INQUIRY: Critical Thinking Across the Disciplines
Vol. 34, No. 2 (Fall 2025): 65–103
DOI: 10.5840/inquiryct202593029
Online First:

Conducting Intersectional Quantitative Analysis with MAIHDA for Education Research

BEN VAN DUSEN
Iowa State University

JAYSON NISSEN
Montana State University

HEIDI CIAN
Maine Mathematics and Science Alliance

LUCY ARELLANO
University of California Santa Barbara

Abstract: Multilevel analysis of individual heterogeneity and discriminatory accuracy (MAIHDA) enables intersectional quantitative educational research with distinct advantages over fixed-effects models. Using data from 9,672 physics students across 40 institutions, we compared MAIHDA to traditional fixed-effects models to assess the two methods' theoretical alignment with intersectionality and ability to model outcomes for diverse social groups. The results indicated that MAIHDA provided more precise measures of outcomes for 95 of the 106 intersectional groups. The manuscript offers guidance for applying MAIHDA in educational research, including R code, and emphasizes the responsibility of researchers to consider critical quantitative theory throughout the research process.

Keywords: MAIHDA, intersectionality, critical quantitative theory, STEM education, quantitative methods

Introduction

Historically, research using a critical perspective has used qualitative methods (Gillborn, 2018). The last decade, however, has seen a growing interest in using critical perspectives in quantitative investigations

(Tabron & Thomas, 2023). While researchers have identified tensions between critical theory and quantitative methods (Garcia et al., 2018), these tensions have supported researchers in reimagining how quantitative studies can promote a more just society (Wells, 2015; Castillo & Strunk, 2024; López et al., 2023). One point of this tension lies in how researchers can use intersectionality in quantitative research. Intersectionality highlights the importance of examining the dynamic interactions between an individual's multi-faceted social identities and society's power structures that lead to oppression or privilege (Cho, 2013; Crenshaw, 1991). Quantitative analyses often require aggregating an individual's social identities for statistical power. This aggregation can lead to the erasure of groups and the differences across aggregated social identities (e.g., heterogeneity). While a single study or analysis cannot consider all identity dimensions, we agree with McCall (2005) and Walter and Andersen (2013) that aggregation decisions should be driven by theoretical considerations rather than the limitations of the methods. Unfortunately, many studies are forced to aggregate students based on the limitations of their data and methods (Van Dusen, 2022).

To reduce the need for researchers to aggregate data to meet the limitations of their methods, Evans et al. (2015; 2018) proposed *multilevel analysis of individual heterogeneity and discriminatory accuracy* (MAIHDA) as a statistical modeling method designed for intersectional research. Evans and colleagues developed MAIHDA in the context of health studies. Subsequent investigations have shown its efficacy in education research (Van Dusen et al., 2024; Keller et al., 2023; Prior et al., 2022). To support researchers in taking up MAIHDA in their investigations, we set out to accomplish four tasks in this manuscript:

1. Provide an overview of MAIHDA and how it differs from intersectional fixed-effect models.
2. Examine the ways that MAIHDA aligns with intersectionality and critical quantitative theory.
3. Provide an empirical example of the benefits of using MAIHDA for intersectional modeling in a science education context.
4. Provide practical guidance on creating MAIHDA models.

In what follows, we will assume that our readers have a basic understanding of multilevel modeling. For more information on multilevel models, we recommend Woltman et al. (2012), Raudenbush (2002), and Peugh (2010).

Definitions

We present a few definitions of terms in Table 1 that we use throughout the manuscript.

Table 1. Definition of Terms.

Term	Definition
Aggregation	Combining members of identity groups into a single group for the purpose of drawing statistical conclusions.
Bayesian statistics	A statistical approach that incorporates researchers' prior knowledge to predict the probability of an outcome. (Lee, 1989)
Critical quantitative (CritQuant) theory	CritQuant integrates critical theory with quantitative research methods to challenge power structures and address social inequalities by analyzing data through a lens of equity, justice, and systemic oppression. (Tabron & Thomas, 2023)
Cross-classified multilevel model	An extension to multilevel models that allows for higher-level units to share a level. For example, a student (level 1) can be nested in both a major (level 2a) and a course (level 2b).
Intersectional identities	An individual can hold multiple memberships in various social identity groups, e.g., race, ethnicity, gender, sexuality, class, religion, ability, etc.
Intersectionality	A theoretical framework named by Crenshaw (1989) that emphasizes that people do not experience discrimination or privilege based solely on one identity category; rather, power structures interact dynamically with multiple facets of identities to create unique forms of advantage or disadvantage.
MAIHDA	Alternative to fixed-effects modeling that leverages multilevel modeling to nest individuals within strata (Evans, 2015).
Strata	A code (typically numeric) that represents each group in the model. These groups are based on social identities, but can also include other model factors that interact with social identity groups. For example, our model included race, gender, and whether it was pretest or posttest scores.
Shrinkage	Bayesian shrinkage is a statistical feature of multilevel models by which the predicted outcomes for a strata are informed by data from other strata. This process, also known as 'borrowing strength,' helps to stabilize estimates, especially when data for some strata are sparse. By incorporating information from the entire dataset, Bayesian shrinkage reduces the variance of estimates and improves their accuracy (Raudenbush & Bryk, 1986; Evans et al. 2024).

Positionality

Ben Van Dusen: I identify as a continuing-generation White cisgender, heterosexual man with a mild visual impairment and partial hearing

disability. I was raised in low-income households, but now earn an upper-middle-class income. People with similar privileges have created and maintained our society's unjust power structures. I believe it is the obligation of those with the privilege to dismantle oppressive systems. My privilege, however, limits my perspective on the lived experiences of marginalized individuals.

Jayson Nissen: My identity as a White, cisgendered, heterosexual, nondisabled man has provided me with an assumed acceptance and opportunities denied to others in the sciences. I focus on science education to better share the financial and intellectual benefits that science has provided me by broadening participation in the sciences. Growing up in a poor home and as a veteran of the all-male submarine service informs my work and the need for quantitative research that emphasizes disaggregation and intersectionality. My work on this project was shaped by the post-positivist scientific culture I was educated in and my questioning of objectivity that lies at the core of that culture.

Heidi Cian: I am drawn to studying intersectionality and the use of social groups in research to understand how learners of all ages can experience STEM in ways that affirm how they identify with their bodies, communities, and histories. I am particularly interested in how intersectional research can consider group membership beyond the traditional demographic categories of gender and race, and enlighten understanding of how individuals experience STEM opportunities in ways that respect or reject their politics and cultural wealth. I identify as a White cisgender heterosexual woman with centrist political views and a rural, lower socioeconomic upbringing.

Lucy Arellano: As a Xicana from East L.A., raised in a low-income household, the first in my family to be born in the U.S., the first to attend college, moving 2,500 miles away from home to enter an undergraduate STEM major at a prestigious university, I utilize intersectionality and quantitative methods to understand my own journey through higher education. Now, utilizing my privilege as a cisgender, heterosexual professor at an R-1 university, I endeavor to spotlight the multiple systems of oppression endured by minoritized student groups to transform postsecondary institutions.

MAIHDA Background

When developing intersectional models, education researchers will often account for the dynamic interactions between facets of an individual's social identity (e.g., gender and race) by including interaction terms between each social identity variable (Van Dusen et al., 2024). For example, if a model included primary terms for White and man, it would also include an interaction term for White*man. In this simple example, only a single

interaction term is required. However, the number of interaction terms can increase geometrically as social identity groups increase. We refer to these models as fixed-effect models as they rely on the fixed-effect coefficients to predict group outcomes.

Like fixed-effect models, MAIHDA models include primary terms for each social identity group of interest. Instead of using interaction terms, however, MAIHDA nests individuals in social identity groups (Figure 1). The nesting creates an error term for each social identity group. This error term accounts for the difference between the actual outcomes for each intersectional group and the predicted outcomes for those groups based on the fixed-effects terms. Researchers can then generate predicted outcomes for each social identity group by combining their relevant primary terms with the error term.

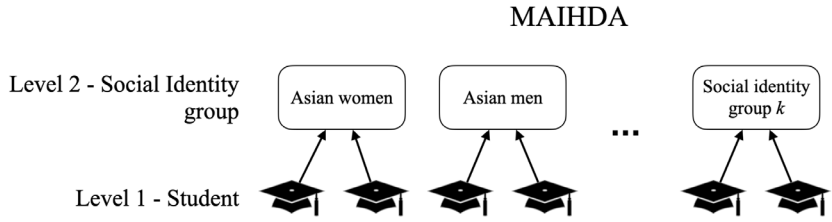

Figure 1. MAIHDA model structure with the study participants nested within their intersectional identity group.

Replacing a large number of fixed-effect interaction terms with error terms allows MAIHDA models to produce more accurate predictions (Bell et al., 2019; Evans et al., 2020; Lizotte et al., 2020), particularly for small-N groups (Van Dusen et al., 2024). The improved performance stems from two places. First, MAIHDA reduces the number of fixed-effect terms, thereby reducing the model's statistical power requirements. Second, the error terms generated by MAIHDA models are unlikely to be extreme because shrinkage pulls estimates toward the value predicted by the fixed-effect terms, preventing overfitting to small, noisy samples. This shrinkage ensures that the model produces more reasonable estimates by balancing the specific data for each group with the overall trend. This results in more stable and reliable predictions, especially when data is sparse or noisy.

Conceptual Framework

We grounded our theoretical perspective in *intersectionality* and *CritQuant*. In this manuscript, we use MAIHDA as a concrete example of how these theories can intersect in practice.

Intersectionality

Intersectionality originated as a legal framework (Crenshaw, 1991) that examines the dynamic interactions of social positions (e.g., race, gender, and class) and their relations to power structures to understand societal inequities (Bauer et al., 2021). For example, the inequities between men and women created by power structures in the science, technology, engineering, and mathematics disciplines are well established (Cheryan et al., 2017; Cimpian et al., 2020). However, the impacts of oppressive systems on intersecting gender identification with other identity factors such as geographic region (Galvin et al., 2024), racial identity (Ireland et al., 2018), socioeconomic status (Hailu, 2022), experience of colonialism (Idahosa & Mkhize, 2021) and religious affiliation (Avraamidou, 2020) have shown that examining STEM-related outcomes for "women" as a whole obscures systemic marginalization of women occupying space not just as women but also as racialized individuals within cultural contexts. As such, educational research has embraced intersectional interpretive lenses in qualitative research, bringing awareness that individuals occupying multiple marginalized identities exist in distinct categories at the intersections of those identities.

Intersectional theory's focus on individuals' unique and dynamic experiences lies in tension with quantitative method's aggregation of individuals into groups for statistical analyses. While this tension can be productive (Cho et al., 2017), historical quantitative research practices have oppressed marginalized groups of people (Strunk, 2023) through erasure (Lopez & Gaskin, 2022).

