariance was supported. The results generally indicated the presence of metric invariance.

On the basis of the assumption that metric invariance holds, we checked for *scalar invariance*, which is the invariance of intercepts for indicator variables. Essentially, this is a test of whether the item intercepts are consistent/invariant across groups. Without further investigation of the estimates and intercepts, the values of ΔCFI, ΔCFI, ΔRMSEA and ΔSRMR were found to be -.002, -.001 and -.007, indicating scalar invariance. Nonetheless, we performed a series test of intercepts by relaxing the equality of the constraints to improve the model by determining the intercepts that were somewhat noninvariant and resulted in values of lower model fit indices. We used chi-square tests for comparisons. However, a significant difference was not obtained. We maintained the baseline scalar invariant model.

We finally tested *residual invariance*. The changes in RMSEA and SRMR were within acceptable ranges (ΔRMSEA = .011 & ΔSRMR = .007), but the change in CFI was slightly outside the range given, yielding ΔCFI = -.027. Consequently, we performed score tests aimed at testing each residual for invariance. The partial invariance model, where the residual variance for 'Exploration items 1 & 4' were freely estimated across the groups, resulted in changes in CFI, RMSEA and SRMR (ΔCFI = .003, ΔRMSEA = -.002 and ΔSRMR = 0). On the basis of the less invariant parameters identified, we examined the fit of a partial model, where all factor loadings, intercepts and residuals were constrained equal except the residual variance related to 'Exploration 1 & 4', against the model supposing full scalar invariance. Compared with the other models, apart from residual invariance, a significant difference was not observed. However, the partial invariance model fit the data better than the full residual model did. This partial invariance model is used as a baseline model for the subsequent invariance tests.

Figure 3.

Diagram of the Residual Invariance Test of the Amharic Version

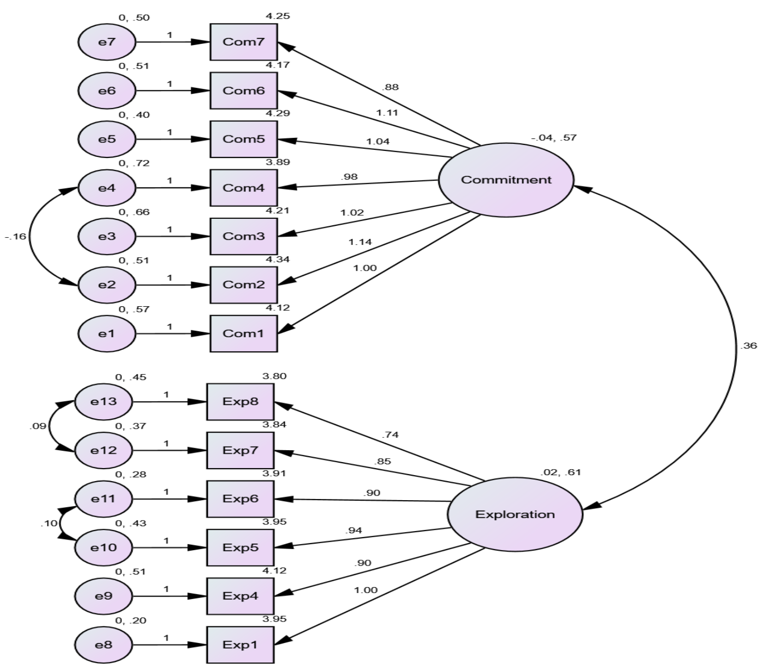
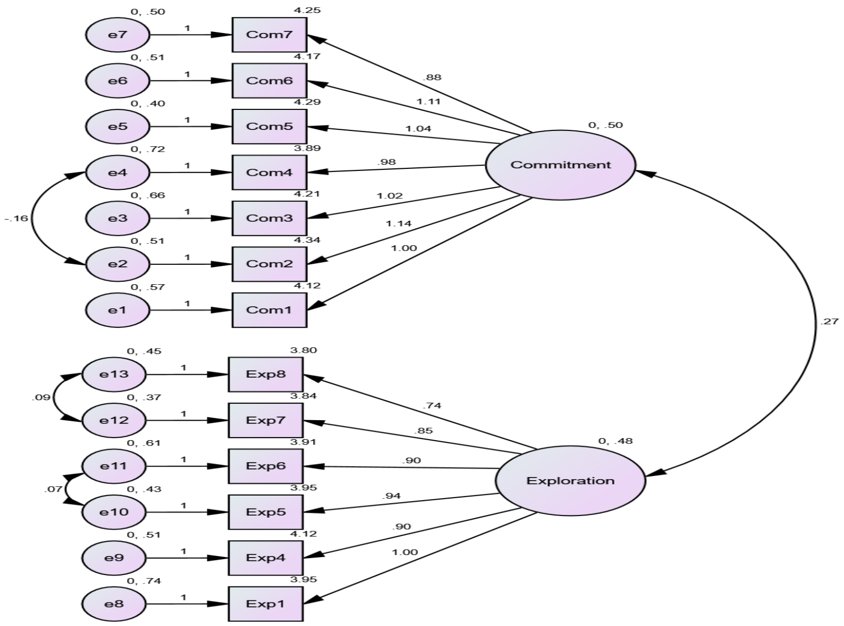


