ANNENBERG
BROWN UNIVERSITY

EdWorkingPaper No. 24-954

The State of State Civics Scores: An Application of Multilevel Regression with Post-Stratification using NAEP Test Scores

Steven Michael Carlo
University of Florida

The National Assessment of Educational Progress (NAEP) has tested the civic, or citizenship knowledge of students across the nation at irregular intervals since its very inception. Despite advancements in reading and mathematics, evidenced by results from the National Assessment of Educational Progress (NAEP), civics proficiency has remained consistently low, which raises concerns among educators and policymakers. This study attempts to provide those educators and policymakers with state-level predictions, not currently provided for the civics assessment. This research addresses this gap in state-level civics education data by applying multilevel regression with poststratification (MRP) to NAEP's nationally representative civics scores, yielding state-specific estimates that account for student demographics. A historical analysis of NAEP's development underscores its significance in national education and highlights the challenges of transitioning to state-level reporting, particularly for civics, which lacks state-level generalizability. Furthermore, this paper evaluates NAEP's frameworks, questioning their alignment with civics education's evolving needs, and investigates the presence of opportunity gaps in civics knowledge across gender and racial/ethnic lines. By comparing MRP estimates with published NAEP results, the study validates the method's credibility and emphasizes the potential of MRP in educational research. The findings reveal persistent racial/ethnic disparities in civic knowledge, with profound implications for civics instruction and policy. The research concludes by stressing the necessity for state-specific data to inform education policy and practice, advocating for teaching methods that enhance civic understanding and engagement, and suggesting future research directions to address the uncovered disparities.

VERSION: April 2024

Suggested citation: Carlo, Steven Michael. (2024). The State of State Civics Scores: An Application of Multilevel Regression with Post-Stratification using NAEP Test Scores. (EdWorkingPaper: 24-954). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/rn72-q717

The State of State Civics Scores: An Application of Multilevel Regression with Post-Stratification using NAEP Test Scores

Steven Michael Carlo

April 2024

Abstract

The National Assessment of Educational Progress (NAEP) has tested the civic, or citizenship knowledge of students across the nation at irregular intervals since its very inception. Despite advancements in reading and mathematics, evidenced by results from the National Assessment of Educational Progress (NAEP), civics proficiency has remained consistently low, which raises concerns among educators and policymakers. This study attempts to provide educators and policymakers with state-level predictions, not currently provided for the civics assessment. This research addresses this gap in state-level civics education data by applying multilevel regression with poststratification (MRP) to NAEP's nationally representative civics scores, yielding state-specific estimates that account for student demographics. A historical analysis of NAEP's development underscores its significance in national education and highlights the challenges of transitioning to state-level reporting, particularly for civics, which lacks state-level generalizability. Furthermore, this paper evaluates NAEP's frameworks, questioning their alignment with civics education's evolving needs, and investigates the presence of opportunity gaps in civics knowledge across gender and racial/ethnic lines. By comparing MRP estimates with published NAEP results, the study validates the method's credibility and emphasizes the potential of MRP in educational research. The findings reveal persistent racial/ethnic disparities in civic knowledge, with profound implications for civics instruction and policy. The research concludes by stressing the necessity for state-specific data to inform education policy and practice, advocating for teaching methods that enhance civic understanding and engagement, and suggesting future research directions to address the uncovered disparities.

Introduction

Over the last 20 years of National Assessment of Educational Program (NAEP) civics assessments, the percentage of the nation's eighth-grade students earning a rating of proficient or higher has been both low and stagnant, ranging from 22-24% (NAEP, 2022). Over the same period, NAEP reading proficiency increased from 32-36% and mathematics from 26-34%, both statistically significant amounts of growth (NAEP, 2022). This lack of growth in civics proficiency has not gone unnoticed by the education community (Bittman & Russell III, 2016; Diamond, 2015; Finn Jr., 2018; Litvinov, 2017; Shapiro & Brown, 2018; Stern et al., 2021) or media outlets (Brown, 2015; Dillon, 2011).

Yet, issues in the design of NAEP civics limit its usefulness for policymakers and practitioners. Unlike NAEP assessments in reading and mathematics which are generalizable at the state and national level, NAEP civics is only generalizable to a national sample of students. As a result, NAEP civics seems to have little bearing on state civics education. In a survey of state education agencies completed in 2021, 14 of the 29 respondents stated that the NEAP civics framework had no influence on their state standards, and another 10 stated it had little or uncertain influence (O'Malley & Norton, 2022). These statements expose a disconnect between one of the assessments' purposes (i.e., to inform policy and practice) and the tangible evidentiary needs of state policymakers to make these reforms.

I address a portion of current policy gaps by applying a multilevel regression with poststratification to the NAEP's nationally representative civics scores to obtain estimates of their respective state-level scores. I also provide these state-level estimates of student performance disaggregated by student demographics. This analysis will help provide a more nuanced and practical set of estimates for state education leaders and policymakers to use in

generating and evaluating their policy decisions. Lastly, I perform these analyses for the two most recently available NAEP civics testing cycles (2014 and 2018) to provide some longitudinal context for the state-level estimates. The state-level results from this study will also provide methodological proof-of-concept for other researchers. The NAEP assesses several other subjects at the national level of generalizability only (i.e., U.S. History, Economics, or the Arts), providing a trove of untapped research potential.

Literature Review

This review takes a detailed look at the journey of the NAEP from its inception to the present. This section provides an overview of NAEP's history in an effort to better explain its relevance to both national and state educational interests. This section also explains the initial and evolving testing frameworks used as responsibilities shift from corporation to corporation. The transition from reporting on national-level testing to providing state-level results has been a multi-decade endeavor, attempting to balance fears of federal overreach and validating its own existence through Congressional mandates. Next, I describe how the civics assessment of the NEAP was designed, the creation of its objectives and questions, as well as plans for future iterations. Lastly, I shine a light on previous research concerning gender, race/ethnicity, and standardized testing, both at large and specific to social studies.

NAEP as the Nation's Report Card

The NAEP is currently seen as the 'gold standard' of nationally representative assessments, affectionately known as the Nation's Report Card (U.S. Department of Education, 2010). In the early 1960s, the U.S. Commissioner of Education Francis Keppel realized that his annual reports on America's education system were filled with inputs, such as per-pupil expenditures, but lacked outputs, such as student learning (Jones & Olkin, 2004). Realizing that the common standardized tests of the time were geared toward ranking achievement and not on

measuring learning itself, he began the process of creating a new national assessment, focused on aggregated results rather than individual ones (Jones & Olkin, 2004, p. 27). NAEP has been assessing students under its federal contracts since April 1969, when it tested citizenship, science, and writing (Jones & Olkin, 2004, p. 11). Over its tenure as the Nation's Report Card, NAEP has administered tests in over 20 subjects on multi-year cycles to students all across the nation (Jones & Olkin, 2004).

The concept of national tests, such as NAEP, was not universally welcomed or trusted in the 1960s. Fears of federal overreach and the potential for the creation of a nationalized curriculum created early barriers to adoption. Many professional organizations, including the National Education Association and the Association of School Administrators, challenged the creation of this new national test and felt uneasy about its potential usage as a comparison tool (Jones & Olkin, 2004). The initial testing implementation contract was given to the Education Commission of the States (ECS) as an olive branch, of sorts, to state and local governments that feared that NAEP was a threat to their autonomy (Finder, 2004; Jones & Olkin, 2004; Stedman, 2009; Vinovskis, 2001).

In the 1970s, early challenges, to the validity of a national assessment were raised, leading to several rounds of improvements in sampling, question generation, analysis, and reporting (Jones & Olkin, 2004; U.S. Department of Education, 2010). In February 1983, the National Institute of Education shifted NAEP's design, analysis, and reporting operations from ECS to the Educational Testing Service (ETS), which "promised a wider range of statistical analysis", "increased roles for the policy-making committee", and to "make the data more useful for the states" (Jones & Olkin, 2004, p. 91; Layton, 1985, p. 275; Stedman, 2009).

In 1988, Congress formalized a NAEP schedule that included minimum testing for reading and math every other year, writing and science every four years, and once every six years for all other subjects, including civics, geography, and arts (H.R.5 Hawkins-Stafford, 1988). It also established the National Assessment Governing Board (NAGB) to set policy for NAEP including testing subjects, questions, and timetables (National Assessment Governing Board, n.d.). The law also opened up NAEP to administering state-level pilot testing, known as Trial State Assessments, from 1990-1994, as compared to previous iterations of NAEP testing that were only generalizable to a national sample (H.R.5 Hawkins-Stafford, 1988). By 1997, NAGB planned to include both national and state-level test results for reading and mathematics (Vinovskis, 2009). However, these state tests were only administered voluntarily. NCLB tied mandatory NAEP participation in reading and mathematics to the acceptance of Title I funds, ending the previous voluntary nature of state-level NAEP testing (H.R.1 No Child Left Behind, 2002). NAEP had not been seen as a state-to-state comparative tool for accountability measurement until this point. This new accountability push initiated by NCLB also forced NAEP results to be released within six months to make the results more current and useful to policymakers and education administrators (Jones & Olkin, 2004).

NAEP-C Frameworks

In one format or another civics-related knowledge and skills have been assessed by NAEP since its inception. Initially, it was referred to as ‘Citizenship’ in 1969/70, 1975/6, and 1981/2 (Jones & Olkin, 2004) but later renamed to the current ‘Civics’ since the passage of the Hawkins-Safford Act in 1988 (National Assessment Governing Board, 2018). The ‘Citizenship’ questions were also embedded as part of the broader ‘Social Studies’ exams in 1971/2, 1975/6, and 1981/2 rather than being a stand-alone subject test (Jones & Olkin, 2004). Aside from its name changes, the NAEP-C has also shifted from focusing on student ages to their grade levels,

from percent correct to scale scores, and has even changed the scales themselves (Stedman, 2009). In this study, NAEP tests containing civics material will be referred to as NAEP-C.

The 1960s NAEP-C objectives were centered on students' "skills, life habits, and attitudes" and would be assessed using standard multiple-choice questioning plus the addition of "constructed responses, self-reporting, and observations in group settings" (Jones & Olkin, 2004, p. 343). The American Institutes for Research (AIR) saw an opportunity to create a more comprehensive testing experience that went beyond classroom instruction. The objectives created for the test assumed that "the responsibility for teaching good citizenship is a joint responsibility of parents, community, schools, and the student" (Jones & Olkin, 2004, p. 344). The following is an excerpt from the NAEP Citizenship exam in 1969/70 showcasing the complexity of the mapping of domains and objectives:

> Goal: Seek community improvement through active democratic participation.
>
> Objective: Applying democratic procedures on a practical level when working in a group.
>
> Age 9: Group situations in which nine-year-olds might be involved including classroom activities, clubs, and recreation teams. In such situations, students try to help the group move toward its goals. They encourage the hearing of different viewpoints before voting on an issue. They abide by democratically determined decisions and follow established procedures for trying to change a decision (persuasion, petition, etc.). They mediate or seek compromise when others disagree. They are willing to give in when the situation calls for some immediate action and their own objections are unimportant. They explore and take turns with various leader and follower roles (Jones & Olkin, 2004, pp. 345-346).

While about 60% of the questions asked of test takers require a multiple-choice response, the question stems were created to require a deeper level of knowledge (NAGB, 2018). The other 40% of the questions were open-ended tasks with either short or extended responses (NAGB, 2018). The non-traditional nature of this test was seen as providing more useful guidance towards improving what is taught and how it is taught, although the costs to train and administer

were much higher than a traditional multiple-choice exam alone (Jones & Olkin, 2004). The following is an excerpt of question stems used for the NAEP Civics exam illustrating how, for students to be successful, they are required to know more than the basic recollection of facts and dates:

- "Identify the civic issue in the photo"
- "Interpret the implications of the following quote"
- "Demographic trend policy implications in this graph"
- "Describe how states raise money"
- "Use quotation to identify the role of associations"

The 1988 edition of NAEP-C for 8th and 12th graders was a series of six fifteen-minute blocks of multiple-choice questions and one fifteen-minute block for a single open-ended question (Weiss et al., 2001, p. 5). In 1994, Congress, in alignment with President Clinton's administration, presented a set of national educational aspirations under the name Goals 2000 which included reemphasizing civics (H.R.1804 - Goals 2000, 1994). These goals motivated NAGB to evaluate their current civics testing materials and objectives leading to a new testing framework for the 1998 edition of NAEP-C (O'Malley & Norton, 2022). For the 1998 version, "nearly half of the assessment would be short- or extended-response items rather than multiple choice" with more "stimulus materials such as political cartoons and documents" (Weiss et al., 2001, p. 4). The 2018 NAEP-C became the first edition to be administered on a digital platform (National Assessment Governing Board, 2018).

After adjusting for the pause in testing caused by the COVID-19 pandemic, NAGB published upcoming testing dates for NAEP-C to be administered nationally in 2022, 2026, and 2030, with a voluntary state-level option for 8th grade starting in 2030. (National Assessment Governing Board, 2021). This iteration will be the first time that the NAEP-C is going to be available at the state level, unlike their mathematics and reading counterparts that have been offering state-level results since the 1990s and requiring them since 2001. With these new dates

set, NAGB will be updating the 2030 NAEP-C framework to replace the curricular standards that had been developed around 1996 (O'Malley & Norton, 2022). The 30-year span between frameworks is another sign that the civics curriculum had laid dormant while other subject materials took center stage. The 2030 retooling of the framework capitalizes on its current reinvigoration on the national stage.

As mentioned in the introduction, unlike the NAEP mathematics and English assessments, the NAEP-C exam was only designed to generalize to a nationally representative sample of students at set age intervals. With numerous education policies being created at the state level, it would only seem natural that policymakers would need data at the state level as well. In NAEP development meetings as early as the mid-1960s, one of the assessment's original subordinate objectives was "to provide comparative data to stimulate competition among the states and local communities" (Jones & Olkin, 2004, p.348). Based on the current national-only model of NAEP-C, a policymaker in Nebraska, for example, cannot determine if their state has made progress from one assessment cycle to the next. This framework means they cannot use NAEP-C to evaluate their state's curricular decisions, partially neglecting the initial mission. They can only know that the national average has been relatively stagnant over time. Without state-level assessment results, it's unknown if a particular state is buoying the nation, dragging it down, or maintaining the status quo.

As a specific example, when comparing 8th-grade NAEP reading and mathematics scores from 1992-2013, "the top-gaining 10 states was 1.6 points per year—double the 0.8-point annual adjusted gain in the bottom-gaining 10 states" (Carnoy et al., 2015, para. 16). At the beginning of this period, Louisiana's students had low adjusted scores while Vermont's students started with a relatively high adjusted score, yet both states were top-gaining states (Carnoy et al., 2015).

Having the ability to compare states that have similar characteristics can provide state policymakers with a litmus test of the effectiveness of their current policies, standards, and instructional strategies. They can also look to these states for examples to follow or cautionary tales to learn from.

Civics Opportunity and Achievement Gaps and Standardized Tests

NAEP reading and mathematics test scores have shown that achievement gaps between White students and students of color, particularly for Black and Hispanic students, are not closing, but rather widening since NCLB and its focus on standardized testing (Fouquet, 2019; Lee, 2006; Lee & Reeves, 2012; Ravitch, 2013). Studies have also shown disparities between male and female students on standardized tests, most prevalent in mathematics (Arias et al., 2023; Fryer Jr. & Levitt, 2010; Sohn, 2012). At the same time, gaps between the rich and poor are found as well (Ladson-Billings, 2006; Ravitch, 2013) and although poverty and race/ethnicity appear strongly linked, poverty alone does not appear to fully explain the racial achievement gaps found (Myers et al., 2004; White et al., 2016). While NAEP reading and mathematics assessment data are used more often, social studies assessments, and civics specifically, have their own concerns regarding racial disparities in achievement.

Studies across the last fifteen years have uncovered evidence of what they term either the 'civics opportunity gap' (Herczog, 2012; Kahne & Middaugh, 2008; Levine, 2009) or the 'civics empowerment gap' (Addington, 2016; Levinson, 2010, 2012; Swalwell, 2015) addressing a pattern of students of color experiencing lower-quality quality civics instruction leading to lower learning outcomes, such as civic knowledge. Student survey responses from the 1998 and 2010 iterations of the NAEP-C have shown disparities in civic knowledge associated with disparities in civics course exposure in locations with higher proportions of minoritized students (Heafner & Fitchett, 2018; Littenberg-Tobias & Cohen, 2016; Niemi & Junn, 1998).

Broadly speaking, a lack of exposure to civics instruction that focuses on established best practices, such as debates on controversial issues, mock trials, discussions of current events, or other simulations of civic engagement, will likely lead to lower civic knowledge overall (Ehrlich, 1999; Feldman et al., 2007; Hepburn, 1997; Stern et al., 2021). While the NAEP-C does not directly connect these practices to specific questions, the exam includes a survey portion that does ask about their learning experiences and the instructional styles used in their civics classrooms. Research shows that White students, compared to their Hispanic or African American counterparts, are more likely to receive instruction using these high-impact strategies (Kahne & Middaugh, 2008; Kawashima-Ginsberg, 2013) and that they score substantially higher on standardized tests in civic skills and knowledge (Fitchett & Heafner, 2017; Furgione et al., 2018; Levinson, 2010). Due to the patterns of disparities found in past iterations of NAEP testing overall, it is important to this study that any continued disparities amongst race/ethnicity are identified, and their potential implications discussed.

There is a dearth of research literature that explores gender and social studies assessments. In a review of the results from the 2010-2012 NAEP economics, civics, geography, and U.S. history assessments, statistically significant differences by gender were only found in economics, which aligns with the extent literature about males outperforming their female counterparts in mathematics content (Bohan, 2017). When looking at AP testing in 2014, there is a prevalence of more female students taking the three exams centered on history while more male students taking the two economics-centered exams (Bohan, 2017). Even though there are preferences in tests chosen by gender, The College Board, which administers the AP exams, expressed that gender-based differences in scores for social studies were significantly correlated

with item content and appear to have stopped reporting results by gender (Buck et al., 2002; College Board, 2014).

Data

The NAEP-C test is not taken annually but, since 2006, every four years. The data used for this study is from the following years: 2014 and 2018. The first set of data was pulled from NAEP's restricted-access data available by request through the National Center of Educational Statistics. It includes each test taker's scale score (in the form of plausible values), state, gender, and race/ethnicity for each of the respective years. Because the test has changed over time, this study focuses on eighth-grade student results, available for each of the years included in the study period. The second set of data was obtained from the Integrated Public Use Microdata Series (IPUMS). For this source, I retrieved the number of eighth-grade students in each state by their gender, race/ethnicity, and public/non-public status. The third set of data is the region of each state, as categorized by the U.S. Census Bureau. Aside from the student-level characteristics, this study also uses select state-level characteristics that may influence a student's overall civics knowledge, creating a third set of data. These state-level characteristics include the governor's political party, political leanings of the state legislature, local area unemployment, median family income, percent of childhood poverty, percentage of students with a disability, percentage of English language learning students, and average per pupil expenditures. Refer to Table 1 for the descriptions of each variable used in this study.

NAEP Sampling and Weights

Sampling for NAEP-C since 1998 has been conducted by Westat, Inc. (Lutkus et al., 1999). The NAEP-C includes 8th-grade students from both public and private schools, students with disabilities (SD), and English language learners (ELL), and identifies students by six

racial/ethnic categories (National Assessment Governing Board, 2018). For more information on the demographic breakdown of the NEAP-C framework across these years, see Table 2.

The process for selecting the sampled students involves three phases (Jones & Olkin, 2004). Phase one begins by breaking the nation into different strata based on the student population of geographic regions, known as primary sampling units (PSUs). Depending on their size, a PSU might only include one large county or several contiguous ones. These PSUs are classified into strata and selected with probability proportional to their student population size. In phase two, public and private schools are chosen from within the PSUs selected in the first phase. If any school refuses to comply with testing, they are replaced by another school within the same PSU. Lastly, in phase three, students are randomly selected from within the schools chosen in phase two. Testing accommodations are provided to SD and ELL students to maximize inclusion. Given the complex sampling design, survey weights are typically used in studies using the NAEP-C to ensure that results are representative of a national sample of eighth graders. As I am aiming to use MRP to generalize to the state level, however, survey weights will not be used in the current study.