Quantitative research can lead to data "erasure" of social identity groups through researchers' aggregation decisions. Aggregating groups who share a single identifying feature (e.g., race) ignores intersectional experiences and can focus on a single cause of marginalization (e.g., "racism" rather than "racism and ableism"). This "erasure" through aggregation can produce conclusions about a broad category (e.g., Asian Americans in STEM; "Indigenous people") that is driven by a subcategory that only partially represents that larger category (e.g., Southeast Asian Americans; Jang, 2018) (Cole, 2009; Shafer, 2021). Hailu (2022) illustrates how, even among a relatively narrowly-defined population (i.e., Ethiopian STEM college women), considerations of economic and geographic background yield disparate experiences of gender marginalization. Quantitative researchers can address data erasure by collecting finer-grained demographic data, collecting data from more participants, and disaggregating that data in the analysis. Executing these aims, however, requires resources. Quantitative researchers can draw on qualitative findings and community participation to target their data collection or to seek additional resources to mitigate data erasure. To address data erasure, demographic surveys can also provide more categorical choices (Rubin et al., 2018), such as identifying with a

home country rather than aggregating "Hispanic American" or providing tribal nation drop-down menus or write-in options for American Indian and Alaskan Native (AIAN) participants.

Multilevel analysis of individual heterogeneity and discriminatory accuracy (MAIHDA) shows promise as an analytical method conducive to applying an intersectional lens (Evans, 2015; Evans et al., 2018; Merlo, 2018; Evans, 2019). MAIHDA leverages multilevel modeling to nest individuals within social identities and allows for smaller-*N* groups (Van Dusen et al., 2024) without meaningfully increasing the statistical power required to include many unique groups. This limits the risk of data erasure by maintaining accurate predictions with smaller grain sizes that account for intersectional groups that are less represented in the population, such as AIAN respondents.

Critical Quantitative Theory

Applying a CritQuant lens requires attending to both theoretical and technical considerations. MAIHDA does not, on its own, call out and disrupt structural inequities in education. In our presentation of MAIHDA, we use a CritQuant perspective to center the responsibility of educational inequities on oppressive power structures (Garcia et al., 2018). Quantitative researchers' responsibility lies in using tools to understand and improve *inequitable systems* instead of placing the onus for those inequities on the individuals *marginalized by those systems.* We align our approach to MAIHDA with the tenets of Quantitative Critical Race Theory (Gillborn et al., 2018) and Critical Race Quantitative Intersectionality (Covarrubias et al., 2017). Our CritQuant perspective combines and expands on these tenets to include other forms of oppression beyond racism: 1) the centrality of oppression, 2) data and methods are not neutral, 3) data cannot speak for itself, 4) groups are neither natural nor inherent, and 5) the importance of intersectionality.

1. The Centrality of Oppression

We take as a fact that structural racism and sexism plague the U.S. economic, political, and educational systems. Inequities in student performance result from systems-wide policies and approaches that implicitly and explicitly disadvantage broad groups of students (Solorzano et al., 2000). Society's power structures oppress learners across a myriad of social identities beyond gender and race, including ableness, religious affiliation, and socioeconomic status. We commit to designing and interpreting anti-racist, anti-sexist quantitative research approaches to disrupt narratives and oppressive systems that frame minoritized students as deficient.

When discussing the inequities we find using MAIHDA, we can name them as racist and sexist outcomes. By a priori stating that oppressive

power structures create inequities in outcomes, we can focus our discussion on identifying the mechanisms and impacts of these oppressive systems.

2. Data and Methods Are Not Neutral

Researchers often fail to adequately discuss the biases introduced in data collection and analytical methods. This leads many researchers and readers to interpret quantitative findings as objective facts (Wyly, 2009) and obscures the reality of culturally-contextualized understandings of quality and ethical research (Chilisa et al., 2016). We acknowledge that all data and analyses introduce biases and strive to minimize and acknowledge these biases.

We strive to avoid biasing findings by critically examining commonly used methods. For example, it is problematic in equity research to use *p*-values as go/no-go tests to identify group differences (Wasserstein et al., 2019). *p*-values depend on sample size. As many minoritized groups are underrepresented in STEM disciplines, collecting sufficiently large sample sizes to find statistically significant differences between them and another group is prohibitive for many research projects. This challenge compounds as research projects consider further disaggregating socially defined groups (e.g., racial groups) to account for intersectional experiences (e.g., Black men). The lack of data from minoritized learners has led many research projects to find that racism's impacts are not statistically significant and conclude that equity was achieved. Instead of using *p*-values to represent uncertainty, we use point estimates and standard errors to distinguish between the meaningfulness and uncertainty of a measurement (Wasserstein et al., 2019; Amrhein et al., 2019).

3. Data Cannot Speak for Itself

When researchers present data or findings without an explicit perspective, readers will likely interpret the data and findings through the dominant perspective, which often leads to racist and sexist interpretations (see Tenet 1; Zuberi, 2008). Such interpretations can reinforce existing deficit narratives about minoritized groups. Researchers must actively speak for their data to counter the default deficit narrative (Covarrubias & Vélez, 2013). The importance of speaking for data increases as the complexity of the statistical model increases, thereby creating more opportunities for misinterpretation.

One way we speak for our data is that, when discussing differences between social identity groups, we don't refer to differences in outcomes as gender or racial gaps. Instead, we draw on Ladson-Billings' (2006) idea of educational debts to frame the inequities as debts that society owes marginalized groups. This framing places the onus of repaying debts on the inequitable social systems that created those debts rather than the marginalized individuals.

When interpreting our model findings, we go beyond looking at the coefficients. We combine coefficients to create full predictions for groups and then visualize these outcomes. The coefficients describe how groups differ from a normative group that is often the privileged group. Combining coefficients, we can create descriptions of groups that do not rely on a normative group. We visualize group outcomes to facilitate the examination and discussion of heterogeneity within the learning context.

4. Groups Are Neither Natural Nor Inherent

In U.S. society, people often take an individual's racial and gender identity as immutable. We instead acknowledge these identity categories to be socially constructed and fluid. Statistical analysis requires aggregation of data to support claims about group outcomes. However, we strive to create models that respect learners' identities, judiciously aggregate data, represent diversity in learner outcomes, and be transparent in our decisions regarding social identity groups. For example, in this analysis, we did not aggregate groups beyond how they self-identified on the demographic questions.

Taken with Tenet 3, we see that these categories do not define *individuals* as much as they help to understand the *shared oppressive experiences* that individuals experience due to assignment to these socially-constructed identity groups. In the work we describe, we therefore draw from the presumption of immutability to use these socially constructed groups as tools to make sense of the centrality of oppression (see Tenet 1).

5. The Importance of Intersectionality

Addressing student experiences and outcomes from an intersectional perspective is critical to understanding the complex interactions between social identity groups and power structures. While the models we examine predict outcomes across social identity groups, the coefficients do not describe the groups but the ways that oppressive power structures impact them. For more details, see the discussion of intersectionality above.

Empirical Example

Van Dusen et al. (2024) found that MAIHDA performs better than fixed-effects models when examining outcomes for multiply marginalized identities. However, because it is a novel method, it is understandable for researchers to want to continue using the methods they are familiar with unless a new method provides clear and compelling reasons to begin using it. In this empirical example, we outline how to create a MAIHDA model and the potential benefits of doing so. Our subsequent practical guidance section provides further advice on engaging in high-quality MAIHDA modeling.

Mahendran et al. (2022a; 2022b) found that MAIHDA models provided robust estimate accuracy with average group sample sizes as low as N=10.4. Van Dusen et al. (2024) found that MAIHDA models could generate more accurate estimates of outcomes for groups with a minimum sample size of 10 than fixed-effect models could do with a minimum sample size of 20. In this empirical example, we explore the practical value of this difference by addressing two questions:

1) How does the shift from a minimum sample size of 20 to 10 students affect the number of groups represented in the model?
2) How does MAIHDA, compare to fixed-effect models in the degree of uncertainty in predicted student outcomes, particularly for small-*N* groups?

Methods

Data Collection, Cleaning, and Imputation

Data for this analysis came from the LASSO platform's research database (Van Dusen, 2018; Van Dusen et al., 2021). LASSO is an online assessment platform that hosts, administers, and analyzes research-based assessments for STEM instructors and students. Data in the LASSO research database is anonymized and only includes students and instructors who consented to share their data.

We analyzed pretest (i.e., first week of class) and posttest (i.e., last week of class) scores on the Force Concept Inventory (FCI; Hestenes et al., 1992) from 9,672 students in 310 college calculus-based physics courses across 40 institutions. All of the instructors in this sample self-report engaging students in collaborative learning. The Force Concept Inventory is a commonly used research-based assessment (Henderson, 2002) designed to assess student's conceptual knowledge of kinematics and forces. The FCI has undergone significant validation work (e.g., Wang & Bao, 2010; Eaton & Willoughby, 2018), including examinations of differential item function (Traxler et al., 2018; Buncher et al., 2021) and construct invariance (Morley et al., 2023).

The FCI typically takes students 15–30 minutes to complete. To clean the data, we removed any scores from students who completed the assessment in under five minutes, indicating a student was not attempting to answer the questions correctly.

Each assessment was administered twice, once as a pretest at the start and once as a posttest at the end of the term. Of the students who completed at least one of the administrations, 84% completed the pretest, 61% completed the posttest, and 45% completed both the pretest and the posttest. To minimize the bias created by this missing data and to maximize the sample's power, we created ten imputed datasets using a

3-level hierarchical multiple imputation model in the mice package (Van Buuren & Groothuis-Oudshoorn, 2011). Our imputation model nested tests in students and students in courses. Visual examination of trace plots is a commonly used technique for checking convergence. Other methods for checking for convergence, such as R-hat scores, could be overly conservative for our application.

Data Analysis

To identify the unique social identity groups in our sample, we combined each of the race and gender options that a student selected or wrote in when completing their LASSO assessment. Our data had 546 unique social identity groups in it. Fifty-three of the social identity groups had ten or more students in them.

In our fixed-effect model, we included all fixed and interaction coefficients required to fully model pretest and posttest scores for the 53 social identity groups with at least ten students. For example, 13 students identified as American Indian or Alaskan Native (AIAN) Hispanic men. To ensure that the fixed-effect model could predict the group's scores, we included the primary coefficients *AIAN*, *Hispanic*, *men*, *posttest*, their six two-way interaction coefficients, four three-way interaction coefficients, and one four-way interaction term. Including the intercept and a coefficient to account for whether they had previously enrolled in the course, the fixed-effect model had 196 coefficients. The complete set of coefficients for both models can be seen in Supplemental Table 1. We visually examined the trace plots to ensure that the model converged.

To construct the MAIHDA model, we created unique numbers for each of the 546 social identity groups during each administration to serve as strata terms. By generating a set of unique numbers to serve as strata for each social identity group on the pretest and another set for the posttest, we created a total of 1092 strata that allowed the model to predict the pretest and posttest scores for each group independently. We then ran the model with all of the primary coefficients present in our 53 social identity groups with at least ten students. For example, to ensure that the MAIHDA model could predict scores for American Indian or Alaskan Native (AIAN) Hispanic men, we included the primary coefficients *AIAN*, *Hispanic*, *men*, *posttest*, and their strata. Including a coefficient to account for whether they had previously enrolled in the course, the MAIHDA model had 27 coefficients.