Figure 3.
Diagram of the Residual Invariance Test of the English Version



Discussion

The results of the comprehensive analysis above revealed that the criteria necessary for establishing measurement invariance—specifically, configural, metric, scalar, and residual invariance—were successfully met for both language adaptations of the instrument. In addition, we conducted a thorough assessment of composite reliability, convergent validity, and discriminant validity. The results confirmed that these metrics met the established standards for reliability and validity, indicating that the measurements used in the study are both consistent and accurately reflecting the intended constructs.

Our analysis corroborated the existence of a two-factor model in both the English and Amharic versions of the MEIM, which is consistent with previous research in the field (Gaines *et al.*, 2013; Maehler *et al.*, 2019; Musso *et al.*, 2017; Phinney and Ong, 2007; Roberts *et al.*, 1999; Schwartz *et al.*, 2012). This finding implies that the participants perceived and interpreted the MEIM items in a similar manner, regardless of the language in which they were administered. It is noteworthy that there was no evidence of cross-loading among any of the items; the majority of correlations between items and their respective factors were found to be robust, further supporting the reliability of the model.

The parallel factor loadings across both groups (indicating metric invariance) suggest that the relationships among the factors and items can be validly compared between the English and Amharic versions. This ensures that the MEIM items operate equivalently for all participants, regardless of the language used for administration.

Additionally, scalar invariance was successfully established, which allows researchers to make meaningful comparisons of mean scores between the two language groups. This level of equivalence implies that the underlying constructs represented by the items in both language versions are inherently aligned. it also offers a solid foundation for comparative analysis.

Furthermore, the establishment of residual invariance suggests that any comparisons based on raw observed scores are valid, as the impact of measurement error remains consistent across both language versions. This consistency in specific variance and measurement error enhances the overall reliability of the MEIM scores derived from both the English and Amharic contexts.

Our results are consistent with previous MI studies involving construct adaptations across diverse language groups (Fisher *et al.*, 2020; Maehler *et al.*, 2019; Mastrotheodoros *et al.*, 2012). The established measurement invariance implies that observed differences in MEIM scores between the two language versions are not attributable to measurement error, facilitating generalizations about the construct with confidence; however, it is essential to advocate for additional research into measurement invariance across a wider array of languages beyond English, which would further bolster the validity of this instrument.
In summary, the study demonstrates that both the English and Amharic versions of the national identity scale exhibit measurement invariance. This confirms the reliability and validity of the current Amharic version of the MEIM for evaluating Ethiopian national identity within a university context. The adapted Amharic instrument offers valuable insights and serves as a foundational tool for future research in this area across various settings.