NAEP-C Scales

NAEP-C does not give every student the full set of civics-related questions. Therefore, a simple percentage correct model is inadequate and leads to biased estimates of the results, especially when looking deeper into student subgroups (National Assessment Governing Board, 2018). NAEP-C creates sets of plausible values for each student rather than providing a single percentage correct. These plausible values represent potential scale scores on a distribution of scale scores for students that match their individual characteristics and answer choices. This process allows NAEP-C to give each student a smaller sample of all the questions and still predict their scale score as if they had taken the full exam. These plausible values are then

inserted into a generalized partial-credit model that allows scaling when there are multi-point (multiple plausible values) ratings. Conventional analyses of the NAEP-C results require the use of all plausible values. In this study, I simply use the first plausible value for each student for the main analysis and the second plausible value for a sensitivity analysis.

NAEP-C Achievement Levels

NAEP-C uses the scale scores to create four achievement levels: below basic (0-133), basic (134-177), proficient (178-212), and advanced (212-300). These levels are cumulative in the sense that earning a 'proficient' level means that you are also expected to have mastered the 'below' level of skills as well as the 'proficient' level. For this study, I focused on students who earned an achievement level of proficiency or higher, a scale score of 178 and above.

Descriptive Details

The NAEP-C sampling process explained above allows for a nationally representative sample to be created each testing cycle but that process also allowed for the inclusion of some states in 2014 that were not in 2018 and vice versa, or to be left out altogether from both tests. For example, there were 34 states selected in 2014 and 42 states selected in 2018. Two states (KS and MT) were selected in 2014 but not in 2018 and nine states (AR, IA, NH, ND, MA, RI, TN, WV, and WY) were selected in 2018 but not in 2014. Seven states (AK, DE, ID, KY, ME, SD, and VT) were not selected for either testing cycle. This means that the MRP predictions created for them relied on their demographic breakdown and state characteristics. The demographic homogeneity of Alaska, for instance, might be the main contributor to both its high estimated means but also the very large credible intervals.

Methods

Who is Mr. P?

Scholars are often faced with nationally representative surveys, such as public opinion polling before an election, that are not representative at the state, county, or municipal level of interest. Multilevel regression with Poststratification (MRP) has become the standard method for using national surveys to generate estimates of subnational, or small area, results (Gelman & Little, 1997; Lobo, 2022; Lopez-Martin et al., 2019; Park et al., 2004). As the name suggests, MRP requires two main analytical procedures: multilevel regression modeling (MLM) followed by poststratification (Gelman & Little, 1997; Park et al., 2004). According to a recent systematic review (Lobo, 2022), MRP studies are most commonly used for social science research, with only 4% of studies in education (e.g., Berkman & Plutzner, 2011; Houston, 2019; Ortagus et al., 2021; Skinner & Doyle, 2024). The vast majority of the studies were in political science (e.g., Ghitza & Gelman, 2013; Ghitza & Gelman, 2020; Hanretty, 2020; Hanretty et al., 2018; Kiewiet de Jonge et al., 2018; Lopez-Martin et al., 2022; Lax & Phillips, 2009a; Lax & Phillips, 2009b).

The first step in MRP is to estimate a two-level multi-level model. The first level uses individual-level predictors, such as their response to a survey, and demographic characteristics, such as gender, age, and race/ethnicity. The second level of the MLM incorporates area-level variables representing the subnational, or smaller area of interest. In the case of this study, the area-level variables represent the states. These variables may include anything that research literature in the field expects to aid in estimating results such as the U.S. region, political partisanship, or median family income. Lastly, a poststratification frame is created using a census, which in my case, is provided by the IPUMS. This frame requires the "joint-distribution proportions for all individual-level categories in each small area", referred to as population cells (Lobo, 2022, p. 14). This joint distribution is the matching of individual-level data to small-area

(state-level) data. An example of this can be matching the number of Alabama's White, public-school girls that obtained a proficient level on the test with the actual number of White girls attending public schools in Alabama according to the census.

Current Study

As mentioned in the Data section, the individual-level data collected included the test taker's state, public/non-public status, gender, race/ethnicity, and scale scores (plausible values) for each year. The state-level data used was the state's location based on the four U.S. regions. The IPUMS provided the public/non-public status, gender, and racial/ethnic breakdown of eighth graders in each state for each year. The size of the poststratification frame is determined by multiplying the individual-level category levels together. This study incorporates 51 states (D.C. included), public and non-public status, two genders, and six races/ethnicities. This frame creates 1224 unique population cells (i.e., 51 x 2 x 2 x 6). Equation 2-1 is expressed as:

$$y_i = \gamma^0 + \alpha_{s[i]}^{state} + \alpha_{r[i]}^{public} + \alpha_{r[i]}^{race} + \alpha_{r[i]}^{male} + \alpha_{g[i],r[i]}^{male.race} + \boldsymbol{\sigma_{s[i]}} \quad (2\text{-}1)$$

Where y_i for student i is a function the grand mean γ^0, random intercepts meant to indicate the state $\alpha_{s[i]}^{state}$, public/non-public status $\alpha_{r[i]}^{public}$, race/ethnicity $\alpha_{r[i]}^{race}$, gender $\alpha_{r[i]}^{male}$, and the interaction of gender and race/ethnicity $\alpha_{g[i],r[i]}^{male.race}$. There are also a series of state-level covariates represented by $\boldsymbol{\sigma_{s[i]}}$.

For Equation 2-2, the predicted probabilities, π_j, are aggregated to the state, θ_S, using:

$$\theta_S = \frac{\sum_{j \in S} N_j \pi_j}{\sum_{j \in S} N_j} \quad (2\text{-}2)$$

In this stage, I reweighted each demographic cell's predicted probability, π_j, based on their respective population counts, N_j, from the IPUMS data set. Equation 2-1 above used the subscripts [i] for individual students, s for state, g for gender, and r for race/ethnicity to better

clarify the creation of each population cell in the model. For Equation 2-2, these subscripts were simplified to use j to represent a population cell.

I choose to use a Bayesian approach to determine the predictors in this MRP study. I employed the *stan_glmer* command from the *rstanarm* package in R (Gabry et al., 2024)). The *stan_glmer* command acts like a traditional generalized linear model except "rather than performing (restricted) maximum likelihood estimation, Bayesian estimation is performed via Markov Chain Monte Carlo" (Gabry et al., 2024, p. 120). Using a Bayesian approach allows for the incorporation of prior information (i.e., demographic characteristics), adding to the robustness and accuracy of the estimates (Gelman & Hill, 2006). Using *stan_glmer*, specifically, allows R to exercise the flexibility of *stan,* a probalistic programming language, to create the posterior distributions for all model parameters (Carpenter et al., 2017). Lastly, using a Bayesian approach to create full posterior distributions for estimates provides assess a range of likely outcomes and their probabilities rather than relying solely on point estimates or standard errors (Lax & Phillips, 2009).

Results

Study Validation

In an effort to validate the accuracy, or fit, of the MRP model used for this study, the national-level predictions from the model were compared with NAEP-C's published results from both 2014 and 2018. I compared the national-level model results overall, by public/non-public status, by gender, and lastly, by race/ethnicity. Before delving into these comparisons, note that NAEP-C estimates are frequentist while the MRP estimates are Bayesian. This means that the two values are not strictly commensurable but are still the most likely comparison available in terms of common estimation practices. Since these two estimates come from differing statistical frameworks and approaches to weighting, I discuss the MRP estimates' using their means and

credible intervals and discuss the NAEP-C's estimates by their means and confidence intervals. The means are expressed in percentages and the intervals in brackets. I also want to reiterate that all results are for the percentage of students estimated to be at the proficiency level or above (scale scores of 170+) on the NAEP-C in their given year.

Starting with the overall national results from 2014, shown in Figure 1, the NAEP-C estimated that 23.37% [21.05, 25.86] of students were proficient and above, while the MRP estimated 24.6% [22.8, 26.6]. The 2018 NAEP-C estimated the overall national proficiency rate was 23.7% [22.05, 25.43] and the MRP estimate was 24.2% [23.2, 25.3]. There is only 1.23 percentage point difference between the two estimates in 2014 and a 0.5 percent point difference in 2018. Aside from the closeness of the percentage point differences for the overall national estimates, both mean proficiency rates fall within the intervals of their corresponding pair, indicating the accuracy of the MRP model being used.

The next set of results refer to the proficiency rates of students based on their public-school status as seen in Figure 2. According to the 2014 NAEP-C results, only 21.99% [19.56, 24.63] of students attending public school were proficient while 38.11% [32.9, 43.61] of non-public students were proficient. The MRP model estimated that 22.8% [21.1, 24.9] of public-school students and 40% [36.2, 43.6] of non-public students were proficient. For the 2014 public school estimates, there is a 0.81 percentage point difference and a 1.89 percentage point difference for non-public students.

The 2018 proficiency rate estimates for public school students on the NAEP-C were 22.73% [21.17, 24.36] and 22.4% [21.4, 23.5] for the MRP model. The non-public student proficiency estimate was 41% (no confidence intervals reported) for the NAEP-C and 39.1% [36.2, 42] for the MRP model. For the 2018 public school estimates, there is a 0.33 percentage

point difference and a 0.9 percentage point difference for non-public students. Once again, all of the mean proficiency rates fell within the intervals of their corresponding pairs.

The next set of results concerns the national level results based on student's gender and can be seen in Figure 3. The proficiency rate estimates for males in 2014 were 24.47% [21.75, 27.41] for the NAEP-C and 25.6% [23.6, 27.8], resulting in a difference of 1.13 percentage points. The proficiency rates for female students in 2014 were 22.22% [19.79, 24.86] for the NAEP-C and 23.6% [21.7, 25.8] for the MRP model, leading to a difference of 1.38 percentage points. The corresponding estimates for male students in 2018 were 23.58% [21.76, 25.51] for the NAEP-C and 24.3% [23.2, 25.4], resulting in a difference of 0.72 percentage points. Lastly, the proficiency rates for female students in 2018 were 23.82% [21.92, 25.82] for NAEP-C and 24.1% [23, 25.2] for the MRP model, leaving a 0.28 percentage point difference.

There are six racial/ethnic categories used in this study: White, Black, Hispanic, Asian/Pacific Islander (API), American Indian/Alaskan Native (Native), and Multiracial. To be more succinct, I only explore four specific instances of inconsistencies between the NAEP-C and MRP model estimates as seen in Figure 4. The first case is the American Indian/Alaskan Native category for 2014 in which NAEP-C does not provide an estimate at all due to the sparsity in the sample. If one uses the 2010 and 2018 versions of the NAEP-C as a range, then a hypothetical value would be between 11-14%. The MRP estimate for 2014 was 12.9% [6.8, 21.2]. While this result would match, the sparsity of this racial category creates very large credible intervals and makes estimates less dependable.

In the other three irregular cases, the NAEP-C provides estimated proficiency rates, but those rates fall outside of the corresponding MRP credible intervals. The estimated proficiency rate for Black students 2014 on the NAEP-C was 8.93% [6.88, 11.52], however, the MRP

estimate was 11.6% [9.7, 13.8]. This means that the MRP model produces estimates in this racial category that are higher than that of the official NAEP estimates. The estimated proficiency rate for Asian/Pacific Islander students in 2018 was 41.19% [35.71, 46.91], while the MRP estimate was 35.7% [32.5, 39.0]. Also in 2018, the proficiency rates of Multiracial students were 27.67% [22.38, 33.66], while the MRP model estimate was 24.1% [21.0, 27.4]. For both these racial categories in 2018, the MRP model produces estimates that are lower than NAEP.

There were 11 estimates compared each year to help validate the MRP model. The NAEP-C estimates were within the MRP credible intervals 91% (10 of 11) of the time in 2014, 82% (9 of 11) in 2018, or 86% (19 of 22) overall. The MRP estimates, on the other hand, were within the NAEP-C confidence intervals 100% of the time in both years (22 of 22). In short, the overall comparability of the proficiency rates across methods helps support the credible estimation of state-level predictions of proficiency rates.

State-Level Results

For the state-level analysis, it was important to look at the MRP estimates for each state individually, but also to be able to compare the states to one another. I first look at the estimated results for 2014, as seen in Figures 5 and 6. The top five ranked states by mean proficiency rates were Alaska, Wyoming, Ohio, Kansas, and Pennsylvania. The bottom five states were Wisconsin, Hawaii, Alabama, Nebraska, and Mississippi. While roughly half of the state means were above the national average, most of their credible intervals included the national average for 2014. Seven states (i.e., FL, IL, NJ, OH, PA, VA, and WA) had credible intervals that fell completely above the national average of 23.37% for 2014. On the other end of the spectrum, there were four states (i.e., AL, LA, MS, and NY) whose credible intervals fell completely below the national average of 23.37% for 2014.

The top five states in 2018, by mean proficiency rates, were Alaska, Virginia, West Virginia, Maryland, and New Jersey. The bottom five states were Louisiana, Mississippi, Hawaii, Vermont, and New Mexico. The 2018 results show six states (i.e., AK, FL, IL, NJ, VA, and WA) with credible intervals landing completely above the national average of 23.70%. There were also nine states (i.e., AZ, CA, HI, LA, MS, NV, NM, OR, PA) with credible intervals falling completely below the national average of 23.70%.

Alaska was estimated to have the highest mean proficiency rates in both testing cycles; however, it is worth noting that no Alaskan students were sampled to take either test. There are five states (i.e., FL, IL, NJ, VA, and WA) that maintained above-average credible intervals. Alabama and New York were both able to make strides from 2014 to 2018 but still maintained means below the national average. Hawaii and Mississippi were the only two states to have a mean proficiency rate in the bottom five for both test cycles.

I also ran a Wilcoxon Sign-Rank Test to determine if the differences in mean proficiency rates across the two testing cycles are statistically different (Rosner et al., 2006; Woolson, 2008). To complete this test, I prepared the data by pairing each state's two mean rates, ranking them by the absolute value of their differences, and then re-applying their respective signs (see Table 3). The Wilcoxon results ($V = 666$, p-value = 0.9813) suggest that there is no statistical difference in the ranking of states between the two testing cycles. For clarity of understanding, Figure 7 provides a visual representation of the signed ranks of the states. As shown in Figure 7, on the far left is New Mexico, representing the largest decrease in mean proficiency rate (-15.30%) between 2014 and 2018. On the far right is West Virginia, representing the largest increase in mean proficiency rate (17.67%) between 2014 and 2018.

State-Level Subgroups

I broke down the state-level predictions into several subgroupings: public school status, gender, and race/ethnicity. As expected, based on the national estimates, non-public school mean proficiency rates were higher than their public school counterparts across all states, as seen in Figure 8. At first glance, Hawaii's non-public schools were projected to have lower mean proficiency rates than the national average in both testing cycles. However, the credible intervals for Hawaii's non-public schools included the national average.

As mentioned above, the literature on civics testing has not shown meaningful gender differences. In 2014 and 2018 respectively, the NAEP-C percentage point difference between male and female students was only 2.00 and 0.20. The MRP projections seen in Figure 9 support this assumption, as I see no significant gender differences in any state for either testing cycle.

The last subgroupings were for students' race/ethnicity across all states for both testing cycles. Since civics assessment literature and national-level NAEP-C results indicate racial/ethnic gaps between White and non-White students, I report state-level comparisons between White students and the other racial/ethnic subgroups independently. Results from both the 2014 and 2018 testing cycles show that API students had a higher mean proficiency rate than their White student counterparts in every state except the District of Columbia. The mean proficiency rate of API students was within the credible interval for White students in both cases, as seen in Figure 10, meaning that this difference was not statistically significant.

The White-Black mean proficiency rate gaps for 2014 can be seen in Figure 11 Panel A. For 19 of the 51 states, the credible intervals of the two races overlapped meaning that the differences were not statistically significant. For the majority of states (32), however, the differences were significant. Also of interest, in 42 of the states, the Black student proficiency rates and their corresponding credible intervals fell completely below the national average,

showing that these rates were, in fact, significantly lower. The 2108 White-Black gaps (Figure 11 Panel B) show larger differences overall. In 2018, all but five states (i.e., Alaska, Maryland, Montana, Nebraska, and Vermont), had mean proficiency rates that were statistically different among White and Black students. In 48 of the states, the Black student mean proficiency rates and their corresponding credible intervals were completely below the national average.

The White-Hispanic gap is the next subgroup to be addressed. As seen in Figure 12, both test cycles showed that Hispanic students had lower mean proficiency rates than their White counterparts in all states. In 2014, the gaps were significantly different from each other in 31 states. For 42 states, the Hispanic students' mean proficiency rates and corresponding credible intervals fell entirely below the national average. The gaps and differences in 2018 were much larger with 38 states showing statistically different proficiency rates between White and Hispanic students. In 47 states, Hispanic students' mean proficiency rates and their corresponding intervals were fully below the national average.

Significant racial/ethnic differences were not found for Multiracial students in 2014 (see Figure 13 Panel A). Although in all but two states, North Dakota and Vermont, Multiracial students' mean proficiency rates were lower than their White student counterparts, they all overlapped with their corresponding credible intervals. There were only three states (i.e., Mississippi, Missouri, and New York) with credible intervals wholly under the national average, however, there were eight states in which the credible intervals were completely above the national average. In 2018, the mean proficiency rates for Multiracial students were lower than their White counterparts in every state, as seen in Figure 13 Panel B. The 2018 results do show significant differences in the White-Multiracial mean proficiency rates in five states.

Coincidentally, there were eight states with credible intervals entirely below the national average and eight others entirely above the national average.

The final racial/ethnic subgroup gap is that of White-Native students (see Figure 14). In both the 2014 and 2018 testing cycles Native students' average lower mean proficiency rates than their White counterparts. In 2014, 25 of the 51 states had overlapping credible intervals between White and Native students. Essentially showcasing half of the states with significant differences between these two subgroups. There were also 22 states with credible intervals for Native students fully below the national average and none fully above. Alaska was the only state with a mean proficiency rate for Native students above the national average, although their corresponding credible intervals included that average. In 2018, only 17 states had overlapping credible intervals between the White and Native student populations. The other 34 states had White credible intervals fully above their Native counterparts. The 2018 predictions also show that the Native students' mean proficiency rates and corresponding credible intervals fell entirely below the national average without any states having fallen completely over the national average. Only three states (i.e., Alaska, District of Columbia, and Maryland), had mean proficiency rates above the average, but in all cases, their credible intervals contained that average.

Basic Achievement Level Validation

In order to further validate the models used in this study, I re-ran the analysis for both testing cycles based on the percentage of students with a 'basic' achievement level and above (basic+), a scale score of 134 and higher. This allowed me to test if my model fit was consistent across multiple achievement levels. The overall national estimated basic+ rates, provided in Figure A1, for 2014 was 74.46% [71.71, 77.04] for NAEP-C and 75.9% [74.2, 77.4] for the MRP model, bringing on a 1.44 percentage point difference. The 2018 corresponding estimates were 72.55% [70.74, 74.29] for NAEP-C and 73.7% [72.8, 74.6] for the MRP model, with a

difference of 1.15 percentage points. The 2018 results show that the MRP model estimates for the overall national basic+ rates are too high (see Figure A-1).

The estimated basic+ rates based on public-school status are shown in Figure A-2 for both testing cycles. For 2014, the estimated basic+ rates for public school students were 73.11% [70.16, 75.87] for NAEP-C and 74.2% [72.3, 75.7] for the MRP model, having a 1.09 percentage point difference. For 2014, the estimated basic+ rates for non-public school students were 88.96% [83.31, 92.86] for NAEP-C and 91.4% [89.5, 92.9] for the MRP model, having a 2.44 percentage point difference. The MRP model estimates for non-public students in 2014 were too high. The estimated basic+ rates for public school students in 2018 were 71.34% [69.47, 73.14] for NAEP-C and 71.8% [70.8, 72.7] for the MRP model, having a 0.46 percentage point difference. The 2018 estimated basic+ rate for non-public students was 91% [no confidence interval reported] for NAEP-C and 90.3% [88.7, 91.7] for the MRP model, resulting in a difference of 0.7 percentage points (see Figure A-2).