We ran the models using Bayesian functions in the brms package (Bürkner, 2017b). We used uninformed priors to simplify the example. Additionally, we are unsure how to create equivalent priors across the two types of models. The fixed-effect model (Figure 2) was a three-level model nesting tests (level 1) within students (level 2) within courses (level 3). The MAIHDA model (Figure 3) was a three-level cross-classified multi-level

model nesting tests (level 1) within students (level 2) within courses (level 3a) and strata (level 3b). The models took between 2 hours (MAIHDA) to 48 hours (fixed-effects) to run on a Mac Studio with 32GB of RAM and an M1 Max processor (See Table 1). We visually examined the trace plots to ensure that the model converged.

For each model, we combined the coefficients and strata (for the MAIHDA model) to create two predicted scores (pretest and posttest) with standard errors for each of the 53 social identity groups with at least ten students. This produced 106 sets of predicted scores and standard errors for each model. Supplemental Table 1 shows the coefficient estimates and standard errors for each model. Supplemental Table 2 shows both models' predicted group scores and standard errors for the pretest and posttest.

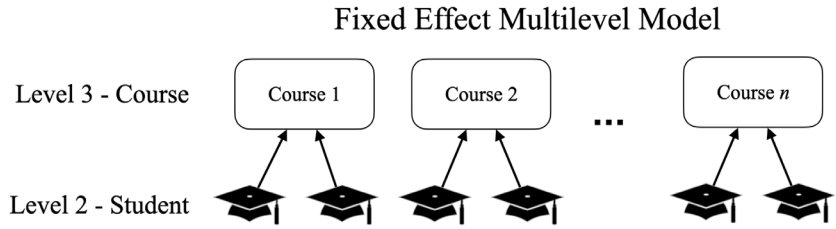

Figure 2. Level 2 (students) and 3 (courses) of the fixed-effect model. For simplicity, level 1 (tests) is not included in the figure.

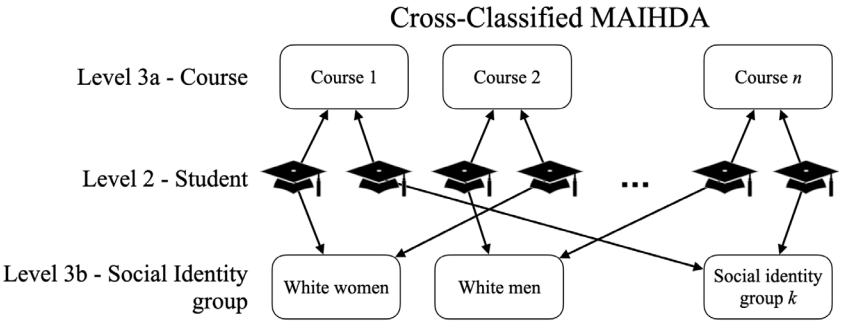

Figure 3. Level 2 (students), 3a (courses), and 3b (social identity groups) of the MAIHDA model. For simplicity, level 1 (tests) is not included in the figure.

Table 1. Data, coefficients, and model run times.

Model	Tests (level 1)	Students (level 2)	Courses (level 3a)	Strata (level 3b)	Soc. Iden. Group (N ≥ 10)	Coeffi-cients	Run time
Fixed Effect	12928	9403	310	-	53	197	48 hours
MAIH-DA	12928	9672	310	1092	53	28	2 hrs

Findings

How does the shift from a minimum sample size of 20 to 10 students impact the number of groups represented?

In our sample, 35 social identity groups had at least 20 students (Supplemental Table 2). Fifty-three social identity groups had at least ten students in them. Shifting the minimum sample size from 20 to 10 students resulted in a 54% increase in the number of social identity groups included.

How does using MAIHDA vs. fixed-effect models impact the uncertainty in student outcomes, particularly for small-N groups?

The MAIHDA model produced a smaller standard error in social identity groups' predicted scores for 95 of the 106 predicted groups compared to the fixed-effect model (Supplemental Table 2). The fixed-effect model produced mean standard errors in students' predicted scores that were 4.94 percentage points compared to 3.82 percentage points (23% smaller) for the MAIHDA model (Table 2). When using a MAIHDA model, the smaller sample size groups have the largest reduction in the mean standard errors (Figure 4). For the 18 social identity groups with sample sizes that ranged from 10 to 19 students, MAIHDA reduced the mean standard error from 6.32 percentage points to 4.41 percentage points (30% smaller).

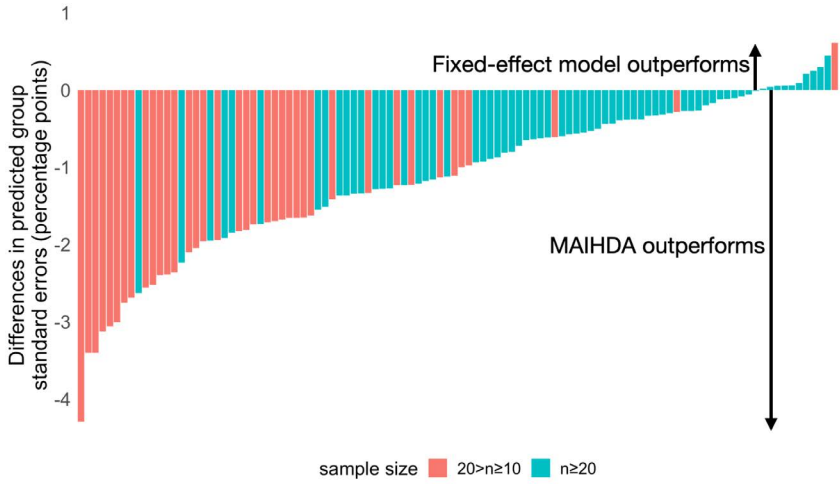

Figure 4. The difference in standard error terms for the 53 social identity groups predicted pretest and post-test scores across the two models. Values were calculated as $S.E._{MAIHDA} - S.E._{fixed\text{-}effect}$*, so negative values indicate smaller standard errors for predicted scores from the MAIHDA model.*

Table 2. The mean standard errors for the 53 social identity groups predicted pretest and post-test scores.

	Total	n≥20	20>n≥10
Model	Mean S.E. (perc. points)	Mean S.E. (perc. points)	Mean S.E. (perc. points)
Fixed Effect	4.94	4.17	6.32
MAIHDA	3.82	3.49	4.41
Difference	-23%	-16%	-30%

The reduced standard error terms, particularly for small-*N* groups, enable differentiating outcomes across groups that quantitative research seldom includes. For example, this dataset had three groups of AIAN men reach our minimum sample size requirement of 10 (AIAN, Hispanic or Latino, man (n=13); AIAN, man (n=11); AIAN, White, man (n=18)). The MAIHDA model created predictions with small enough uncertainty that the differences in outcomes are distinguishable (i.e., the confidence intervals across the groups have little overlap; see Figure 5). This degree of uncertainty allowed us to not only represent these commonly ignored groups, but also to disaggregate them to see intersectional impacts across multi-racial AIAN student groups.

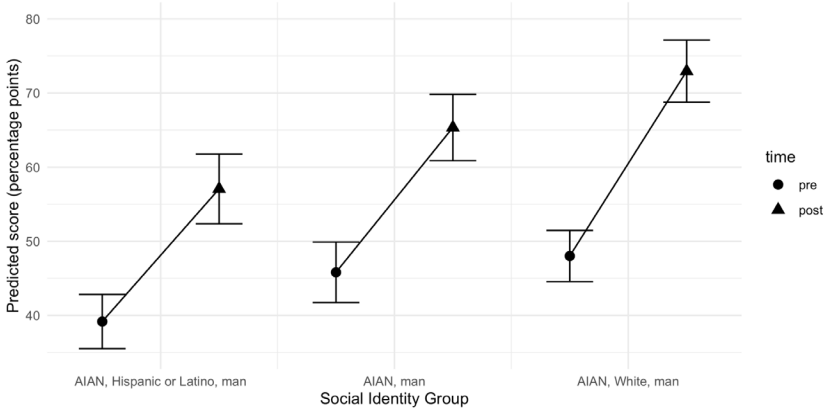

Figure 5. Predicted pretest and posttest scores for AIAN students from the MAIHDA model.

Discussion

We found MAIHDA to create a model with fewer fixed-effect terms that computed more quickly than fixed-effect models and produced more

precise predictions. While researchers should be vigilant to guard against artificially small error terms, Van Dusen et al. (2024) demonstrated that MAIHDA's smaller error terms were warranted. The smaller error terms for the MAIHDA model than the fixed-effect model follow from the decrease in the number of primary coefficients by 84% (from 196 to 27) when using MAIHDA.

MAIHDA models conserve statistical power by using fewer coefficients while leveraging Bayesian shrinkage to use information about related groups to produce more accurate estimates, even for small *N*-groups (Evans, 2024). One of MAIHDA's strengths lies in its ability to share information across related groups, allowing social identity groups to proliferate instead of being aggregated. By supporting researchers in moving beyond highly aggregated models, MAIHDA has the potential to improve our understanding of outcomes for communities that have been historically ignored and to identify interventions that create a more just and equitable society. For these reasons, we recommend MAIHDA over fixed-effect models for intersectional models that include 20 or more groups.

Practical Guidance

We have argued how MAIHDA can advance quantitative methodologies using critical theories such as intersectionality (Collins, 2015). We also provided an empirical example utilizing national data of physics students to illustrate the benefits of using MAIHDA over fixed-effects models, particularly for predicting outcomes for small-*n* social identity groups. We now offer guidance for employing MAIHDA in ways grounded in the tenets noted above–particularly being mindful of MAIHDA's opportunities to acknowledge multi-faceted identities. Some steps we share below are required for MAIHDA, while others will likely strengthen a MAIHDA analysis.

For each step, we refer the reader to resources that elaborate on the ideas that we present here. We intend our outline to be a foundation that researchers can start with as they consider using MAIHDA in ways aligned with Critical quantitative theory and encourage readers to consult these recommended resources for deeper explorations.

To support readers in following our recommended steps, we have annotated the R code from our analysis and made it and the data available at https://github.com/benvandusen/MAIHDA_Example .

Data Preparation

When engaging in intersectional modeling–regardless of the methodology–the size and diversity of the dataset and the quality of its social identity data lay the foundation for rigorous analytical approaches. MAIHDA

lends itself to analyzing large datasets that include 20 or more social identity groups (Evans, 2019). Those social identifiers, however, are social constructs that are neither natural nor inherent. Researchers can apply a critical framework to design their social identity data collection or to work with extant datasets that may have limited or problematic social identity data.