Conclusion

In the study, 343 and 286 participants have participated in the Amharic and in the English versions consecutively. Along with testing the pattern of factors and factor loadings via, the four types of MI (configural, metric, scalar and residual) were examined. The result of the CFA showed all the observed variables were significantly represented by their latent variables, and that good model fitness was attained. The variance explained by the observed variables (R^2) for the Amharic version, for example, ranged from .415 (exploration item 13) to .766 (exploration item 1). This means that the explained variance ranged from 41.5% to 76.6%. For the English version, the variance ranged from .362 (exploration item

9) to .571 (commitment 11). The four basic steps of test of measurement invariance were examined. The results showed that the fit indices for testing the configural, metric, scalar, and residual invariance were found to be adequate to show invariance. The model for both versions yielded values of good model fit indices (x^2 = 140.250; $p < .05$, CFI = .966; IFI = .966; TLI = .956; RMSEA = .062; and .048 for the Amharic version) and (x^2 = 140.942; $p < .05$, CFI = .945; IFI = .945; TLI = .933; RMSEA = .055; and .051 for the English version. However, the partial invariance model fit the data better than the full residual model did. This partial invariance model is therefore used as a baseline model for the subsequent invariance tests.

The study has effectively demonstrated the functional universality of the Multidimensional Ethnic Identity Measure (MEIM) through a comprehensive process of adaptation, translation, and validation specifically tailored to the Ethiopian context. This validation was conducted among Amharic-speaking and English-speaking public university students, employing a rigorous measurement invariance test as a key method for establishing the tool's reliability and applicability.
While this study is believed to provide a robust foundation for understanding national identity within this particular population, it also paves the way for further research opportunities for other scholars. Future investigations could explore the functional universality of the measurement factors in greater depth, particularly by examining how national identity interacts with various outcome variables such as academic performance, psychological well-being, and social integration.

It is important to note, however, that the research was limited to a sample of public university students. Consequently, the findings may not fully represent or be applicable to other demographic groups or settings within Ethiopia or beyond, underscoring the need for further research to draw more comprehensive conclusions.

We assume that we introduced a new adapted measurement of Ethiopian national identity, presenting a contextualized and tailored measurement specifically for students in the Ethiopian public university setting. We hope that this work will encourage researchers to further examine this topic by aiming at upholding togetherness and peaceful coexistence in general and fostering a favorable learning environment for students' effectiveness in particular. Besides, the present study's result can be taken as an integral part of the knowledge domain of social identity, adding a methodological body of literature to the field.

Limitations and Future Recommendations

While the study offers valuable contributions to the field at large and provides specific insights for the country, it is important to recognize its limitations. Firstly, the response rate for the English version of the questionnaire was notably lower, with only 71% of the distributed surveys being accurately completed and returned. In contrast, the Amharic version achieved a higher response rate of 85%. Although this discrepancy does not compromise the integrity of the statistical analyses or the conclusions drawn, it highlights the unequal sample sizes utilized in the research. Secondly, the findings of this study are not corroborated by local literature, largely due to the absence of previously conducted research that employs the same specific tool. These factors pose limitations to the current study's overall scope and applicability. Hence, we recommend that future researchers examine the construct on a large sample size. Furthermore, we recommend that future researchers carry out meta-analyses by collecting multiple studies of the construct via different languages to address similar MI assumptions.

Data Availability Statement

The data that support the findings of this study are available upon request from the corresponding author.

Competing Interests

The authors declare that no competing interests exist. The authors have no funding to disclose either.

Compliance with Ethical Standards

The principle of informed consent was obtained from all the participants, as per the ethical standards aligned with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards in APA. We got an ethical clearance letter from the Research Ethics Committee of Addis Ababa University, College of Education and Behavioral Studies, School of Psychology (Date: 01/02/2023, No. SoP–Eth/005/23).

Authors' Contribution

Both authors contributed to the study. They formulated the problem together. The first author carried out tasks such as data collection, analysis and manuscript preparation. The second author reviewed the final manuscript.

Acknowledgements

We would like to express our gratitude to the experts who involved in the adaptation, translation, and content validity assessment processes. Additionally, we extend our sincere appreciation to the study participants, whose contributions were invaluable to our research.

