**Online Technical Appendix for What Works for Whom: Exploring the students, settings, and outcomes in What Works Clearinghouse study data**

This online technical appendix is presented as a Google Drive document to facilitate an anonymous review. The materials will ultimately be available on GitHub.

This technical appendix contains five parts:

**Part A** provides more information about the extraction and cleaning of the WWC study for this paper.

**Part B** includes information about the degree to which the study data are missing contextual information about the underlying studies by the WWC’s purpose of review.

**Part C** provides more information about the National Center for Education Statistics and Office of Civil Rights data used as a point of comparison to calculate descriptives for all schools and students in the U.S.

**Part D** includes additional graphs.

**Part E** includes additional references for the technical appendix.

Technical Appendix Part A: Data Extraction and Cleaning

Data extraction

WWC study data were extracted in November 2022 from <https://ies.ed.gov/ncee/wwc/StudyFindings>. The merged file was used because it contains detailed information on the findings in the studies that meet WWC standards. The study data were restricted to include only studies that were reviewed with WWC Handbooks Version 2.1 or higher. Version 2.1 of the WWC Handbook was the first to include pilot standards for studies with regression discontinuity and single case designs (WWC, 2011). Previous versions of the WWC Handbook included only standards for studies with randomized controlled trial and quasi-experimental designs (WWC, 2008).

Data deduplication

Because studies may have been reviewed by the WWC more than once, the study data were deduplicated. First, the review under the most recent review standards version was retained. Next, for the studies that were reviewed multiple times under the same standards version, reviews were prioritized based on the purpose of review: 1) intervention report; 2) single study review; 3) IES-funded study, grant competition, or other WWC report; and 4) practice guide or quick review. The remaining duplicates were manually checked. Either the duplicates were retained because they involved multiple contrasts and different interventions or different grade bands within the same study, or they were deduplicated retaining the most recent review.

Outcome domains

The WWC defines and names outcome domains in [review protocols](https://ies.ed.gov/ncee/wwc/handbooks#protocol), and domain names and definitions vary from protocol to protocol (WWC, 2022d). One challenge with using WWC data is aligning outcome domains over time. In some cases, outcome domains became more narrowly focused over time. For example, “geometry & measurement” were included in the general math domains in older review protocols, whereas “geometry & measurement” is a separate domain in the newer protocols. In other cases, the name of the outcome domain changed, but the scope of the outcome domain did not change substantially. For example, the problem behavior, external behavior, and student behavior domains contain similar constructs. For the purposes of this paper, domains that are conceptually similar were aligned when possible. In addition, for simplicity, multiple domains were combined when the domains could be considered as subdomains. For example, “physical science” and “life science” were combined in “science” and “writing quality” and “writing conventions” were combined in “writing”. The mappings from the original to revised outcome domains can be found in the [data](https://github.com/betsyjwolf/evidence-maps?tab=readme-ov-file).

Outcome domains were flagged as being relevant to specific student subpopulations and/or grade levels based on [review protocols](https://ies.ed.gov/ncee/wwc/handbooks#protocol). For example, domains listed in protocols focused on students with disabilities were coded as being relevant for students with disabilities. Moreover, all potentially relevant domains were preserved, even though the WWC may not have identified or reviewed research in all relevant domains. Therefore, the evidence maps capture both research reviewed by the WWC as well as research deemed relevant for the systematic reviews but not included in the WWC data, because either no research was identified or the identified research did not meet WWC standards.

Effect sizes

The WWC typically calculates the study effect sizes based on information provided in the study and obtained through author queries (WWC, 2022c). In some cases, however, the WWC can use the study-reported effect sizes (WWC, 2022c). Because it is not possible to determine from the public extract data file which effect size is the official one for the study review, the analysis used the WWC calculated effect size when it was available and the study-reported effect size otherwise. Findings with missing effect sizes were excluded from the analysis because the size and direction of the effect are unknown. Effect sizes were missing for 5% of the study findings.

Some effect sizes, particularly in behavioral domains, reflect reductions in negative behaviors and were recorded as being negative when they should have been recorded as positive. In these cases, negative effect sizes were changed to positive, indicated by a positive improvement index.