No single "best practice" for collecting social identity information exists. The context of the data collection and how it will be used should inform the methods used. Creating an inclusive set of social identity options and allowing students to select multiple choices or write their answers may gather more authentic responses (Abboud et al., 2019). In other contexts, however, these choices may lead to less authentic responses due to eliciting protest responses (Jaroszewski et al., 2018). Similarly, queries that ask for self-identification versus ethnic ancestry yield different response groupings and interpretations (Walter & Andersen, 2013). Researchers should consider research goals and their contexts when designing social identity questions.

Secondary data analysis limits researchers' control over how the data was collected. Yet, it is essential to consider how and what social identity data was collected and how these choices might bias findings. For example, institutional research datasets can be large but often constrain student's gender and racial identification. Researchers should contextualize the data collection and its limitations when interpreting their findings in these cases.

Find additional guidance in *An ethics and social-justice approach to collecting and using demographic data for psychological researchers* (Call et al., 2023) *More comprehensive and inclusive approaches to demographic data collection* (Fernandez et al., 2016), and *Indigenous Statistics: A Quantitative Research Methodology* (Walter & Andersen, 2013).

Data Preparation Step 2: Address Missing Data

Almost all datasets will have missing data. Improper handling of missing data can lead to biased and misleading results (Arellano, 2022). Most researchers use complete case analysis to handle missing data; this is not a neutral decision and can hide or bias differences or relationships across groups (Shafer, 1999). Unless very little data is missing and the researcher can argue that it is missing completely at random, the missingness will bias the findings (Baraldi & Enders, 2010). In educational research, the institutional injustices that contribute to inequities in student outcomes may also contribute to factors that influence participation in a study. For example, a student experiencing housing insecurity may miss more classes and be more likely to miss a day of data collection. In this case, using complete case analysis could obscure the outcomes of students experiencing housing insecurity. To mitigate bias in findings and retain statistical

power, researchers should address missing data using a principled method, such as multiple imputation (Rubin, 1996). Figure 6 shows the workflow for generating and analyzing multiply imputed datasets.

In our analysis for this publication, we performed hierarchical multiple imputation using the mice (Van Buuren & Groothuis-Oudshoorn, 2011) package in R. However, many other R packages and statistical programs can perform multiple imputations (e.g., blimp, SPSS, and SAS). Multiple imputation takes the dataset and uses statistical models and an element of randomness to impute the missing data. That imputation occurs several times, ten in the present study, to allow the modeling process to account for the uncertainty introduced by the missing data. Pooling the analyses from each imputed data set produces more accurate and often more statistically powerful results than complete case analysis (Shafer, 1999). Including all of the informative variables available in a missing data model tends to improve those models (Shafer, 1999). Hierarchical data requires hierarchical multiple imputation models. Woods and colleagues (2021) offer a decision tree for when and how to perform multiple imputations.

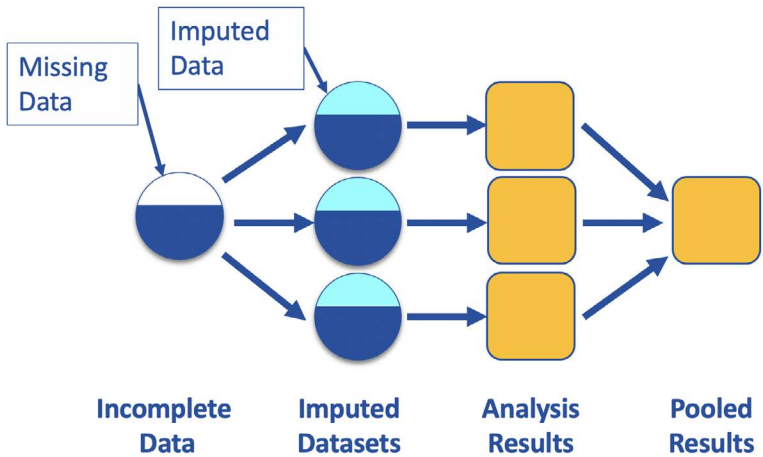

Figure 6. Multiple imputation involves imputing missing data to create a complete data set several times, analyzing those datasets independently, and then pooling the results to develop more accurate models.

When performing multiple imputation on social identity groups for a MAIHDA model, the imputation can create a social identity group/strata that exists in one imputation but not another. In our example, the data included race, gender, and first-generation college status. Some groups had an N of 1 and were missing the first-generation data. Some imputations categorized that strata as first-generation, others as continuing-generation college students. When this happened, an imputed dataset could have a strata that was not present in all of the other imputed datasets. This

prevented pooling the results across the imputed datasets. To address this issue, we removed any strata that do not exist in each imputed dataset. This strategy worked for our analysis but may not work for every case, and additional research can better inform how to handle missing data for MAIHDA models.

Find additional guidance in *Missing data and multiple imputation decision tree* (Woods et al., 2024), *What is the difference between missing completely at random and missing at random?* (Bhaskaran & Smeeth, 2014), and *Missing data and bias in physics education research: A case for using multiple imputation* (Nissen et al., 2019).

Data Preparation Step 3: Determine the Social Identity Groups to Model and Present in the Findings

Building intersectional models requires collecting and manipulating social identity data. Researchers have extensive degrees of freedom over what data they collect, what data they include in their models, and how they interpret those models. None of those decisions are neutral nor an inherent characteristic of the data. The social identity groups to include in a model should follow from the theory, research questions, and study design.

Research questions should motivate the social identity groups to include in a model. MAIHDA allows modeling students with all of the available social identity data combinations. In large data sets, this MAIHDA feature can prevent the need for creating aggregated groups, such as 'underrepresented minorities (URM)'. Despite this feature, researchers must make many decisions about handling social identity data.

When determining the minimum sample size for a group's predicted outcome to be included in the findings, researchers must balance the desire to represent marginalized groups with the model's statistical power. We have followed the guidance of a minimum size of 20 students in a strata when performing fixed-effect models (Simmons et al. 2011). MAIHDA models with a minimum size of 10 students in a strata can perform similarly well to fixed-effects models with a minimum size of 20 students in a strata (Van Dusen et al., 2024). So, we have used ten as our minimum group sample size. This is not a universal minimum group sample size. Every research project should establish its own cutoff value and articulate its rationale.

Find additional guidance in *On over-fitting in model selection and subsequent selection bias in performance evaluation* (Cawley et al., 2010), *Model selection in ecology and evolution* (Johnson & Omland, 2004); *Impact of broad categorization on statistical results: How underrepresented minority designation can mask the struggles of both Asian American and African American students* (Shafer et al., 2021), and *How statistical model development can obscure inequities in STEM student outcomes* (Van Dusen & Nissen, 2022).

Data Preparation Step 4: Defining the Strata for the MAIHDA Model

As noted in the example, using strata distinguishes MAIHDA from fixed-effects models. While MAIHDA requires creating strata, the prior three data preparation steps set the research up to include ones reflecting the four tenets.

Strata are a unique number for each social identity group. For example, Asian men's strata could be 1, Asian women's strata could be 2, and so forth until all the unique strata are represented. A versatile method for creating strata uses a series of 1s and 0s to represent all indicator variables for a student's social identity. For example, if the first variable is Asian, the second variable is women, and the third variable is men, then Asian men's strata would be "101" and Asian women's strata would be "110". Variables that interact with social identity groups must be integrated into the strata. For example, in our student pretest and post-scores model, we included an indicator term as the end of the strata for the pretest (0) and post-test (1). This allowed the model to predict each group's pretest and posttest scores without modeling each group as gaining the same amount from pretest to posttest.

Find additional guidance in *A tutorial for conducting intersectional multilevel analysis of individual heterogeneity and discriminatory accuracy (MAIHDA)* (Evans et al., 2024) and *Math versus meaning in MAIHDA: A commentary on multilevel statistical models for quantitative intersectionality* (Lizotte et al., 2020).

Modeling

Modeling Step 1: Select Coefficients for the Model

Researchers should include coefficients based on their research questions and the predicted group outcomes they will include in their findings. Including covariates is not a neutral decision as they change the meaning of the model coefficients. Including variables that control for prior educational debts (e.g., high school GPA, pretest scores, or SAT scores) can clarify relationships and differences or obscure the impact of oppressive power structures on students (Stewart, 2008).

We focus on social identity and contextual coefficients. The research question will determine the contextual coefficients (e.g., assessment administration, whether a student has previously taken the course, and intervention status) included in the model. The social identity coefficients (e.g., race, gender, and socio-economic status) should be the primary terms for the groups that will have predicted outcomes. For example, if the findings included Black women, Black men, Asian women, and Asian men, then the primary coefficients would include Black, Asian, men, and women. The model would not, however, include any of the interaction coefficients between race and gender. If the model measures a

relationship that varies across social identity coefficients, it will include it as both a primary term and an indicator term in the strata. For example, we modeled the pretest and the posttest scores by including a primary term for administration timing in the model and an indicator for it in our strata. Excluding the indicator in the strata would have created a model that assumed all groups had the same gain in scores from the pretest to the posttest.

Find additional guidance in *Critical issues in statistical causal inference for observational physics education research* (Adlakha & Ko, 2023), *Causation and race* (Holland et al., 2008), and *How statistical model development can obscure inequities in STEM student outcomes* (Van Dusen & Nissen, 2022).

Optional Modeling Step 2: Set Priors

In Bayesian modeling, priors allow researchers to account for their initial assumptions and knowledge about the study. For example, an analysis of the effectiveness of an intervention in repaying educational debts could use results from a meta-analysis of those debts as priors. These priors would allow the model to account for the findings from prior investigations and the researcher's assumption that those investigations applied to the current study. Using priors forces researchers to make explicit assumptions about an investigation and speak for their data.

Use prior publications or related datasets to generate estimates of what the coefficients should be and how confident you are in those values. Input the estimates into the Bayesian model. Running the model with and without priors can provide a sensitivity analysis on the impact of the priors on the model's coefficients.

Find additional guidance in *The use of Bayesian priors in ecology: The good, the bad and the not great* (Banner et al., 2020) and *Prior specification in Bayesian statistics: Three cautionary tales* (Van Dongen, 2006).

Modeling Step 3: Run the Model

MAIHDA models can use Bayesian or frequentist methods. Several aspects of Bayesian modeling can align with a critical quantitative perspective (Frisby, 2024). For example, Bayesian models provide a means of integrating a critical epistemology into models by including priors that reflect the impact of oppressive systems on student outcomes. Bayesian models also offer more accurate uncertainties when combining coefficients and error terms to create predicted outcomes (Evans et al., 2024). R packages (e.g., Bürkner, 2017b) have functions that can create Bayesian models from non-imputed data (e.g., brm) or imputed data (e.g., brms). The functions for creating models from imputed data often automate the process by pooling the model results.

Find additional guidance in *A tutorial for conducting intersectional multilevel analysis of individual heterogeneity and discriminatory accuracy*

(MAIHDA) (Evans et al., 2024), *brms: An R package for Bayesian multilevel models using Stan* (Bürkner, 2017b) and *An introduction to Bayesian multilevel models using brms: A case study of gender effects on vowel variability in standard Indonesian* (Nalborczyk et al., 2019).