References

Adamu, A. Y. (2020). Ethnic violence challenging Ethiopian universities. Available at https://www.insidehighered.com/blogs/world-view/ethnic-violence-challenging-ethiopian

Akaike, H. (1987). *Factor analysis and AIC. Psychometrika, 52*(3), 317–332. https://doi.org/10.1007/BF02294359

Ashdown, B.A., Homa, N., & Brown, C.M. (2014). Measuring gender identity and religious identity with adapted versions of the Multigroup Ethnic Identity Measure—Revised. *Journal of Educational and Developmental Psychology, 4*(1), 226–237. https://doi.org/10.5539/JEDP.V4N1P226

Bellamy, R. (2008). *Citizenship: A very short introduction*. Oxford: Oxford University Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.). *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance. *Psychological Bulletin, 105*(3), 456–466. https://doi.org/10.1037/0033-2909.105.3.456

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology, 95*(5), 1005–1018. https://doi.org/10.1037/a0013193

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-to-fit indexes for testing measurement invariance. *Structural Equation Modeling: A multidisciplinary Journal, 9*(2), 233–255. https://doi.org/10.1207/ S15328007SEM0902_5

Collier, J. E. (2020). *Applied structural equation modeling using AMOS: Basic to advanced techniques*. Routledge.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*(1), 16-29.
https://doi.org/10.1037/1082-989X.1.1.16

Demlie, D., & Andualem, T. (in press). Adaptation and Validation of the Multigroup Ethnic

Identity Measure (MEIM) to the Ethiopian National Identity Measure. *Abyssinia Journal of Business and Social Science*.

Diane, L., & Putnick, M. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Espinosa, A., Schmitz, M., & Rottenbacher, J. (2016). Ideological bases of national identity and their influence on subjective, psychological and social well-being in Peru. *Network of Scientific Journals from Latin America, the Caribbean, Spain and Portugal, 11*(35), 18–32.

Fisher, S., Zapolskib, T.B., Wheeler, L., Arora, P.G., Barnes-Najore, J. (2020). Multigroup Ethnic Identity Measurement invariance across adolescence and diverse ethnic groups. *Journal of Adolescence, 83*, 42–51. https://doi.org/org/10.1016/j.adolescence.2020.07.006

Fisseha, Y. (2016). *Conceptualizations and impacts of multiculturalism in the Ethiopian education system*. A PhD Thesis: University of Toronto.

Gaines, S.O., Marelich, W., Bunce, D., Robertson, T., & Wright, B. (2013). Multigroup Ethnic Identity Measure (MEIM) expansion: Measuring racial, religious, and national aspects of sense of ethnic identity within the United Kingdom. Identity: *An International Journal of Theory and Research, 13*(4), 289–317. https://doi.org/10.1080/15283488.2013.780973

Habibi, M., Etesami, M.S., Taghizadeh, M.A., Akrami, F.S., Garcia, D. (2021). Psychometric properties of the Multi-group Ethnic Identity Measure (MEIM) in a sample of Iranian young adults. *PeerJ9*:e10752. http://doi.org/10.7717/peerj.10752

Hair, Jr. J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning; EMEA.

Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations across Nations* (2nd ed.). Sage, Thousand Oaks, CA. https://doi.org/10.1016/S0005-7967(02)00184-5

Hutchinson, J. (2019). Nations and nationalism: the first twenty-five years. *Journal of the Association for the Study of Ethnicity and Nationalism, 25*(1), 5–15. https://doi.org/10.1111/nana.12489

Maehler, D. B., Zabal, A., & Hanke, K. (2019). Adults' Identity in Acculturation Settings: The Multigroup Ethnic & National Identity Measure (MENI). *Identity: An International Journal of Theory and Research, 19*(4), 245-257. https://doi.org/10.1080/15283488.2019.1641408

Mastrotheodoros, S., Dimitrova, R., Motti-Stefanidi, F., Abubakar, A., & Schoot, R. (2012). Measurement invariance of the Multigroup Ethnic Identity Measure (MEIM) across Bulgarian, Dutch and Greek samples. *European Journal of Development al Psychology, 9*(4), 508–515. https://doi.org/10.1080/17405629.2012.688099

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93*(3), 568–592. https://doi.org/10.1037/0021-9010.93.3.568