Subgroup and supplemental findings

In addition to the full sample (“main”) findings, the WWC may also review findings for specific subgroups or at additional time points, and these are referred to as “supplemental” findings. Although the WWC database contains a flag for subgroup findings, in reality, these findings could be either subgroup findings or other supplemental findings, such as findings from follow-up time periods (WWC, 2022e). These findings were dropped from the sample because the WWC generally reviews the full sample findings by default, and the students in the subgroup and supplemental findings will be represented in the full sample findings (WWC, 2022c). However, some subgroup findings may have been reviewed as main (non-subgroup) findings if relevant to the systematic review and specified in the review protocol (WWC, 2022d). For example, a systematic review focused on English learners may have reviewed the English learner subgroup as the main findings. In these cases, subgroup findings are analyzed in this report.

A limitation to the WWC study data is that characteristics of subgroup and supplemental findings are not coded systematically and are instead manually entered by WWC reviewers, which makes analyzing these findings more challenging. The WWC could increase the utility of subgroup and supplemental findings if these findings were more systematically coded.

Variables for intervention reports versus study reviews

The WWC study data contain “i\_variables,” which are populated for intervention reports, and “s\_variables,” which are populated for all study reviews, including study reviews included in intervention reports. However, in some cases, the s\_variables regarding demographic and contextual information are missing, but the i\_variables are populated. When there was only one study meeting WWC standards in the intervention report, the i\_variables are essentially the same as the s\_variables, In these cases, the analysis used the i\_variables to preserve the maximum information about the study reviews.

Grade levels

WWC study data contain an indicator at the study level for each grade level of students in the sample, ranging from preschool to post-secondary education. Grade level was missing for only 1% of studies initially, and the missing grade levels were manually imputed by examining the outcome measures, intervention names, and the WWC study review webpages. Grade levels were then combined into five bands:

* Early childhood: Grades PK and K
* Lower elementary: Grades 1–3 or grades K–3 with no PK
* Upper elementary: Grades 4–5 or grades 3–6 with no lower grades or older grades
* Middle school: Grades 6–8 or grades 6–9 with no older grades
* High school: Grades 9–12
* Post-secondary: All levels post high school

Seventy-seven percent of studies were identified as having one grade band, 18% as having two, 3% as having three (e.g., K–8), 2% as having 4 (e.g., K–12), and less than a percent as having all five (e.g., PK–12). Therefore, about 23% of studies were tagged as having more than one grade band.

Student versus teacher characteristics

Although the WWC typically reviews studies in which the study samples consist of students, in some instances, study samples can represent other populations, such as educators. The “f\_l1\_unit\_of\_analysis” variable is helpful to flag findings where the demographic variables describe educators, as opposed to students. In a few cases, the “f\_l1\_unit\_of\_analysis” variable may incorrectly indicate that the demographic characteristics are for the student samples when the variable instead reflects the characteristics of educators. The analysis did not use the demographic information or sample sizes when the value of the “f\_l1\_unit\_of\_analysis” variable indicated the information described educators as opposed to students. Only 1% of findings were flagged as reflecting educators, not students.

Race and ethnicity

The race and ethnicity variables for the student sample were modified to correct errors. First, race and ethnicity were coded as missing (as opposed to percentages of zero) when there was no race or ethnicity information, respectively, for the study.

Second, studies where the new total sum of the race categories exceeded 110% were manually checked against the study pages on the website. There were obvious errors in the race variables in a few studies, but when the percentages were plausible (because students can belong to more than one race category), the study data were preserved. Care was taken to avoiding changing the review data if possible, but data errors are possible because the race variables are manually entered in by WWC reviewers. After these revisions, only 1% of studies had percentages that summed to over 110% across the different race categories. Third, given the very small counts of studies that reported race separately for native American and Pacific Islander students, these categories were combined in an “indigenous race” category.

Socio-economic status

The WWC database contains a variable for the percent of students who qualified for free- and reduced-price lunch in the study. However, this variable may reflect other indicators of low socio-economic status, such as the percent eligible for Pell grants in studies of post-secondary interventions. Therefore, this variable represents low socio-economic status more generally.

Other student characteristics

The WWC data also include the percentages of female students, students with disabilities or individual education plans (IEPs), and English learners.