The national basic+ rates for male students in 2014 and 2018 shown in Figure A-3, make clear that the MRP model estimates were higher than NAEP's estimates. In 2014, the NAEP-C estimate was 73.97% [70.96, 76.76] and for the MRP model, 75.9% [74.1, 77.4] for a 1.93 percentage point difference. In 2018, the NAEP-C estimate was 70.51% [68.24, 72.68] and for the MRP model, 71.9% [70.8, 73.1] for a 1.39 percentage point difference. The national basic+ rates for female students in 2014 had a difference of 0.92 percentage points while the 2018 cycle had a difference of 1.0 percentage points. All female basic+ rates also fell within their corresponding pairs' intervals (see Figure A-3).

The last set of national basic+ rates relate to the racial/ethnic breakdown of students, as available in Figure A-4. For the sake of brevity, I again only explore the four inconsistences

between NAEP-C and MRP model estimates. The estimated basic+ rates for Black students in 2014 were 54.65% [46.06, 62.96] for NAEP-C and 59% [55.9, 62.2] for the MRP model, providing a difference of 4.35 percentage points. The estimated basic+ rates for Black students in 2018 were 52.08% [48.46, 55.68] for NAEP-C and 54.9% [52.4, 57.3] for the MRP model, providing a difference of 2.82 percentage points. There is a relatively large difference between the two basic+ rates for Black students in both years. The MRP model appears to be producing estimated basic+ rates for Black students that are consistently higher than the NAEP estimates.

In 2014, the estimated basic+ rate for Multiracial students were 80.35% [72.4, 86.45] for NAEP-C and 75.8% [71.4, 79.6] for MRP, resulting in a 4.55 percentage point difference in the estimates. For Multiracial students in 2014, the basic+ rate that the MRP model produces is lower and created a relatively large difference between the estimates. The final example for basic+ rates was in 2018 concerning the Asian/Pacific Islander students. The NAEP-C estimate was 87.3% [83.47, 90.35] while the MRP model estimates were 84.4% [81.8, 86.6], leaving a difference of 2.9 percentage points. It is another example of the MRP model producing estimates that are lower than the NAEP estimates.

Overall, of the 11 paired estimates for basic+ each year, only 64% (14/22) of the NAEP-C estimated means fall within their corresponding MRP credible intervals. On the other hand, all 22 of the MRP mean estimates fell within their corresponding NAEP-C confidence intervals. That being said, there were only two cases out of twenty-two in which the differences between the estimates reached four percentage points or higher. In the majority of cases, the difference between the estimates was below two percentage points.

Sensitivity Analysis

Lastly, I re-ran the analysis for both testing cycles, based on proficiency level again, but using the second plausible value instead of the first. This test allowed me to determine if my

model fit holds across multiple plausible values, not just multiple achievement levels of the same plausible value. For clarity, I refer to the estimated proficiency rates using the second plausible value as 'proficiency2' rates. Interestingly in all 11 of the paired estimates for 2014, the NAEP-C mean estimates fell within their corresponding MRP credible intervals as well as all MRP mean estimates fell within their corresponding NAEP-C confidence intervals. The only inconsistencies (6) between the estimates came from the 2018 testing cycle. As such I only explore those pairs of estimates.

The national overall estimated proficiency2 rate was 23.7% [22.05, 25.43] for NAEP-C and 24.9% [24.1, 26.1] for the MRP model, resulting in a 1.2 percentage point difference (see Figure A-5). In this case, the NAEP-C mean proficiency2 rate falls below the MRP credible interval. The public and non-public school student results, available in Figure A-6, showed no inconsistencies. The estimated proficiency2 rates for male students were 23.58% [21.76, 25.52] for NAEP-C and 24.9% [24.1, 26] for the MRP model, showing a difference of 1.32 percentage points. For female students, the estimates were 23.82% [21.92, 25.82] for NAEP-C and 24.9% [24.1, 26.1] for the MRP model, resulting in a 1.08 percentage point difference. The MRP model produced estimates that were slightly higher for both genders in 2018 (see Figure A-7).

There were also three inconsistent cases in 2018 with regard to race/ethnicity as seen in Figure A-8. In the first two cases, White and Black students, the MRP model estimates were slightly higher, leaving the NAEP-C mean estimate outside the respective credible intervals. For White students, the NAEP-C estimate was 31.5% [28.93, 34.18] and the MRP estimate was 32.9% [31.7, 34.7] with a difference of 1.4 percentage points. For Black students, the NAEP-C estimate was 9.54% [7.78, 11.65] and the MRP estimate was 11.6% [11.1, 12.1] with a difference of 2.06 percentage points. The final case is for 2018's Multiracial students in which

the MRP model produced estimates that were lower than the NAEP estimates. For Multiracial students the NAEP-C estimate was 27.67% [22.38, 33.66] and the MRP estimate was 24.3% [23.2, 25.8] with a difference of 3.37 percentage points.

Overall, the NAEP-C mean estimated proficiency2 rate fell within the corresponding MRP credible intervals 73% (16 of 22) of the time. The MRP mean estimated proficiency2 rates fell within their corresponding NAEP-C credible intervals 100% (22 of 22) of the time.

Discussion and Implications

The purpose of this study was to provide state-level proficiency rate estimates using the NAEP-C's national-level testing results from 2014 and 2018. This study contributes to the policy and research fields by providing nationally comparable estimates of how students in each state are performing in civics. From a research lens, it reinforces the work of previous MRP, or small area estimation, studies by showcasing a novel case use for this methodology (Ortagus et al., 2021; Skinner & Doyle, 2024). This study also allows for a more nuanced discussion concerning the potential impact of civics education policies and practice. Providing state-level assessment predictions creates the opportunity to use NAEP-C as an evaluation tool and policy springboard for policymakers.

One major concern that the MRP predictions showcased was the racial/ethnic opportunity gaps between White students and their Black, Hispanic, and Native counterparts. As Milner posited (2012), "Are we focusing on too much testing and not enough teaching?" (p. 694). These testing results are the clue, or red flag, that education and policy experts can use as a starting point for investigation into the practices used to teach civics education. With civics education and the ideals of citizenship overlapping, it is concerning that some racial/ethnic groups may not see themselves as full actors in our society. Could these disparities in civics attainment be a window into similar racial/ethnic patterns seen with long-term civic engagement such as

registering to vote, voting, and running for office (Jacobsen & Linkow, 2012; Littenberg-Tobias & Cohen, 2016; Plummer et al., 2022)?

Limitations and Future Studies

One of the challenges in creating this MRP study is finding matching information about the students. This study used the student's state, public school status, gender, and race in order to create state-level estimates. Perhaps other student-level characteristics could have been more informative to their potential success on a civics assessment (i.e., number of books in the home) that was not selected or was not available for both the IPUMS and CCD data sources. I also chose to include specific state-level characteristics that I felt would be informative based on my understanding of the content area and educational assessment research more broadly. Other researchers may choose to create their own variation on the MRP model that includes different characteristics (i.e., including their state's English Language Arts assessment scores).

While a strength of the MRP methodology is that I can create a prediction for states where no students were tested, those predictions have relatively large credible intervals. In some instances, such as Alaska, the estimated NAEP scores appear to reflect the demographic characteristics of students in the state rather than some distinct practice related to civics education. Alaska lacks a stand-alone civics course or a required civics exam, which signals that the state is not expected to be a leader in civics education (Stern et al, 2021). Instead, the model is likely picking up the large fraction of White students that make up Alaska's student enrollment, and that, at the aggregate level, White students tend to perform higher on the NAEP-C overall.

The current scope of this study was restricted to two NAEP-C testing cycles, however, the MRP model developed here can be used to provide estimates for all previous NAEP-C cycles. This would allow researchers to analyze the long-term trends in civics. It would also

allow for the analysis of future NAEP-C testing cycles, until state-level results come to fruition, currently planned to begin as early as 2030.

Another opportunity for future research would be to deepen the exploration of states with the highest mean proficiency rates to look for trends. These trends may show up in the form of instructional practices in the classroom, curricular materials assigned, alignment between state standards and practice, or even similarities in their professional development opportunities.

With obvious proficiency gaps between White students and several minority subgroups, it may be beneficial to look for similar racial/ethnic gaps across other subject areas. Are there any major differences in the standards, teaching approaches, course offerings, or class times devoted to civics education in states with larger racial/ethnic gaps than those with smaller ones?
From a policy perspective, future research can be done to look for any association between differing state-level mandates (i.e., more years of social studies, stand-alone civics courses, or civics exams) and higher levels of NAEP-C proficiency. Researchers could also explore if there are any associations between higher NAEP-C proficiency rates and the state's civics-related allocation and expenditures.

Summary

By disaggregating student performance by demographics at the state level, this study informs state education leaders and policymakers, allowing for the creation and evaluation of practice and policy. The findings of racial/ethnic opportunity gaps shine a light on systemic issues that contribute to disparities in civic knowledge and engagement. It also underscores a pressing need for reform in civics education which emphasizes teaching practices that promote equity and inclusion to cultivate a more engaged and informed citizenry. One of these potential reforms could be the nurturing of the principles of civil discourse in order to elicit more fruitful

policy debates rather than the divisiveness of mere screaming matches. Paired with the follow-up of focused inquiries concerning teaching practices, curriculum alignment, and policy effects on civics education across the nation, this study has the potential to begin to bridge the gap between assessment data and civics education practices.

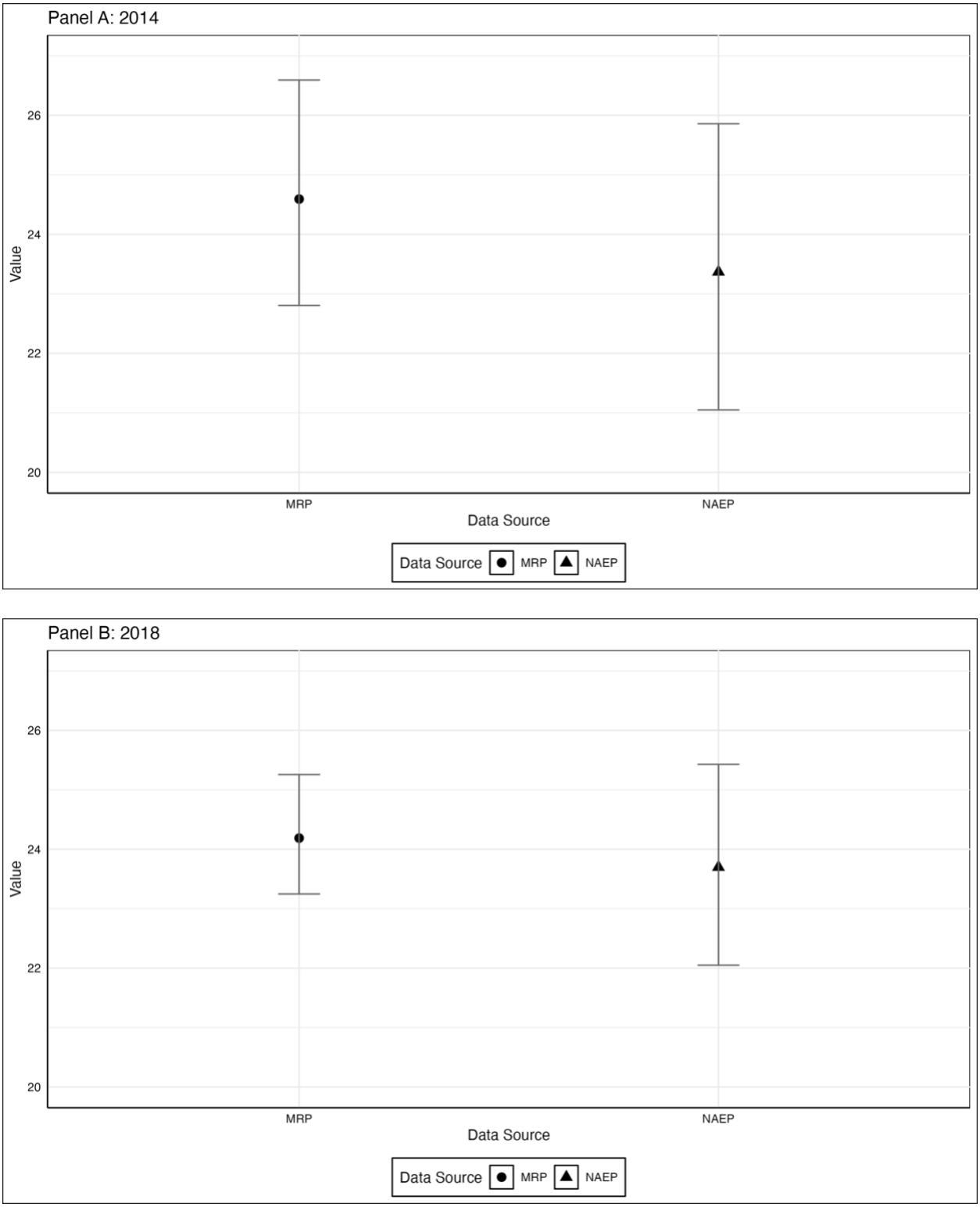
Table 1 Variable types and sources

Name	Type	Source	Level
State	Categorical (1-51)	NAEP	First – Student Level
Race/Ethnicity	Categorical (1-6)	NAEP	First – Student Level
Gender	Binary (Male = 0)	NAEP	First – Student Level
Public/Non-Public	Binary (Public = 0)	NAEP	First – Student Level
Plausible Values	Continuous	NAEP	First – Student Level
Region	Categorical (1-4)	U.S. Census	First – State Level
Governor's Party	Binary (Dem = 0)	Book of the States	First – State Level
Legislature's Leaning	Continuous	Book of the States	First – State Level
Local Area Unemployment	Continuous	U.S. Bureau of Labor and Statistics	First – State Level
Median Family Income	Continuous	Small Area Income and Poverty Estimates	First – State Level
Childhood Poverty	Continuous	Small Area Income and Poverty Estimates	First – State Level
Per Pupil Expenditures	Continuous	Common Core of Data	First – State Level
English Language Learners	Continuous	Common Core of Data	First – State Level
Students with a Disability	Continuous	Common Core of Data	First – State Level
State	Categorical (1-51)	IPUMS	Second – Area Level
Race/Ethnicity	Categorical (1-6)	IPUMS	Second – Area Level
Gender	Binary (Male = 0)	IPUMS	Second – Area Level
Public/Non-Public	Binary (Public = 0)	IPUMS	Second – Area Level

Table 2 NAEP-C demographics (2014 and 2018)

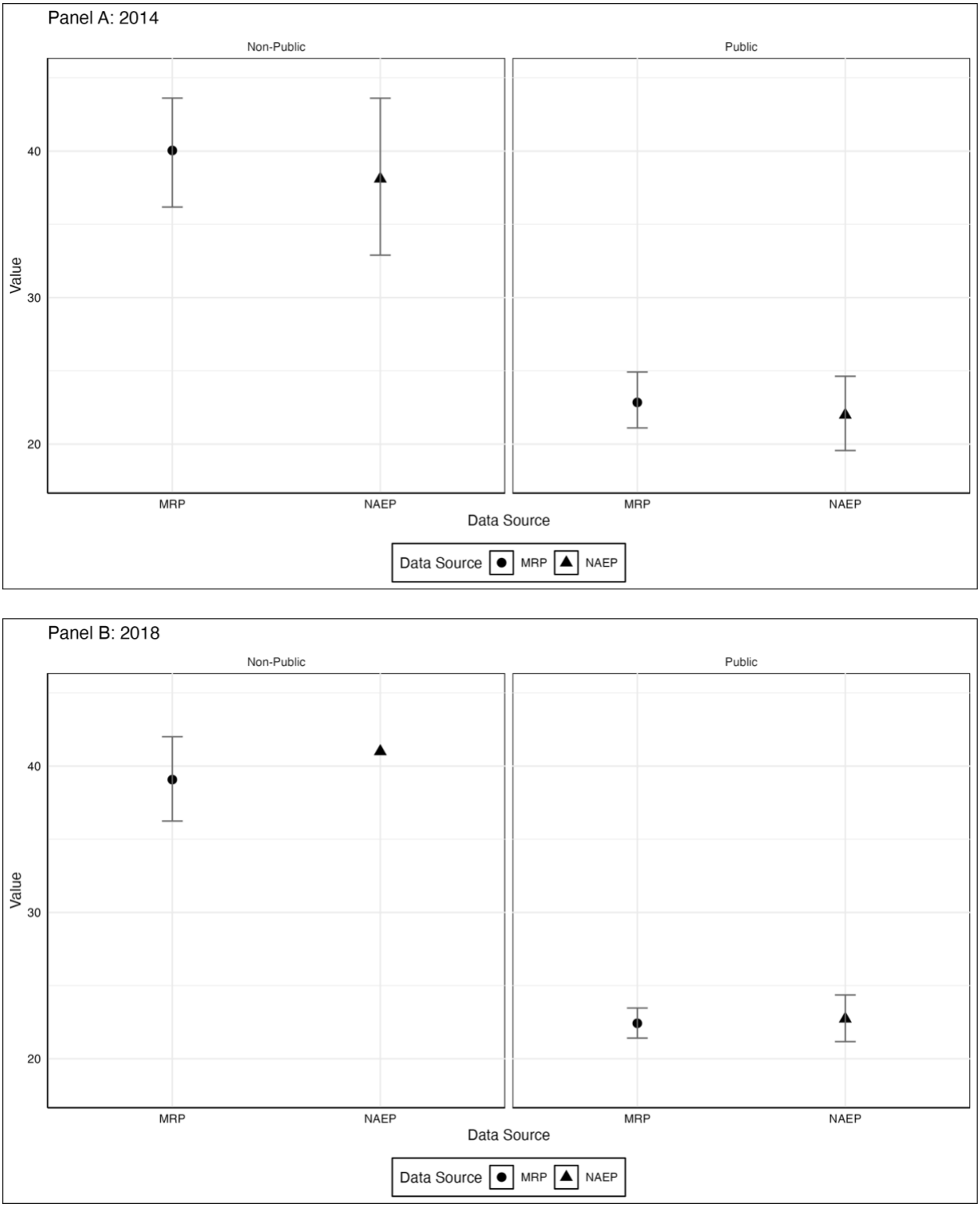
NAEP-C		2014	2018
Schools*	Total	410	780
	Public	330	640
	Private	80	140
Students**	Total	9,100	13,400
	Public	8,100	12,000
	Private	1,000	1,400
	Percent SD	11	13
	Percent ELL	5	7
Race by Percentage	White	50	49
	Black	15	14
	Hispanic	26	27
	Asian/ Pacific Islander	6	5
	American Indian/ Alaska Native	1	1
	Two or More	2	3
Gender by Percentage	Male	51	51
	Female	49	49

NOTE: * Rounded to the nearest ten; ** Rounded to the nearest hundred; # Rounds to zero
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.



