SC EDUCATION OVERSIGHT COMMITTEE
Reporting facts. Measuring change. Promoting progress.
PO Box 11867 | 227 Blatt Building
Columbia SC 29211 | WWW.SCEOC.ORG

AGENDA
Academic Standards and Assessments Subcommittee Meeting

Monday, September 21, 2026
10:00 A.M.
Room 433, Blatt Building

I. Welcome and Introductions ..... Dr. Patty Tate

II. Approval of Minutes of May 18, 2026 ..... Dr. Patty Tate

III. Action Items:

Education Scholarship Trust Fund (ESTF), School Year 2024-25:
A Formative Review of Data Quality and Transparency ..... Dana Yow

Alternative Instructional Methods Study,
School Year 2024-25 ..... Amina Asghar,
EOC Data Engineer

Psychometric Review of WIN Work-Ready Assessments .....
Dr. Matthew Madison,
Assistant Professor,
Department of Educational Psychology; &
Heather Bolinger, Research & Public Service Manager;
Dr. Stephanie A. Lai, Content Evaluator;
K-12 Assessment Solutions,
Mary Frances Early College of Education,
University of Georgia

IV. Adjournment

Academic Standards and Assessments Subcommittee

Dr. Patty Tate, Chair	Sidney Locke
Tammy Achziger	Laura McKinney
Rep. Terry Alexander	Melissa Pender
Rep. Bill Hager	Sen. Ross Turner
Barbara Hairfield	

SOUTH CAROLINA EDUCATION OVERSIGHT COMMITTEE

Academic Standards & Assessments Subcommittee

Minutes of the Meeting

May 18, 2026

Members Present (in-person or remote): Dr. Patty Tate, Barbara Hairfield, Melissa Pender, Tammy Achziger, Sidney Locke, Laura McKinney and Sen. Ross Turner.

Special Guests: Dr. Brian Newsome; Commander, Lt. Col Gerald Isabelle; Command Senior Chief Anthony Scott; Sr. Enlisted Advisor Mike Pruitt; and Test Control Officer Scott Burgess, Fort Jackson MEPS;

Heather Bolinger, Stephanie A. Lai, and Dr. Matthew Madison - K-12 Assessment Solutions University of Georgia,

Debbie Elmore, SCSBA, Ivy Coburn, SREB and Daniel Cammisa, SCDE

EOC Staff Present: Crystal Garcia, Tenell Felder, Gabrielle Fulton, Hope Johnson-Jones, Dr. Jenny May, Dr. Matthew Lavery, Dr. Rainey Knight, and Dana Yow.

Dr. Patty Tate opened the meeting and asked for a motion to approve the ASA subcommittee minutes from the March 16, 2026 meeting. The minutes were approved by motion, second, and unanimous vote. After the minutes were approved, Dr. Tate introduced Lt. Col. Gerald Isabelle, Commander of Fort Jackson MEPS, along with Command Senior Chief Anthony Scott, Senior Enlisted Advisor Mike Pruitt, and Test Control Officer Scott Burgess to provide an overview of the Armed Services Vocational Aptitude Battery (ASVAB) and discuss test security concerns. MEPS explained its dual mission of processing military applicants and supporting schools through administration of the ASVAB as a recognized Career and College Readiness (CCR) indicator. Fort Jackson MEPS reported reaching more than 22,000 high school students during the past year, the highest number among the nation's 65 Military Entrance Processing Stations. Presenters emphasized that the ASVAB serves not only as a military qualification test but also as a career exploration tool, providing students access to resources that help identify career interests, educational pathways, colleges, tuition information, admissions requirements, and job outlook data. A score of 31 or higher is recognized by the state as meeting the career readiness benchmark.

MEPS also clarified the difference between the official ASVAB Computer Adaptive Test (CAT), which is administered in a secure testing environment and can be used for CCR purposes, and the PiCAT (Pending Internet Computerized Adaptive Test), a recruiter-issued assessment that may be taken remotely and is not valid for career readiness reporting. Presenters described a new pilot program that allows trained school personnel to administer the ASVAB independently, reducing the need for MEPS staff to travel to schools while maintaining access to testing opportunities. Schools participating in the pilot receive training and oversight but are responsible for following established testing protocols and maintaining test integrity.

The primary focus of the presentation centered on concerns regarding test security and the validity of ASVAB scores. MEPS officials noted that while ACT and SAT scores have generally declined or remained flat in recent years, ASVAB career-ready scores have continued to increase. The most notable concern was the increase in perfect scores of 99, which rose from 1 student in 2023 to 7 in 2024, 23 in 2025, and 39 so far this year with additional testing sessions still scheduled. MEPS expressed concern that the absence of a lockdown browser and limited monitoring capabilities may allow some students to access calculators, internet resources, or artificial intelligence tools during testing, although no specific schools or districts were identified.

To address these concerns, MEPS reviewed current security measures, including the use of trained test administrators, military recruiters serving as additional proctors, detailed testing scripts, and training materials that explicitly prohibit calculators, internet searches, and AI tools during testing. Presenters emphasized the challenges of monitoring large testing groups and discussed potential solutions such as implementing lockdown browsers, utilizing screen-monitoring software, strengthening school-level oversight, and providing additional training for school staff. The presentation concluded with a discussion on how the state, districts, and MEPS can work together to improve test security and preserve the integrity of ASVAB scores used for both career readiness reporting and military qualification purposes.

Laura McKinney asked how the ASVAB is used beyond military recruitment, how it supports career exploration, and how it differs from other aptitude assessments that focus on personal interests rather than right and wrong answers. She also questioned whether the ASVAB could be used more broadly within the state's Career and College Readiness (CCR) framework, noting that some students avoid taking the assessment because of its military association and that schools may not consistently provide follow-up guidance on interpreting career exploration results. She shared a personal example in which her daughter took the ASVAB but did not receive information about the career exploration component until much later.

In response, Lt. Col. Isabelle explained that the ASVAB is maintained by the Department of Defense and consists of multiple aptitude-based subtests designed to measure a student's potential to learn and succeed in different career fields rather than simply assessing what they have already learned. He noted that the assessment includes areas such as reading, mathematics, mechanical knowledge, and electrical concepts, although a condensed version is administered in schools. He further explained that revisions to the ASVAB are currently being considered to better align with modern workforce needs, including potential additions such as cybersecurity-related content and updates to language and comprehension sections. According to Lt. Col. Isabelle, the purpose of the assessment is to identify strengths and aptitudes that may indicate suitability for various career paths, whether in logistics, mechanics, technical fields, or other occupations.

Lt. Col. Isabelle also emphasized the value of the Career Exploration Program (CEP), describing it as a robust career-planning tool that helps students connect their ASVAB results to potential occupations, educational pathways, and career opportunities. He acknowledged that more work is needed to ensure students receive meaningful post-test interpretation and guidance. Over the past two years, MEPS has trained more than 100 school counselors, JROTC instructors, and other personnel to conduct post-test interpretation sessions and help students understand their results. He stated that his goal is for every ASVAB administration to be followed by a structured interpretation session so students can fully utilize the career exploration resources available through the program, regardless of whether they have an interest in military service.

Melissa Pender asked whether there is a significant difference between the ASVAB administered to students in 11th or 12th grade at their schools and the version administered at an official MEPS testing site. In response, MEPS officials explained that the assessments are essentially the same test, with only different versions being used. In fact, the school-administered version currently in use was previously used at Fort Jackson MEPS for enlistment testing. They noted that all versions have been standardized and normalized to ensure comparable results. The primary distinction is that students who take the ASVAB through their school gain access to the Career Exploration Program (CEP) and related career-planning resources, while individuals who take the test solely for enlistment purposes at a MEPS location do not receive that career exploration component. MEPS emphasized that this added career guidance and exploration benefit is a key advantage of offering the ASVAB in schools.

Dana Yow asked how many states currently use the ASVAB as part of their College and Career Readiness (CCR) accountability systems. In response, Lt. Col. Isabelle stated that only about 10 states explicitly recognize the ASVAB as a CCR indicator, making South Carolina part of a relatively small group of states utilizing the assessment in this way. He emphasized that, beyond representing Fort Jackson MEPS, he also has an interest in supporting South Carolina's education system and ensuring that CCR indicators remain accurate and credible. He noted that maintaining the integrity of ASVAB results is important because the data are used to inform educational accountability measures and should provide a truthful representation of student readiness.

Next, Dr. Tate introduced the committee's first action item, the Technical Review of SC READY English Language Arts (ELA) and the English 2 End-of-Course (EOC) assessment. The EOC is required to periodically review state assessments to evaluate their alignment with South Carolina's academic standards, level of difficulty, validity, and ability to differentiate student achievement levels. As part of this process, the EOC develops recommendations for any necessary assessment changes. The completed review must be provided to the State Board of Education, the South Carolina Department of Education, the Governor, the Senate Education Committee, and the House Education and Public Works Committee. Following receipt of the review, the Department of Education is required to report back to the EOC within one month regarding any assessment changes made in response to the recommendations. Dr. Tate welcomed Heather Bolinger, Stephanie A. Lai, and Dr. Matthew Madison from K-12 Assessment Solutions at the University of Georgia, who joined the meeting virtually to present the findings of the review conducted under contract with the EOC.

The review was guided by the Standards for Educational and Psychological Testing and included analyses of test blueprints, content alignment, Depth of Knowledge (DOK), psychometric properties, and performance-level classifications. Subject matter experts reviewed all assessment items for alignment to state standards and cognitive rigor, while psychometric analyses examined reliability, dimensionality, item functioning, and fairness across student groups. Overall, the review found that the assessments are well aligned with the 2024 South Carolina College- and Career-Ready English Language Arts Standards and provide valid and reliable measures of student achievement.

Reviewers reported that the test blueprints for SC READY ELA and EOCEP English 2 appropriately represent most standards and emphasize DOK Level 2 skills, with increasing proportions of DOK Level 3 items in higher grades. While overall alignment between blueprint

specifications and operational items was found to be acceptable, a small number of standards at each grade level are not currently assessed. The review recommended examining these unassessed standards to determine whether future operational forms should include them. Item-level analyses found that most items aligned appropriately with their intended standards and DOK levels, although a limited number of items—two in Grade 3 and one in Grade 7—may require further review for standard alignment. Additionally, a small number of items across grade levels could benefit from updates to DOK classifications or minor revisions to strengthen content validity and rigor.

The psychometric evaluation found strong evidence supporting the quality and fairness of the assessments. Classical Test Theory and Rasch analyses indicated that items functioned well, covered an appropriate range of difficulty levels, and effectively differentiated among students. Principal Component Analysis supported the conclusion that each assessment measures a single dominant construct, while reliability estimates were strong across grades, test forms, and student subgroups. Differential Item Functioning analyses showed minimal evidence of bias, with 98 percent of item comparisons demonstrating no or negligible differential functioning across demographic and administration groups. One recommendation was to consider adding more difficult items to the Grade 3 ELA assessment to improve measurement precision for higher-performing students.

Regarding performance-level classifications, reviewers found that the current standard-setting process and cut scores provide reliable distinctions among broad performance categories. While classification agreement naturally decreases when differentiating among more detailed performance levels, the results remain consistent with industry expectations. The review recommended that performance-level results continue to be used alongside other indicators of student achievement and suggested exploring future psychometric approaches, such as cognitive diagnosis models, that may further strengthen classification accuracy and reporting. Overall, the review concluded that the assessments demonstrate strong technical quality while identifying targeted opportunities for continued improvement.

Barbara Hairfield asked which English Language Arts standards were not included on the assessments. In response, the presenters explained that there are approximately 25 ELA standards, and the relatively small number of standards not assessed on a given operational form are distributed across multiple reporting domains rather than concentrated within a single area. They noted that roughly 80 percent of the omitted standards span different reporting categories, ensuring that no one domain is disproportionately affected.

Daniel Cammisa, SCDE Team Lead for Test Development and Administration in the Office of Assessment and Standards, further clarified that assessment development is driven by the test blueprint's required number of items within each reporting category rather than by ensuring every individual standard is assessed during every administration. He explained that the Department's primary commitment is to maintain the prescribed item distribution across reporting categories. When certain standards are not included on a specific operational form, the Department follows a long-term rotation plan to ensure those standards are assessed in future administrations. As a result, while every standard may not appear on a test each year, all standards are addressed over time through a planned cycle of assessment coverage.

A motion was made to accept the recommendations, and it was accepted.

Next, Dr. Matthew Lavery provided an informational update on accountability system improvements and introduced the proposed revisions to the Multilingual Learners' Progress (MLP) Indicator. He noted that 2025 has been a significant year for accountability in South Carolina, marked by the second legislatively required cyclical review of the accountability system under Section 59-18-910. The review involved a 21-member Accountability Advisory Committee and four focused stakeholder convenings addressing public expectations, the high school experience, multilingual learners' progress, and awarding performance. These efforts resulted in 19 recommendations for improving the accountability system. The EOC and SCDE conducted a systematic review that led to enhanced cross-functional collaboration, improved documentation, more resilient workflows, and strengthened accountability reporting processes.

Dr. Lavery explained that a new cross-functional team structure has been established to oversee accountability and School Report Card development. The team's mission is to ensure that South Carolina's accountability system clearly defines excellence, measures meaningful student outcomes, and provides families, educators, and policymakers with understandable and actionable information. Three primary goals guide this work: improving accountability measures, enhancing reporting and analytics, and strengthening data collection and quality. The effort has resulted in a comprehensive understanding of the data elements, timelines, responsibilities, and dependencies involved in producing School Report Cards. Dr. Lavery emphasized that accountability decisions must be made well in advance to allow sufficient time for policy updates, federal approvals, data system modifications, stakeholder communication, and implementation. He also highlighted ongoing efforts related to Recommendation 15 from the cyclical review, including investments in dashboards, data visualizations, GPS Navigator tools, report card

calculations, and integration of school climate survey data into the state's Ed-Fi infrastructure to improve transparency and support continuous improvement.

Turning to the action item, Dr. Lavery explained that the Multilingual Learners' Progress Indicator is required under the Every Student Succeeds Act (ESSA), which mandates that states measure English language learners' progress toward English proficiency. South Carolina currently uses the WIDA ACCESS assessment, defines proficiency as a composite score of 4.4, and establishes a five-year timeline for achieving proficiency. Feedback gathered during the multilingual learners focused convening identified several strengths of the current indicator, including the ability to know targets in advance and the appropriateness of the five-year proficiency timeline. However, participants also identified challenges, including the highly mobile nature of multilingual learners, complexities in record keeping and coding, and concerns that the current linear growth expectations do not accurately reflect actual patterns of language acquisition.

Based on these findings, staff proposed several revisions to the indicator. The revised model would establish annual growth targets based on a student's most recent English proficiency score rather than a multi-year progression table and would reflect the nonlinear nature of language acquisition, recognizing that students typically demonstrate more rapid growth early in the learning process and slower growth as they approach proficiency. The proposal would also separate students into two groups: Multilingual Learners Included in the Indicator (ML SII), consisting of students continuously enrolled from Day 45 through the first day of the testing window, and Multilingual Learners in the Test Participation Rate (ML TPR), consisting of students enrolled for at least 20 instructional days during the testing window. This distinction is intended to ensure schools are held accountable only for students whose academic growth they have had a reasonable opportunity to influence while still maintaining expectations for test participation. Under the proposed model, indicator points would be based on students included in the accountability calculation, while participation rates would be used to adjust earned points. Staff recommended adopting annual growth targets based on the most recent assessment results, aligning targets with observed English language acquisition patterns in South Carolina and nationally, and decoupling growth calculations from participation calculations to better reflect school responsibility and improve fairness in accountability determinations.

A motion was made to accept the recommendations. The motion was seconded and carried.