School setting variables

The WWC study database contains information about school types (public, private, parochial, and charter), urbanicities (urban, rural, suburban), and geographic location (states where studies were conducted). Studies may include settings with more than one school type, urbanicity, or geographic location. For each variable, such as school type - public, the variable is coded as 1 if the setting is included in the study, and the variable is left as missing if the setting is not included in the study.

Because the school type variables do not include a separate variable to reflect a postsecondary institution, some reviewers leave the school type variables blank when the studies include only postsecondary institutions. Other reviewers complete the school type variables to indicate whether the postsecondary institution was public, private, or parochial (charter is not applicable).

Geographic location, or states included in the studies, may be missing for several reasons. First, study authors may not report the specific state locations. Second, the studies may have been conducted outside of the U.S., although this is a small proportion of studies. Finally, WWC reviewers may not have recorded this information in the study database. When analyzing the states included in the studies, only states included in more than 50 studies were singled out in this report due to the large number of states. The remaining states were combined into another category.

Intervention type and delivery method variables

WWC study data also include intervention type (policy, practice, curriculum, supplement, school-level, teacher-level) and intervention delivery method (individual, school, small group, whole class). These variables are not mutually exclusive because study settings may include multiple schools, and interventions may contain multiple components. These variables were coded as missing only when no category was selected for the variable.

Evidence tiers

The WWC assigns evidence tiers to favorable findings in its study reviews according to the U.S. Department of Education’s regulatory evidence definitions. However, errors in the underlying data may result in a finding not being assigned an evidence tier in the public extract file. Additional findings were flagged as earning an evidence tier when each of the following conditions were true:

* The finding was statistically significant (“f\_is\_statistically\_signifcant” was True)
* The effect size was positive, or the WWC-calculated effect size was negative but the improvement index was positive (see Effect size section)
* There were no statistically significant and negative findings within the same outcome domain in the study review

Findings with the above characteristics that also had a multisite sample (s\_multisite was True) and a sample size of 350+ participants were flagged as earning a Tier 1 strong or Tier 2 moderate evidence tier depending on the rating for the finding (f\_finding\_rating). Findings with the above characteristics that did not have a multisite sample (s\_multisite was False) or had a sample fewer than 350 participated was flagged as earning a Tier 3 promising evidence tier. This process resulted in an additional 13 findings being flagged as Tier 1 strong evidence, 5 findings as Tier 2 moderate evidence, and 92 findings as Tier 3 promising evidence. Note that this paper does not distinguish among the three evidence tiers, but rather flags findings if they have any evidence tier.

Other changes to the data

Three studies were dropped from the sample because they either did not meet WWC standards or were ineligible for WWC review. The study design was missing from one study and was imputed based on information in the original report.

Technical Appendix Part B: Missingness by the Purpose of the Study Review

The paper previously discusses that contextual information may not have been recorded by WWC reviewers for all study reviews, and that the degree of missingness is likely related to the purpose of the WWC study review. Table B.1 below shows the percent of studies missing each contextual field by the WWC purpose of review. Studies reviewed for grant competitions have some of the highest percentages of missingness on student demographic variables, which is expected given that these reviews are time-sensitive and prioritize internal validity according to the WWC’s research standards. Conversely, studies reviewed for intervention reports have some of the lowest percentages of missingness on student demographic variables, which is also expected because intervention reports include descriptive information about the study samples. Still, student demographic variables may be missing for studies reviewed for intervention reports if the demographic information was not reported in the original studies or was not recorded by WWC reviewers.

**Table B.1. Missing contextual fields by purpose of review**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Purpose of review** | | | | | | |
|  | Grant competition  N (%) |  | ED or IES funded  N (%) | Single study review  N (%) | Other  N (%) | Practice guide  N (%) | Intervention report  N (%) |
| *Setting* |  |  |  |  |  |  |  |
| School type | 58 (37%) |  | 44 (30%) | 99 (46%) | 27 (84%) | 127 (49%) | 107 (42%) |
| Urbanicity | 51 (33%) |  | 54 (36%) | 76 (36%) | 20 (63%) | 98 (38%) | 81 (32%) |
| State | 39 (25%) |  | 34 (23%) | 70 (33%) | 17 (53%) | 118 (45%) | 58 (23%) |
| *Sample* |  |  |  |  |  |  |  |
| Race | 69 (44%) |  | 57 (38%) | 49 (23%) | 5 (16%) | 57 (22%) | 62 (25%) |
| Ethnicity | 78 (50%) |  | 62 (41%) | 72 (34%) | 5 (16%) | 69 (26%) | 105 (42%) |
| Low income | 85 (54%) |  | 90 (60%) | 125 (59%) | 21 (66%) | 132 (51%) | 136 (54%) |
| SWD | 98 (71%) |  | 98 (69%) | 113 (69%) | 4 (50%) | 104 (49%) | 140 (67%) |
| English learner | 88 (64%) |  | 71 (50%) | 82 (50%) | 5 (63%) | 81 (38%) | 129 (61%) |