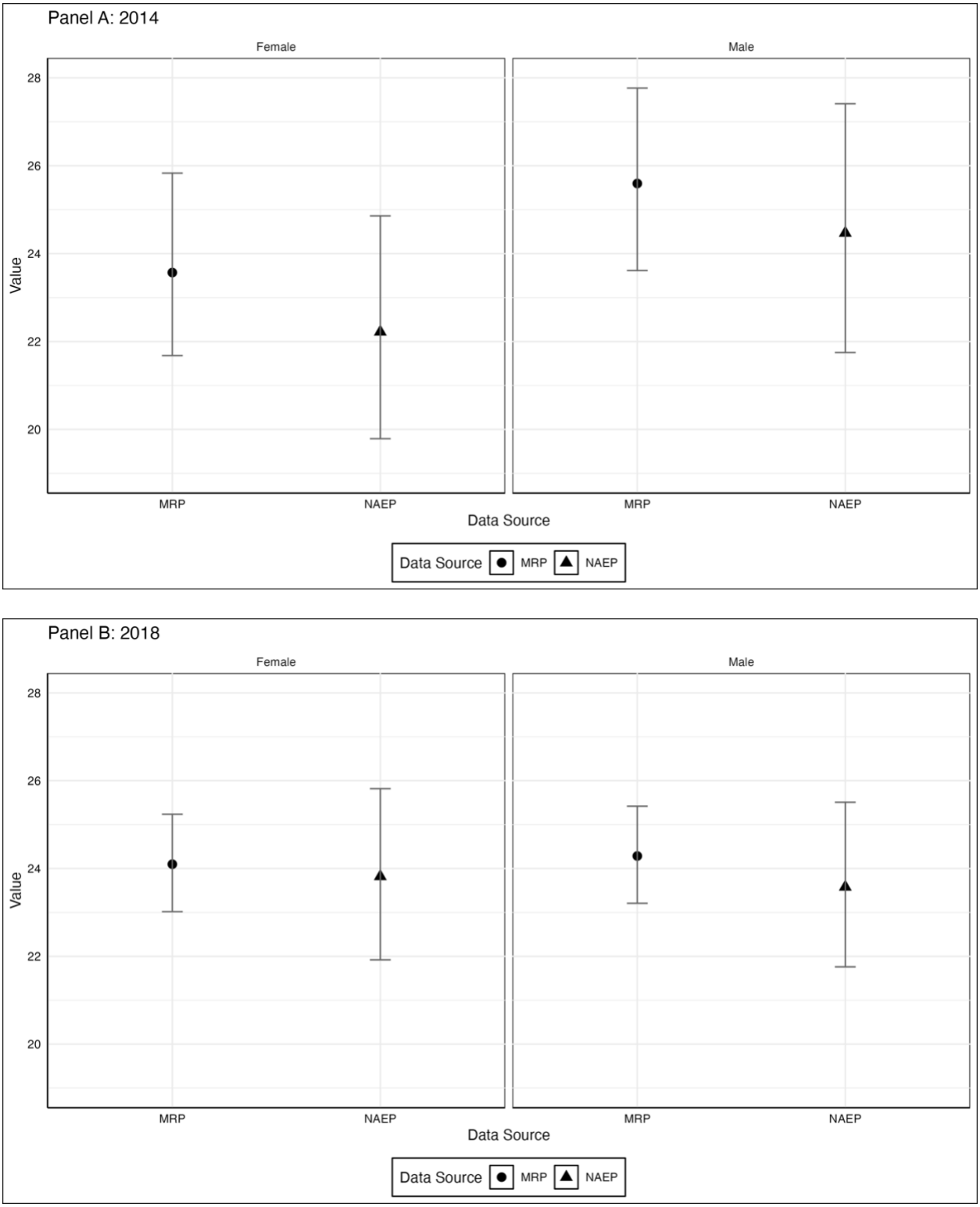
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.

Figure 1 National comparison of NAEP and MRP results (2014 and 2018)



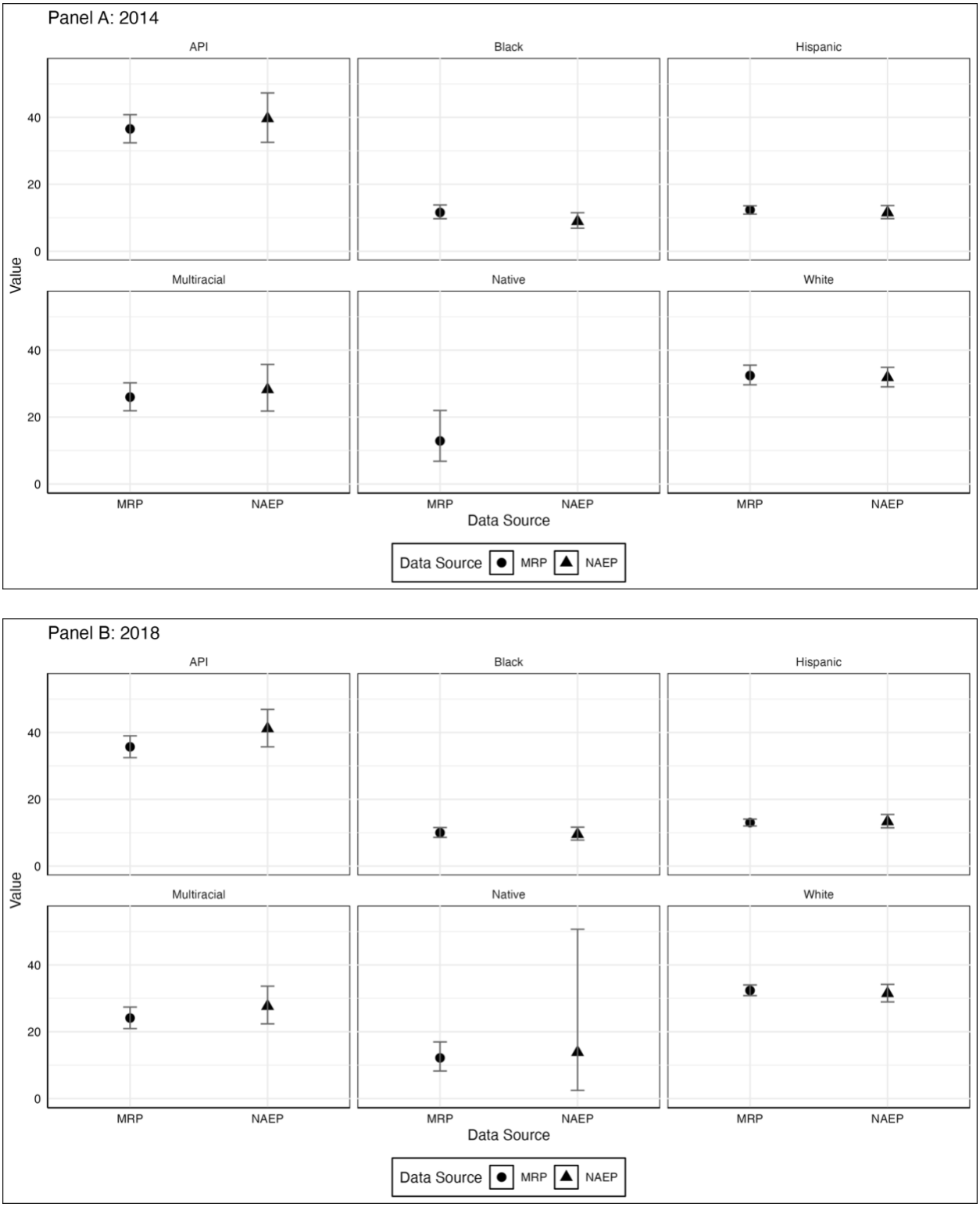
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.

Figure 2 National comparison of NAEP and MRP results by public status (2014 and 2018)



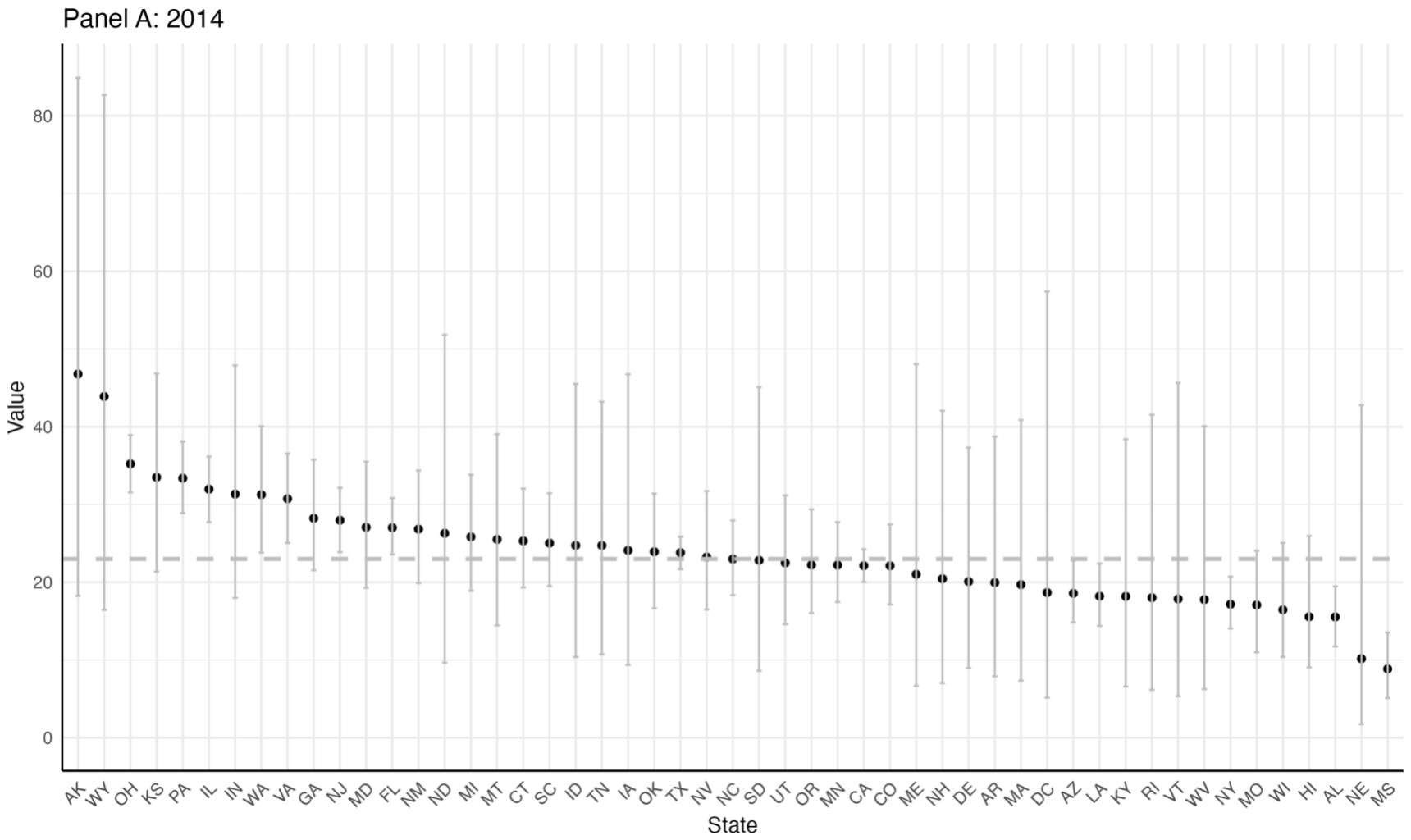
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.

Figure 3 National comparison of NAEP and MRP results by gender (2014 and 2018)



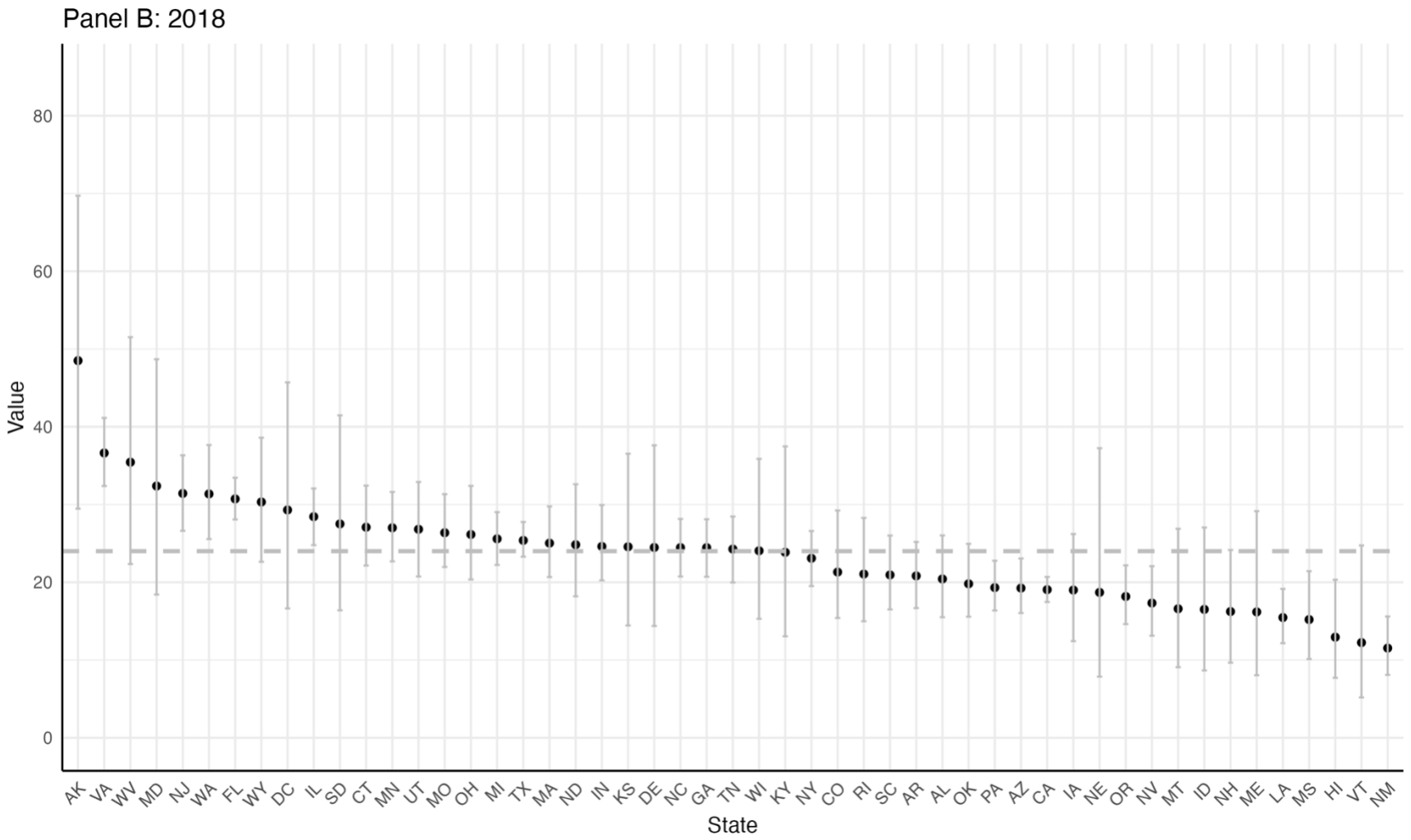
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.

Figure 4 National comparison of NAEP and MRP results by race/ethnicity (2014 and 2018)



SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.

Figure 5 State MRP estimates and 95% credible intervals for 2014



SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.

Figure 6 State MRP estimates and 95% credible intervals for 2018

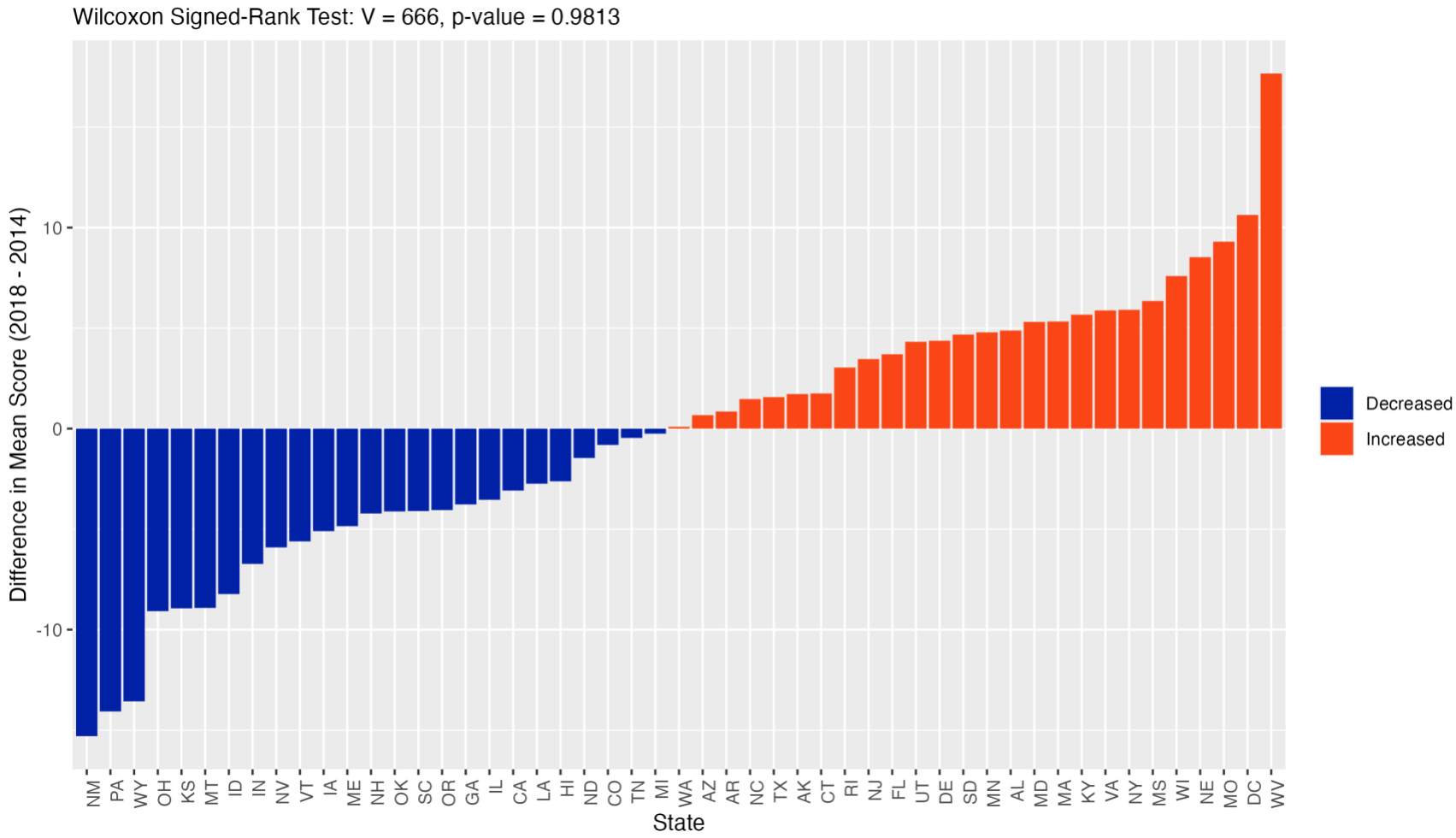
Table 3 Wilcoxon signed-rank test data

State	2014 Mean	2018 Mean	2018 Mean - 2014 Mean			
			Sign	Absolute Difference	Rank	Signed Rank
WA	31.27	31.36	1	0.09	1	1
MI	25.83	25.58	-1	0.25	2	-2
TN	24.73	24.27	-1	0.46	3	-3
AZ	18.58	19.25	1	0.67	4	4
CO	22.12	21.31	-1	0.81	5	-5
AR	19.97	20.82	1	0.85	6	6
ND	26.29	24.83	-1	1.46	7	-7
NC	22.99	24.46	1	1.47	8	8
TX	23.81	25.38	1	1.57	9	9
AK	46.78	48.5	1	1.72	10	10
CT	25.32	27.07	1	1.75	11	11
HI	15.56	12.94	-1	2.62	12	-12
LA	18.2	15.46	-1	2.74	13	-13
RI	18.01	21.05	1	3.04	14	14
CA	22.13	19.05	-1	3.08	15	-15
NJ	27.97	31.43	1	3.46	16	16
IL	31.97	28.43	-1	3.54	17	-17
FL	27.02	30.72	1	3.7	18	18
GA	28.23	24.46	-1	3.77	19	-19
OR	22.22	18.17	-1	4.05	20	-20
SC	25.04	20.94	-1	4.1	21	-21
OK	23.92	19.8	-1	4.12	22	-22
NH	20.45	16.23	-1	4.22	23	-23
UT	22.49	26.81	1	4.32	24	24
DE	20.1	24.47	1	4.37	25	25
SD	22.82	27.5	1	4.68	26	26
MN	22.21	27	1	4.79	27	27
ME	21.03	16.18	-1	4.85	28	-28
AL	15.54	20.42	1	4.88	29	29
IA	24.1	19	-1	5.1	30	-30
MD	27.07	32.38	1	5.31	31	31
MA	19.7	25.03	1	5.33	32	32
VT	17.85	12.24	-1	5.61	33	-33
KY	18.18	23.85	1	5.67	34	34
VA	30.74	36.62	1	5.88	35	35
NY	17.17	23.08	1	5.91	36.5	36.5