Musso, P., Moscardino, U., & Inguglia, C. (2017). The Multigroup Ethnic Identity Measure–Revised (MEIM-R): Psychometric evaluation with adolescents from diverse ethnocultural groups in Italy. *European Journal of Developmental Psychology,* 1–16. https://doi.org/10.1080/17405629.2016.1278363

Nigbur, D., & Cinnirella, M. (2007). National identification, type and specificity of comparison and their effects on descriptions of national character. *European Journal of Social Psychology, 37*(4), 672–691. https://doi.org/10.1002/ejsp.382

Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. Routledge. 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. https://doi.org/10.1177/074355489272003

Phinney, J. S., & Ong, A. (2007). Conceptualization and measurement of ethnic identity: current status and future directions. *Journal of Counseling Psychology, 54*(3), 271–281. https://doi.org/10.1037/0022-0167.54.3.271

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: two approaches for exploring measurement invariance. *Psychological Bulletin, 114*(3), 552–566. https://doi.org/10.1037/0033-2909.114.3.552

Roberts, R. E., Phinney, J. S., Masse, L. C., Chen, Y., Roberts, C. R., & Romero, A. (1999). The structure of ethnic identity of young adolescents from diverse ethnocultural groups. *The Journal of Early Adolescence, 19*(3), 301-322. https://doi.org/10.1177/0272431699019003001

Schwartz, S. J., Park, I. J. K., Huynh, Q.-L., Zamboanga, B. L., Umaña-Taylor, A. J., Lee, R., Rodriguez, L. M., Kim, S. Y., Whitbourne, S. K., Castillo, L. G., Weisskirch, R. S., Vazsonyi, A. T., Williams, M. K., & Agocha, V. B. (2012). The American Identity Measure: Development and validation across ethnic group and immigrant generation. *Identity: An International Journal of Theory and Research, 12*(2), 93–128. https://doi.org/10.1080/15283488.2012.668730

Schwartz, S.J., Unger, J.B., Des Rosiers, S.E., Lorenzo-Blanco, E.I., Zamboanga, B.L., Huang, S., Baezconde-Garbanati, L., Villamar, J.A., Soto, D.W., Pattarroyo, M., & Szapocznik, J. (2014). Domains of acculturation and their effects on substance use and sexual behavior in recent Hispanic immigrant adolescents. *Prevention Science, 15*(3), 385–396. https://doi.org/10.1007/s11121-013-0419-1

Smith, P. B., Giannini, M., Helkama, K., Maczynski, J., & Stumpf, S. (2005). Positive auto-stereotyping and Self-construal as predictors of national identification. *International Review of Social Psychology, 18*(1), 65–90.

Storm, E. (2022). The rise of the nation-state during the Age of Revolution: Revisiting the debate on the roots of nations and nationalism. *Journal of the Association for the Study of Ethnicity and Nationalism, 28*(4), 1137–1151. https://doi.org/10.1111/nana.12857

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: Nelson-Hall.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–70. https://doi.org/10.1177/109442810031002

Vedder, P. H. (2005). Language, ethnic identity, and the adaptation of immigrant youth in the Netherlands. *Journal of Adolescent Research, 20*(3), 396-416. https://doi.org/10.1177/0743558405274853

Widaman, K. F., & Reise, S. P. (1997). *Exploring the measurement invariance of psychological instruments: applications in the substance use domain.* In K. J. Bryant, M. E. Windle, & S. G. West (Eds.), The science of prevention: Methodological advances from alcohol and substance abuse research (pp. 281–324).Washington, DC: American Psychological Association.

Windari, T. (2021). National Identity Attachment and Its Variables. *Journal of International Women's Studies, 22*(3), 81–95. https://vc.bridgew.edu/jiws/vol22/iss3/9

Woods, E.T., & Debs, M. (2013). Towards a cultural sociology of nations and nationalism. *Nations and Nationalism, 19*(4), 607–614. https://doi.org/10.1111/nana.12036

Zekarias, B. (2020). Ethnic-based conflicts in Ethiopia's higher education: The case of Debre Berhan University.