Next, Dr. Matthew Lavery presented a proposed change to the High School Student Success Indicator that would allow the High School Employability Credential to be recognized within the

state's “On-Track” measure. The recommendation emerged from the 2025 cyclical review of accountability and was reinforced by a South Carolina School Boards Association resolution requesting that students pursuing the employability credential be included in the measure. Dr. Lavery explained that while the credential is already recognized as a Career-Ready indicator and included in the Five-Year Student Success Rate, it is not currently reflected in the On-Track to Graduate metrics because the credential pathway does not require the same English and mathematics coursework needed for a regular diploma.

To provide context, Dr. Lavery reviewed the state's accountability framework and noted that research shows students who are on track at the end of ninth grade are nearly four times more likely to graduate on time than students who are off track. South Carolina's current on-track measures require students to earn specific numbers of credits, including English and mathematics credits, each year of high school. However, because students pursuing the employability credential follow a different pathway, they may be successfully progressing toward their intended outcome while still being classified as off track under the current system. He noted that approximately 57 districts award the employability credential each year and that credential earners demonstrate positive workforce outcomes, earning approximately 8 percent more in wages than comparable non-graduates without the credential.

Staff recommended renaming the existing “On-Track to Graduate” metrics as simply “On-Track” metrics and allowing a limited percentage of students pursuing the employability credential to be counted as on track. To qualify, students would need to be identified in PowerSchool as seeking the credential and successfully complete appropriate English and mathematics coursework each year. Schools would receive allowable on-track credential counts proportional to the number of credential-seeking students they serve. Dr. Lavery emphasized that the proposal is intended to better recognize valid pathways to postsecondary success while maintaining the regular diploma as the primary goal and preserving the integrity of the accountability system.

A motion was made to accept the recommendations. The motion was seconded and carried.

Next, Dr. Matthew Lavery presented a proposal to include the South Carolina Seal of Biliteracy as an additional indicator within the College and Career Readiness (CCR) metric. He explained that students earn the Seal of Biliteracy by maintaining at least a 3.0 GPA in required English Language Arts courses (or achieving English proficiency on ACCESS for multilingual learners) and demonstrating intermediate-level proficiency or higher in a language other than English through a state-approved assessment. Participation in the program has steadily increased

across South Carolina, and its inclusion in CCR was recommended during the 2025 cyclical review of accountability as Recommendation 8. Dr. Lavery noted that the seal aligns closely with South Carolina's Profile of the Graduate by promoting communication skills, global perspectives, and world-class knowledge.

To evaluate the proposal, EOC staff analyzed postsecondary outcomes for students earning the Seal of Biliteracy. The analysis found that while many seal recipients already meet other college-readiness criteria, students who earned the seal were significantly more likely to enroll in a postsecondary institution within one year of high school graduation than students who did not earn the seal. Approximately 75 percent of seal recipients enrolled in college compared to about 57 percent of graduates without the seal. Based on these findings, staff concluded that earning the Seal of Biliteracy is a strong predictor of college enrollment and therefore supports its inclusion as a college-readiness indicator.

Staff recommended adding the South Carolina Seal of Biliteracy as an additional pathway for demonstrating college readiness within the CCR indicator and treating Bronze, Silver, and Gold seal levels equally for accountability purposes. If approved, the change would be incorporated into the 2027 Accountability Manual and reflected in 2027 School Report Cards.

Barbara Hairfield expressed reservations about including the Seal of Biliteracy as a College and Career Readiness indicator. She questioned whether earning the seal is truly an indicator of a student's readiness for college or their likelihood of succeeding with college-level coursework. Hairfield noted that while the Seal of Biliteracy demonstrates language proficiency, she was uncertain that it serves as a strong measure of college readiness and raised concerns about whether it should be used as a predictor of postsecondary success within the accountability system. Ms. Hairfield questioned whether enrollment alone is sufficient evidence that the seal predicts student success in college and suggested that persistence in postsecondary education may be a more meaningful indicator. In response, Dr. Matthew Lavery explained that the EOC's approach is to evaluate college-readiness indicators by examining whether they predict college enrollment and, ultimately, college persistence, just as career-readiness indicators are evaluated through employment and earnings outcomes. He acknowledged that persistence is a stronger measure of postsecondary success but noted that sufficient longitudinal data are not yet available to determine whether Seal of Biliteracy recipients persist in college at higher rates. He emphasized that the indicator would continue to be evaluated over time as additional outcome data become available.

Dr. Lavery further explained that the current analysis focuses on whether students enroll in a postsecondary institution within one year of high school graduation. Using National Student Clearinghouse data matched with accountability records, the EOC can track enrollment patterns for students regardless of whether they graduate early, on time, or later than expected. Dana Yow suggested that persistence into the sophomore year of college might provide stronger evidence of readiness and asked when such data would become available. Dr. Lavery responded that while reliable persistence data are not yet available for recent graduates, future data collections will allow the EOC to analyze whether students earning the Seal of Biliteracy remain enrolled beyond their first year of college and thereby provide a stronger measure of college success.

A motion was made to accept the recommendations. The motion was seconded and carried.

Dr. Matthew Lavery presented a proposed revision to the process used for approving dual enrollment courses that count toward the College and Career Readiness (CCR) indicator. He explained that students currently earn a college-ready designation by successfully completing at least six credit hours of approved dual enrollment coursework with a grade of C or higher. While the EOC established approval criteria in 2017, the approval process itself was not formally defined until 2023, when a review process involving the South Carolina Department of Education, the EOC, and the Commission on Higher Education (CHE) was implemented. Through that review, the list of approved courses was reduced from 702 to 473, largely due to the removal of courses that were no longer being offered or taken by students.

Dr. Lavery noted that findings from the 2023 South Carolina Transfer and Articulation Action Plan raised additional questions about the effectiveness of dual enrollment as a college-readiness measure. The report found that many courses accepted for transfer were counted only as elective credit rather than fulfilling general education or degree requirements, and a significant number of transfer students reported having to retake coursework after enrolling in college. These findings prompted a broader discussion about whether approved dual enrollment courses are consistently supporting successful postsecondary outcomes and whether current approval procedures clearly define how and when course approvals should apply.

Staff recommended revising the approval process to clarify that newly approved courses would apply prospectively rather than retroactively, strengthen the emphasis on transferability to degree programs and fulfillment of general education or program requirements, and establish a formal review process every five years. The periodic review would evaluate whether approved

courses are regularly offered and taken by students and whether students who successfully complete those courses demonstrate higher rates of postsecondary enrollment and persistence. Dr. Lavery emphasized that the goal is to ensure that dual enrollment courses included in the CCR indicator remain meaningful predictors of college readiness and student success.

A motion was made to accept the recommendations. The motion was seconded and carried.

Laura McKinney sought clarification on whether the purpose of dual enrollment within the College and Career Readiness (CCR) indicator is strictly to measure college readiness rather than career readiness, noting that some approved courses appear to be Career and Technical Education (CTE) related. Dr. Matthew Lavery responded that this distinction was intentional and stems from the EOC's 2017 decision to recognize dual enrollment courses in English, mathematics, science, engineering, technology, and social studies as indicators of college readiness. Ms. McKinney also questioned why an additional review process involving the Commission on Higher Education (CHE) was necessary if the courses are already approved and taught by postsecondary institutions for college credit. Ms. Yow explained that when the course list was reviewed, many of the 702 previously approved courses did not meet the EOC's established criteria. The review process was therefore intended both to align approved courses with the original policy and to ensure that courses counted toward CCR are genuinely transferable and meaningful indicators of college readiness. She noted that, unlike Advanced Placement courses, dual enrollment does not have a standardized external assessment that independently verifies student mastery of the content, making transferability and postsecondary applicability important considerations. Ms. McKinney expressed concern that challenges with dual enrollment may stem less from the courses themselves and more from differences in implementation across districts. She observed that dual enrollment opportunities vary considerably, with some courses taught in high school classrooms, others offered online, and some requiring students to attend courses on college campuses. She suggested that these variations may influence student experiences and outcomes, raising broader questions about consistency and access within dual enrollment programs.

A motion was made to accept the recommendations. The motion was seconded and carried.

Next, Dana Yow presented an information item on the ongoing Industry Certification Review process. Four priority TAC meetings have been scheduled for Advanced Manufacturing, Supply Chain, Logistics and Transportation, Marketing and Sales, and Public Service and Safety. These committees include career and technical educators, postsecondary representatives, and

business and industry partners who will evaluate certifications and make recommendations regarding their inclusion as career-ready indicators. Ms. Yow noted that recommendations from these groups are expected to be presented at the full EOC committee meeting on June 15. Ms. Yow also provided an update on the EOC retreat, announcing that it will be held in Camden from August 9-10. She encouraged members to save the date and noted that lodging accommodations have been arranged at the SpringHill Suites by Marriott in Camden, with additional details to be distributed by email. The retreat will take place at the Revolutionary War Visitor Center, and members will be hosted for dinner at Sidney Locke's home on the first evening.

After this update, the meeting was adjourned.

EDUCATION OVERSIGHT COMMITTEE

DATE: September 21, 2026

SUBCOMMITTEE:
Academic Standards & Assessments Subcommittee

ACTION ITEM:
ESTF Year 1 (School Year 2024-25) Formative Review

PURPOSE/AUTHORITY
§SECTION 59-8-150 Requirements for education service providers, department, and Education Oversight Committee.

(E) The Education Oversight Committee shall:

(1) comply with all student privacy laws;

(2) report on and publish associated learning gains and graduation rates to the public by means of a state website with data aggregated by grade level, gender, family income level, number of years participating in the program, and race and a report for any participating school if at least fifty-one percent of the total enrolled students in the private school participated in the program in the prior school year or if there are at least thirty participating students who have scores for tests administered. If the Education Oversight Committee determines that the thirty participating-student cell size may be reduced without disclosing the personally identifiable information of a participating student, the Education Oversight Committee may reduce the participating-student cell size, but the cell size may not be reduced to fewer than ten participating students;

(3) evaluate and report the academic performance of scholarship students compared to similar public school populations; and

(4) collaborate with the department to develop and administer an annual parental satisfaction survey for all parents of scholarship students on issues relevant to the program, to include effectiveness and length of the program participation. Results of this survey must be provided to the General Assembly by December thirty-first of each year.

CRITICAL FACTS
The EOC staff sent the report to the SCDE staff on Wednesday, September 9, 2026 in advance of publication in the packet. A response memo was sent back to the EOC on September 11; that memo is included in the subcommittee meeting packet and is publicly posted on September 14.

TIMELINE/REVIEW PROCESS
- **September 21, 2026:** anticipated ASA subcommittee approval of first-year report of Year 1 formative report.
- **October 26, 2026:** Year 1 report goes to full EOC
- **December 14, 2026:** Full EOC to receive report of assessment results from ESTF Years 1 and 2.

ECONOMIC IMPACT FOR EOC
Cost: no impact

ACTION REQUEST

☒ **For approval** ☐ **For information**

ACTION TAKEN

☐ **Approved** ☐ **Amended**
☐ **Not Approved** ☐ **Action deferred (explain)**

SC EDUCATION OVERSIGHT COMMITTEE

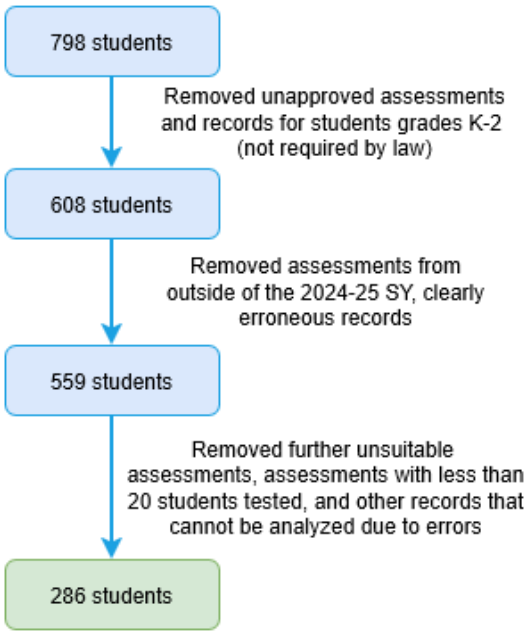
Reporting facts. Measuring change. Promoting progress.

Education Scholarship Trust Fund, Year One (School Year 2024-25)

A Formative Review of Data Quality and Transparency

Submitted to the EOC Academic Standards Subcommittee, September 21, 2026

The Education Scholarship Trust Fund (ESTF) completed its first operational year in 2024-25. Under Act 8 of 2023, as amended by Act 11 of 2025, the Education Oversight Committee (EOC) is charged with reporting the learning gains and graduation rates of participating students and comparing their academic performance to similar students in South Carolina's public schools.



This first-year report is formative in purpose. Its aim is to strengthen the foundation on which future evaluations will rest by examining how *complete, accurate,* and *transparent* the program's assessment data proved to be in its first year. That examination found that the data received cannot yet support a valid account of how participating students performed. Reportable results survived for only 286 of the 2,879 students admitted to the program, about one in ten, a sample too limited and too unlikely to represent all participants to serve as a reliable basis for conclusions about student outcomes. Analyses of participant achievement are therefore deferred to a future report to be released by December 31, 2026, once data from the 2025-26 School Year (the second operational year) allow them to be conducted reliably. What follows is what this first year does make clear: where the data fell short, and what would place the next evaluation on firmer ground.

Data Completeness and Quality

The EOC received assessment records for 798 unique students, 13 fewer than the 811 SCDE reported as having taken approved assessments. Preparing those records for analysis required removing data at several stages, for reasons that recurred throughout the file: assessments not approved for the program, scores for students in grades not required to test, tests taken outside the 2024-25 school year, records with clear errors, and assessments taken by too few students to report. The figure above traces that reduction from the 798 students received to the 286 whose records were both usable and reportable.

Beyond the records set aside, three gaps limited what any first-year analysis could accomplish. The gaps are:

1. No prior-year achievement was collected, so year-over-year learning gains could not be measured, and no comparison could support cause-and-effect conclusions.
2. Incomplete student demographic information was provided to the EOC, preventing reliable reporting of outcomes by grade, gender, family income, and race.
3. No graduation data were reported to the EOC, so the graduation-rate reporting the law requires could not be completed.

Two further patterns bear on data quality directly. A small number of interim and benchmark assessment records, all from online charter schools, showed within-year growth so large that it contradicted the same students' state test results, a sign that some interim scores are not being administered or recorded reliably. And several of the assessments submitted, though nationally normed, lacked evidence of alignment to South Carolina's academic standards or the linking studies needed to relate them to the state tests. Both are preconditions indicated in the law for using an assessment to gauge progress against state expectations.

Viewed in context, these conditions are not unexpected for the first year of a program of this complexity, particularly one whose authorizing law was altered by the courts and the General Assembly during the year it was first being implemented. They are, however, the most important finding this evaluation can offer. Before the program's effect on student learning can be appropriately reported, the data describing that learning must be more complete, more accurate, and more transparent. The recommendations that follow are directed at that goal.

EOC Staff Recommendations

- **Require State-approved, standards-aligned assessments.** Limiting participants to the state summative assessments, or to interim and benchmark assessments approved for use under Section 59-18-310 of the SC Code, would ensure results are comparable across students and aligned to the standards against which their progress is measured. The use of many different, unaligned assessments was the single largest obstacle to analyzing outcomes this year.
- **Deliver assessment results on a fixed, timely schedule.** Set delivery deadlines for program participants so that the first transmission of data to the EOC coincides with the statewide public release of end-of-year summative results (SC READY and EOCEP), and so that all requested evaluation data reach the EOC no later than the public release of School Report Cards. Such a schedule would allow the annual evaluation of ESTF outcomes to be published by the end of each calendar year.
- **Route assessment data through the State's interoperable data system.** Hand-entered and manually retrieved data introduced avoidable error. Routing ESTF assessment data through the Ed-Fi enabled system already used for public schools would improve accuracy.
- **Collect prior-year achievement.** Without a baseline, neither year-over-year learning gains nor valid comparisons to similar students are possible. Prior-year scores are the prerequisite for the growth analyses the program is designed to support.
- **Provide graduation data in National Student Clearinghouse format.** This would satisfy the graduation-rate reporting required by Act 11 and allow postsecondary outcomes to be compared to similar students. Participating education service providers could be offered the same postsecondary reporting the EOC provides to all South Carolina public high schools.
- **Give evaluators appropriate access to fund-use records.** Access to the ClassWallet marketplace backend, subject to applicable privacy and security requirements, would allow the EOC to connect how funds are spent to the outcomes students achieve.