Optional Modeling Step 4: Predict Outcomes

Researchers often use model coefficients as their primary findings. While the model coefficients can answer many research questions, they are typically insufficient for determining the confidence intervals for groups or the inequities between them. MAIHDA and similar complex models require the combination of multiple terms to predict a group outcome. While a typical coefficient table has the information necessary to combine coefficients to get a point estimate, it does not have the error terms required for MAIHDA models or the necessary information to calculate the confidence interval for that estimate. Statistical packages, however, can provide accurate point estimates with error terms.

Predicting each group's outcome begins by producing the posterior predictions for each coefficient in the Bayesian model. This creates a set of possible values for each coefficient. The next step adds the coefficients for each group together. For example, Black women's predicted outcomes might be: intercept + Black + women + $\text{error}_{\text{Black women strata}}$. In the case of the brms package, it will generate the group outcomes and their uncertainties.

Modeling Step 5: Interpret Outcomes

Model interpretation often relies on p-values exceeding a cutoff, such as $p < 0.05$. Researchers, however, should not "...conclude anything about scientific or practical importance based on statistical significance (or lack thereof)" (Wasserstein et al., 2019, p. 1).

Instead, researchers should interpret effect sizes and uncertainties in those effect sizes. These interpretations should rely on something other than common rules of thumb for effect sizes (Kraft, 2020). Amrhein et al. (2019) recommend interpreting the practical significance of the estimated value and both confidence limits to assess the meaningfulness of the model outcomes and the uncertainty in those outcomes. For example, a model may estimate the Cohen's *d* effect size for an educational debt to be 0.4 (a meaningful difference in most circumstances) with confidence intervals that range from 0 (unimportant) to 0.8 (very important). Researchers may then incorrectly dismiss this as insignificant because the lower uncertainty limit is zero and the p-value is above 0.05. This educational debt is likely to be meaningful despite having a large uncertainty ranging from unimportant to very important. Groups with small samples, such as those historically marginalized in education systems, will have larger uncertainties in their estimates. Dismissing differences as insignificant due to uncertainty when they may be meaningful and harmful is bad science and furthers student marginalization and injustice.

Find additional guidance in *Critical race quantitative intersectionality: A toiling movement-building paradigm that refuses to "let the numbers speak for themselves"* (Covarrubias et al., 2021) and *Moving to a world beyond "p< 0.05"* (Wasserstein et al., 2019).

Discussion and Implications

Our example illustrates that using MAIHDA to model intersectional outcomes is simpler and more accurate than fixed-effect models. Using the example of a small-*n* group (i.e., AIAN Hispanic men), we demonstrate how MAIHDA can allow researchers to disaggregate their data further, representing outcomes for commonly overlooked social identity groups. We present this example to show how disaggregation can be driven by the need for research to understand intersectional forces of oppression rather than by methodological limits. However, we do not suggest that researchers must always disaggregate to the smallest *n* size possible. Theory and research questions should determine the groups of interest and minimum group sample sizes. MAIHDA enables researchers to make more of these decisions driven by the inequities of interest and the capacity of their data rather than the limitations of their methods.

In providing practical guidance for using MAIHDA, we outline approaches researchers can take to account for the interactions between multi-faceted identities and power structures in their modeling. This outline mirrors approaches we have taken in analyzing the physics student data presented in the case study. Using MAIHDA, we integrate five tenets grounded in Quantitative Critical Race Theory (Gillborn et al., 2018) and Critical Race Quantitative Intersectionality (Covarrubias et al., 2017) to inform our decisions. For instance, we considered social identity groups to include in the model (Data Preparation Stage 3) in light of tenets that acknowledge the social construction of "groups" and that defining and drawing conclusions about social groups in quantitative research should acknowledge the researcher agency–and responsibility–in understanding these groups as historical and social constructions. This is where theoretical and conceptual frameworks (such as intersectionality) can aid the researcher in interpreting outcomes and understanding contexts yielding those results. However, we acknowledge that conducting analyses using MAIHDA without this critical lens is mechanically possible.

We further believe that the tenets we use in this paper describe a mindset that guides decision-making about quantitative research–not a checklist of what should be done to achieve equity-minded work. For readers familiar with science education standards in the Next Generation Science Standards, this can be likened to how many lessons touch on various competencies. Still, single lessons rarely address one indicator in isolation and in full. In this way, we encourage readers to take our outline

and the connections we drew to CritQuant as an example rather than an exemplar. By referring to the tenets that organized our work, we mean to express how these tenets can guide an equity-oriented mindset using methods that support drawing conclusions that involve multi-faceted identities. We do *not* believe there is a singular "how-to" for carrying out critical quantitative research.

References

Abboud, S., Chebli, P., & Rabelais, E. (2019). The contested Whiteness of Arab identity in the United States: Implications for health disparities research. *American Journal of Public Health, 109*(11), 1580–1583. https://doi.org/10.2105/AJPH.2019.305285

Adlakha, V., & Kuo, E. (2023). Critical issues in statistical causal inference for observational physics education research. *Physical Review Physics Education Research, 19*(2), 020160. https://doi.org/10.1103/PhysRevPhysEducRes.19.020160

Amrhein, V., Trafimow, D., & Greenland, S. (2019). Inferential statistics as descriptive statistics: There is no replication crisis if we don't expect replication. *The American Statistician*, 73(sup1), 262–270. https://doi.org/10.1080/00031305.2018.1543137

Arellano, L. (2022). Questioning the science: How quantitative methodologies perpetuate inequity in higher education. *Education Sciences, 12*(2), 116. https://doi.org/10.3390/educsci12020116

Avraamidou, L. (2020). "I am a young immigrant woman doing physics and on top of that I am Muslim": Identities, intersections, and negotiations. *Journal of Research in Science Teaching*, *57*(3), 311–341. https://doi.org/10.1002/tea.21593

Banner, K. M., Irvine, K. M., & Rodhouse, T. J. (2020). The use of Bayesian priors in Ecology: The good, the bad and the not great. *Methods in Ecology and Evolution, 11*(8), 882–889. https://doi.org/10.1111/2041-210X.13407

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology, 48*(1), 5–37. https://doi.org/10.1016/j.jsp.2009.10.001

Bauer, G. R., Churchill, S. M., Mahendran, M.., Walwyn, C., Lizotte, D., & Villa-Rueda, A. A. (2021). Intersectionality in quantitative research: A systematic review of its emergence and applications of theory and methods. *SSM-Population Health, 14*, 100798. https://doi.org/10.1016/j.ssmph.2021.100798

Bell, A., Holman, D., & Jones, K. (2019). Using shrinkage in multilevel models to understand intersectionality. *Methodology*. https://doi.org/10.1027/1614-2241/a000167

Bhaskaran, K., & Smeeth, L. (2014). What is the difference between missing completely at random and missing at random?. *International Journal of Epidemiology, 43*(4), 1336–1339. https://doi.org/10.1093/ije/dyu080

Buncher, J. B., Nissen, J. M., Van Dusen, B., Talbot, R. M., & Huvard, H. (2021, October). Bias on the Force Concept Inventory across the intersection of gender and race. In 2021 Physics Education Research Conference Proceedings. https://doi.org/10.1119/perc.2021.pr.Buncher

Bürkner, P. C. (2017a). Advanced Bayesian multilevel modeling with the R package brms. arXiv preprint arXiv:1705.11123. https://doi.org/10.32614/RJ-2018-017

Bürkner, P. C. (2017b). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Call, C. C., Eckstrand, K. L., Kasparek, S. W., Boness, C. L., Blatt, L., Jamal-Orozco, N., ... & Scholars for Elevating Equity and Diversity (SEED). (2023). An ethics and social-jus-

tice approach to collecting and using demographic data for psychological researchers. *Perspectives on Psychological Science, 18*(5), 979–995. https://doi.org/10.1177/17456916221137350

Carbado, D. W. (2013). Colorblind intersectionality. *Signs: Journal of Women in Culture and Society, 38*(4), 811–845. https://doi.org/10.1086/669666

Castillo, W., & Strunk, K. K. (2024) *How to QuantCrit: Applying Critical Race Theory to Quantitative Data in Education. Taylor and Francis*. https://doi.org/10.4324/9781003429968

Cawley, G. C., & Talbot, N. L. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *The Journal of Machine Learning Research, 11*, 2079–107.

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin, 143*(1), 1–35. https://doi.org/10.1037/bul0000052

Chilisa, B., Major, T. E., Gaotlhobogwe, M., & Mokgolodi, H. (2016). Decolonizing and indigenizing evaluation practice in Africa: Toward African relational evaluation approaches. *Canadian Journal of Program Evaluation*, *30*(3), 313–328. https://doi.org/10.3138/cjpe.30.3.05

Cho, S., Crenshaw, K. W., & McCall, L. (2013). Toward a field of intersectionality studies: Theory, applications, and praxis. Signs: Journal of women in culture and society, 38(4), 785–810. https://doi.org/10.1086/669608

Cimpian, J. R., Kim, T. H., & McDermott, Z. T. (2020). Understanding persistent gender gaps in STEM. *Science*, *368*(6497), 1317–1319. https://doi.org/10.1126/science.aba7377

Cole, E. R. (2009). Intersectionality and research in psychology. *American Psychologist, 64*(3), 1700–180. https://doi.org/10.1037/a0014564

Collins, P. H. (2015). Intersectionality's definitional dilemmas. *Annual Review of Sociology, 41*(1) 1–20. https://doi.org/10.1146/annurev-soc-073014-112142

Collins, P. H., & Bilge, S. (2020). *Intersectionality*. John Wiley & Sons.

Covarrubias, A., Lara, A., Nava, P., & Burciaga, R. (2021). Critical Race Quantitative Intersectionality: A Toiling Movement-Building Paradigm That Refuses to "Let the Numbers Speak for Themselves". In *Handbook of Critical Race Theory in Education* (pp. 279–295). Routledge. https://doi.org/10.4324/9781351032223-24

Covarrubias, A., & Vélez. V.N. (2013). Critical Race Quantitative Intersectionality: An Anti-Racist Research Paradigm that Refuses to "Let the Numbers Speak for Themselves." In *Handbook of Critical Race Theory in Education*, (pp. 270–285). New York: Routledge.

Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review 43*(6):1241–99. https://doi.org/10.2307/1229039

Eaton, P., & Willoughby, S. D. (2018). Confirmatory factor analysis applied to the Force Concept Inventory. *Physical Review Physics Education Research, 14*(1), 010124. https://doi.org/10.1103/PhysRevPhysEducRes.14.010124

Evans, C. R. (2015). Innovative approaches to investigating social determinants of health-social networks, environmental effects and intersectionality. Harvard University.