*Notes*. The grant competition purpose of review includes reviews that were conducted as part of Department of Education grant competitions that required a strong or moderate level of evidence. The ED or IES funded purpose of review includes reviews that were conducted as part of IES performance measures or other studies flagged by the Department as being funded by the Department. The single study purpose of review includes other stand-alone study reviews, including “single study reviews” and “quick reviews”. The other purpose of review includes reviews for publications outside of intervention reports and practice guides.

School type includes public, private, charter, and parochial. Urbanicity includes urban, rural, and suburban. Study location is based on the state(s) where the study was conducted.

SWD=students with disabilities. The percentage missing for SWD and English learners was calculated for the PK–12 study sample only.

Technical Appendix Part C: Data on All Schools and Students in the U.S.

To provide a point of comparison for how the WWC study samples compare to all PK–12 public schools and students in the U.S., data were pulled from the National Center of Education Statistics’ (NCES) [ElSi table generator](http://nces.ed.gov/ccd/elsi/) from 2021–22, with the exception of the percentages of students with disabilities and English learners, which were pulled from the [Civil Rights Data Collection](https://civilrightsdata.ed.gov/data) (CRDC) from 2017–18.

The data for all PK–12 public schools and students were manipulated in similar ways to how the WWC study data were manipulated for comparability. First, all schools not including any grade in PK–12 were dropped. Then, duplicates were identified by retaining the entry with the largest student enrollment for schools with the same name, NCES school ID, state, and locale. Next, grade level bands were created, as described above, which means that schools could have more than one grade level band. Last, schools were categorized into one of five quintiles for each student characteristic: white, Black, Latino, Asian, indigenous (native American and pacific islander were combined due to very low counts), female, and low-income. There were missing data on these characteristics either because it was missing or because NCES suppressed the data because it did not meet its quality standards. The percent missing for each characteristic was as follows: low-income (19%), female (3%), Asian (13%), Latino (4%), Black (8%), white (3%), indigenous (18%).