- **Collect and report reasons for program withdrawal.** Reporting why students and families leave the program would provide context for patterns of attrition and missing assessment data and help identify opportunities for program improvement.
- **Examine interim and benchmark assessment integrity statewide.** The implausible and contradictory interim scores found here are unlikely to be unique to ESTF. A statewide review would gauge the scope of the issue and inform more secure administration of these assessments.
- **Require sufficient alignment evidence to support every linking study.** A linking study relates scores on a non-state assessment to the state test, but it cannot show that the assessment measures the content South Carolina's standards define. Only an alignment study can establish that, and it is the foundation on which a valid linking rests. Before a nationally normed assessment is approved as a measure of progress toward South Carolina standards, it should carry a technically sound alignment study to those standards, documented transparently enough for the EOC to verify its quality independently. However rigorous its statistical methods, a linking study cannot substitute for that alignment evidence; the two are complementary and non-substitutable requirements.

The first year of the ESTF program did not yet yield a dependable picture of how participating students are performing. It did yield a clear account of what stands between the program and that picture, along with a concrete path toward closing the gap. The EOC offers these findings and recommendations in that constructive spirit and looks forward to publication of a fuller evaluation in December 2026 as the data mature.

SOUTH CAROLINA
DEPARTMENT OF EDUCATION

MEMORANDUM

TO: Education Oversight Committee

FROM: South Carolina Department of Education

DATE: September 11, 2026

RE: Education Scholarship Trust Fund, Year One (SY 2024-25) Report

The South Carolina Department of Education appreciates the Education Oversight Committee's formative review of the first year of the Education Scholarship Trust Fund program and shares the EOC's interest in ensuring that future program evaluations are supported by accurate, complete, and useful data.

SCDE has been transparent in providing the EOC the program data available to the Department. At the same time, we readily acknowledge the data-quality challenges associated with the ESTF's first operational year. While the Department clearly communicated the statutory assessment requirement and the assessments approved for participating students, implementation in the first year of any new and complex program presents challenges. Those challenges were compounded significantly when, during the program's first year, the South Carolina Supreme Court's *Eidson* decision removed independent schools, which had been the primary full-time education service providers participating in the program, from ESTF eligibility.

Following that decision, the Department worked directly with participating families to ensure students could satisfy the program's assessment requirements and to collect the information necessary to comply with the law. As SCDE communicated at the time, portions of that information necessarily depended on parent reporting and manual collection processes. Those processes allowed the Department to meet the immediate needs of a rapidly changing program, but they were never intended to represent the mature data infrastructure for ESTF going forward.

Context for the data the EOC received. During the 2024–25 school year, 2,879 students participated in the program at some point; however, many exited during the year for routine reasons such as voluntary withdrawal, return to public school, relocation, or other eligibility changes. At year's end, 1,440 students remained enrolled. Of those, 452 were exempt from testing under state law, including students in grades K–2 and six students with documented disabilities. An additional 136 did not ultimately provide usable assessment data due to circumstances such as graduation or data transfer issues, and a

subset of students was later removed from the program for failure to submit a required test. As a result, 798 students submitted assessment data to SCDE — the number upon which the EOC's analysis is based. The difference between the original enrollment figure and the number of reportable results is therefore largely explained by program attrition, statutory testing exemptions, and the EOC's own analytical screening criteria, not by a systemic failure of data collection.

The Department therefore views many of the issues identified in the EOC's report not as new discoveries, but as useful confirmation of process improvements SCDE had already identified and, in many cases, implemented for the 2025-26 school year. Notably, the Department has already established an earlier and more structured timeline for transmitting 2025-26 assessment data to the EOC, consistent with the timeline recommended in this report.

Among the additional process improvements already implemented or underway are:

Assessment collection infrastructure. A dedicated digital platform now receives and stores student assessment documentation, replacing Year One's manual processes and including standardized templates, approved-assessment guidance, and a set collection timeline.

Standardized enrollment, graduation, and residency data. Dropdown menus for district and school of residence replace the former free-text fields, the Department is procuring school attendance-area data to enable automated address-to-school matching, and education service providers are now required to report graduation information through a structured collection process.

Quarterly enrollment verification. Families now complete verification forms each quarter under a formal operating procedure. Accounts are suspended for non-completion, and address documentation is actively reviewed through program integrity processes.

Systematic cross-office process improvement. An operations team organized around assessments, address verification, and data reporting has documented first-year pain points, built standard operating procedures for each core function, and is developing a consolidated data platform as the program's authoritative reporting source.

As this work continues, SCDE welcomes recommendations that strengthen the program and is committed to acting on those that fall within the Department's authority under Title 59, Chapter 8. Several of the EOC's recommendations align with improvements the Department has already begun or can readily advance. Where the EOC’s recommendations contemplate new requirements or data collection beyond what is

currently established in law, additional legal, technical, or policy consideration will be needed to ensure consistency with the governing statute and the respective roles of SCDE, the EOC, education service providers, and participating families. With that framework in mind, SCDE has implemented or plans to implement the following initiatives addressed in the report:

Encourage use of the state summative or formative assessment where practical, while recognizing that Section 59-8-150 expressly provides families and education service providers with multiple assessment pathways and expressly prohibits the state from requiring independent schools to change their curriculum. SCDE should actively communicate these statutory options to education service providers and families and should develop guidance clarifying logistical procedures for ESTF students who elect to take SC READY at a public school site. The Department worked closely with the Charter Institute at Erskine in 2025-26 to offer SC READY to ESTF students at sites across the state.

Require independent test administration for non-state assessments. Testing fidelity is a prerequisite for score validity. Any approved non-state assessment used to satisfy § 59-8-150(C)(1)(b) should be administered by a department-approved third-party test administrator to provide a controlled proctored environment, standardized protocol, and consistent data reporting.

Route ESTF program data through the State's interoperable data system. Hand-entered and manually retrieved data introduced avoidable error. ESTF assessment data, student demographic information, and third-party administrator data will be routed through the Ed-Fi enabled system already used for public schools. Consolidating program data in a single interoperable system will improve accuracy and place ESTF evaluation data on the same infrastructure that supports statewide accountability reporting.

Refine the assessment approval and linking framework for the alternative testing pathway. For ESTF students who do not take SC READY or a state-approved formative assessment per 59-8-150(C), SCDE will define an administrative review process under which test publishers submit content crosswalks and blueprint comparisons against the SC College- and Career-Ready Standards, and SCDE staff evaluate submissions for completeness, plausibility, and the absence of disqualifying content gaps. For each assessment approved through this process, SCDE will continue to apply its developed equi-percentile concordance methodology, the most appropriate and cost-effective type of linking study, using South Carolina student data to establish score relationships with SC READY. SCDE will also continue to publish annually a list of all approved assessments, including SC READY, state-approved formative assessments, and approved norm-referenced

instruments, together with the alignment and concordance methodologies and interpretive guidance for families and education service providers.

Most importantly, the Department believes the appropriate focus is forward. The first year of ESTF was truly a "Year Zero": a new program implemented under extraordinary circumstances and substantially disrupted midway through its first operational year. The 2025-26 program is operating under a different statutory framework, with stronger processes, better data collection, and lessons learned from that first year already incorporated into implementation.

SCDE looks forward to continuing to work collaboratively with the EOC as the data mature and as South Carolina develops a sound, sustainable framework for evaluating the program's effectiveness.

EDUCATION OVERSIGHT COMMITTEE

DATE: September 21, 2026

SUBCOMMITTEE:
Academic Standards & Assessments Subcommittee

ACTION ITEM:
Evaluation of Alternative Instruction Methods

PURPOSE/AUTHORITY
Proviso 1A.66 of 2025-26 Appropriation Act. (SDE-EIA: Evaluation of Alternative Instruction Methods) With funds appropriated, the Education Oversight Committee is responsible for evaluating the impact of alternative methods of instruction on student learning and working with other agencies to expand access to quality remote instruction which can be dispatched if necessary. Alternative methods of instruction may include, but are not limited to, online or virtual instruction, remote learning, and hybrid models. The Department of Education and school districts providing alternative methods of instruction must provide data as requested by the committee to evaluate the effectiveness of the instruction. The Education Oversight Committee shall report annually to the Governor, the General Assembly, the Department of Education, and the State Board of Education.

CRITICAL FACTS
This study presents a comprehensive analysis of student outcomes associated with various instructional delivery methods across South Carolina public schools during the 2024-2025 academic year. The study focuses on two major statewide assessments:
EOCEP (End of Course Examination Program) for high school students (Grades 9-12)
SC READY for elementary and middle school students (Grades 3-8)

TIMELINE/REVIEW PROCESS
2024-2025 academic year

ECONOMIC IMPACT FOR EOC
Cost: no impact

ACTION REQUEST

☒ **For approval** ☐ **For information**

ACTION TAKEN

☐ **Approved** ☐ **Amended**
☐ **Not Approved** ☐ **Action deferred (explain)**

Alternative Instructional Methods Study

Student Outcomes Across Instructional Delivery Types

Prepared by Amina Asghar

Data Engineer, EOC

SC EDUCATION OVERSIGHT COMMITTEE

Reporting facts. Measuring change. Promoting progress.

Alternative Instructional Methods Study

Student Outcomes Across Instructional Delivery Types

School Year 2024-2025

SC READY Grades 3-8 | EOCEP High School Courses

Prepared by:

Amina Asghar

Data Engineer, EOC

Contents

1. Executive Summary 1
2. Study Context, Scope, and Objectives 3
3. Findings: EOCEP High School Courses 4
   - Finding 1: District Virtual and Hybrid students have lower pass rates in all four EOCEP subjects, with the largest gaps in Algebra 1 4
   - Finding 2: Large differences remain after accounting for school, grade, semester, and poverty 5
   - Finding 3: Poverty is strongly related to EOCEP outcomes but does not account for the delivery-type gaps 6
   - Finding 4: Only about one-third of VirtualSC enrollments have a recorded EOCEP score 7
   - Finding 5: Unknown reflects missing reporting information, not an instructional method 8
4. SC READY, Grades 3 to 8 9
   - Finding 6: Differences remain among students who started at the same score, especially in Math 9
   - Finding 7: Raw score gaps narrow in ELA by grade 8 but never close in Math 10
5. Limitations and Data Quality 11
6. Implications and Recommendations 12
7. Appendix A: EOCEP Tables and Charts 0
8. Appendix B: SC READY Tables and Charts 11
9. Appendix C: Data Preparation, Methods, and Sensitivity Checks 17

1. Executive Summary

This study examines student outcomes across instructional delivery types in South Carolina public schools during the 2024-25 school year. It covers 183,612 students in grades 7 to 12 with End-of-Course Examination Program (EOCEP) eligible course enrollments in Algebra 1, English 2, Biology 1, and U.S. History, including some grade 7 and 8 students taking these courses early, and 313,727 students in grades 3 to 8 assessed through SC READY in English Language Arts (ELA) and Math. The delivery types compared are Face-to-Face, VirtualSC, District Virtual, Out-of-State Virtual, Hybrid, Distance Learning, and Unknown, a label used when the enrollment records do not identify how a course was delivered. Achievement results are based on students with a recorded test score. Test participation is examined separately; EOCEP participation is counted at the student-subject level.

Table 1. Instructional Delivery Type Definitions

Instructional Delivery Type	Definition
Face-to-Face	Teacher is physically present in the classroom providing in-person instruction to students (also referred to as Instructor Led or Instruction Type = A in PowerSchool).
VirtualSC*	Teacher provides instruction to students through the VirtualSC program (also referred to as SC Virtual School Program or Instruction Type = B in PowerSchool).
District Virtual*	Virtual instruction is provided by a teacher as part of an SC school/district/state entity (also referred to as Online In-State or Instruction Type = C in PowerSchool).
Out-of-State Virtual*	The teacher who provides instruction to students is not part of an SC school/district/state (also referred to as Online Out-of-State or Instruction Type = E in PowerSchool). This includes any course that is provided by a vendor (e.g., EdOptions, Edmentum, Florida Virtual).
Hybrid*	The teacher provides in-person and virtual instruction in some capacity as part of an SC school/district/state entity (also referred to as Instruction Type = F in PowerSchool). When the in-person instructional time does not meet the seat time requirement for the initial offering of a credit-bearing course, the section should be marked as Hybrid. This is not referring to dual modality.
Distance Learning*	The teacher provides instruction for students physically located at another school location. For example, the teacher may be located physically at a television studio, yet teaching several groups of students housed in different schools or districts.
Unknown	The instruction type is unknown and the enrollment records do not identify how a course was delivered.

*Denotes Alternative Instruction Method. For more information, see Student Information System (SIS) Data Entry Manual.

Principal Findings

EOCEP pass rates are substantially lower for District Virtual and Hybrid students, with the largest differences in Algebra 1. In Algebra 1, 55.4 percent of Face-to-Face students passed, against 20.2 percent of District Virtual students and 17.4 percent of Hybrid students. Every subject

shows a gap of this kind. Algebra 1 is also where delivery type is most strongly connected to scores, associated with 3.9 percent of score variation, the largest share in the study.

The large differences remained after accounting for school, grade, semester, and poverty. Depending on the subject, adjusted pass rates for District Virtual students were 18.9 to 32.1 percentage points lower than for Face-to-Face students; for Hybrid students, 20.3 to 35.3 points lower. These are strong associations, not proof that delivery type caused the differences.

Poverty is strongly related to achievement, but it does not account for the delivery-type differences. The instructional groups serve different populations: pupils in poverty make up 57.9 percent of tested Face-to-Face EOCEP records, 69.7 percent of Hybrid, 83.0 percent of VirtualSC, and 86.6 percent of Distance Learning. Yet the delivery gaps hold when students are compared within the same poverty group, and in models that account for poverty. Poverty carries its own weight too: adjusted pass rates run 11.4 to 21.2 percentage points lower for pupils in poverty, depending on the subject.

EOCEP participation is an important limitation for VirtualSC. Only 224 of 710 VirtualSC student-subject enrollments, 31.5 percent, have a recorded EOCEP score; the Face-to-Face figure is 96.2 percent. The available data do not show why the remaining records have no score, so VirtualSC's EOCEP results describe only the subset with a recorded score. SC READY shows nothing similar: there, VirtualSC participation is 96 percent or higher in both subjects.

Unknown is a data-quality category, not an instructional method. In EOCEP, ten districts account for 86.1 percent of the 61,203 enrollment records with no reported delivery type; Greenville 01 alone accounts for about 42 percent. In SC READY, about 11 percent of student-subject assignments are Unknown, either because the delivery type is missing or because no course for the subject is on file. In the fully adjusted EOCEP models, Unknown is not statistically different from Face-to-Face in any subject. In SC READY, its large raw advantage shrinks to about 1 point higher in ELA and 2 points lower in Math once prior achievement is taken into account, differences that are very small in practical terms. The Unknown category is not used to draw conclusions about instructional delivery.

In grades 3 to 8, after controlling for prior achievement, students receiving alternative instruction methods underperform compared to students who started at the same score, particularly in Math. District Virtual students scored about 24 points lower in Math and 11 points lower in ELA than Face-to-Face students with the same prior-year score and grade; Hybrid students scored about 23 and 8 points lower, respectively. Adding poverty as a covariate changed these estimates very little. In the raw results, the ELA gap between Face-to-Face and District Virtual is essentially gone by grade 8, while the Math gap remains substantial at every grade.