Table 3 Continued

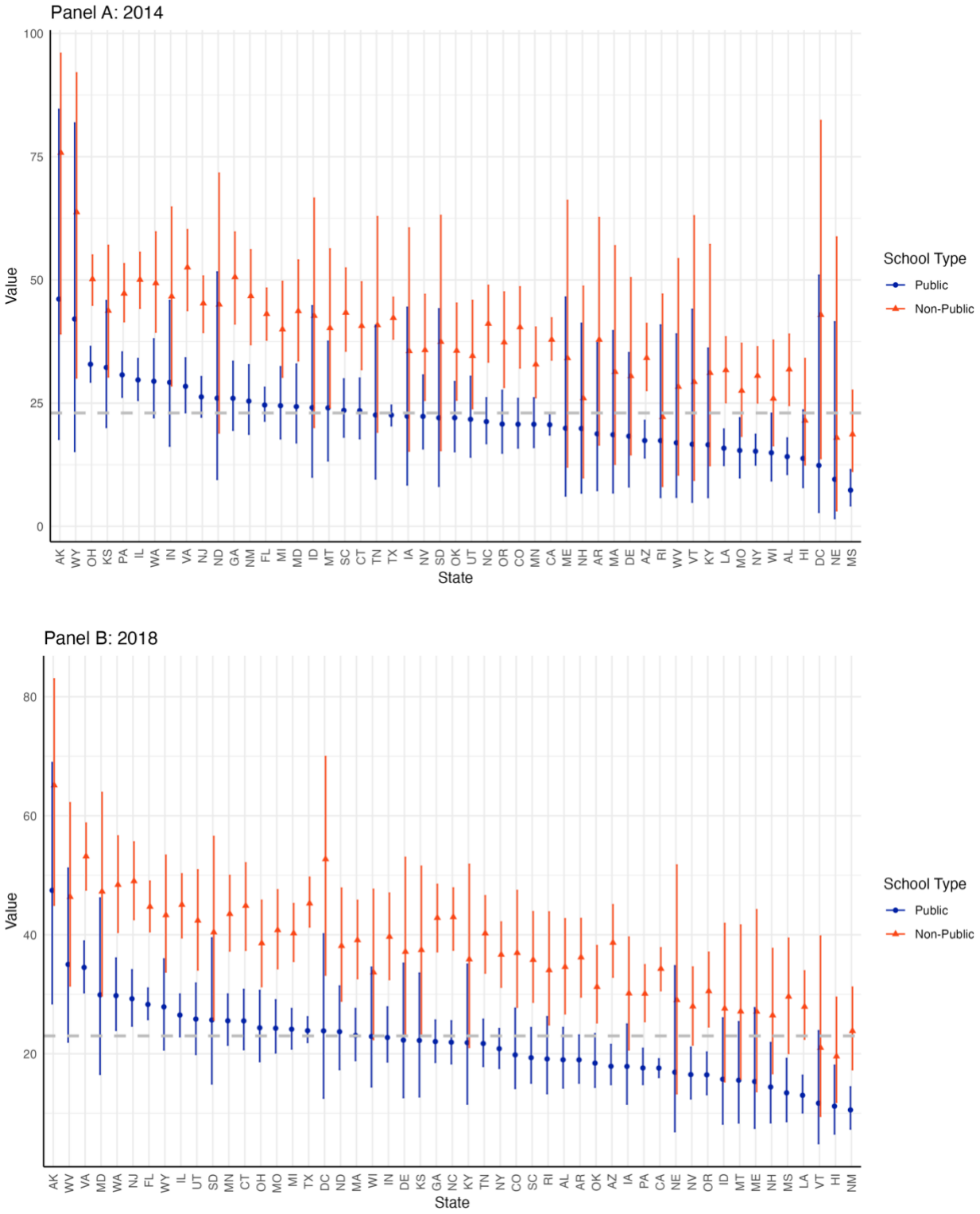
State	2014 Mean	2018 Mean	2018 Mean - 2014 Mean			
			Sign	Absolute Difference	Rank	Signed Rank
NV	23.24	17.33	-1	5.91	36.5	-36.5
MS	8.85	15.2	1	6.35	38	38
IN	31.34	24.61	-1	6.73	39	-39
WI	16.46	24.05	1	7.59	40	40
ID	24.74	16.51	-1	8.23	41	-41
NE	10.17	18.7	1	8.53	42	42
MT	25.51	16.59	-1	8.92	43	-43
KS	33.5	24.56	-1	8.94	44	-44
OH	35.22	26.14	-1	9.08	45	-45
MO	17.07	26.37	1	9.3	46	46
DC	18.67	29.3	1	10.63	47	47
WY	43.89	30.32	-1	13.57	48	-48
PA	33.38	19.31	-1	14.07	49	-49
NM	26.83	11.53	-1	15.3	50	-50
WV	17.77	35.44	1	17.67	51	51

NOTE: Wilcoxon Signed-Rank Test: V = 666, p-value = 0.9813; expressing no statistical difference between the mean values across both NAEP-C testing cycles.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.



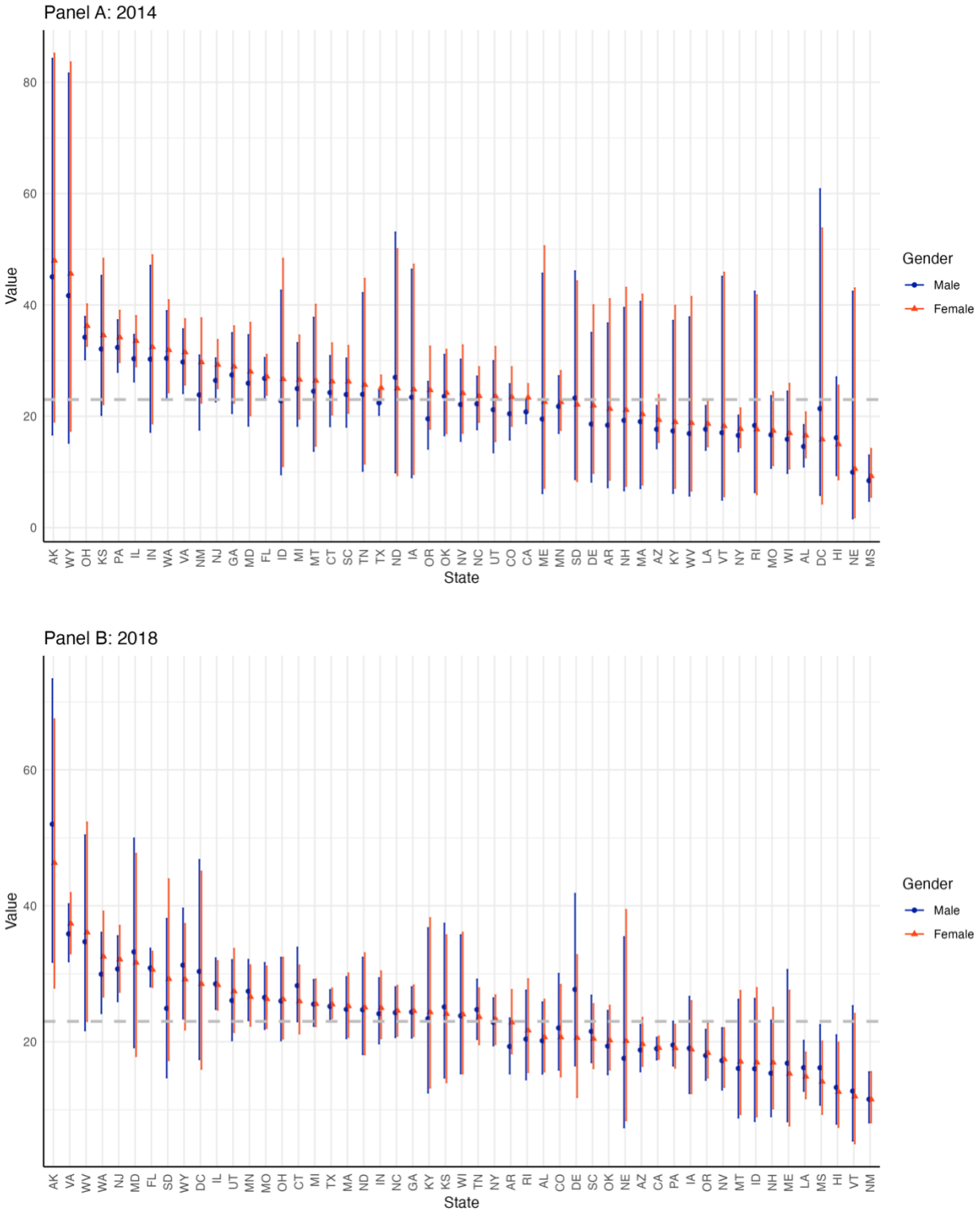
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.

Figure 7 Difference in NAEP-C scores by state (2018 vs 2014)



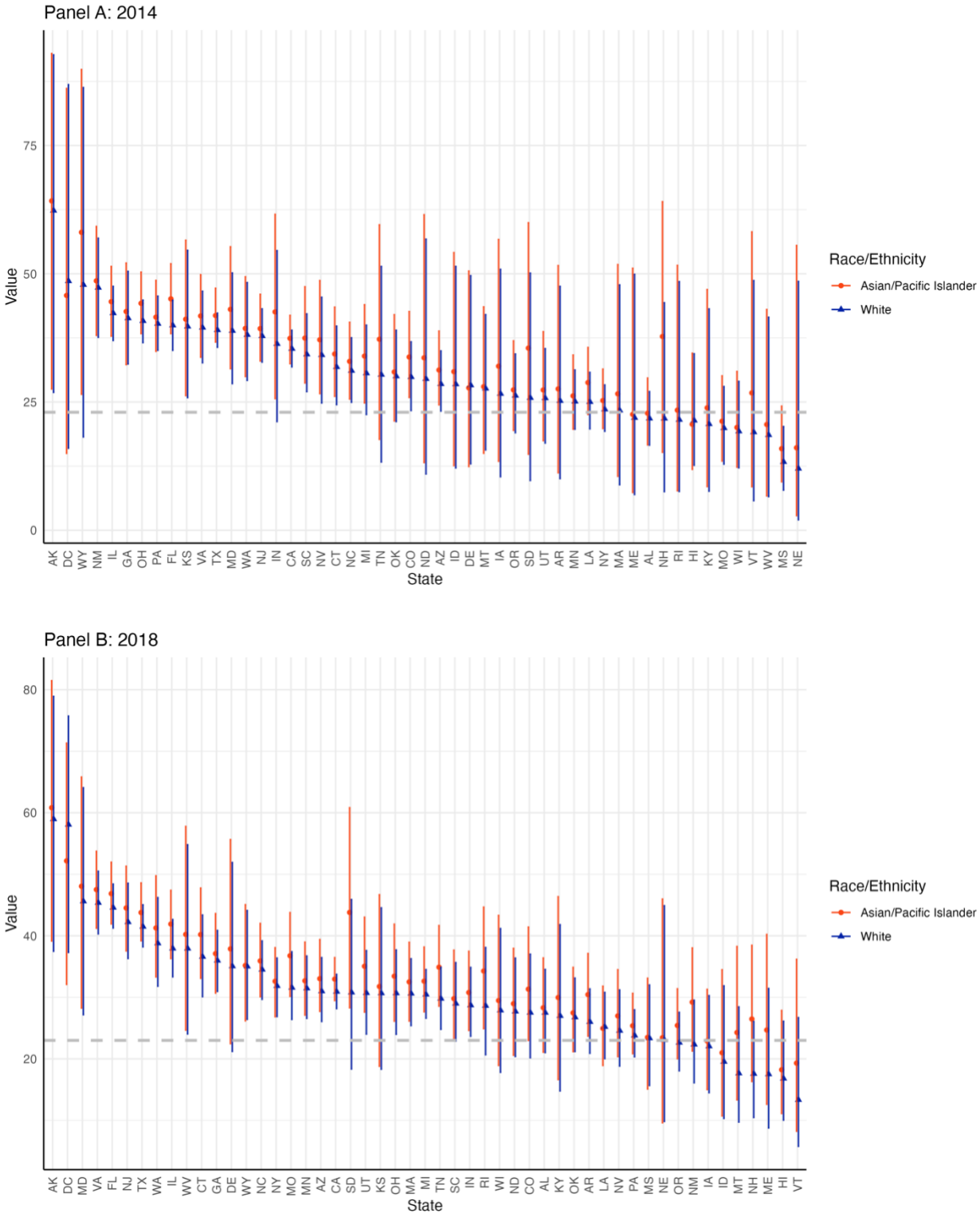
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.

Figure 8 State-level comparison of NAEP and MRP results by public school status (2014 and 2018)



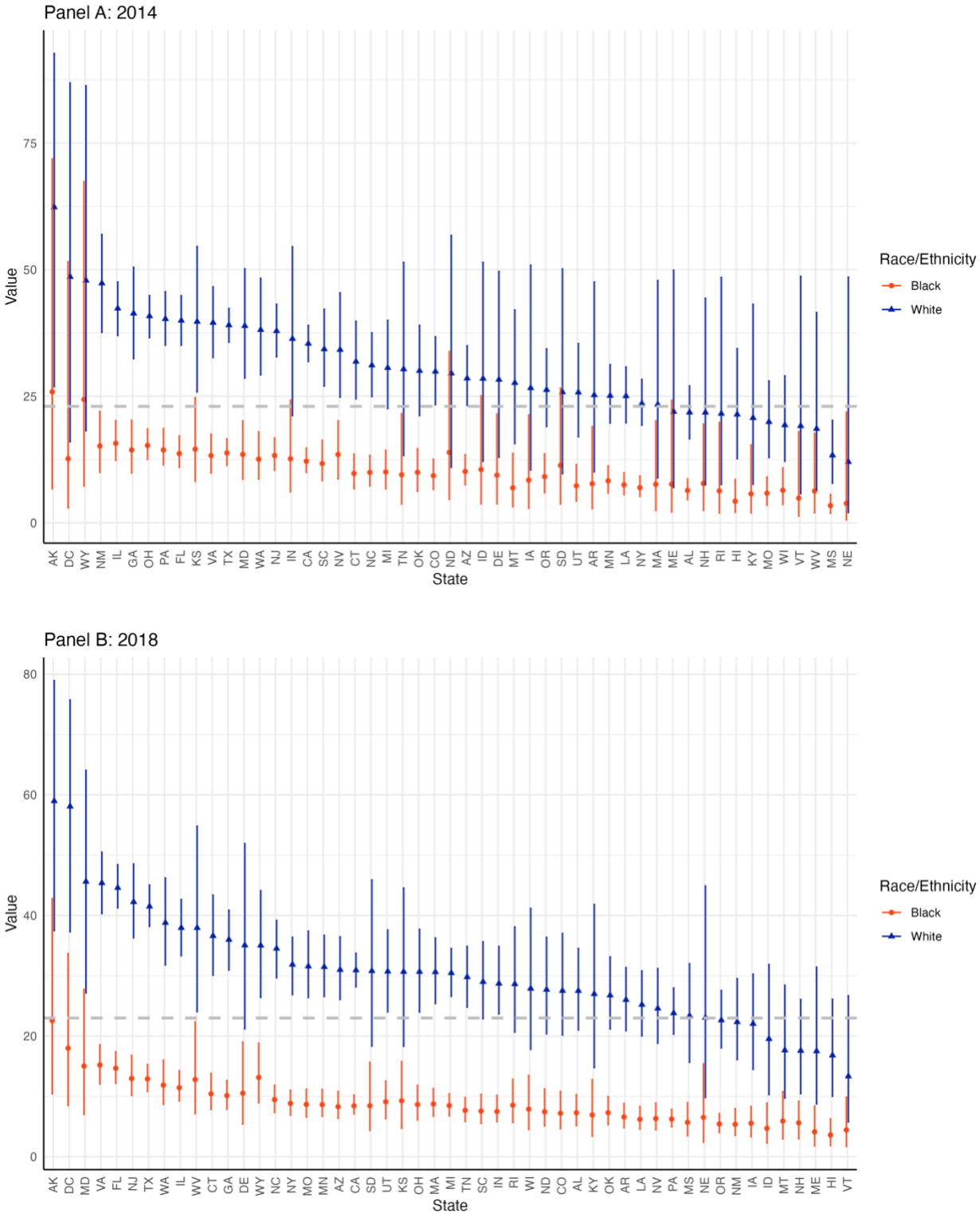
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.

Figure 9 State-level comparison of NAEP and MRP results by gender (2014 and 2018)



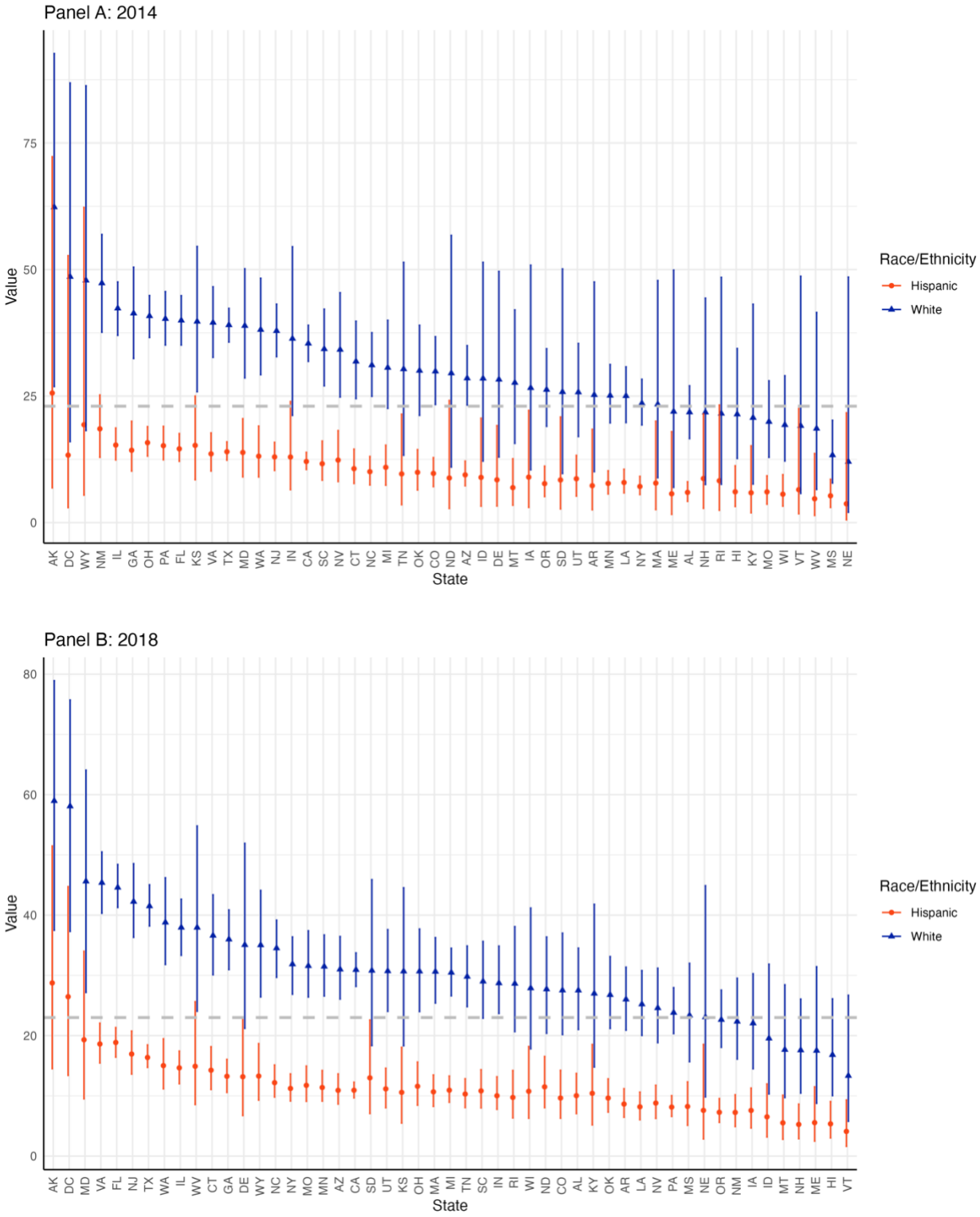
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.