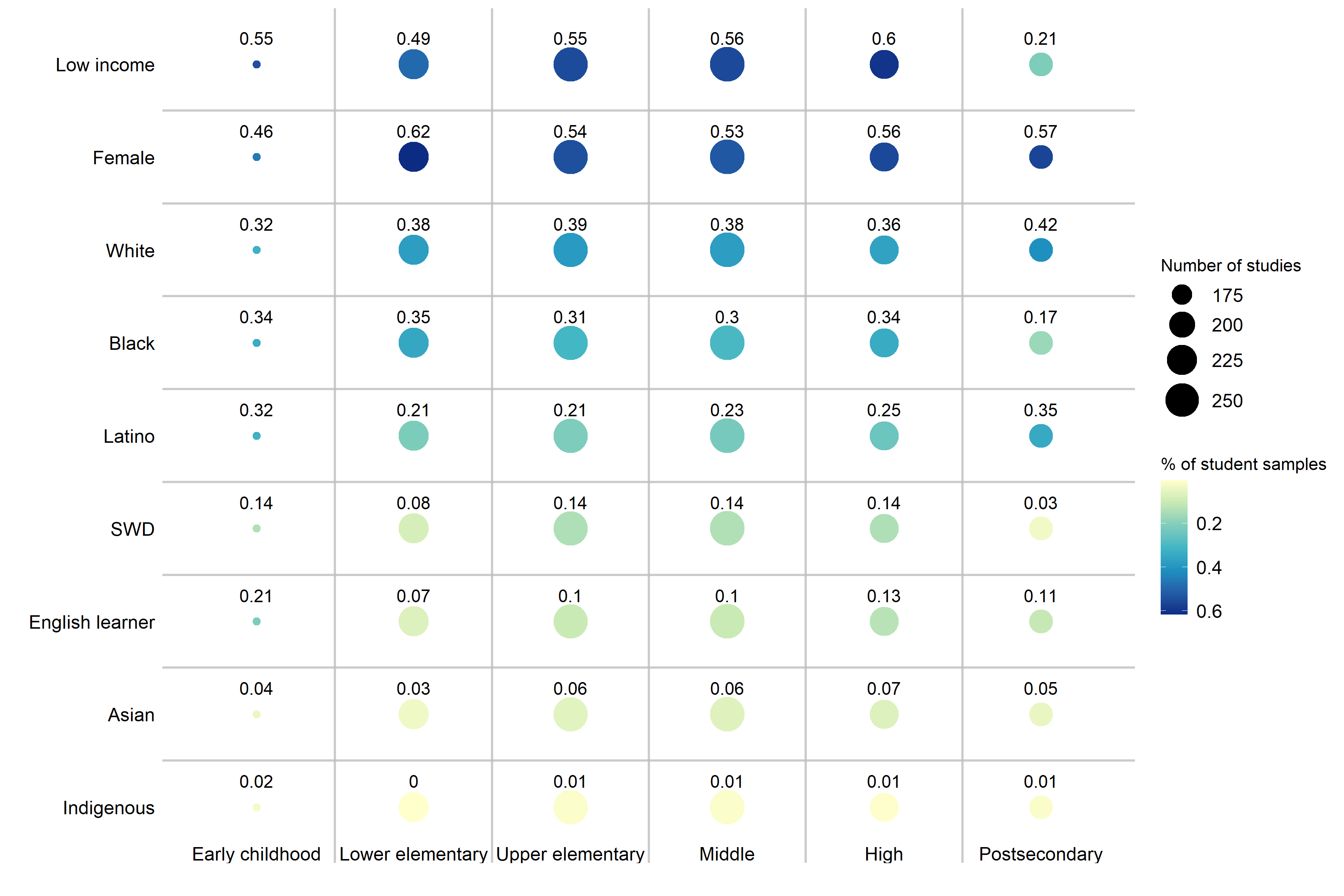
The percentages of students with disabilities (SWDs) and English learners were pulled from the CRDC from the 2017-18 school year because the total enrollment was not available in the most recent school year. The count of the total number of schools in the CRDC data was similar to that in the NCES data. The percentages of students with SWDs and English learners were calculated in the same way as described above.

Data for postsecondary students were obtained from NCES’s [Integrated Postsecondary Education Data System](https://nces.ed.gov/ipeds/datacenter/InstitutionList.aspx?goToReportId=5&sid=f783143e-dbd5-4ace-87f7-977ff3fe6574&rtid=5). All degree-granting public and private institutions with undergraduates were retained. Student race/ethnicity and gender came from the undergraduate enrollment estimates from fall 2022. The percentage of undergraduate students awarded Pell grants (which is a proxy for low-income status) came from the financial aid data from 2021-22. The percent of students with disabilities was estimated from the percentage of undergraduate students who were formally registered as students with disabilities in fall 2021.

When analyzing the study sample separately for PK–12 and postsecondary, studies with grade levels in both PK–12 and postsecondary were included in PK–12, as it is not uncommon to examine longer term outcomes when students matriculate into postsecondary. Studies where the only grade levels were postsecondary were included in the postsecondary sample.