Readers should note that this study is observational. Students are not randomly assigned to instructional delivery types, so the findings are associations rather than causal effects. SC READY provides a prior-year achievement control for grades 4 to 8; EOCEP has no comparable prior-year test, so its models compare students within schools while accounting for grade, semester the test was taken, and poverty. Full limitations are in Section 5. The recommendations in Section 6 focus on improving delivery-type reporting, investigating the VirtualSC participation gap, giving particular attention to Math and Algebra 1, and comparing like with like when judging delivery types.

2. Study Context, Scope, and Objectives

This study examines student academic performance in South Carolina public schools in the 2024-25 school year across multiple instructional settings, including traditional in-person classes, VirtualSC, district and out-of-state online courses, hybrid instruction, and distance learning options. School districts record the instructional delivery type on each course enrollment when reported, allowing comparison of student outcomes by setting.

The primary objective is to assess differences in student outcomes by instructional delivery type and determine if these differences remain after adjusting for prior academic performance, poverty status, grade level, school context, and semester the test was taken. The study also examines whether these patterns are consistent across subjects, grade levels, and poverty groups.

The analysis focuses on five main questions:

1. Which students participate in each instructional delivery type, and how do these groups differ, particularly regarding poverty status?
2. How do student academic outcomes compare across instructional delivery types?
3. Do differences in student outcomes persist after accounting for prior achievement, poverty, grade, school context, and semester the test was taken, where available?
4. Are these patterns consistent across subjects, grade levels, and student groups defined by poverty status?
5. What data quality or reporting issues should be considered when interpreting the results?

The study focuses on two assessment groups: the End-of-Course Examination Program (EOCEP), covering Algebra 1, English 2, Biology 1, and U.S. History for 183,612 students in grades 7 to 12 (grade 7 and 8 test takers are middle school students taking these courses early), and SC READY, which measures English language arts (ELA) and Math for 313,727 students in grades 3 to 8. Course enrollment records indicate instructional delivery type, assessment records provide student outcomes, and 2025 Report Card data identify pupil-in-poverty (PiP) status.

For analysis, each student's coursework is classified into one instructional delivery type per subject. Every result in this report is based on that one-record-per-student-per-subject approach. For EOCEP, the classification reflects the delivery type of the student's courses in the tested subject. For SC READY, ELA and Math are classified separately, each from that subject's own courses. Some middle school students take an EOCEP course early, typically Algebra 1. Those courses appear in the high school enrollment file, so both course files are used in the classification.

Students are grouped as Unknown when enrollment records do not provide enough information to identify how the course was delivered. For SC READY, this may occur because a subject course has no reported delivery type or because no course is found for that subject. For EOCEP, it usually means the course is present, but the delivery type is missing. Unknown is treated as a data-quality category, not as an instructional method.

For EOCEP, passing is defined as performance level A, B, or C. To protect student privacy and ensure reliability, results are not reported for groups with fewer than 20 tested students. Appendix C provides additional details on data preparation, instructional type assignment, statistical methods, and sensitivity analyses.

3. Findings: EOCEP High School Courses

Finding 1: District Virtual and Hybrid students have lower pass rates in all four EOCEP subjects, with the largest gaps in Algebra 1

EOCEP pass rates differed sharply by instructional delivery type. In Algebra 1, 55.4 percent of Face-to-Face students passed, compared to 20.2 percent of District Virtual and 17.4 percent of Hybrid students. Only 2.2 percent of the 45 VirtualSC students with an Algebra 1 score passed. These students are the minority of VirtualSC enrollees who tested at all (Finding 4). Algebra 1 had the largest pass rate differences, while English 2 had the smallest. Biology 1 and U.S. History fell in between (Figure 1, Table 2). The Unknown group is a reporting category rather than an instructional method and is examined separately in Finding 5.

Statistical testing confirms real differences among delivery types in every subject and in 47 of the 55 head-to-head comparisons between pairs of delivery types. The pattern is not an Algebra problem; it is present in all four subjects. Virtual and hybrid students pass at substantially lower rates across the board. In English 2, the narrowest case, Face-to-Face students pass at 72.7 percent, compared with 50.4 percent for Hybrid. In Biology 1 and U.S. History, District Virtual and Hybrid students pass at roughly half the Face-to-Face rate. Algebra 1 shows the most extreme pattern. It is also the subject where delivery type is most strongly connected to scores overall, associated with 3.9 percent of score variation. Test details and full results are in Appendix Table A3.

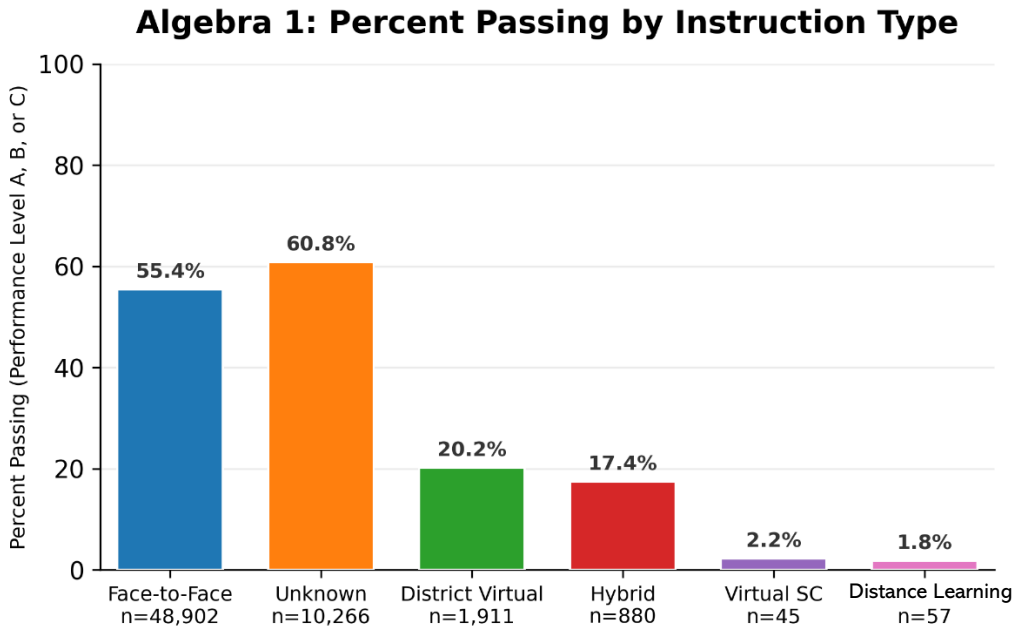
Table 2. Percent passing by subject and instruction type, among tested students

Instruction Type	Algebra 1	English 2	Biology 1	U.S. History
Face-to-Face	55.4	72.7	47.0	49.5
Unknown	60.8	69.6	48.7	48.9
District Virtual	20.2	60.9	27.1	26.8
Hybrid	17.4	50.4	19.3	20.1
VirtualSC	2.2	9.7	24.1	11.6
Distance Learning	1.8	29.6	0.0	*
Out-of-State Virtual	*	*	*	*

** Fewer than 20 tested students, suppressed. Tested counts and mean scale scores: Appendix Tables A1 and A2.*

Figure 1. Algebra 1 pass rates are much lower for District Virtual and Hybrid students

Companion charts for the other three subjects: Appendix Figures A1 to A3.



Note: Among tested students; n = tested students in the group. Formats with fewer than 20 tested students are not shown. Virtual SC results describe its tested minority (31.5 percent of enrollees).

Finding 2: Large differences remain after accounting for school, grade level, semester, and poverty

Even after adjusting for school, grade level, semester the test was taken, and student poverty, substantial differences in EOCEP pass rates remained. In Algebra 1, District Virtual students had an adjusted pass rate 32.1 percentage points lower than Face-to-Face students, while Hybrid students were 23.2 points lower. For VirtualSC students with an Algebra 1 score, the adjusted pass rate was 48.1 percentage points lower. Because fewer VirtualSC students took the EOCEP test, this estimate applies only to those with a recorded test score (Finding 4).

Similar gaps appeared in the other EOCEP subjects. District Virtual pass rates were 18.9 to 32.1 percentage points lower than Face-to-Face, while Hybrid pass rates were 20.3 to 35.3 points lower. VirtualSC also had lower adjusted pass rates in three subjects, with the largest difference in English 2 at 59.2 percentage points. In Biology 1, the VirtualSC estimate was not statistically different from Face-to-Face because of the uncertainty around the estimate (Figure 2, Table 2). Unknown did not differ statistically from Face-to-Face in any of the four subjects after adjustment. Given that Unknown represents records where instructional delivery could not be identified, it should not be interpreted as a separate instructional method (Finding 5).

Accounting for school, grade level, semester the test was taken, and poverty did not eliminate the differences in pass rates. Those factors therefore do not explain the gaps on their own. Other differences between students may still be present in the data but not measured in the analysis. For that reason, the findings show a persistent relationship between instructional delivery type and pass rates, but they do not show that delivery type caused the differences. In this analysis, VirtualSC refers to students enrolled in regular public schools who take classes through the state virtual program. These students are not enrolled in virtual charter schools.

To ensure reliable comparisons, model results are reported only when a group includes enough schools and tested students. Face-to-Face, Unknown, District Virtual, Hybrid, and VirtualSC met

this requirement. Out-of-State Virtual did not, and Distance Learning met it only for Algebra 1. Additional details are provided in Appendix Table A5.

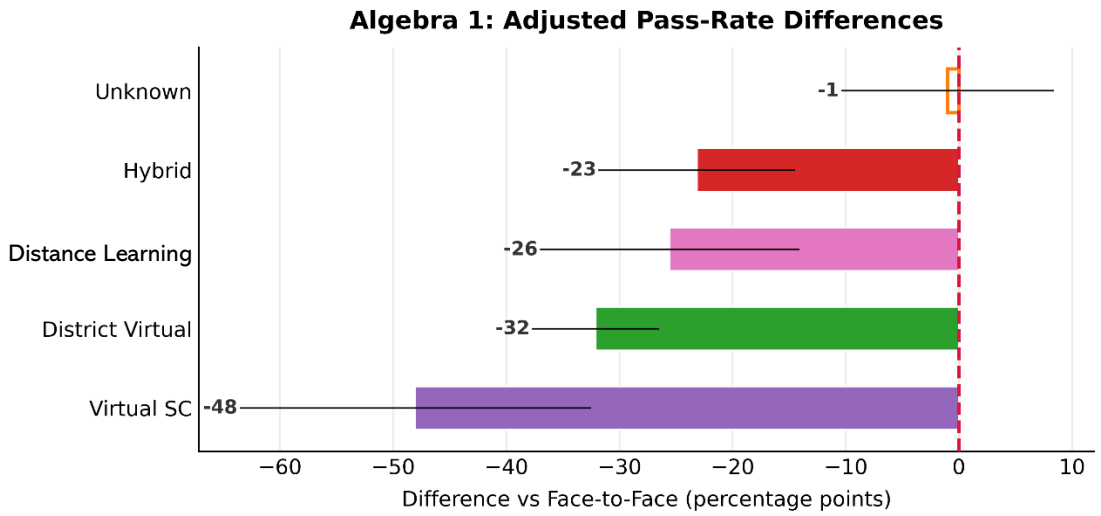
Table 2. Adjusted differences in passing, poverty adjusted (percentage points versus Face-to-Face, within school, grade level, and testing window)

Model Term	Algebra 1	English 2	Biology 1	U.S. History
District Virtual	-32.1	-26.7	-18.9	-19.2
Hybrid	-23.2	-28.9	-35.3	-20.3
VirtualSC	-48.1	-59.2	-17.7 (ns)	-30.6
Distance Learning	-25.6	*	*	*
Unknown	-1.0 (ns)	+0.9 (ns)	+11.3 (ns)	+2.3 (ns)
Pupils in poverty (within school and type)	-11.4	-13.3	-20.0	-21.2

*Linear probability models; school, grade, and test administration window fixed effects; school-clustered standard errors; ns = not significant. Estimates from groups spanning too few schools within a subject are excluded per the reliability standard described in Appendix Table A5. Confidence intervals are reported in Appendix Table A5. * = reliability standard not met in that subject.*

Figure 2. Adjusted differences in Algebra 1 pass rates by instructional delivery type

Percentage-point differences versus Face-to-Face after adjusting for school, grade, test administration window, and poverty. Whiskers show 95% confidence intervals; open bars indicate estimates that are not statistically distinguishable from Face-to-Face.



Finding 3: Poverty is strongly related to EOCEP outcomes but does not account for the delivery-type gaps

Poverty rates vary across instructional settings. Among tested EOCEP records, pupils in poverty accounted for 57.9 percent of Face-to-Face records, 69.7 percent of Hybrid, 83.0 percent of VirtualSC, and 86.6 percent of Distance Learning (Appendix Table A6). If poverty accounted for the observed delivery differences, those gaps should narrow substantially within poverty groups. They do not.

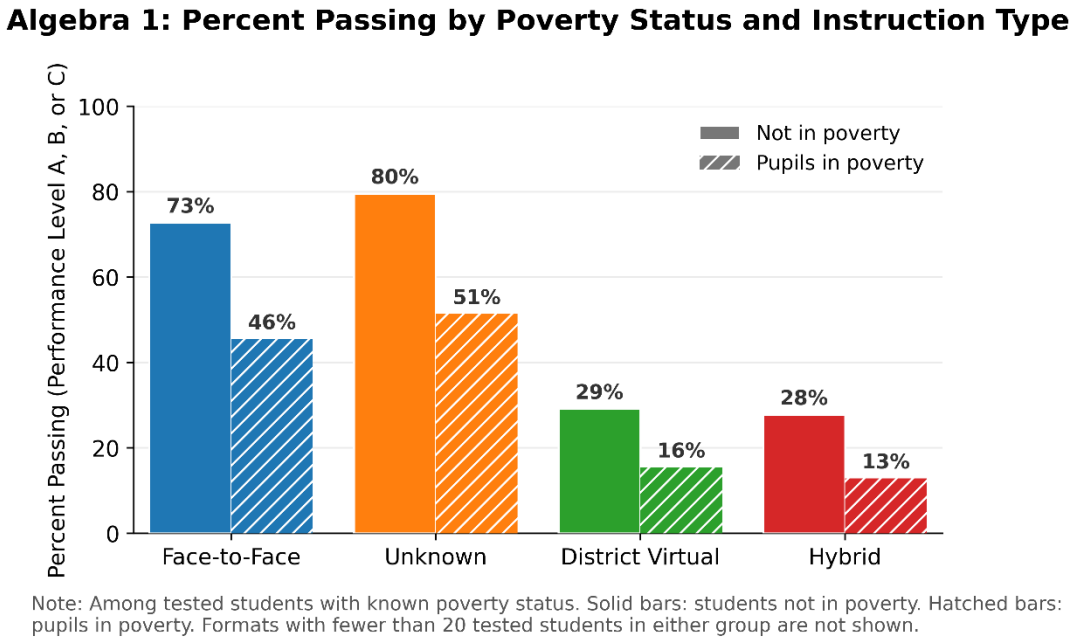
In Algebra 1, 45.5 percent of Face-to-Face students in poverty passed, compared with 15.5 percent of District Virtual students and 12.9 percent of Hybrid students. Among students not in poverty, the corresponding pass rates were 72.8 percent for Face-to-Face, 29.2 percent for District Virtual, and 27.8 percent for Hybrid (Figure 3). The same general pattern appears in the other EOCEP subjects (Appendix Figures A7 to A9).

Adjusted models show similar results. Estimates with and without poverty adjustment generally differ by less than three points for the larger groups, and by up to about five points for VirtualSC (Appendix Tables A4 and A5). Therefore, higher poverty concentrations do not explain the delivery-type differences. However, poverty remains strongly associated with EOCEP performance: after controlling for school, grade, semester the test was taken, and delivery type, students in poverty had pass rates 11.4 to 21.2 percentage points lower, depending on the subject. This association is strongest in Biology 1 and U.S. History, while Algebra 1 showed the strongest overall relationship between instructional delivery type and performance.

Poverty matters, but it does not explain the differences across instructional delivery types. The gaps remain after accounting for poverty.