Evans, C. R. (2019). Adding interactions to models of intersectional health inequalities: comparing multilevel and conventional methods. *Social Science and Medicine*, *221,* 95–105. https://doi.org/10.1016/j.socscimed.2018.11.036

Evans, C. R., Leckie, G., & Merlo, J. (2020). Multilevel versus single-level regression for the analysis of multilevel information: the case of quantitative intersectional analysis. *Social Science and Medicine*, *245,* 112499. https://doi.org/10.1016/j.socscimed.2019.112499

Evans, C. R., Leckie, G., Subramanian, S. V., Bell, A., & Merlo, J. (2024). A tutorial for conducting intersectional multilevel analysis of individual heterogeneity and discriminatory accuracy (MAIHDA). *SSM-Population Health*, 101664. https://doi.org/10.1016/j.ssmph.2024.101664

Evans, C. R., Williams, D. R., Onnela, J-P, & Subramanian, S.V. (2018). A multilevel approach to modeling health inequalities at the intersection of multiple social identities. *Social Science and Medicine, 203,* 64–73. https://doi.org/10.1016/j.socscimed.2017.11.011

Fernandez, T., Godwin, A., Doyle, J., Verdin, D., Boone, H., Kirn, A., ... & Potvin, G. (2016). More comprehensive and inclusive approaches to demographic data collection. *School of Engineering Education Graduate Student Series*. Paper 60.

Frisby, M. B. (2024). A Case for Bayesian Statistics in Critical Quantitative Research in Education. https://doi.org/10.31234/osf.io/x6evk

Galvin, Danielle J., Susan C. Anderson, Chelsi J. Marolf, Nikole G. Schneider, and Andrea L. Liebl. (2024). Comparative analysis of gender disparity in academic positions based on US region and STEM discipline. *Plos One 19*(3), e0298736. https://doi.org/10.1371/journal.pone.0298736

Garcia, N. M., López, N., & Vélez, V. N. (2018). QuantCrit: Rectifying quantitative methods through critical race theory. *Race Ethnicity and Education, 21*(2), 149–157. https://doi.org/10.1080/13613324.2017.1377675

Gillborn, D., Warmington, P., & Demack, S. (2018). QuantCrit: education, policy, 'Big Data' and principles for a critical race theory of statistics. *Race Ethnicity and Education, 21*(2), 158–179. https://doi.org/10.1080/13613324.2017.1377417

Hailu, M. F. (2022). "I don't think it makes the difference": An intersectional analysis of how women negotiate gender while navigating STEM higher education in Ethiopia. *Comparative Education Review*, *66*(2), 321–341. https://doi.org/10.1086/718931

Henderson, C. (2002). Common concerns about the force concept inventory. *The Physics Teacher, 40*(9), 542–547. https://doi.org/10.1119/1.1534822

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher, 30*(3), 141–158. https://doi.org/10.1119/1.2343497

Holland, P. W., Zuberi, T., & Bonilla-Silva, E. (2008). Causation and race. White logic, white methods: *Racism and Methodology, 4*, 93–109. https://doi.org/10.5040/9798216386278.ch-005

Idahosa, G. E. O., & Mkhize, Z. (2021). Intersectional experiences of black South African female doctoral students in STEM: Participation, success and retention. *Agenda*, *35*(2), 110–122. https://doi.org/10.1080/10130950.2021.1919533

Ireland, D. T., Freeman, K. E., Winston-Proctor, C. E., DeLaine, K. D., McDonald Lowe, S., & Woodson, K. M. (2018). (Un) hidden figures: A synthesis of research examining the intersectional experiences of Black women and girls in STEM education. *Review of Research in Education*, *42*(1), 226–254. https://doi.org/10.3102/0091732X18759072

Jang, S. T. (2018). The implications of intersectionality on Southeast Asian female students' educational outcomes in the United States: A critical quantitative intersectionality analysis. American Educational Research Journal, 55(6), 1268–1306. https://doi.org/10.3102/0002831218777225

Jaroszewski, S., Lottridge, D., Haimson, O. L., & Quehl, K. (2018, April). " Genderfluid" or" Attack Helicopter" Responsible HCI Research Practice with Non-binary Gender Variation in Online Communities. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1–15). https://doi.org/10.1145/3173574.3173881

Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution, 19*(2), 101–8. https://doi.org/10.1016/j.tree.2003.10.013

Keller, L., Lüdtke, O., Preckel, F., & Brunner, M. (2023). Educational Inequalities at the Intersection of Multiple Social Categories: An Introduction and Systematic Review of the Multilevel Analysis of Individual Heterogeneity and Discriminatory Accuracy (MAIHDA) Approach. Educational Psychology Review, 35(1), 31. https://doi.org/10.1007/s10648-023-09733-5

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241–253. https://doi.org/10.3102/0013189X20912798

Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in US schools. *Educational Researcher*, *35*(7), 3–12. https://doi.org/10.3102/0013189X035007003

LASSO Platform. (n.d.). *LASSO*. https://lassoeducation.org/

Lee, P. M. (1989). *Bayesian statistics* (pp. 54–5). London:: Oxford University Press.

Lizotte, D. J., Mahendran, M., Churchill, S. M., & Bauer, G. R. (2020). Math versus meaning in MAIHDA: A commentary on multilevel statistical models for quantitative intersectionality. *Social Science & Medicine*, 245, 112500. https://doi.org/10.1016/j.socscimed.2019.112500

Lopez, J. D., & Gaskin, F. T. (2022). Whiteness and the erasure of indigenous perspectives in higher education. In *Critical Whiteness praxis in higher education* (pp. 245–255). New York: Routledge. https://doi.org/10.4324/9781003443919-15

López, N., Erwin, C., Binder, M., & Chavez, M. J. (2023). Making the invisible visible: Advancing quantitative methods in higher education using critical race theory and intersectionality. In N. M. Garcia, N. López, & V. N. Vélez (Eds.) *QuantCrit* (pp. 32–59). Routledge. https://doi.org/10.4324/9781003380733-3

Mahendran, M., Lizotte, D., & Bauer, G. R. (2022a). Describing intersectional health outcomes: an evaluation of data analysis methods. Epidemiology, 33(3), 395–405. https://doi.org/10.1097/EDE.0000000000001466

Mahendran, M., Lizotte, D., & Bauer, G. R. (2022b). Quantitative methods for descriptive intersectional analysis with binary health outcomes. *SSM-Population Health, 17*, 101032. https://doi.org/10.1016/j.ssmph.2022.101032

Mayhew, M. J., & Simonoff, J. S. (2015). Non-White, no more: Effect coding as an alternative to dummy coding with implications for higher education researchers. *Journal of College Student Development, 56*(2), 170–175. https://doi.org/10.1353/csd.2015.0019

McCall, L. (2005). The complexity of intersectionality. *Signs: Journal of Women in Culture and Society, 30*(3), 1771–1800. https://doi.org/10.1086/426800

Merlo, J. (2018). Multilevel analysis of individual heterogeneity and discriminatory accuracy (MAIHDA) within an intersectional framework. *Social Science and Medicine, 203,* 74–80. https://doi.org/10.1016/j.socscimed.2017.12.026

Morley, A., Nissen, J. M., & Van Dusen, B. (2023). Measurement invariance across race and gender for the Force Concept Inventory. *Physical Review Physics Education Research, 19*(2), 020102. https://doi.org/10.1103/PhysRevPhysEducRes.19.020102

Nalborczyk, L., Batailler, C., Lœvenbruck, H., Vilain, A., & Bürkner, P. C. (2019). An introduction to Bayesian multilevel models using brms: A case study of gender effects on vowel variability in standard Indonesian. *Journal of Speech, Language, and Hearing Research, 62*(5), 1225–1242. https://doi.org/10.1044/2018_JSLHR-S-18-0006

Nissen, J., Donatello, R., & Van Dusen, B. (2019). Missing data and bias in physics education research: A case for using multiple imputation. *Physical Review Physics Education Research, 15*(2), 020106. https://doi.org/10.1103/PhysRevPhysEducRes.15.020106

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of school psychology*, *48*(1), 85–112. https://doi.org/10.1016/j.jsp.2009.09.002

Prior, L., Evans, C., Merlo, J., & Leckie, G. (2022). Sociodemographic inequalities in student achievement: An intersectional multilevel analysis of individual heterogeneity and discriminatory accuracy (MAIHDA). Sociology of Race and Ethnicity, 23326492241267251.

Raudenbush, S., and Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 1–17. https://doi.org/10.2307/2112482

Raudenbush, S. W. (2002). Hierarchical linear models: Applications and data analysis methods. *Advanced Quantitative Techniques in the Social Sciences Series/SAGE*.

Ro, H. K., & Bergom, I. (2020). Expanding our methodological toolkit: Effect coding in critical quantitative studies. *New Directions for Student Services*, *169*, 87–97. https://doi.org/10.1002/ss.20347

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association, 91*(434), 473–489. https://doi.org/10.1080/01621459.1996.10476908

Rubin, V., Ngo, D., Ross, A., Butler, D., & Balaram, N. (2018). Counting a diverse nation: Disaggregating data on race and ethnicity to advance a culture of health. *PolicyLink*. Retrieved from https://www.policylink.org/sites/default/files/Counting_a_Diverse_Nation_08_15_18.pdf

Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical Methods in Medical Research*, 8(1), 3–15. https://doi.org/10.1191/096228099671525676

Shafer, D., Mahmood, M. S., & Stelzer, T. (2021). Impact of broad categorization on statistical results: How underrepresented minority designation can mask the struggles of both Asian American and African American students. *Physical Review Physics Education Research, 17*(1), 010113. https://doi.org/10.1103/PhysRevPhysEducRes.17.010113

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Solorzano, D., Ceja, M., & Yosso, T. (2000). Critical race theory, racial microaggressions, and campus racial climate: The experiences of African American college students. *The Journal of Negro Education, 69*(1/2), 60–73.

Stage, F. K. (2007). Answering critical questions using quantitative data. *New Directions for Institutional Research, 133*, 5–16. https://doi.org/10.1002/ir.200

Stewart, Q. T. (2008). Swimming upstream. White logic, White methods: *Racism and methodology*, 111–26. https://doi.org/10.5040/9798216386278.ch-006

Strunk, K. K. (2023). Critical apporaches to quantitative research. In M. D. Young & S. Diem (Eds) *Handbook of Critical Education Research* (pp. 56–74). Routledge. https://doi.org/10.4324/9781003141464-5

Tabron, L. A., & Thomas, A. K. (2023). Deeper than wordplay: A systematic review of critical quantitative approaches in education research (2007–2021). *Review of Educational Research, 93*(5), 756–786. https://doi.org/10.3102/00346543221130017

Traxler, A., Henderson, R., Stewart, J., Stewart, G., Papak, A., & Lindell, R. (2018). Gender fairness within the force concept inventory. *Physical Review Physics Education Research, 14*(1), 010103. https://doi.org/10.1103/PhysRevPhysEducRes.14.010103

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*, 1–67. https://doi.org/10.18637/jss.v045.i03

Van Dongen, S. (2006). Prior specification in Bayesian statistics: three cautionary tales. *Journal of Theoretical Biology, 242*(1), 90–100. https://doi.org/10.1016/j.jtbi.2006.02.002

Van Dusen, B. (2018). LASSO: A new tool to support instructors and researchers. American Physics Society Forum on Education Fall 2018 Newsletter.