Figure 10 State-level comparison of MRP results by race/ethnicity gap White/API (2014 and 2018)



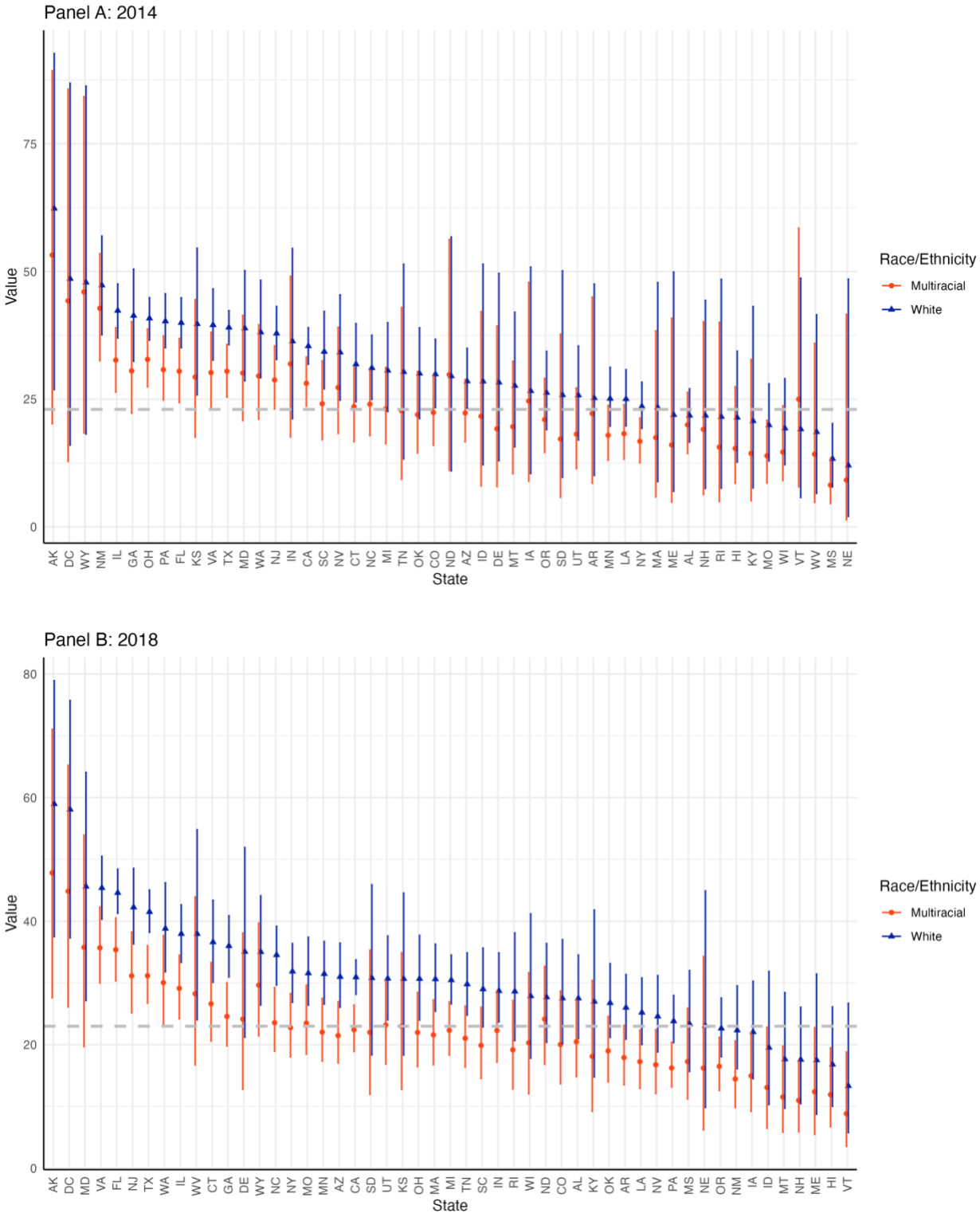
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.

Figure 11 State-level comparison of MRP results by race/ethnicity gap White/Black (2014 and 2018)



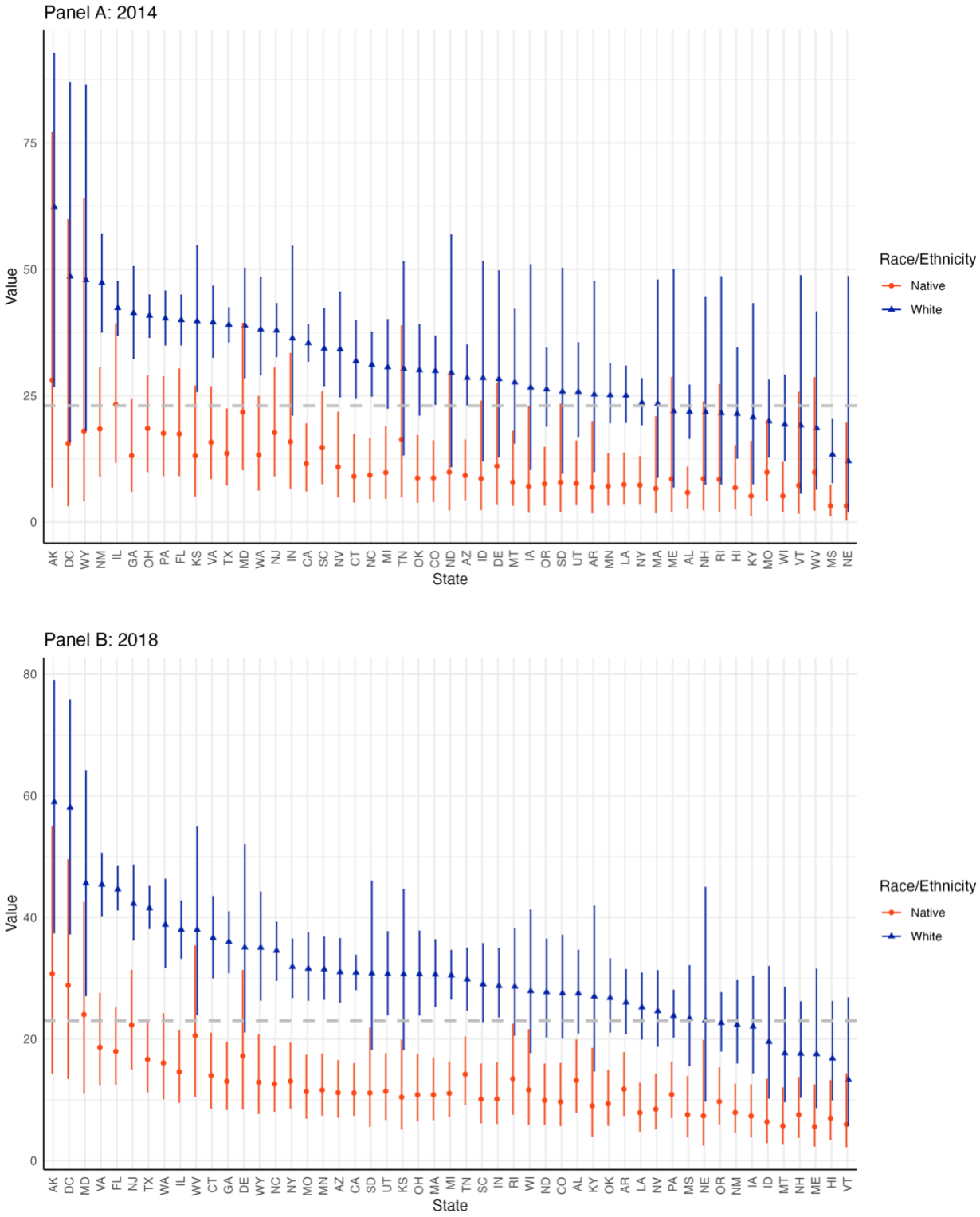
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.

Figure 12 State-level comparison of MRP results by race/ethnicity gap White/Hispanic (2014 and 2018)



SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.

Figure 13 State-level comparison of MRP results by race/ethnicity gap White/Multiracial (2014 and 2018)



SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.

Figure 14 State-level comparison of MRP results by race/ethnicity gap White/Native (2014 and 2018)

References

Addington, L. A. (2016). Using constitutional law classes to address the "civic empowerment gap" among inner-city public high school students. *Education and Urban Society*, *48*(9), 807–826. https://doi.org/10.1177/0013124516630597

Arias, O., Canals, C., Mizala, A., & Meneses, F. (2023). Gender gaps in Mathematics and Language: The bias of competitive achievement tests. *PLOS ONE*, *18*(3), e0283384. https://doi.org/10.1371/journal.pone.0283384

Berkman, M. B., & Plutzer, E. (2011). Local autonomy versus state constraints: Balancing evolution and creationism in U.S. high schools. *Publius: The Journal of Federalism*, *41*(4), 610–635. https://doi.org/10.1093/publius/pjr008

Bittman, B. L., & Russell III, W. B. (2016). Civic education in United States: A multiple regression of civic education scores from the National Assessment of Educational Progress. *Research in Social Sciences and Technology*, *1*(2), 1–16.

Bohan, C. H. (2017). Gender and feminist scholarship in social studies research. In *The Wiley Handbook of Social Studies Research*.

Brown, J. (2015, April 29). Eighth graders' flatline on NAEP U.S. history, civics, and geography tests. *Education Week*. https://www.edweek.org/leadership/eighth-graders-flatline-on-naep-u-s-history-civics-and-geography-tests/2015/04

Buck, G., Kostin, I., & Morgan, R. (2002). *Examining the relationship of content to gender-based performance differences in advanced placement exams* (2002–12; pp. i–34). The College Board. https://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.2002.tb01892.x

Carnoy, M., Garcia, E., & Khavenson, T. (2015). *Bringing it back home: Why state comparisons are more useful than international comparisons for improving U.S. education policy* (Briefing Paper #410). Economic Policy Institute. https://www.epi.org/publication/bringing-it-back-home-why-state-comparisons-are-more-useful-than-international-comparisons-for-improving-u-s-education-policy/

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*, 1. https://doi.org/10.18637/jss.v076.i01

College Board. (2014). *The 10th annual AP report to the nation* (pp. 1–44). The College Board.

Diamond, L. (2015). Facing up to the democratic recession. *Journal of Democracy*, *26*(1), 141–155. https://doi.org/10.1353/jod.2015.0009

Dillon, S. (2011, May 4). Failing grades on civics exam called a 'crisis.' *The New York Times*. https://www.nytimes.com/2011/05/05/education/05civics.html

Dinour, I. (2022, November 21). *Investing in civic education to safeguard our democracy*. Einhorn Collaborative. https://einhorncollaborative.org/investing-in-civic-education-to-safeguard-our-democracy/

Ehrlich, T. (1999). Civic education: Lessons learned. *PS: Political Science & Politics*, *32*(2), 245–250. https://doi.org/10.2307/420559

Feldman, L., Pasek, J., Romer, D., & Jamieson, K. H. (2007). Identifying best practices in civic education: Lessons from the Student Voices program. *American Journal of Education*, *114*(1), 75–100. https://doi.org/10.1086/520692

Finder, M. (2004). *Educating America: How Ralph W. Tyler taught America to teach*. Praeger.

Finn Jr., C. E. (2018, July 18). *The failure of civics education—And the Brown Center*. The Thomas B. Fordham Institute. https://fordhaminstitute.org/national/commentary/failure-civics-education-and-brown-center

Fitchett, P. G., & Heafner, T. L. (2017). *An Analysis of Predictors of History Content Knowledge: Implications for Policy and Practice*. *25*(65).

Fouquet, J. M. (2019). *What really happened in the NCLB era? A state-level analysis of achievement as influenced by critical school and non-school factors*. https://kuscholarworks.ku.edu/handle/1808/31550

Furgione, B., Evans, K., Walker, I., & Russell, I. W. B. (2018). The elephant in the classroom: A comparative study of civics end-of-course assessment. *Social Studies Research and Practice*, *13*(2), 168–184. https://doi.org/10.1108/SSRP-08-2017-0043

Fryer Jr., R. G., & Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, *2*(2), 210–240. https://doi.org/10.1257/app.2.2.210

Gabry, J., Ali, I., Brilleman, S., Buros Novak, J., AstraZeneca, Trustees of Columbia University, Wood, S., R Core Development Team, Bates, D., Maechler, M., Bolker, B., Walker, S., Ripley, B., Venables, W., Burkner, P.-C., & Goodrich, B. (2024). *Bayesian applied regression modeling via Stan* (2.32.1) [Computer software]. https://mc-stan.org/rstanarm/, https://discourse.mc-stan.org

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gelman, A., & Little, T. C. (1997). *Poststratification into many categories using hierarchical logistic regression*. 17.

Ghitza, Y., & Gelman, A. (2013). Deep interactions with MRP: Election turnout and voting patterns among small electoral subgroups. *American Journal of Political Science*, *57*(3), 762–776. https://www.jstor.org/stable/23496652

Ghitza, Y., & Gelman, A. (2020). Voter registration databases and MRP: Toward the use of large-scale databases in public opinion research. *Political Analysis*, *28*(4), 507–531. https://doi.org/10.1017/pan.2020.3

Hanretty, C. (2020). An introduction to multilevel regression and post-stratification for estimating constituency opinion. *Political Studies Review*, *18*(4), 630–645. https://doi.org/10.1177/1478929919864773

Hanretty, C., Lauderdale, B. E., & Vivyan, N. (2018). Comparing strategies for estimating constituency opinion from national survey samples. *Political Science Research and Methods*, *6*(3), 571–591. https://doi.org/10.1017/psrm.2015.79

Heafner, T. L., & Fitchett, P. G. (2018). US history content knowledge and associated effects of race, gender, wealth, and urbanity: Item Response Theory (IRT) modeling of NAEP-USH achievement. *The Journal of Social Studies Research*, *42*(1), 11–25. https://doi.org/10.1016/j.jssr.2017.01.001

Hepburn, M. A. (1997). Service learning in civic education: A concept with long, sturdy roots. *Theory Into Practice*, *36*(3), 136–142. https://doi.org/10.1080/00405849709543759

Herczog, M. M. (2012). It's up to us: Bridging the civic opportunity gap. *Leadership*, *41*(4), 20–22, 24, 26

Houston, D. M. (2019). Schoolhouse democracy: Public opinion and education spending in the states. *Educational Researcher*, *48*(7), 438–451. https://doi.org/10.3102/0013189X19867948

H.R.1 - No child left behind act of 2001, Pub. L. No. PL 107-110 (2002). http://www.congress.gov/

H.R.5 - Augustus F. Hawkins-Robert T. Stafford elementary and secondary school improvement amendments of 1988, Pub. L. No. PL 100-297 (1988). http://www.congress.gov/

H.R.1804 - Goals 2000: Educate America act, Pub. L. No. PL 103-227 (1994). http://www.congress.gov/

Jacobsen, R., & Linkow, T. W. (2012). *The engaged citizen index: Examining the racial and ethnic civic and political engagement gaps of youth adults* (CIRCLE Working Paper #74; pp. 1–41). Center for Information & Research on Civic Learning & Engagement.

Jones, L. V., & Olkin, I. (Eds.). (2004). *The nation's report card: Evolution and perspectives*. Phi Delta Kappa Educational Foundation.

Kahne, J. (2015, April 22). Why are we teaching democracy like a game show? *Education Week*. https://www.edweek.org/policy-politics/opinion-why-are-we-teaching-democracy-like-a-game-show/2015/04

Kahne, J., & Middaugh, E. (2008). *Democracy for some: The civic opportunity gap in high school* (CIRCLE Working Paper 59; pp. 1–30). CIRCLE. https://doi.org/10.2307/j.ctv16f6dbw

Kawashima-Ginsberg, K. (2013). *Do discussion, debate, and simulations boost NAEP civics performance?* (pp. 1–16). CIRCLE

Kiewiet de Jonge, C. P., Langer, G., & Sinozich, S. (2018). Predicting state presidential election results using national tracking polls and multilevel regression with poststratification (MRP). *Public Opinion Quarterly*, *82*(3), 419–446.

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

Lax, J. R., & Phillips, J. H. (2009a). How should we estimate public opinion in the states? *American Journal of Political Science*, *53*(1), 107–121. https://doi.org/10.1111/j.1540-5907.2008.00360.x

Lax, J. R., & Phillips, J. H. (2009b). Gay rights in the states: Public opinion and policy responsiveness. *The American Political Science Review*, *103*(3), 367–386. https://www.proquest.com/docview/214444506/abstract/B95D469E94B84986PQ/1

Layton, D. H. (1985). ECS at 20: New vitality and new possibilities. *The Phi Delta Kappan*, *67*(4), 272–276. https://www.jstor.org/stable/20387614

Lee, J. (2006). *Tracking achievement gaps and assessing the impact of NCLB on the gaps: An in-depth look into national and state reading and math outcomes trends*. The Civil Rights Project at Harvard University.

Lee, J., & Reeves, T. (2012). Revisiting the impact of NCLB high-stakes school accountability, capacity, and resources: State NAEP 1990–2009 reading and math achievement gaps and trends. *Educational Evaluation and Policy Analysis*, *34*(2), 209–231. https://doi.org/10.3102/0162373711431604

Levine, P. (2009). The civic opportunity gap. *Teaching Social Responsibility*, *66*(8), 20–25.

Levinson, M. (2010). The civic empowerment gap: Defining the problem and locating solutions. In L. R. Sherrod, J. Torney-Purta, & C. A. Flanagan (Eds.), *Handbook of Research on Civic Engagement in Youth* (pp. 331–361). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470767603.ch13

Levinson, M. (2012). *No citizen left behind* (Vol. 13). Harvard University Press.

Littenberg-Tobias, J., & Cohen, A. K. (2016). Diverging paths: Understanding racial differences in civic engagement among White, African American, and Latina/o adolescents using structural equation modeling. *American Journal of Community Psychology*, *57*(1–2), 102–117. https://doi.org/10.1002/ajcp.12027

Litvinov, A. (2017, March 16). *Forgotten purpose: Civics education in public schools*. National Education Association. https://www.nea.org/advocating-for-change/new-from-nea/forgotten-purpose-civics-education-public-schools

Lobo, B. (2022). *Can we improve multilevel regression and poststratification (MRP) through new ways to leverage information?* [Ph.D., Royal Holloway, University of London]. https://rhul.elsevierpure.com/en/publications/5ee61c7c-cc3a-4801-a0b8-717bd2315b32

Lopez-Martin, J., Phillips, J. H., & Gelman, A. (2022). *Multilevel regression and poststratification case studies*. https://bookdown.org/jl5522/MRP-case-studies/

Lutkus, A. D., Weiss, A. R., Campbell, J. R., Mazzeo, J., & Lazer, S. (1999). *NAEP 1998 civics report card for the nation* (p. 196). National Center for Education Statistics.

Myers, S. L., Kim, H., & Mandala, C. (2004). The effect of school poverty on racial gaps in test scores: The case of the Minnesota basic standards tests. *The Journal of Negro Education*, *73*(1), 81–98. https://doi.org/10.2307/3211261

National Assessment Governing Board. (n.d.). *The board's role*. Retrieved December 8, 2022, from https://www.nagb.gov/about-us/the-boards-role.html

National Assessment Governing Board. (2018). *Civics framework for the 2018 National Assessment of Educational Progress* (p. 108). National Assessment Governing Board.