Figure 3. Algebra 1 pass rates differ by both poverty status and instructional delivery type

Percent passing by poverty status within instructional delivery type. Results for other EOCEP subjects are shown in Appendix Figures A7 to A9; detailed results are in Appendix Table A7.



Note: Among tested students with known poverty status. Solid bars: students not in poverty. Hatched bars: pupils in poverty. Formats with fewer than 20 tested students in either group are not shown.

Finding 4: Only about one-third of VirtualSC enrollments have a recorded EOCEP score

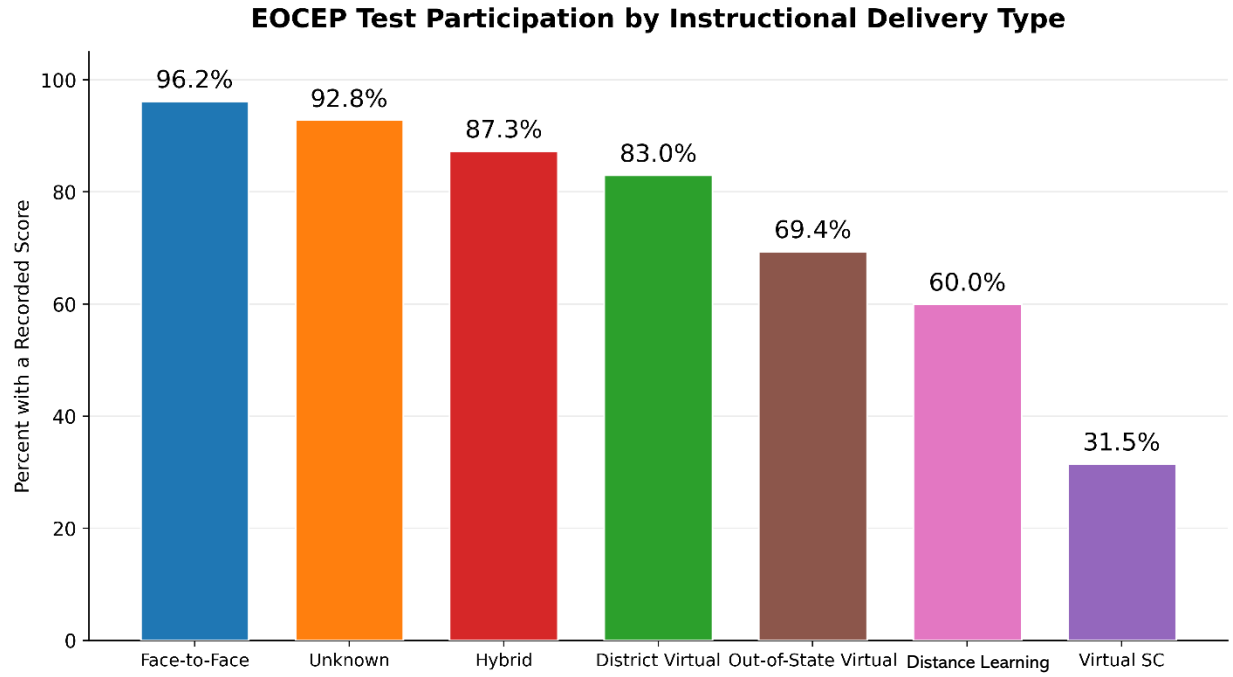
Before comparing EOCEP outcomes, test participation differs substantially across instructional delivery types. Counting one record per student per subject, Face-to-Face had the highest participation rate at 96.2 percent, followed by Unknown at 92.8 percent, Hybrid at 87.3 percent, and District Virtual at 83.0 percent. Participation was far lower for VirtualSC: of 710 student-subject enrollments, only 224 (31.5 percent) had a recorded score (Figure 4, Appendix Table A1). Distance Learning and Out-of-State Virtual also had lower participation, but both groups were very small.

This matters when interpreting the VirtualSC results elsewhere in the report. Pass rates and score comparisons are based only on records with a test score. For VirtualSC, that means the results describe roughly one-third of its student-subject enrollment, not all VirtualSC coursework. The available data do not show why the remaining records have no score; course withdrawal, exemption, absence, and other reasons cannot be distinguished.

The low participation rate is specific to EOCEP. For SC READY, VirtualSC participation was 96 percent or higher in both ELA and Math, comparable to other major delivery groups (Appendix Table B1, Figure B7). Therefore, EOCEP participation is a key limitation when interpreting VirtualSC results.

Figure 4. EOCEP test participation is much lower for VirtualSC

Percent of student-subject records with a recorded EOCEP score, by instructional delivery type. Counts: Appendix Table A1.



Note: Percent of student-subject records with a recorded EOCEP score, SY 2024-25.

Finding 5: Unknown reflects missing reporting information, not an instructional method

Unknown does not represent a separate instructional method. In EOCEP, it mainly comes from course enrollment records where the instructional delivery type was not reported. About one in five EOCEP enrollment records lacked this information, and the missing reporting was heavily concentrated in a small number of districts. Ten districts accounted for 86.1 percent of the 61,203 uncoded EOCEP enrollment records. Greenville 01 alone accounted for about 42 percent; 98.6 percent of its own EOCEP enrollments were uncoded. Greenwood 50 and Marlboro 01 also had uncoded rates above 98 percent (Table 3).

Unknown has a somewhat different meaning in SC READY. About 11 percent of student-subject assignments were classified as Unknown, either because a subject course was listed without a delivery type or because no course for that subject could be found in the enrollment records. The no-course cases accounted for 5,075 ELA and 2,471 Math assignments, and these students scored well above average. This mix is why Unknown is treated as a data-quality category rather than as another instructional delivery type.

The results support this conclusion. In EOCEP, any differences associated with Unknown in the raw data disappear in the fully adjusted models: no subject shows a statistically meaningful gap between Unknown and Face-to-Face (Finding 2). SC READY shows a similar pattern. Unknown initially appears to have a large raw advantage, but after accounting for prior achievement, the difference nearly disappears, with Unknown about 1 point above Face-to-Face in ELA and 2 points below in Math. Although these small differences are statistically detectable due to the large sample sizes, they are very small in magnitude (Finding 6).

Overall, the reporting patterns and adjusted results indicate that Unknown outcomes primarily reflect missing delivery information and the characteristics of students in this category, rather than a distinct instructional approach. As a result, Unknown is reported throughout the analysis but is not used to draw conclusions about the relative performance of instructional delivery types. Section 6 recommends follow-up with districts where missing delivery reporting is concentrated.

Table 3. Districts with the largest numbers of uncoded EOCEP enrollment records

District	EOCEP enrollments	Uncoded	Percent of own enrollments uncoded
Greenville 01	26,184	25,824	98.6
York 03	8,952	7,410	82.8
Florence 01	4,632	4,047	87.4
Sumter 01	7,107	3,735	52.6
Greenwood 50	3,103	3,045	98.1
Cherokee 01	2,628	2,126	80.9
Aiken 01	7,909	2,050	25.9
Charleston 01	18,954	1,672	8.8
Horry 01	16,127	1,467	9.1
Marlboro 01	1,352	1,329	98.3

These ten districts account for 86.1 percent of all uncoded EOCEP enrollment records; 106 of 110 districts had at least some.

4. SC READY, Grades 3 to 8

Finding 6: Differences remain among students who started at the same score, especially in Math

One possible reason for the raw differences is that students in different instructional settings may have entered the year at different achievement levels. To account for this, the analysis linked grade 4 to 8 students to their prior-year SC READY scores and compared outcomes after adjusting for prior score and grade. Prior-year scores were available for 92 to 94 percent of students in each grade.

The differences remained after accounting for prior achievement. In Math, District Virtual students scored 23.8 points lower than Face-to-Face students with the same prior score and grade, and Hybrid students scored 22.5 points lower. Distance Learning students scored 28.5 points lower, although this estimate rests on a much smaller group, and the same group's ELA estimate is too imprecise to interpret. In ELA, the adjusted differences were smaller: 10.8 points lower for District Virtual and 8.0 for Hybrid. VirtualSC did not differ statistically from Face-to-Face in either subject among the students available for this analysis (Table 4, Figure 5). Unknown's large raw advantage nearly vanishes here as well, to about 1 point above Face-to-Face in ELA and 2 points below in Math, consistent with Finding 5's reading of the category.

Prior achievement alone does not account for the remaining differences. Adding poverty to the models changed the estimates by only tenths of a point, consistent with prior-year achievement

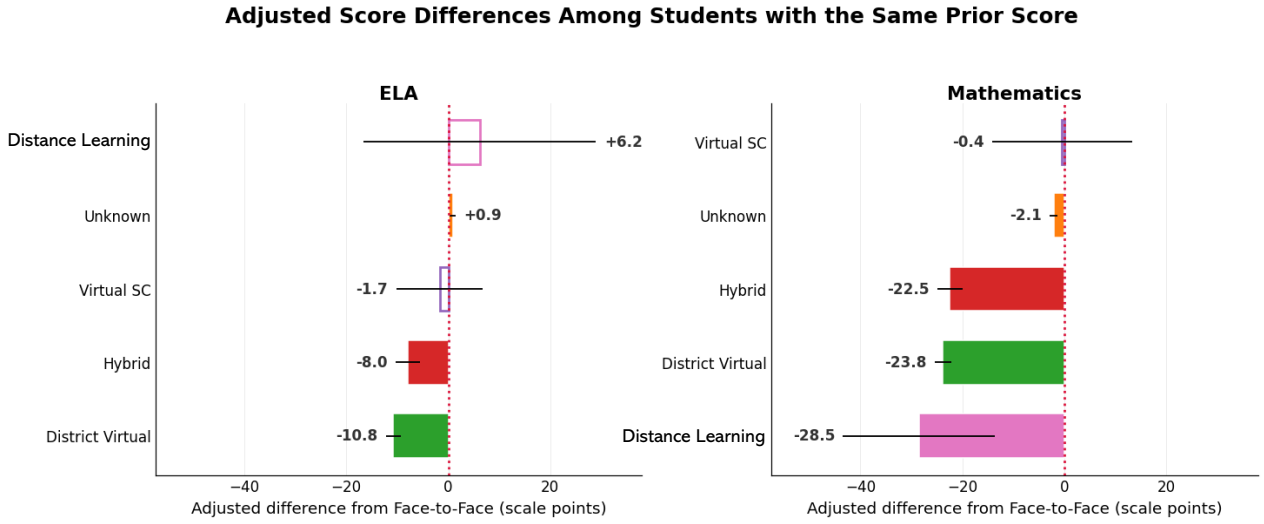
already capturing much of the difference associated with poverty (Appendix Table B4). The main differences therefore remain after accounting for both where students started academically and their poverty status.

Table 4. Adjusted differences in SY 2024-25 score among students with the same prior year score and grade (scale points versus Face-to-Face)

Instruction type	ELA adjusted difference	ELA 95% interval	Math adjusted difference	Math 95% interval
District Virtual	-10.8	-12.2 to -9.3	-23.8	-25.4 to -22.2
Hybrid	-8.0	-10.3 to -5.6	-22.5	-24.9 to -20.0
Distance Learning	+6.2 (ns)	-16.6 to +29.0	-28.5	-43.4 to -13.7
VirtualSC	-1.7 (ns)	-10.2 to +6.8	-0.4 (ns)	-14.1 to +13.4
Unknown	+0.9	+0.3 to +1.5	-2.1	-2.8 to -1.4

Analysis of covariance on 240,601 (ELA) and 240,834 (Math) Grade 4 to 8 students with a prior score; adjusts for prior score and grade; heteroskedasticity-robust intervals; ns = not statistically significant. Grade 3 has no prior test and is excluded; groups with fewer than 20 students are excluded. Full model output: Appendix Table B4.

Figure 5. District Virtual and Hybrid students scored about 22 to 24 points lower in Math after accounting for prior achievement.



Note: Grades 4 to 8 students with a prior year SC READY score; adjusts for prior score and grade; robust 95 percent intervals (black lines). Solid bars are statistically significant; open bars are not distinguishable from Face-to-Face. Grade 3 (no prior test) and groups with fewer than 20 students with a prior year score excluded.

Finding 7: Raw score gaps narrow in ELA by grade 8 but never close in Math

Raw SC READY results generally show Face-to-Face and Unknown students scoring above District Virtual and Hybrid students, although the size of the differences changes by grade and subject (Appendix Tables B2 and B3, Figures B1 and B2). The clearest contrast is between the two subjects. In ELA, the gap between Face-to-Face and District Virtual students shrinks across grades and is essentially gone by grade 8, when District Virtual students averaged 585.5 compared with 581.1 for Face-to-Face. In Math, the gap also narrows with grade but never closes: 82 points in grade 3 (463.0 versus 380.5) and still 29 points in grade 8 (552.8 versus 524.1). This mirrors the adjusted results in Finding 6, where the Math differences were roughly two to three times the ELA differences.

The same general pattern appears when students are compared separately by poverty status. At grade 8, for example, District Virtual and Face-to-Face students have very similar ELA scores within each poverty group, while sizeable Math differences remain (Table 5). Differences in poverty composition alone therefore do not account for the ELA and Math patterns. Poverty composition also varies much less across delivery types here than in high school: the share of students in poverty ranges from about 61 to 72 percent across the grades 3 to 8 delivery types (Appendix Table B5), compared with 58 to 87 percent in EOCEP (Finding 3).

Instructional delivery type is associated with a relatively small share of overall score variation in most grades. In grades 3 through 6, the share ranges from 0.12 to 0.31 percent in ELA and 0.65 to 0.85 percent in Math (Appendix Table B6). Grade 8 ELA is an exception at 2.0 percent, largely because of the unusually high-scoring Unknown group. As shown in Findings 5 and 6, much of that group's raw advantage reflects who is in the category and becomes very small once prior achievement is taken into account.

Table 5. Grade 8 mean scores by poverty status and instructional delivery type

Instruction type	ELA, not in poverty	ELA, in poverty	Math, not in poverty	Math, in poverty
Face-to-Face	604.8	569.4	580.7	538.7
Unknown	635.6	593.9	600.6	545.9
District Virtual	604.4	575.6	544.0	513.4
Hybrid	593.4	570.6	530.5	510.7

Grade 8 tested students with known poverty status; groups under 20 in either poverty cell omitted. Poverty-stratified results for all grades are shown in Appendix Figures B5 and B6.

5. Limitations and Data Quality

Selection. Students were not randomly assigned to instructional delivery types. The results in this report are associations, not causal effects. For SC READY, prior-year score is the main control: the model compares grade 4 to 8 outcomes at the same prior score and grade. That control cannot capture what happened between the two tests (a family move, an illness, a schedule change) or why a student ended up in a virtual, hybrid, or Face-to-Face course. Grade 3 is excluded because there is no prior-year SC READY score.

EOCEP is harder to adjust for because there is no prior-year test in the same subject. The best comparison available is within schools, while also accounting for grade, semester the test was taken, and poverty. That gets closer to comparing similar students, but it still leaves differences the data cannot measure. The model therefore shows what remains after those adjustments; it does not show that delivery type caused the gap.

One methodological note. Delivery type was assigned once per student per subject using course records. In an earlier version, some records with more than one delivery type could be counted in more than one group. Fixing that had little effect on the SC READY adjusted results, but the EOCEP adjusted differences became somewhat larger. A second check used a narrower set of EOCEP-eligible courses; participation and pass-rate results were similar. Appendix C provides the details.

Participation. Score and pass-rate results reflect students with a recorded score. SC READY participation is high, 92 to 97 percent for the major delivery groups, lower only in the small Other Virtual group. EOCEP is different: only 31.5 percent of VirtualSC student-subject enrollments have a

recorded score, and the records do not say why the rest are missing. VirtualSC's EOCEP results therefore describe only the subset with a recorded score.

Prior-year SC READY scores were available for 92 to 94 percent of students in grades 4 to 8, with somewhat lower coverage in the virtual groups. Adjusted SC READY results apply to students with an in-state prior testing history.