Van Dusen, B., & Nissen, J. (2022). How statistical model development can obscure inequities in STEM student outcomes. *Journal of Women and Minorities in Science and Engineering, 28*(3). https://doi.org/10.1615/JWomenMinorScienEng.2022036220

Van Dusen, B., Shultz, M., Nissen, J. M., Wilcox, B. R., Holmes, N. G., Jariwala, M., ... & Pollock, S. (2021). Online administration of research-based assessments. *American Journal of Physics, 89*(1), 7–8. https://doi.org/10.1119/10.0002888

Walter, M., & Andersen, C. (2013). *Indigenous statistics: A quantitative research methodology*. Taylor & Francis.

Wang, J., & Bao, L. (2010). Analyzing force concept inventory with item response theory. *American Journal of Physics, 78*(10), 1064–1070. https://doi.org/10.1119/1.3443565

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond "p< 0.05". *The American Statistician, 73*(1), 1–19. https://doi.org/10.1080/00031305.2019.1583913

Wells, R. S., & Stage, F. K. (2015). Past, present, and future of critical quantitative research in higher education. New Directions for Institutional Research, 2014(163), 103–112. https://doi.org/10.1002/ir.20089

Woltman, H., Feldstain, A., MacKay, J. C., & Rocchi, M. (2012). An introduction to hierarchical linear modeling. *Tutorials in quantitative methods for psychology*, *8*(1), 52–69. https://doi.org/10.20982/tqmp.08.1.p052

Woods, A. D., Davis-Kean, P., Halvorson, M. A., King, K., Logan, J., Xu, M., ... & Vasilev, M. R. (2021). Missing data and multiple imputation decision tree. https://doi.org/10.31234/osf.io/mdw5r

Woods, A. D., Gerasimova, D., Van Dusen, B., Nissen, J., Bainter, S., Uzdavines, A., ... & Elsherif, M. M. (2024). Best practices for addressing missing data through multiple imputation. *Infant and Child Development, 33*(1), e2407. https://doi.org/10.1002/icd.2407

Wyly, E. (2009). Strategic positivism. *The Professional Geographer, 61*(3), 310–322. https://doi.org/10.1080/00330120902931952

Zuberi, T., & Bonilla-Silva, E. (Eds.). (2008). White logic, white methods: Racism and methodology. Rowman & Littlefield. https://doi.org/10.5040/9798216386278

Supplemental Material

Supplemental Table 1. Model coefficient estimates and estimated errors.

	MAIHDA		Fixed Effect	
	Estimate	Est. Error	Estimate	Est. Error
Intercept	43.8	8.6	-	-
americanindianoralaskannative	0.7	2.3	1.3	6.2
anasianracenotlisted	3.5	2.7	33.5	19.1
asian	2.5	1.9	42.7	16
asianindian	4.6	2.3	-19.2	28.8
black	-4.8	2.4	17.3	10.1
chinese	4.3	2.1	-30.4	22.6
colombian	-0.5	5	-1327.5	8814
cuban	-2.3	4.7	-842.2	7147.6
filipino	-1.9	2.4	24.4	11.8
genderqueergendernonconforming	-3.3	2.9	18.2	6.1
hispanicorlatino	-4.8	1.4	17.5	8.5
japanese	1.6	2.8	-9.3	9.1
korean	2.8	2.9	2198.4	8326.6
man	-3.2	2.4	44.3	3.2
mexicanmexicanamericanchicanochicana	-2.8	1.5	30.5	19.1
middleeasternornorthafrican	1.2	2.6	9.2	19.4
nativehawaiianorotherpacificislander	-1	3	45.5	22.3
other	-2.1	3.9	61.9	18.7
other_1	-6	3.1	80.8	26.8
othermiddleeastern	-7	5.4	-6.3	23.2
puertorican	-3.4	2.7	11.8	8
test	20.1	1.4	54	5.3
vietnamese	3.2	2.6	1153.8	7768.9
white	4.9	1.9	52.1	3
woman	-11.1	2.3	32.4	2.5
retake	1.7	1.7	1.9	1.8
americanindianoralaskannative:hispanicorlatino	-	-	-3.5	11.8

	MAIHDA		Fixed Effect	
	Estimate	Est. Error	Estimate	Est. Error
americanindianoralaskannative:man	-	-	0.2	7.6
hispanicorlatino:man	-	-	-22.4	9
americanindianoralaskannative:white	-	-	0.3	9.1
man:white	-	-	-48	4.1
man:anasianracenotlisted	-	-	-31.7	19.2
anasianracenotlisted:woman	-	-	-26.9	19.5
man:asian	-	-	-43.4	15.8
woman:asian	-	-	-40.8	16.2
man:asianindian	-	-	23	29.1
woman:asianindian	-	-	23.7	28.6
hispanicorlatino:black	-	-	-6.4	9.3
man:black	-	-	-25.2	11
white:black	-	-	-3.2	6.9
woman:black	-	-	-21.6	10.1
man:chinese	-	-	35.6	23
white:chinese	-	-	0.3	8
woman:chinese	-	-	37.7	22
man:colombian	-	-	1320.4	8813.8
white:colombian	-	-	1328.7	8818.5
man:cuban	-	-	866.7	7147.5
white:cuban	-	-	846.6	7147.3
man:filipino	-	-	-27.9	12.6
white:filipino	-	-	-3.6	11.3
woman:filipino	-	-	-26	11.8
white:genderqueergendernonconforming	-	-	-21	6.6
hispanicorlatino:mexicanmexicanamericanchi-canochicana	-	-	7.6	12
man:mexicanmexicanamericanchicanochicana	-	-	-36.9	19.3
hispanicorlatino:other	-	-	-89.6	6253.6
man:other	-	-	-67.5	19.3
hispanicorlatino:white	-	-	-24.4	15.6

	MAIHDA		Fixed Effect	
	Estimate	Est. Error	Estimate	Est. Error
hispanicorlatino:woman	-	-	-22.5	9.1
woman:other	-	-	-67.2	21.1
white:woman	-	-	-49.9	3.8
man:japanese	-	-	19.2	9.7
white:japanese	-	-	14.9	11.9
man:korean	-	-	-2198.5	8326
woman:korean	-	-	-2193	8326.1
white:mexicanmexicanamericanchicanochicana	-	-	-45.6	21.1
man:middleeasternornorthafrican	-	-	-13	20.6
white:middleeasternornorthafrican	-	-	1.2	13.7
man:nativehawaiianorotherpacificislander	-	-	-52	22.6
man:other_1	-	-	-88.2	26
white:other_1	-	-	-73.7	45.6
man:othermiddleeastern	-	-	-4	26.2
man:puertorican	-	-	-8.8	11.7
white:puertorican	-	-	-17	11.9
man:vietnamese	-	-	-1154.9	7769.3
woman:mexicanmexicanamericanchicanochicana	-	-	-32.1	19
woman:middleeasternornorthafrican	-	-	-7.1	17.2
woman:nativehawaiianorotherpacificislander	-	-	-44.5	22.7
woman:other_1	-	-	-87.7	24.9
woman:vietnamese	-	-	-1147.9	7769.1
test:americanindianoralaskannative	-	-	-7.2	9.6
test:hispanicorlatino	-	-	-7.6	13.4
test:man	-	-	-37.8	9.5
test:white	-	-	-37	7.6
test:anasianracenotlisted	-	-	-23.5	29.4
test:woman	-	-	-31	6.6
test:asian	-	-	-33.6	19.3
test:asianindian	-	-	54.8	37.2

	MAIHDA		Fixed Effect	
	Estimate	Est. Error	Estimate	Est. Error
test:black	-	-	-24.6	14
test:chinese	-	-	30.4	28.4
test:colombian	-	-	327.4	6663.1
test:cuban	-	-	-3259	6528.9
test:filipino	-	-	-23.8	16.3
test:genderqueergendernonconforming	-	-	-19.7	8
test:mexicanmexicanamericanchicanochicana	-	-	-46.8	23.8
test:other	-	-	-32.5	22.8
test:japanese	-	-	15.6	10.4
test:korean	-	-	-1010.1	4887.4
test:middleeasternornorthafrican	-	-	8.8	21.8
test:nativehawaiianorotherpacificislander	-	-	-19.9	25.7
test:other_1	-	-	-77.3	49.5
test:othermiddleeastern	-	-	24.2	26.7
test:puertorican	-	-	-2.7	9.7
test:vietnamese	-	-	1822.4	8485.6
americanindianoralaskannative:hispanicorlatino:man	-	-	-0.9	14.3
americanindianoralaskannative:man:white	-	-	1.3	11.2
hispanicorlatino:man:black	-	-	13.3	11.5
man:white:black	-	-	9.5	9.7
man:white:chinese	-	-	-1.7	10.3
man:white:colombian	-	-	-1328.1	8817.5
man:white:cuban	-	-	-878	7146.5
man:white:filipino	-	-	9.2	13.4
hispanicorlatino:man:mexicanmexicanamericanchicanochicana	-	-	-5.4	13.3
hispanicorlatino:man:other	-	-	90.5	6253.9
hispanicorlatino:man:white	-	-	22.2	16.1
hispanicorlatino:woman:other	-	-	91.8	6254.1
hispanicorlatino:white:woman	-	-	27.1	15.8
man:white:japanese	-	-	-22.1	13.5

	MAIHDA		Fixed Effect	
	Estimate	Est. Error	Estimate	Est. Error
man:white:mexicanmexicanamericanchicano-chicana	-	-	46	21.5
man:white:middleeasternornorthafrican	-	-	4.8	15.6
man:white:other_1	-	-	68.3	45.1
man:white:puertorican	-	-	4.4	15.6
white:woman:mexicanmexicanamericanchi-canochicana	-	-	44.3	21.1
white:woman:other_1	-	-	73.8	45.4
test:americanindianoralaskannative:hispanicor-latino	-	-	-0.5	15.4
test:americanindianoralaskannative:man	-	-	8.5	11.3
test:hispanicorlatino:man	-	-	9.2	14.3
test:americanindianoralaskannative:white	-	-	-0.1	13.2
test:man:white	-	-	41.2	10.2
test:man:anasianracenotlisted	-	-	23.6	29.6
test:anasianracenotlisted:woman	-	-	15	30.4
test:man:asian	-	-	35.6	19.2
test:woman:asian	-	-	33.9	19.4
test:man:asianindian	-	-	-56.4	38.2
test:woman:asianindian	-	-	-57.5	36.2
test:hispanicorlatino:black	-	-	-2.6	11.9
test:man:black	-	-	26.5	19.2
test:white:black	-	-	9	9.9
test:woman:black	-	-	24	15.6
test:man:chinese	-	-	-32.4	29.4
test:white:chinese	-	-	2.2	11.1
test:woman:chinese	-	-	-34	27.3
test:man:colombian	-	-	-319.6	6664.1
test:white:colombian	-	-	-324.3	6665.7
test:man:cuban	-	-	3237.7	6525.8
test:white:cuban	-	-	3241.6	6532.4
test:man:filipino	-	-	23.2	17.2