National Assessment Governing Board. (2021, November 22). *NAEP assessment schedule*. https://www.nagb.gov/naep/assessment-schedule.htm

National Assessment of Educational Progress. (2022, April 28). *Informing education policy and practice.* https://nces.ed.gov/nationsreportcard/about/policy_practice.aspx

National Assessment of Educational Progress. (2023). *NAEP civics: Achievement-Level results*. https://www.nationsreportcard.gov/civics/results/achievement/

Niemi, R., & Junn, J. (1998). *Civic education: What makes students learn*. Yale University Press.

O'Malley, F., & Norton, S. (2022). *Maintaining the Validity of the NAEP Frameworks and Assessments in Civics and U.S. History* (p. 58). NAEP Validity Studies (NVS) Panel.

Ortagus, J. C., Skinner, B. T., & Tanner, M. J. (2021). Investigating why academically successful community college students leave college without a degree. *AERA Open*, *7*, 233285842110657. https://doi.org/10.1177/23328584211065724

Park, D. K., Gelman, A., & Bafumi, J. (2004). Bayesian multilevel estimation with poststratification: State-level estimates from national polls. *Political Analysis*, *12*(4), 375–385. https://www.jstor.org/stable/25791784

Plummer, J. A., Wray-Lake, L., Alvis, L., Metzger, A., & Syvertsen, A. K. (2022). Assessing the link between adolescents' awareness of inequality and civic engagement across time and racial/ethnic groups. *Journal of Youth and Adolescence*, *51*(3), 428–442. https://doi.org/10.1007/s10964-021-01545-6

Ravitch, D. (2013). *Reign of error: The hoax of the privatization movement and the danger to America's public schools*. Vintage.

Shapiro, S., & Brown, C. (2018). *The state of civics education* (p. 9). Center for American Progress. https://www.americanprogress.org/wp-content/uploads/2019/12/020618_CivicsEducation-UPDATED.pdf

Skinner, B. T., & Doyle, W. R. (2024). Predicting postsecondary attendance by family income in the United States using multilevel regression with poststratification. *Economics of Education Review*, *99*, 102508. https://doi.org/10.1016/j.econedurev.2024.102508

Sohn, K. (2012). A new insight into the gender gap in math. *Bulletin of Economic Research*, *64*(1), 135–155. https://doi.org/10.1111/j.1467-8586.2010.00358.x

Stedman, L. C. (2009). *The NAEP long-term trend assessment: A review of its transformation, use, and findings* (p. 52). Teaching, Learning and Educational Leadership Faculty Scholarship. https://orb.binghamton.edu/education_fac/2

Stern, J. A., Brody, A. E., Gregory, J. A., Griffith, S., & Pulvers, J. (2021). *The state of state standards for civics and U.S. history in 2021* (p. 377). The Thomas B. Fordham Institute.

Swalwell, K. (2015). Mind the civic empowerment gap: Economically elite students and critical civic education. *Curriculum Inquiry*, *45*(5), 491–512.

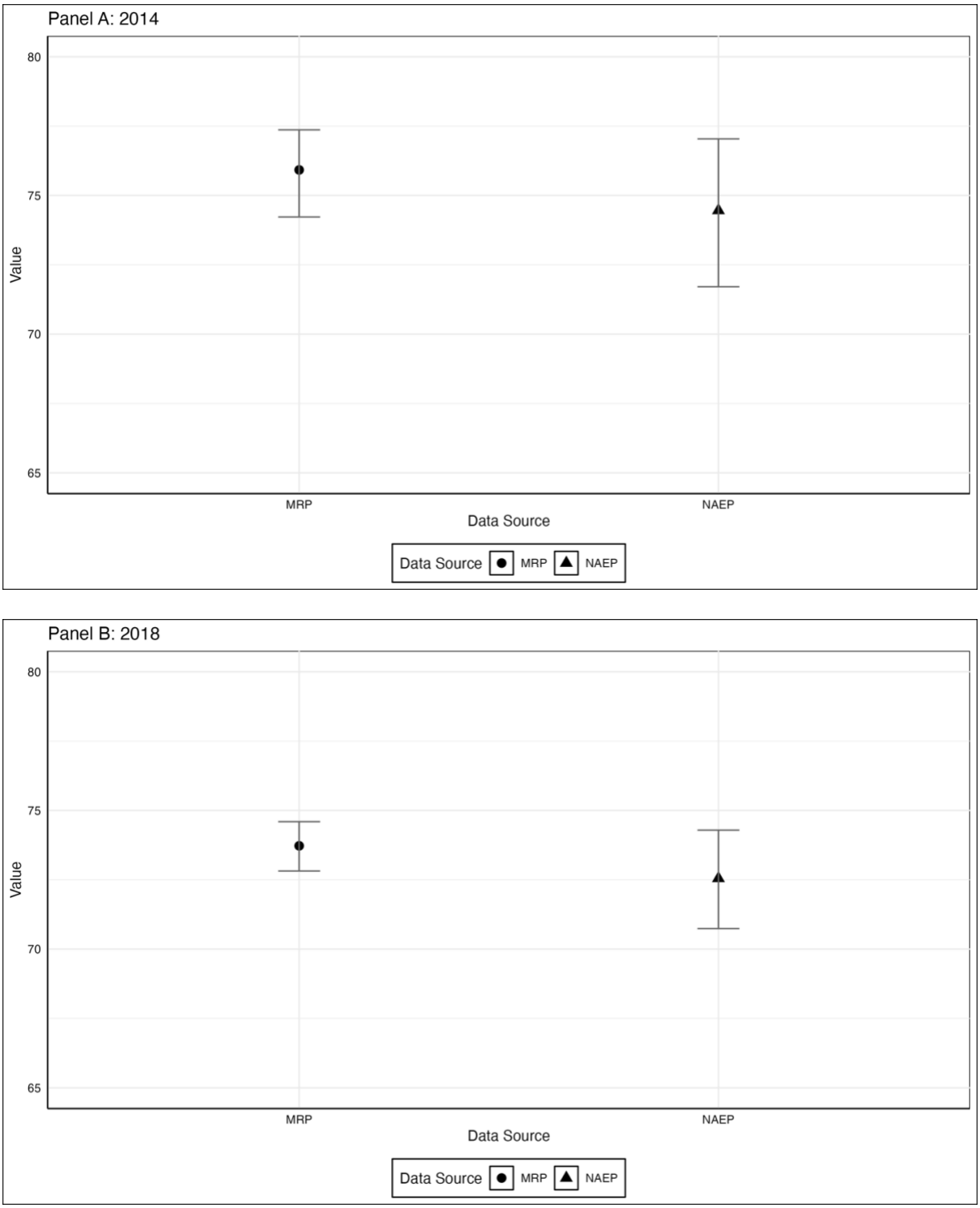
U.S. Department of Education. (2010). *An introduction to NAEP: National Assessment of Educational Progress* (NCES 2010-468; p. 20). National Center for Education Statistics. https://nces.ed.gov/nationsreportcard/pdf/parents/2010468.pdf

Vinovskis, M. A. (2001). *Overseeing the nation's report card: The creation and evolution of the National Assessment Governing Board (NAGB)* (pp. 1–92). National Assessment Governing Board. https://files.eric.ed.gov/fulltext/ED457245.pdf

Weiss, A. R., Lutkus, A. D., Grigg, W. S., & Niemi, R. G. (2001). *The next generation of citizens: NAEP civics assessments—1988 and 1998* (p. 85). National Center for Education Statistics.

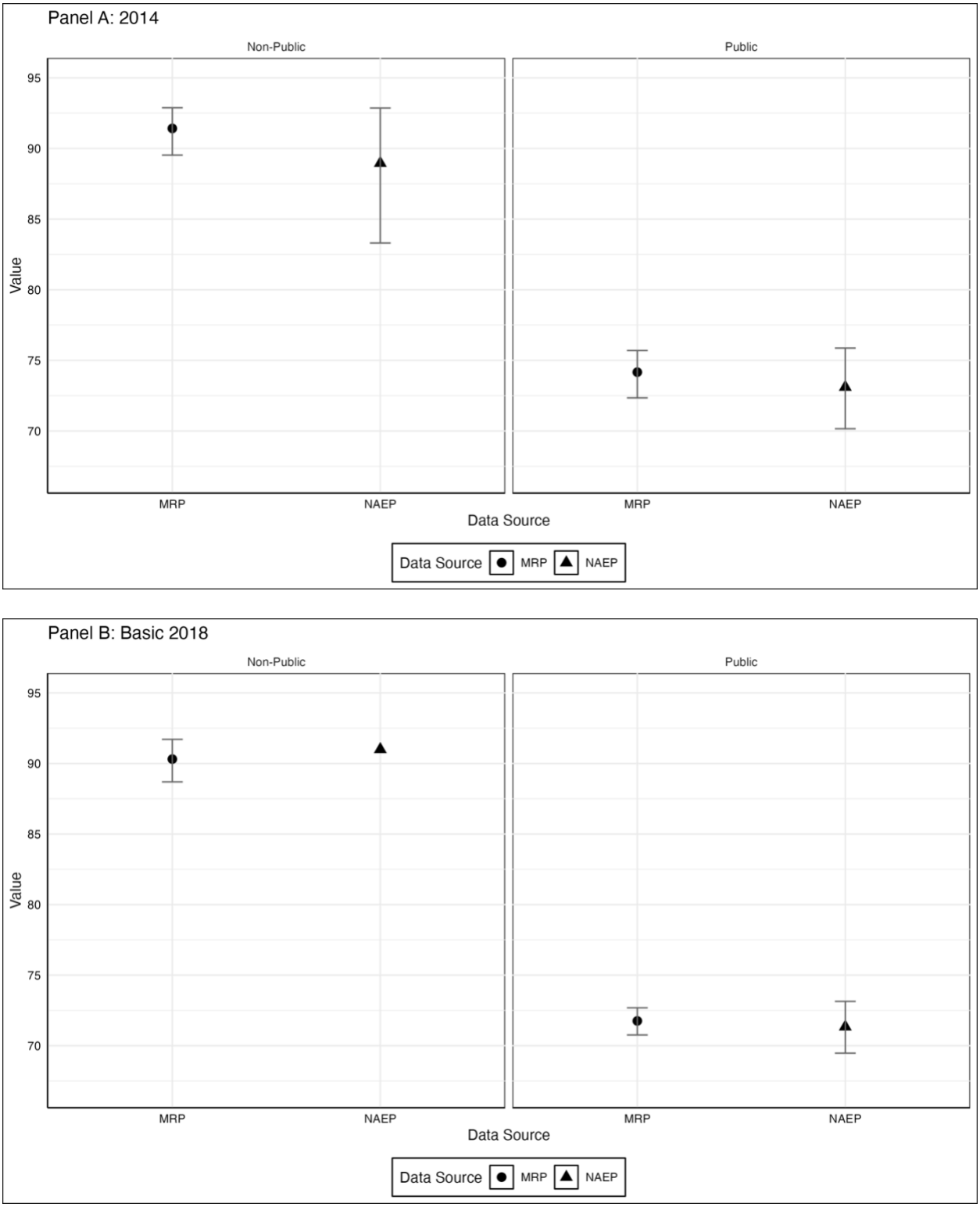
White, G. W., Stepney, C. T., Hatchimonji, D. R., Moceri, D. C., Linsky, A. V., Reyes-Portillo, J. A., & Elias, M. J. (2016). The increasing impact of socioeconomics and race on standardized academic test scores across elementary, middle, and high school. *American Journal of Orthopsychiatry*, *86*, 10–23. https://doi.org/10.1037/ort0000122

APPENDIX A



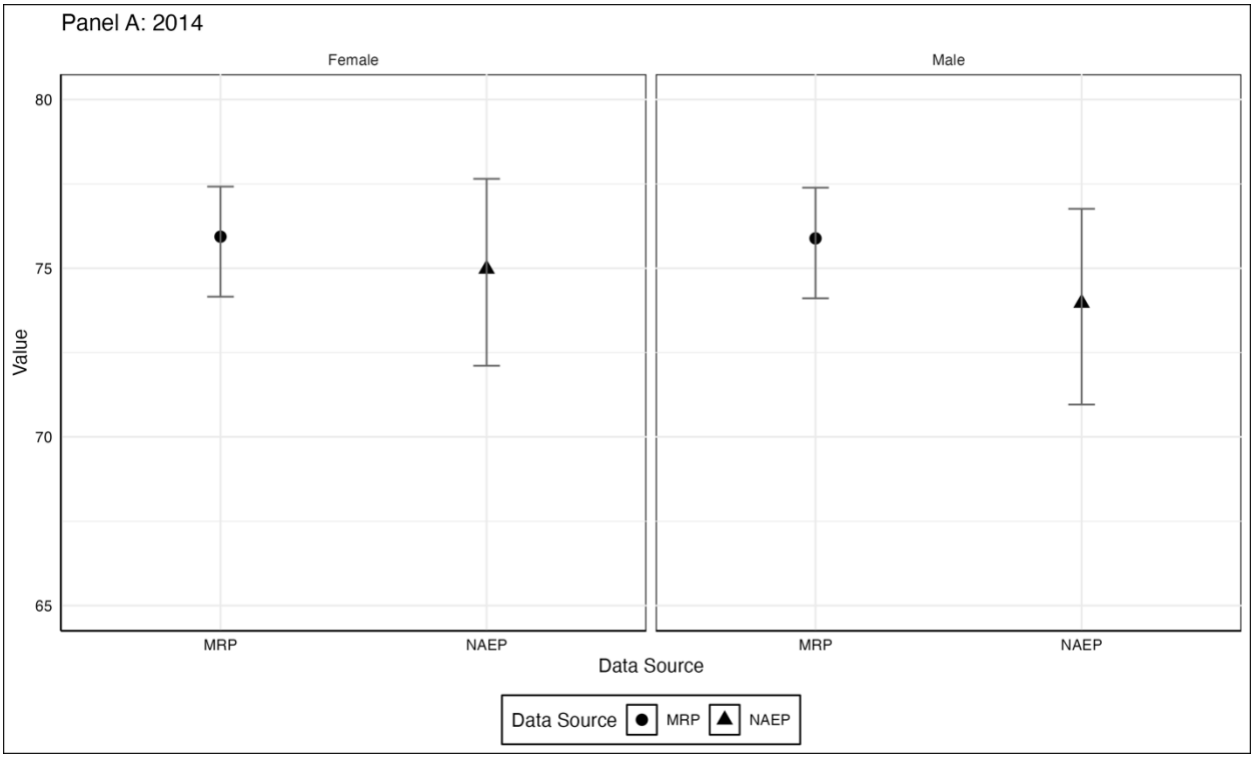
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.

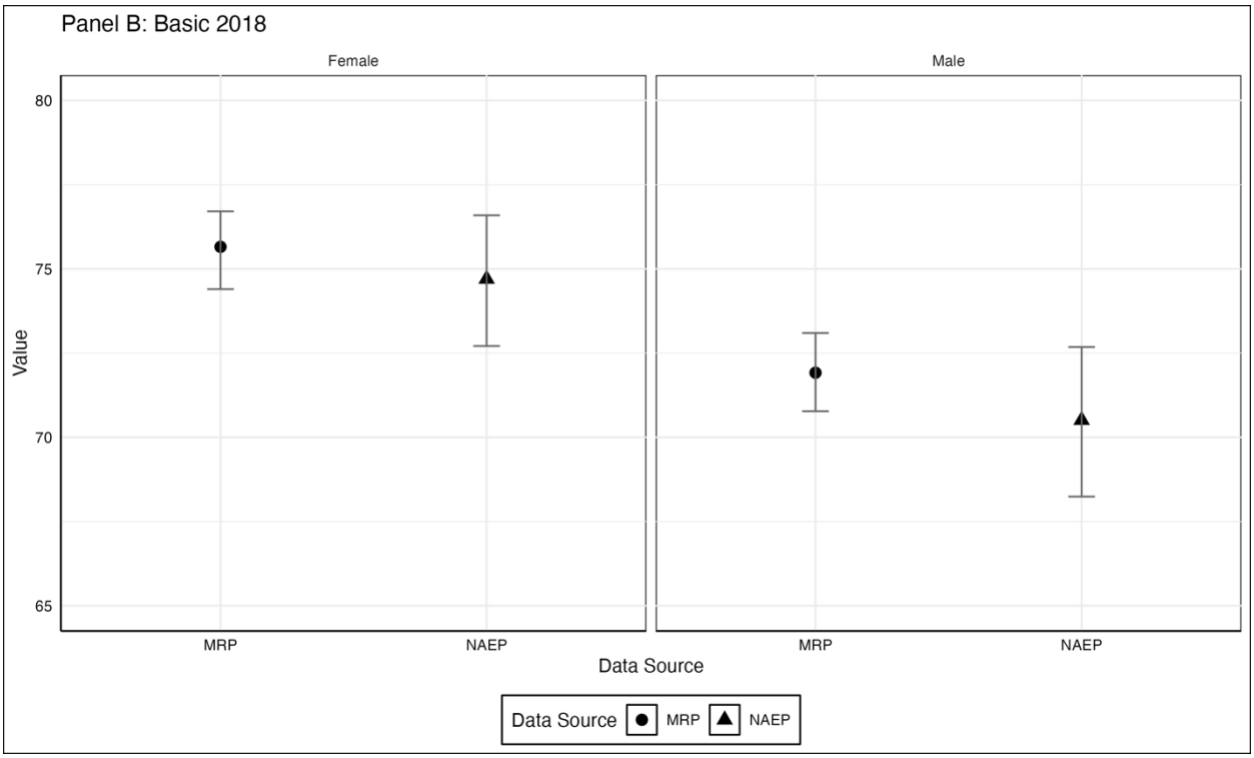
Figure A-1 National comparison of NAEP and MRP results for basic and above (2014 and 2018)



SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.