Small groups and school coverage. Distance Learning and Out-of-State Virtual include relatively few students in both assessments, so many results are suppressed or excluded from the models. Any group with fewer than 20 tested students is not reported. EOCEP models also require enough schools to support the estimate. Out-of-State Virtual did not meet that requirement in any subject, and Distance Learning met it only in Algebra 1. Results from these small groups are therefore limited and should not be generalized broadly.

One note when comparing counts across grades: the grade 8 enrollment file is smaller than the grade 6 and 7 files. That reflects the source data as received, not the analysis.

Population definition and data quality. The list of EOCEP-eligible courses grew from 18 to 23 codes this year; the additions are advanced course variants, so comparisons with prior years should be made carefully. The EOCEP file arrived already limited to eligible courses, which means EOCEP's share of statewide enrollment cannot be determined from these data.

Missing delivery information is the other major data-quality issue. About 20 percent of EOCEP enrollment records lack a reported delivery type, and about 11 percent of SC READY student-subject assignments are Unknown, as detailed in Finding 5. These records stay in the analysis but are not treated as a distinct instructional method.

Additional notes: duplicate test records were resolved by keeping the highest score, and records with invalid identifiers were removed. Poverty status was available for 98.3 percent of the EOCEP analysis population and 97.7 percent of the SC READY analysis population. Where a student's poverty records conflicted, the student was classified as in poverty if any record said so.

6. Implications and Recommendations

- Follow up with districts where instructional delivery type is frequently missing. Ten districts account for 86 percent of uncoded EOCEP enrollment records, with Greenville 01 alone at about 42 percent (Finding 5, Table 3). Improved course and delivery reporting could reduce the Unknown category, which covers about one-fifth of EOCEP enrollment records and about one-tenth of SC READY student-subject assignments.

- Investigate the low EOCEP participation rate for VirtualSC. Only 31.5 percent of VirtualSC student-subject enrollments have a recorded EOCEP score, and the available data do not show why the remaining records have no score (Finding 4). Distinguishing course withdrawal, exemption, absence, and other reasons would clarify what is behind the missing test results.

- Give particular attention to mathematics in future reviews of virtual and hybrid instruction. In SC READY, the largest adjusted differences appear in Math (Finding 6). In EOCEP, Algebra 1 shows the strongest overall relationship between instructional delivery type and performance (Findings 1 and 2).

- Compare like with like (e.g., comparing students taking courses for credit recovery with other course retakes) when judging delivery types. Poverty composition differs across instructional groups, and EOCEP participation is lower in several virtual and hybrid groups, especially VirtualSC (Findings 3 and 4). Raw comparisons mix those differences with differences across instructional delivery types. The adjusted analyses account for measured differences where possible, and sizable associations remain.

Appendices

Appendix A: EOCEP Tables and Charts

Table A1. Student-subject records, tested counts, and participation by instructional delivery type

Instruction type	Student-subject records	With a score	Percent tested
Face-to-Face	194,146	186,743	96.2
Unknown	47,681	44,242	92.8
District Virtual	8,841	7,341	83.0
Hybrid	3,985	3,479	87.3
VirtualSC	710	224	31.5
Distance Learning	200	120	60.0
Out-of-State Virtual	36	25	69.4

One record per student per subject; a student appears once for each subject taken.

Table A2. Mean scale scores by subject and instructional delivery type, among tested students (tested count in parentheses)

Instruction type	Algebra 1	English 2	Biology 1	U.S. History
Face-to-Face	72.1 (48,902)	78.9 (46,705)	68.4 (47,254)	69.6 (43,882)
Unknown	74.5 (10,266)	77.8 (11,707)	69.3 (10,581)	69.1 (11,688)
District Virtual	59.2 (1,911)	74.4 (1,808)	59.7 (1,647)	58.4 (1,975)
Hybrid	58.2 (880)	71.0 (873)	55.9 (776)	54.9 (950)
VirtualSC	53.6 (45)	59.8 (31)	56.3 (79)	50.0 (69)
Distance Learning	51.0 (57)	62.1 (27)	48.6 (21)	*
Out-of-State Virtual	*	*	*	*

** fewer than 20 tested students, suppressed. Pass rates: Table 2 in the main report.*

Table A3. Statistical tests of score differences by instructional delivery type

Subject	Welch F	p	Equal variances	Percent of variation associated with delivery type
Algebra 1	712.9	under .001	rejected	3.91
English 2	124.3	under .001	rejected	0.85
Biology 1	188.9	under .001	rejected	1.18
U.S. History	316.6	under .001	rejected	1.67

Welch's analysis of variance, used because group variances are unequal in every subject. Pairwise: Games-Howell finds 47 of 55 comparisons significant, Tukey 43; the two procedures agree on 51 of 55, with all four disagreements in small-group pairs, treated as not established.

Table A4. Adjusted differences in passing without the poverty term (percentage points versus Face-to-Face)

Subject	Instruction type	Tested	Adjusted difference	95% interval	Significant
Algebra 1	District Virtual	1,911	-31.2	-36.8 to -25.7	Yes
Algebra 1	Hybrid	880	-23.4	-31.9 to -14.8	Yes
Algebra 1	Distance Learning	57	-28.8	-39.8 to -17.8	Yes
Algebra 1	VirtualSC	45	-49.7	-65.4 to -34.0	Yes
Algebra 1	Unknown	10,266	-0.8	-10.3 to +8.7	No

English 2	District Virtual	1,808	-27.3	-34.0 to -20.6	Yes
English 2	Hybrid	873	-28.9	-42.2 to -15.6	Yes
English 2	VirtualSC	31	-62.2	-77.4 to -46.9	Yes
English 2	Unknown	11,707	+1.7	-9.1 to +12.5	No
Biology 1	District Virtual	1,647	-17.5	-29.7 to -5.3	Yes
Biology 1	Hybrid	776	-38.0	-44.4 to -31.7	Yes
Biology 1	VirtualSC	79	-19.8	-51.4 to +11.8	No
Biology 1	Unknown	10,581	+12.6	+0.5 to +24.8	Yes
U.S. History	District Virtual	1,975	-19.0	-29.0 to -9.1	Yes
U.S. History	Hybrid	950	-18.9	-30.6 to -7.1	Yes
U.S. History	VirtualSC	69	-35.9	-51.6 to -20.2	Yes
U.S. History	Unknown	11,688	+1.5	-11.2 to +14.2	No

Linear probability models; school, grade, and test administration window fixed effects; school clustered errors. Tested = students with a score in the group and subject. Groups with too few schools or tested students within a subject are excluded per the reliability standard (Table A5 note).

Table A5. Adjusted differences in passing with the poverty term (percentage points versus Face-to-Face; source of Table 2)

Subject	Model term	Tested	Adjusted difference	95% interval	Significant
Algebra 1	District Virtual	1,911	-32.1	-37.7 to -26.5	Yes
Algebra 1	Hybrid	880	-23.2	-31.9 to -14.5	Yes
Algebra 1	Distance Learning	57	-25.6	-37.0 to -14.1	Yes
Algebra 1	VirtualSC	45	-48.1	-63.6 to -32.5	Yes
Algebra 1	Unknown	10,266	-1.0	-10.4 to +8.4	No
Algebra 1	Pupil in poverty		-11.4	-12.7 to -10.2	Yes
English 2	District Virtual	1,808	-26.7	-33.5 to -19.9	Yes
English 2	Hybrid	873	-28.9	-41.9 to -16.0	Yes
English 2	VirtualSC	31	-59.2	-74.4 to -43.9	Yes
English 2	Unknown	11,707	+0.9	-9.8 to +11.6	No
English 2	Pupil in poverty		-13.3	-14.4 to -12.1	Yes
Biology 1	District Virtual	1,647	-18.9	-30.7 to -7.1	Yes
Biology 1	Hybrid	776	-35.3	-41.2 to -29.4	Yes
Biology 1	VirtualSC	79	-17.7	-47.2 to +11.7	No
Biology 1	Unknown	10,581	+11.3	-0.3 to +22.8	No
Biology 1	Pupil in poverty		-20.0	-21.5 to -18.5	Yes
U.S. History	District Virtual	1,975	-19.2	-28.9 to -9.6	Yes
U.S. History	Hybrid	950	-20.3	-32.1 to -8.4	Yes
U.S. History	VirtualSC	69	-30.6	-44.1 to -17.1	Yes
U.S. History	Unknown	11,688	+2.3	-8.8 to +13.4	No
U.S. History	Pupil in poverty		-21.2	-22.6 to -19.8	Yes

Same models as Table A4 with a student poverty indicator added. Reliability standard: at least 10 schools and 20 tested students per group within a subject; Out-of-State Virtual fails in every subject, Distance Learning in all but Algebra 1.

Table A6. Poverty composition by instructional delivery type among tested EOCEP records

Instruction type	Tested student-subject records	Percent pupils in poverty
Distance Learning	120	86.6
VirtualSC	224	83.0
Out-of-State Virtual	25	76.0

Hybrid	3,479	69.7
District Virtual	7,341	61.6
Unknown	44,242	61.5
Face-to-Face	186,743	57.9

Poverty percentages are calculated among records with known poverty status.

Table A7. Pass rates by poverty status within instructional delivery type, all subjects

Subject	Poverty status	Instruction type	Tested	Percent passing
Algebra 1	In poverty	Face-to-Face	28,513	45.5
Algebra 1	In poverty	Unknown	6,131	51.4
Algebra 1	In poverty	District Virtual	1,231	15.5
Algebra 1	In poverty	Hybrid	606	12.9
Algebra 1	In poverty	VirtualSC	41	2.4
Algebra 1	In poverty	Distance Learning	35	2.9
Algebra 1	In poverty	Out-of-State Virtual	*	*
Algebra 1	Not in poverty	Face-to-Face	18,942	72.8
Algebra 1	Not in poverty	Unknown	3,740	79.6
Algebra 1	Not in poverty	District Virtual	661	29.2
Algebra 1	Not in poverty	Hybrid	270	27.8
Algebra 1	Not in poverty	VirtualSC	*	*
Algebra 1	Not in poverty	Distance Learning	*	*
Algebra 1	Not in poverty	Out-of-State Virtual	*	*
English 2	In poverty	Face-to-Face	26,336	64.8
English 2	In poverty	Unknown	6,981	61.9
English 2	In poverty	District Virtual	1,112	56.1
English 2	In poverty	Hybrid	618	46.9
English 2	In poverty	VirtualSC	28	10.7
English 2	In poverty	Distance Learning	21	33.3
English 2	Not in poverty	Face-to-Face	19,350	86.2
English 2	Not in poverty	Unknown	4,404	85.1
English 2	Not in poverty	District Virtual	681	69.3
English 2	Not in poverty	Hybrid	255	58.8
English 2	Not in poverty	VirtualSC	*	*
English 2	Not in poverty	Distance Learning	*	*
Biology 1	In poverty	Face-to-Face	26,819	34.5
Biology 1	In poverty	Unknown	6,306	38.0
Biology 1	In poverty	District Virtual	1,001	22.0
Biology 1	In poverty	Hybrid	558	16.7
Biology 1	In poverty	VirtualSC	60	16.7
Biology 1	In poverty	Distance Learning	*	*
Biology 1	In poverty	Out-of-State Virtual	*	*
Biology 1	Not in poverty	Face-to-Face	19,456	65.9
Biology 1	Not in poverty	Unknown	3,959	68.2
Biology 1	Not in poverty	District Virtual	641	34.9
Biology 1	Not in poverty	Hybrid	218	26.1
Biology 1	Not in poverty	VirtualSC	*	*
Biology 1	Not in poverty	Distance Learning	*	*
Biology 1	Not in poverty	Out-of-State Virtual	*	*
U.S. History	In poverty	Face-to-Face	24,137	36.7
U.S. History	In poverty	Unknown	6,962	38.4
U.S. History	In poverty	District Virtual	1,153	21.1
U.S. History	In poverty	Hybrid	640	16.4
U.S. History	In poverty	VirtualSC	56	7.1

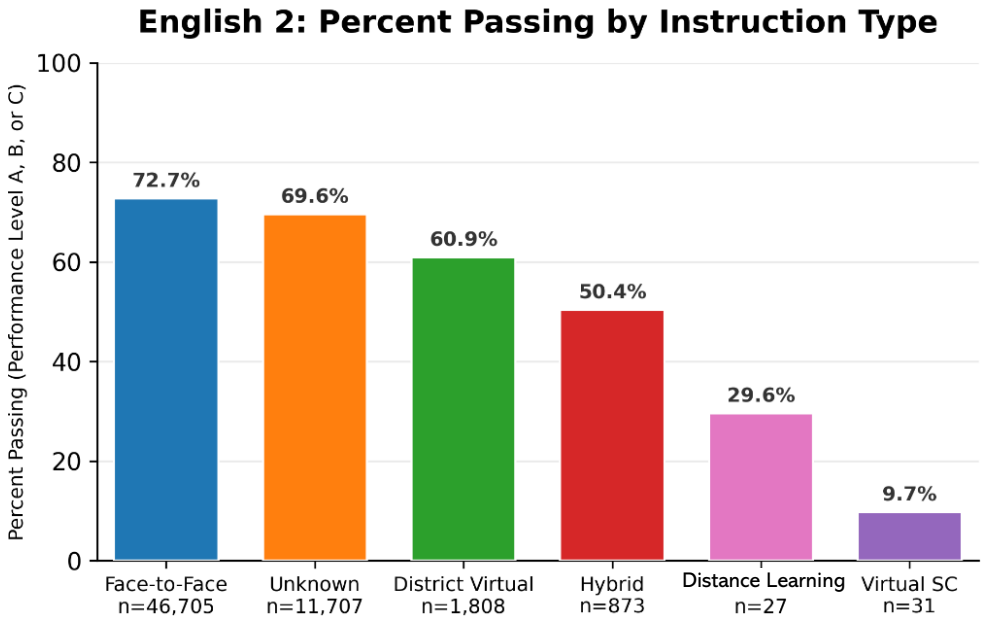
U.S. History	In poverty	Distance Learning	*	*
U.S. History	In poverty	Out-of-State Virtual	*	*
U.S. History	Not in poverty	Face-to-Face	19,186	66.6
U.S. History	Not in poverty	Unknown	4,410	67.6
U.S. History	Not in poverty	District Virtual	816	35.0
U.S. History	Not in poverty	Hybrid	309	27.8
U.S. History	Not in poverty	VirtualSC	*	*
U.S. History	Not in poverty	Distance Learning	*	*

** fewer than 20 tested students in the cell, suppressed. Source of Figure 3.*

(Appendix A continued on next page)

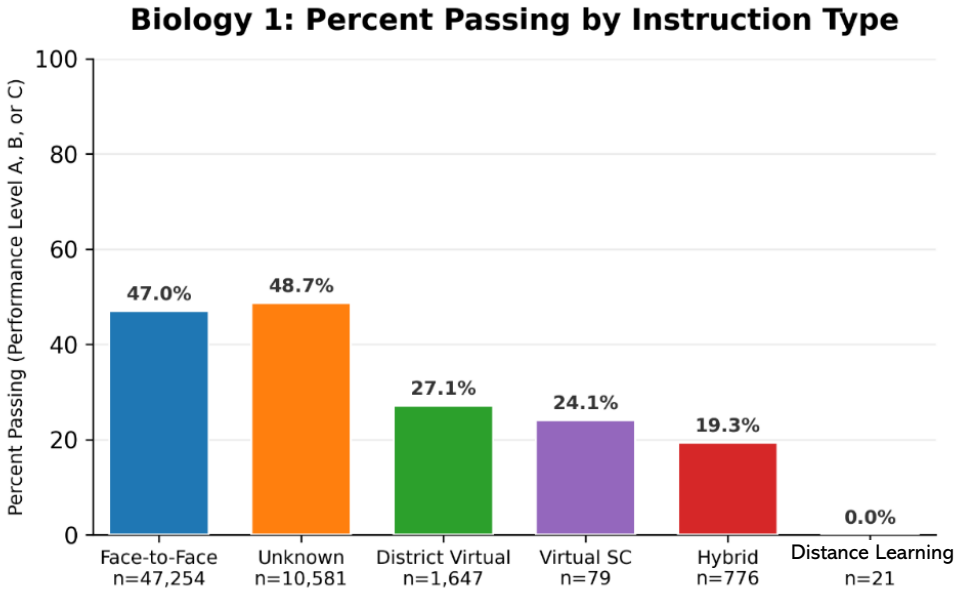
Appendix A figures

Figure A1: Percent passing by delivery type, English 2.