	MAIHDA		Fixed Effect	
	Estimate	Est. Error	Estimate	Est. Error
test:white:filipino	-	-	14.1	15.4
test:woman:filipino	-	-	23.2	16.6
test:white:genderqueergendernonconforming	-	-	20.5	9.6
test:hispanicorlatino:mexicanmexicanamerican-chicanochicana	-	-	18.3	14.3
test:man:mexicanmexicanamericanchicanoch-icana	-	-	51	24.3
test:hispanicorlatino:other	-	-	860	9238.8
test:man:other	-	-	36.9	23.4
test:hispanicorlatino:white	-	-	13.7	22
test:hispanicorlatino:woman	-	-	6.8	14.6
test:woman:other	-	-	36.4	27.1
test:white:woman	-	-	38.8	8.9
test:man:japanese	-	-	-34.3	13.5
test:white:japanese	-	-	-12.9	13.7
test:man:korean	-	-	1012.6	4886.2
test:woman:korean	-	-	1009.6	4887.5
test:white:mexicanmexicanamericanchicano-chicana	-	-	56.7	27.1
test:man:middleeasternornorthafrican	-	-	-3.2	22.7
test:white:middleeasternornorthafrican	-	-	-16.3	16.5
test:man:nativehawaiianorotherpacificislander	-	-	26.7	26.2
test:man:other_1	-	-	87	48.2
test:white:other_1	-	-	76	71.1
test:man:othermiddleeastern	-	-	-23.9	30.1
test:man:puertorican	-	-	-4.5	15.2
test:white:puertorican	-	-	-0.9	14.9
test:man:vietnamese	-	-	-1818.9	8485.4
test:woman:mexicanmexicanamericanchicano-chicana	-	-	42.9	24.2
test:woman:middleeasternornorthafrican	-	-	-4	21.3
test:woman:nativehawaiianorotherpacificis-lander	-	-	19.4	26.9
test:woman:other_1	-	-	85.8	43

	MAIHDA		Fixed Effect	
	Estimate	Est. Error	Estimate	Est. Error
test:woman:vietnamese	-	-	-1822.2	8485.9
test:americanindianoralaskannative:hispanicorlatino:man	-	-	-3.8	18.4
test:americanindianoralaskannative:man:white	-	-	5.8	16
test:hispanicorlatino:man:black	-	-	-4	16.4
test:man:white:black	-	-	-18	20.7
test:man:white:chinese	-	-	-1.9	13.8
test:man:white:colombian	-	-	327.4	6666.3
test:man:white:cuban	-	-	-3217.7	6529.3
test:man:white:filipino	-	-	-21	18.5
test:hispanicorlatino:man:mexicanmexicanamericanchicanochicana	-	-	-18.4	15.8
test:hispanicorlatino:man:other	-	-	-862.1	9238.4
test:hispanicorlatino:man:white	-	-	-13.4	22.5
test:hispanicorlatino:woman:other	-	-	-865.7	9238.5
test:hispanicorlatino:white:woman	-	-	-16.7	23.2
test:man:white:japanese	-	-	30.1	19.4
test:man:white:mexicanmexicanamericanchicanochicana	-	-	-58.9	27.4
test:man:white:middleeasternornorthafrican	-	-	12.1	19.5
test:man:white:other_1	-	-	-83.7	70.2
test:man:white:puertorican	-	-	9.4	20.7
test:white:woman:mexicanmexicanamericanchicanochicana	-	-	-50.6	27.1
test:white:woman:other_1	-	-	-87.2	64.9

Supplemental Table 2. Predicted group scores and standard errors for the pretest and posttest on both models.

		Pretest				Post-test			
		MAIHDA		Fixed effect		MAIHDA		Fixed effect	
group	N	Score	SE	Score	SE	Score	SE	Score	SE
man, White	4318	50.0	1.1	48.4	1.0	70.2	4.0	68.7	3.7
White, woman	1407	36.0	2.2	34.5	1.8	60.7	3.9	59.2	4.0
man, Asian	372	44.9	2.2	43.5	2.6	63.4	3.1	61.7	3.4
man, Black	288	37.8	2.6	36.3	2.3	56.5	5.7	54.5	6.0
Hispanic or Latino, man, White	282	42.8	1.7	41.2	1.8	65.0	3.9	63.5	3.8
woman, Black	239	29.7	1.8	28.1	1.8	52.3	2.8	50.5	3.3
woman, Asian	209	35.8	1.9	34.3	1.8	59.0	2.4	57.6	2.3
man, White, Mexican, Mexican American, Chicano, or Chicana	163	43.9	2.1	42.5	2.2	65.8	4.2	64.7	4.3
man, other	137	40.5	4.8	38.7	7.3	61.0	5.1	59.3	7.5
Hispanic or Latino, White, woman	119	34.0	2.6	32.1	2.6	54.7	3.7	53.0	4.0
Hispanic or Latino, man, other	103	36.4	4.0	34.6	5.2	56.6	6.5	54.8	7.6
Hispanic or Latino, man	86	39.0	2.9	39.3	3.0	59.8	3.4	57.2	4.0
man, Asian Indian	84	48.8	2.8	48.0	3.2	64.4	4.0	62.7	4.4
man, Chinese	82	51.0	3.2	49.4	3.8	66.2	2.8	63.6	3.4
man	81	46.7	3.0	88.5	5.9	62.4	3.5	60.5	3.8

		Pretest				Post-test			
		MAIHDA		Fixed effect		MAIHDA		Fixed effect	
White, woman, Mexican, Mexican American, Chicano, or Chicana	71	34.6	3.3	31.7	3.6	58.9	3.3	58.5	3.7
man, Mexican, Mexican American, Chicano, or Chicana	64	39.3	3.0	37.9	3.3	60.5	4.1	58.2	4.7
woman	52	34.9	2.7	64.8	4.9	55.3	2.9	55.4	3.1
man, Vietnamese	47	45.5	3.4	43.2	4.3	65.2	4.2	63.0	5.4
man, Middle Eastern or North African	40	43.8	4.7	40.5	6.0	64.1	3.8	62.3	5.2
Hispanic or Latino, woman	37	29.7	3.2	27.4	3.9	52.0	3.1	49.5	3.7
man, an Asian race not listed	36	47.5	3.1	46.1	3.5	64.3	4.2	62.4	5.6
woman, Asian Indian	34	37.7	3.0	36.9	3.4	59.6	3.3	57.2	4.1
Hispanic or Latino, woman, other	33	27.5	5.1	24.2	7.0	47.7	6.7	44.5	9.3
man, Native Hawaiian or other Pacific Islander	31	40.1	3.8	37.7	4.7	62.0	3.5	60.6	4.1
man, Filipino	30	43.8	3.4	40.7	4.0	60.4	4.4	56.3	5.6
woman, Chinese	30	41.1	3.0	39.7	4.2	59.3	3.4	59.0	4.8
man, White, Puertorican	29	42.1	3.2	38.8	4.1	63.3	4.7	60.4	6.2
Hispanic or Latino, man, Mexican, Mexican American, Chicano, or Chicana	27	36.0	3.1	35.2	4.3	56.9	4.0	57.0	5.9
White, genderqueergendernonconforming	27	51.1	3.6	49.3	3.6	69.2	4.2	67.1	5.1
woman, Mexican, Mexican American, Chicano, or Chicana	26	32.1	3.0	30.8	4.1	52.5	3.5	49.8	4.1
man, Korean	25	47.4	3.6	44.1	4.1	65.1	4.2	62.9	6.0

		Pretest				Post-test			
		MAIHDA		Fixed effect		MAIHDA		Fixed effect	
man, White, Filipino	21	48.5	3.7	50.4	5.3	65.9	5.5	63.3	7.7
woman, Vietnamese	21	38.4	3.4	38.4	4.7	60.3	3.7	61.6	5.4
man, White, Black	20	46.0	3.8	46.7	4.3	63.3	3.7	60.1	4.5
American Indian or Alaskan Native, man, White	18	48.0	3.5	51.4	4.1	73.0	4.2	78.8	5.8
woman, Korean	18	39.0	3.6	37.8	4.6	59.6	4.0	60.2	5.6
White	17	51.3	3.7	104.1	6.1	70.8	5.0	69.0	5.3
woman, Middle Eastern or North African	17	34.9	3.9	34.6	6.9	60.7	3.9	62.4	5.3
man, White, Middle Eastern or North African	14	47.8	4.1	50.6	5.2	70.9	4.6	72.3	6.7
American Indian or Alaskan Native, Hispanic or Latino, man	13	39.2	3.7	36.4	6.2	57.1	4.7	51.2	8.1
woman, an Asian race not listed	13	41.4	4.0	39.0	4.9	55.3	4.3	53.4	5.4
man, other Middle Eastern	13	36.1	5.5	33.9	7.3	54.6	5.5	50.5	7.2
Hispanic or Latino, man, Black	12	35.4	3.9	38.3	7.3	53.3	3.7	51.6	6.1
woman, Filipino	12	31.8	3.9	30.8	5.9	54.2	4.4	53.2	6.0
woman, Native Hawaiian or other Pacific Islander	12	34.7	4.3	33.3	5.6	55.0	4.2	55.8	5.8
American Indian or Alaskan Native, man	11	45.8	4.1	45.7	5.9	65.4	4.5	63.2	6.2
man, White, Chinese	11	52.7	4.1	52.1	6.2	72.3	5.4	70.8	8.5
man, White, Colombian	11	46.2	5.5	41.9	6.8	69.8	5.5	73.2	7.5
man, White, Cuban	11	45.1	4.8	41.5	6.4	66.0	5.3	64.5	8.4

		Pretest				Post-test			
		MAIHDA		Fixed effect		MAIHDA		Fixed effect	
man, White, Japanese	11	49.0	4.1	51.2	5.8	70.6	4.3	70.1	5.5
man, White, other_1	11	40.4	4.1	35.5	6.8	62.4	4.2	57.8	6.6
White, woman, other_1	10	32.9	4.2	27.8	7.0	53.3	5.0	49.7	9.3

Author Information

Ben Van Dusen, Associate Professor, Iowa State University School of Education; https://orcid.org/0000-0003-1264-0550, bvd@iastate.edu.

Jayson Nissen, Assistant Professor, Montana State University; https://orcid.org/0000-0003-3507-4993, jayson.nissen@montana.edu.

Heidi Cian, Maine Mathematics and Science Alliance; https://orcid.org/0000-0003-3510-2712, hcian@mmsa.org.

Lucy Arellano, University of California Santa Barbara; https://orcid.org/0000-0002-2410-0984, lucya@ucsb.edu.