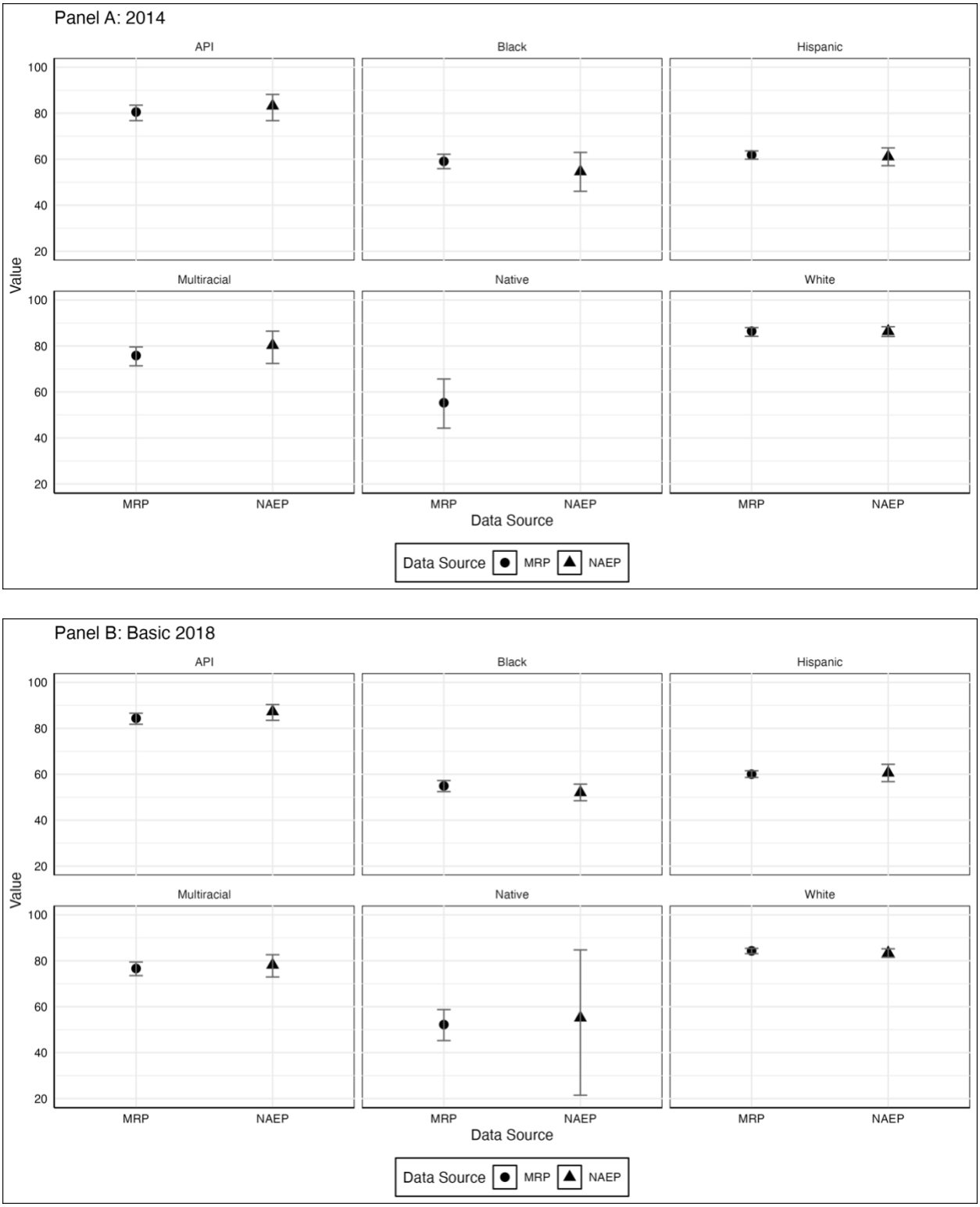
Figure A-2 National comparison of NAEP and MRP results by public status for basic and above (2014 and 2018)





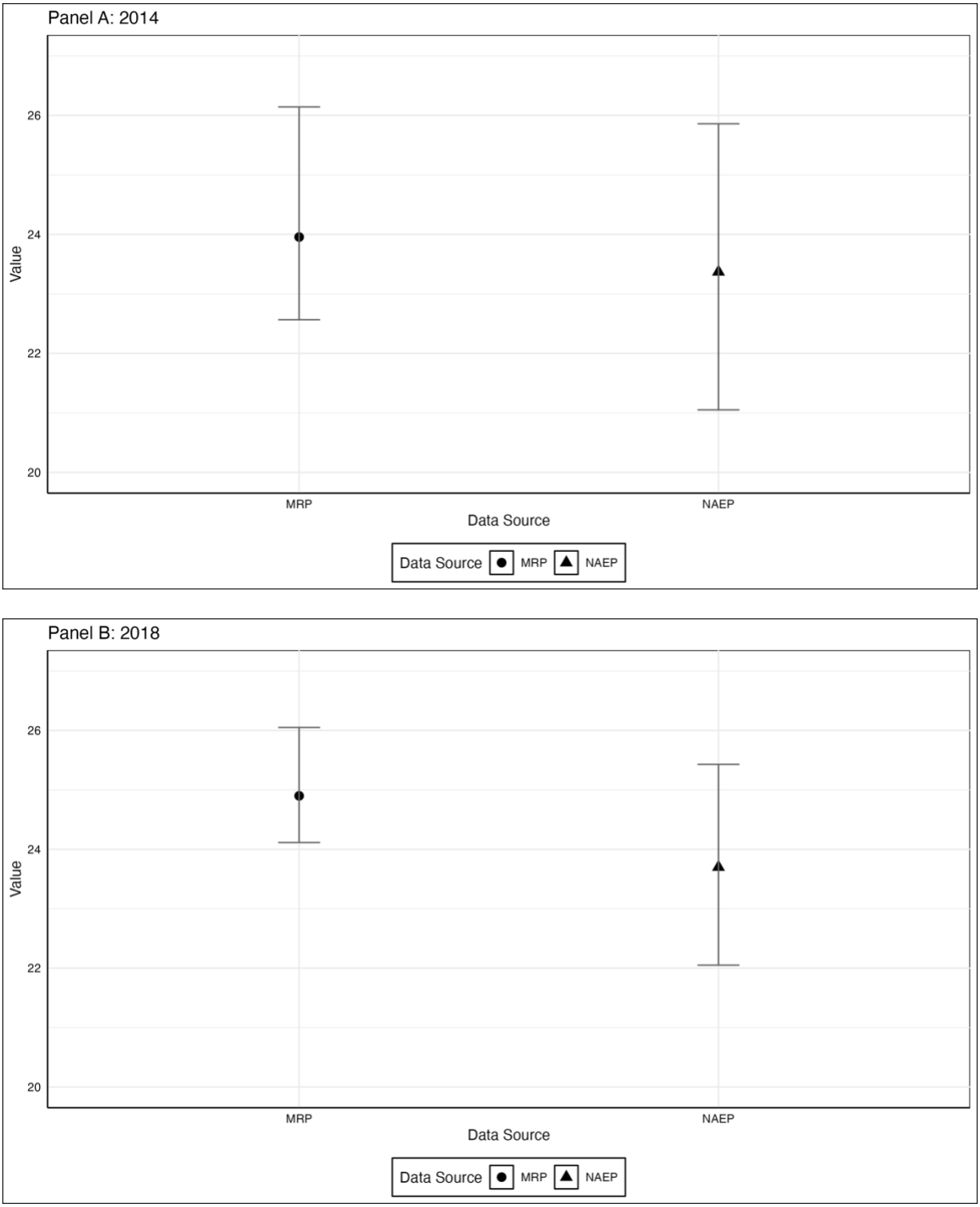
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.

Figure A-3 National comparison of NAEP and MRP results by gender for basic and above (2014 and 2018)



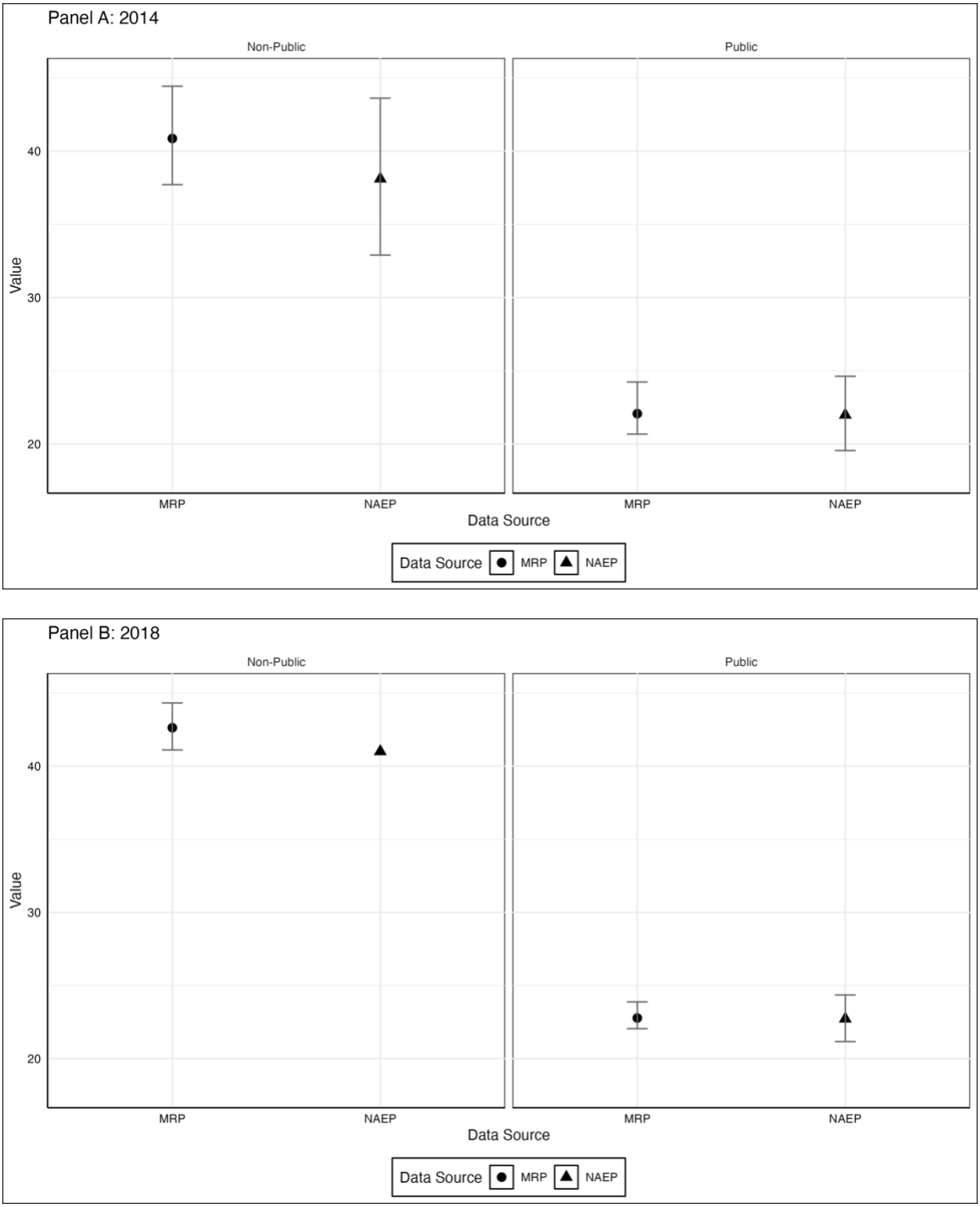
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.

Figure A-4 National comparison of NAEP and MRP results by race/ethnicity for basic and above (2014 and 2018)



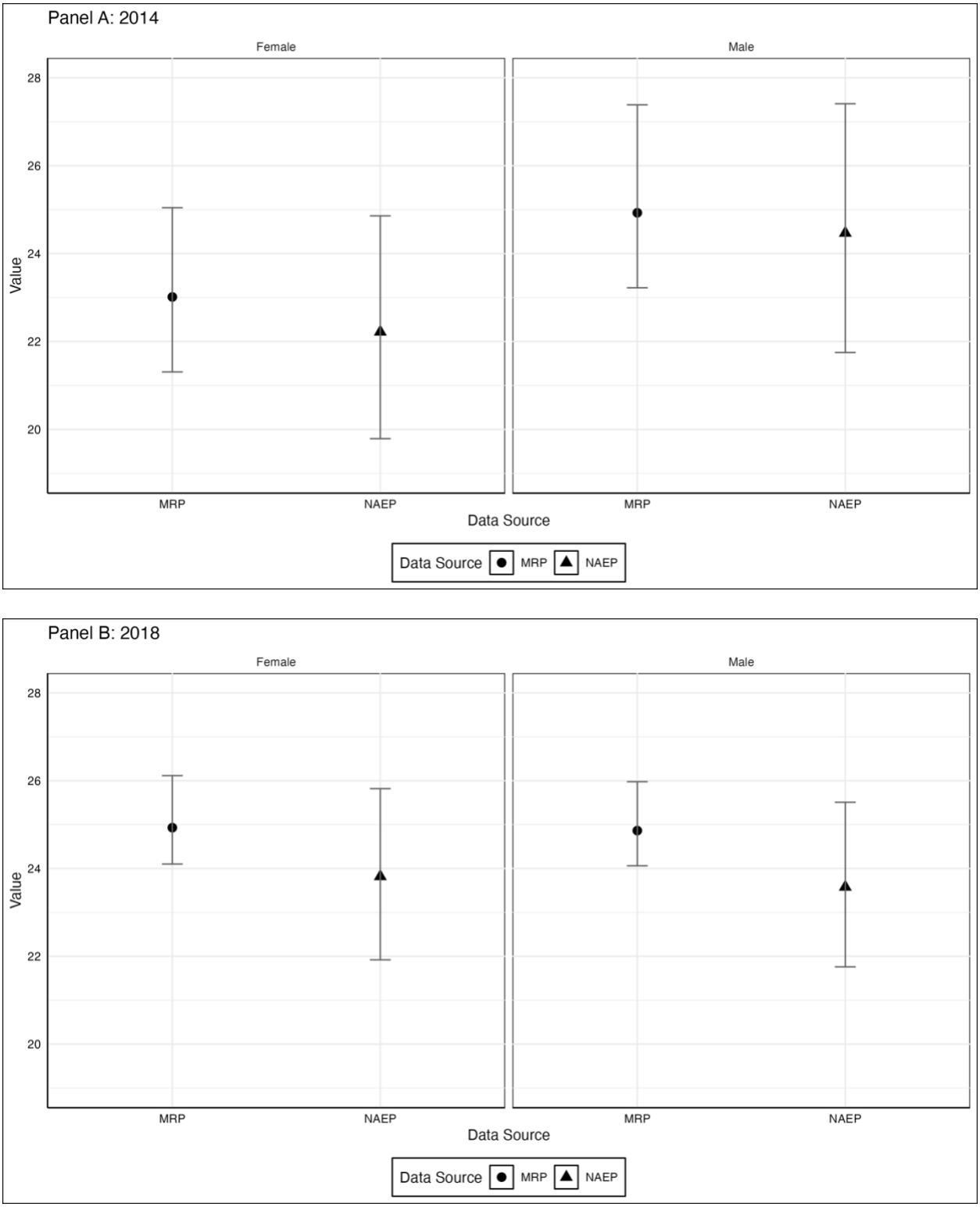
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.

Figure A-5 National comparison of NAEP and MRP results overall for plausible value 2 (2014 and 2018)



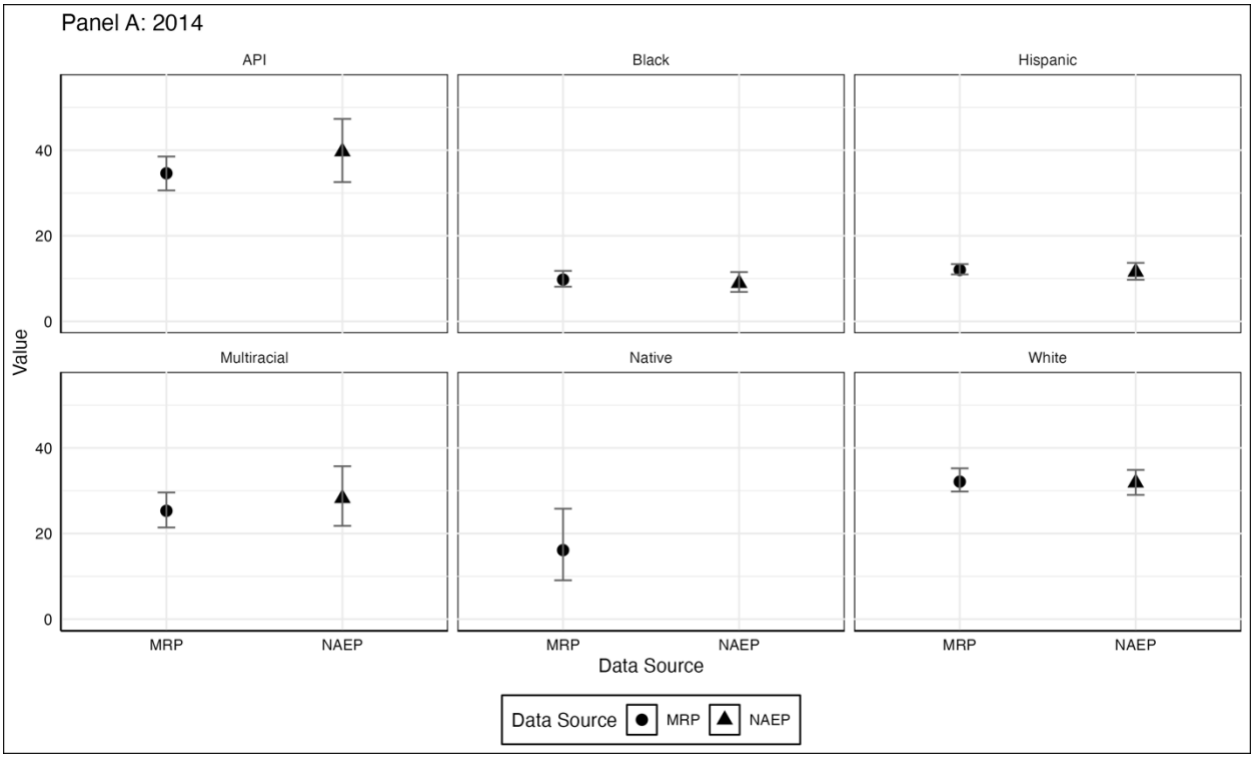
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.

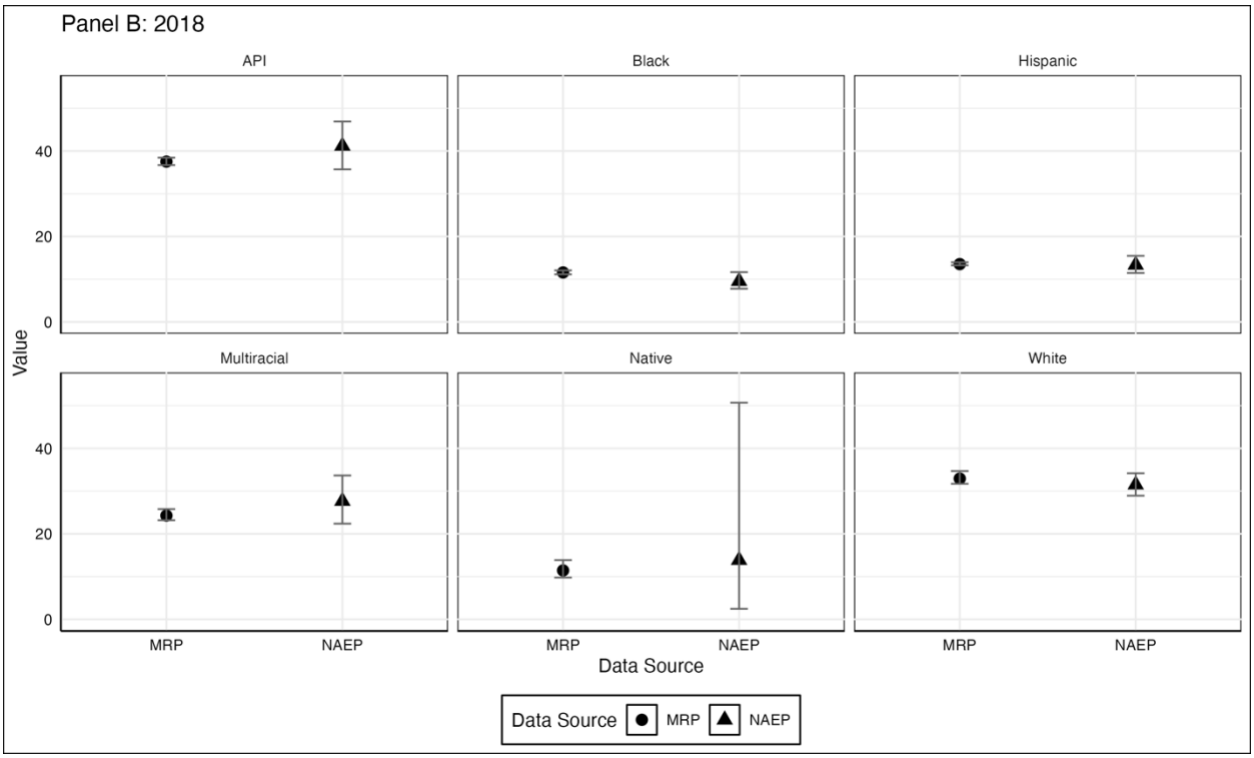
Figure A-6 National comparison of NAEP and MRP results by public status for plausible value 2 (2014 and 2018)



SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.

Figure A-7 National comparison of NAEP and MRP results by gender for plausible value 2 (2014 and 2018)





SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) 2014 and 2018 Civics Assessments.

Figure A-8 National comparison of NAEP and MRP results by race/ethnicity for plausible value 2 (2014 and 2018)