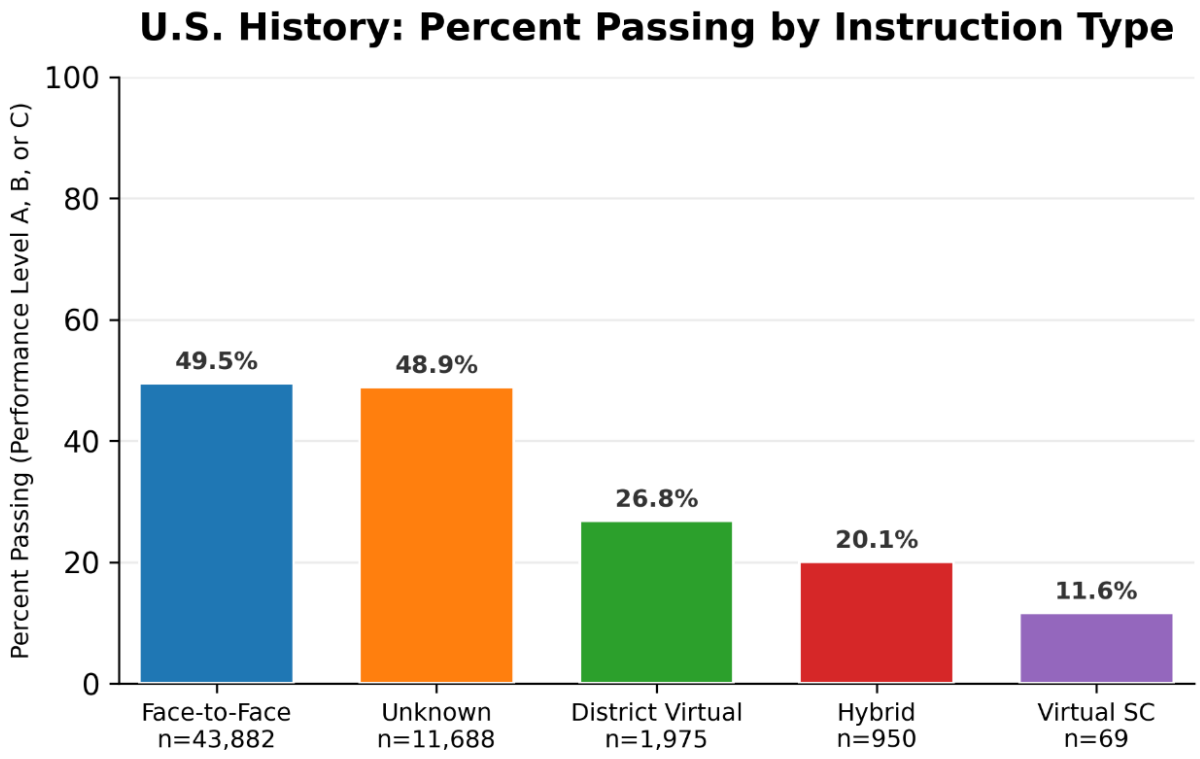
Note: Among tested students; n = tested students in the group. Formats with fewer than 20 tested students are not shown. Virtual SC results describe its tested minority (31.5 percent of enrollees).

Figure A2: Percent passing by delivery type, Biology 1.



Note: Among tested students; n = tested students in the group. Formats with fewer than 20 tested students are not shown. Virtual SC results describe its tested minority (31.5 percent of enrollees).

Figure A3: Percent passing by delivery type, U.S. History.



Note: Among tested students; n = tested students in the group. Formats with fewer than 20 tested students are not shown. Virtual SC results describe its tested minority (31.5 percent of enrollees).

Figure A4: Adjusted differences in passing versus Face-to-Face students in the same school, English 2 (companion to Figure 2).

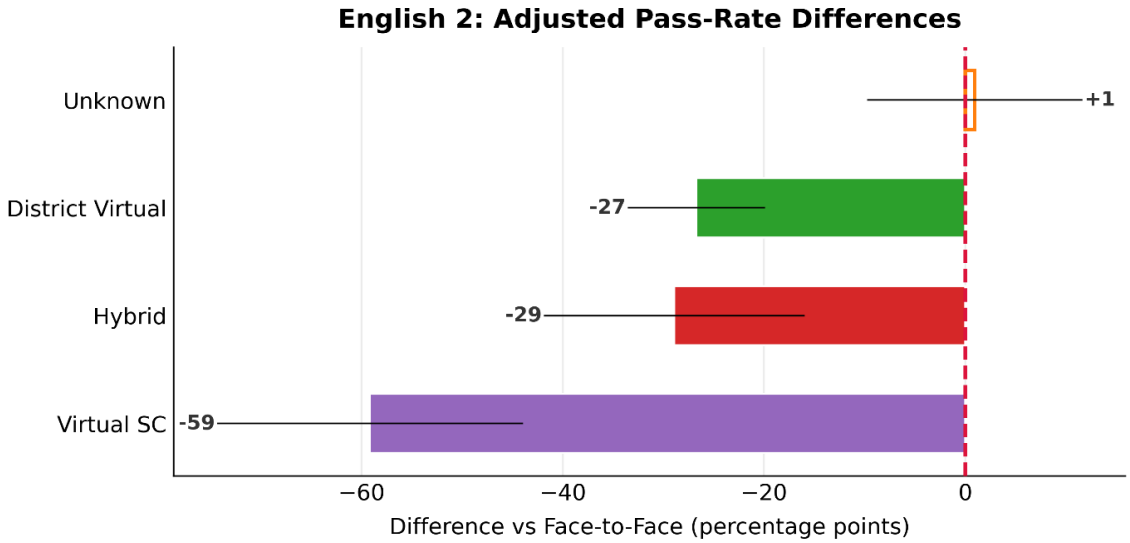


Figure A5: Adjusted differences in passing, Biology 1.

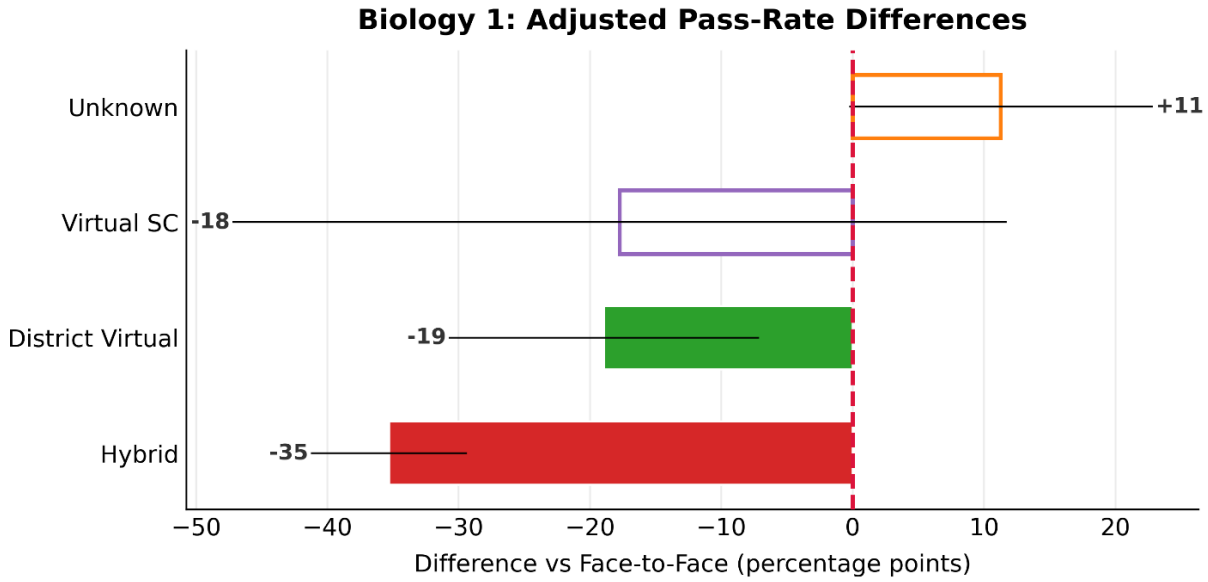


Figure A6: Adjusted differences in passing, U.S. History.

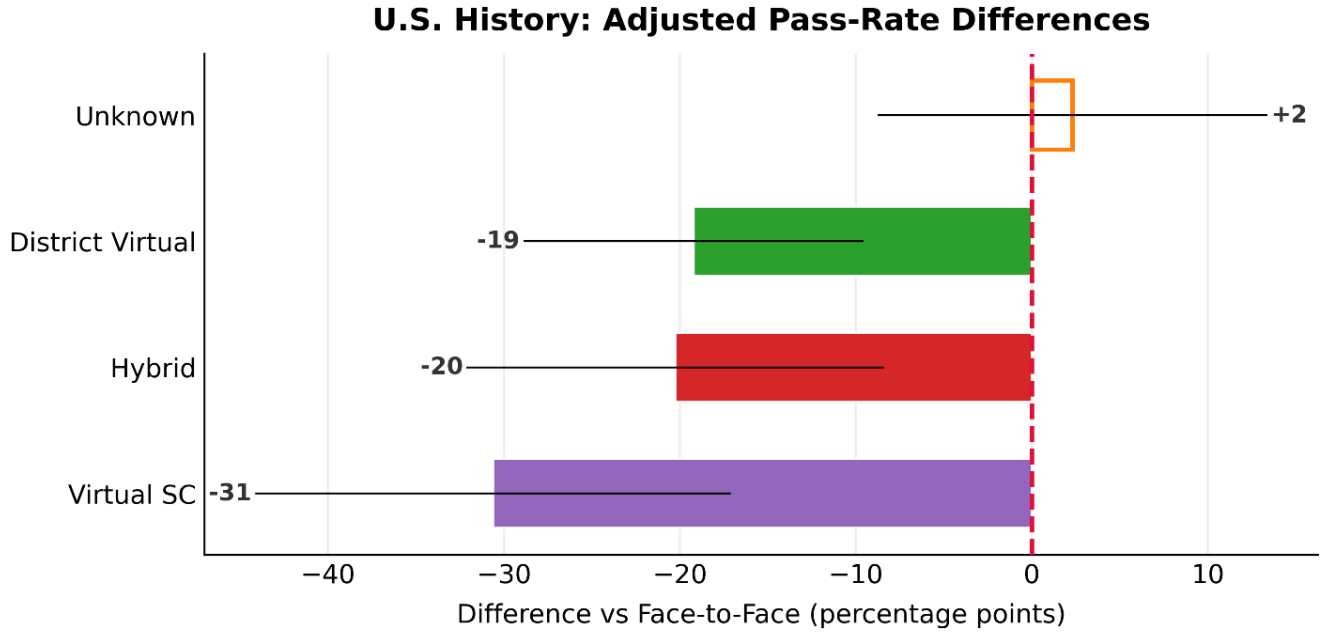
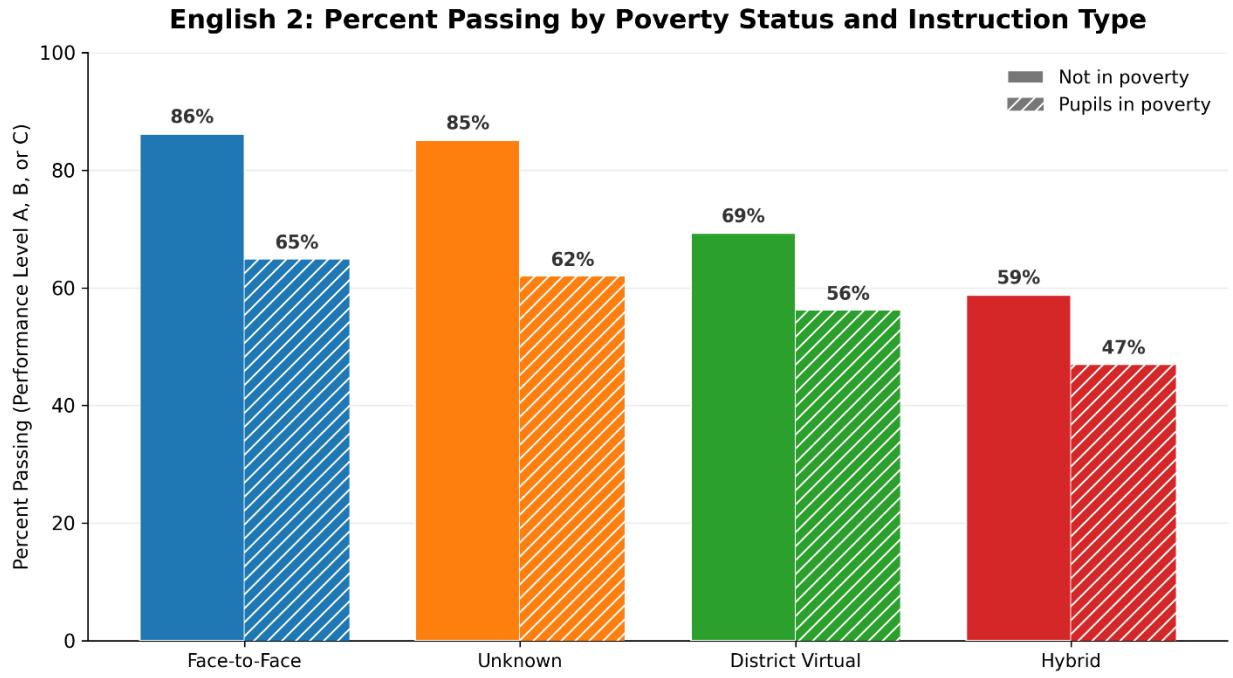
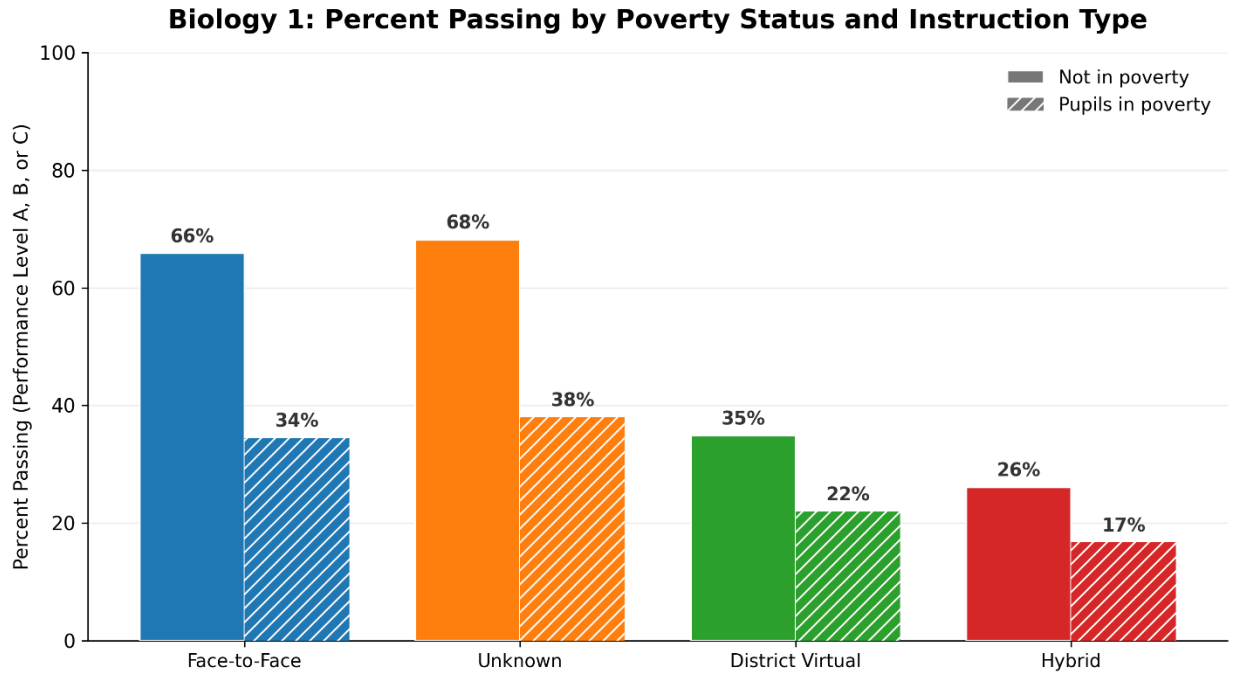


Figure A7: Pass rates by poverty status within delivery type, English 2 (companion to Figure 3).