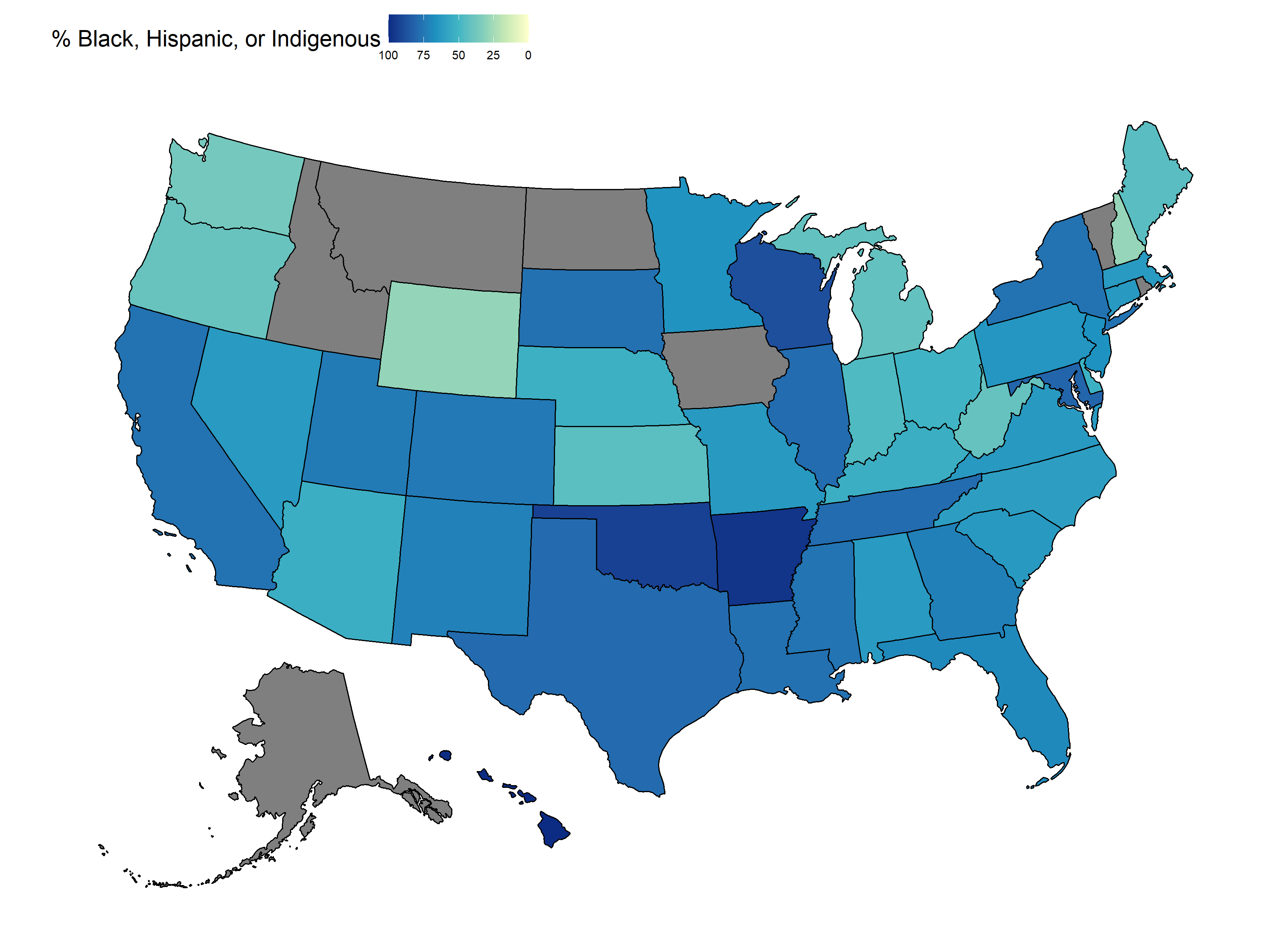
Technical Appendix Part D: Additional Graphs

**Figure D1. Evidence map of the characteristics of student samples by grade band for studies in the WWC database**



*Notes.* The size of the circles reflects the number of studies in the grade band (column). Studies may include more than one grade band so may be included in more than one column. The numbers above the circles and the color indicate the percentages of the student samples with the characteristic (rows). The percentages were calculated only for studies with non-missing data on the characteristic. Therefore, the percentages may reflect student samples in few studies.

**Figure D2. Percent of Black, Hispanic, or Indigenous students in WWC study samples for studies that included at least one urban site by state**



*Note.* Gray means that either there were no studies that included at least one urban site in the state or the sample demographic information was missing for the studies that included at least one urban site. This graph includes only studies in PK–12.

Part E: Additional References

What Works Clearinghouse. (2011). *Procedures and standards handbook (version 2.1).* What Works Clearinghouse, Institute of Education Sciences. U.S. Department of Education. <https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v2_1_standards_handbook.pdf>

What Works Clearinghouse. (2008). *Procedures and standards handbook (version 2.0).* What Works Clearinghouse, Institute of Education Sciences. U.S. Department of Education. <https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_v2_standards_handbook.pdf>