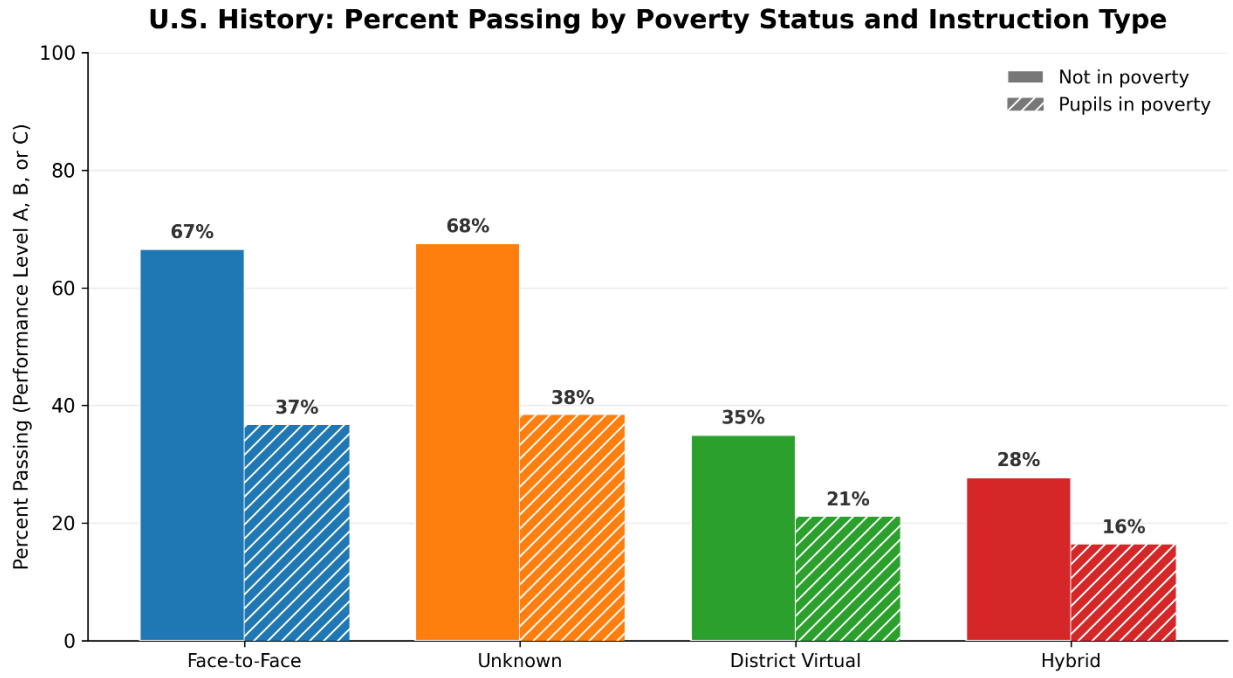
Note: Among tested students with known poverty status. Solid bars: students not in poverty. Hatched bars: pupils in poverty. Formats with fewer than 20 tested students in either group are not shown.

Figure A8: Pass rates by poverty status within delivery type, Biology 1.



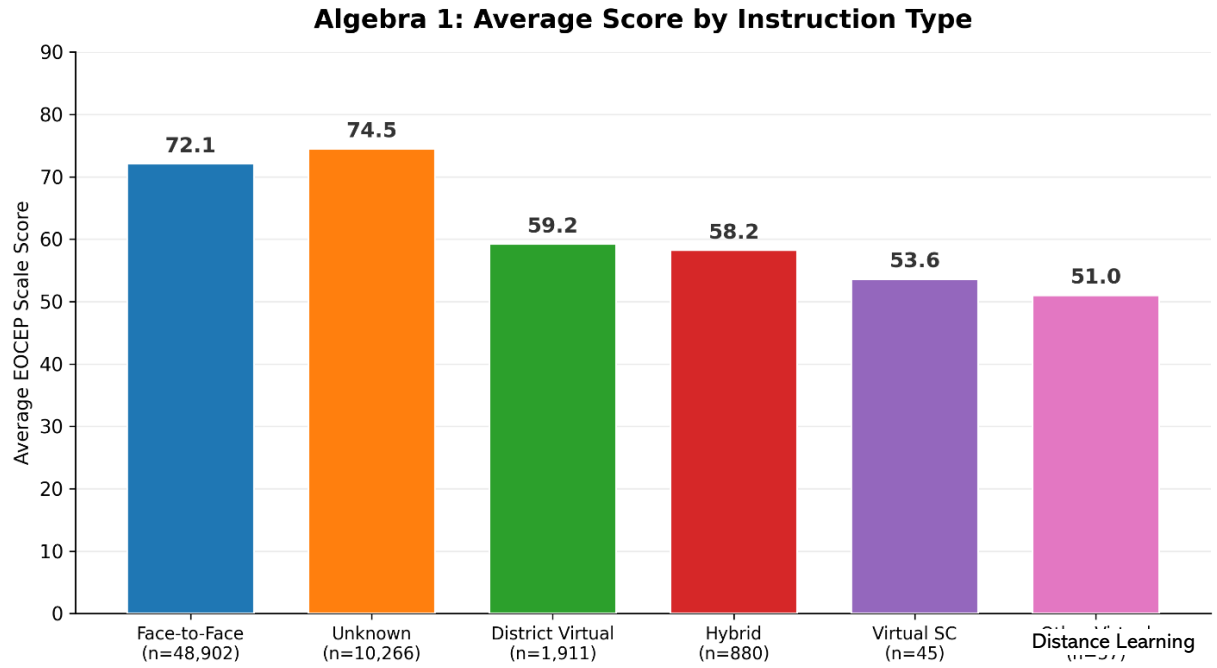
Note: Among tested students with known poverty status. Solid bars: students not in poverty. Hatched bars: pupils in poverty. Formats with fewer than 20 tested students in either group are not shown.

Figure A9: Pass rates by poverty status within delivery type, U.S. History.



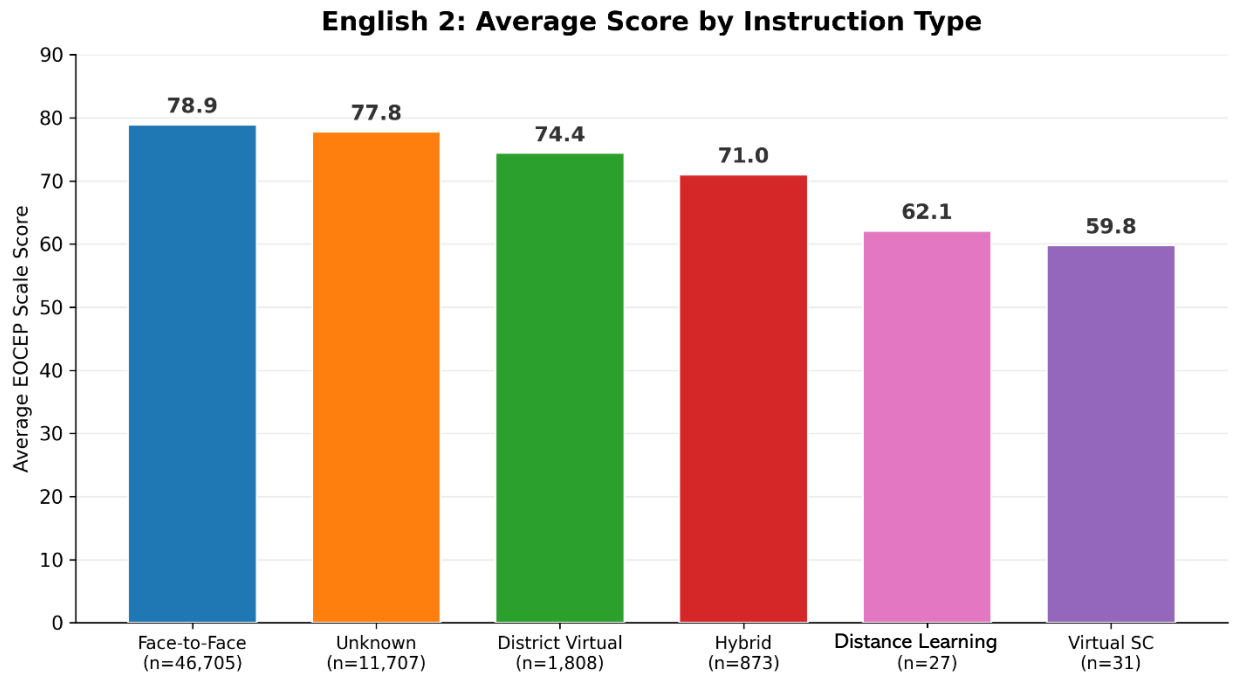
Note: Among tested students with known poverty status. Solid bars: students not in poverty. Hatched bars: pupils in poverty. Formats with fewer than 20 tested students in either group are not shown.

Figure A10: Average scale score by delivery type, Algebra 1 (source data: Table A2).



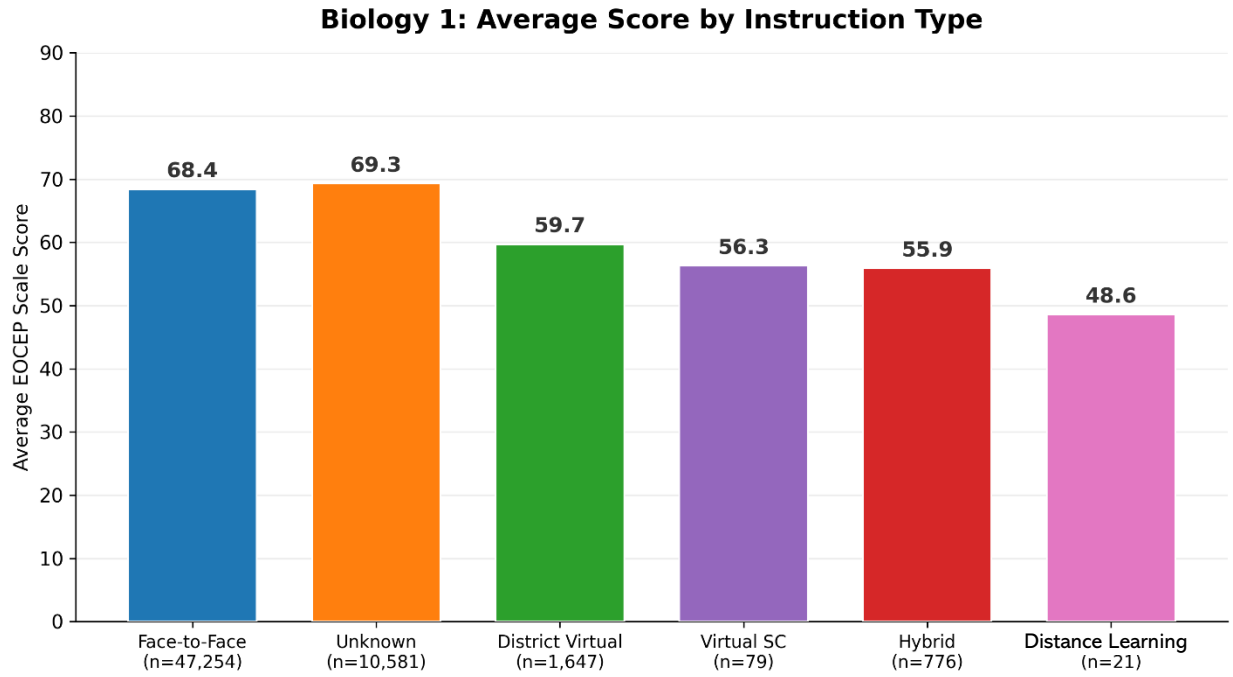
Note: Among tested students; n = tested students in the group. Formats with fewer than 20 tested students are not shown. Virtual SC results describe its tested minority (31.5 percent of enrollees).

Figure A11: Average scale score by delivery type, English 2.



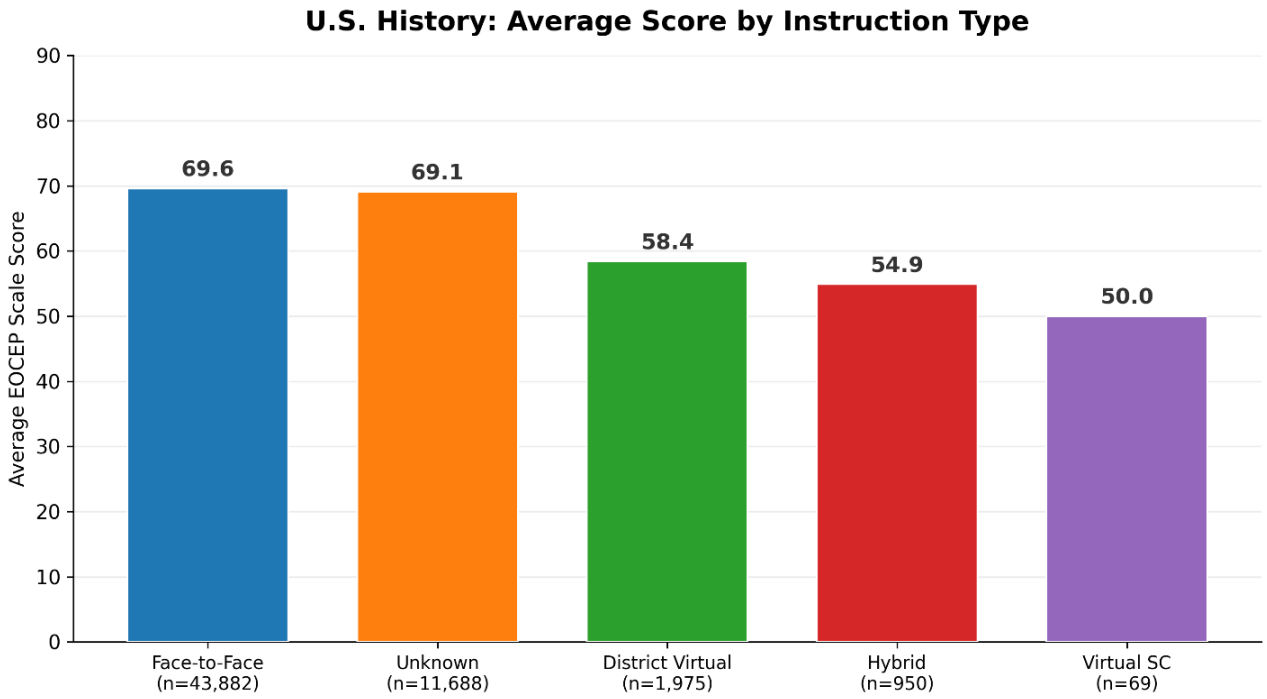
Note: Among tested students; n = tested students in the group. Formats with fewer than 20 tested students are not shown. Virtual SC results describe its tested minority (31.5 percent of enrollees).

Figure A12: Average scale score by delivery type, Biology 1.



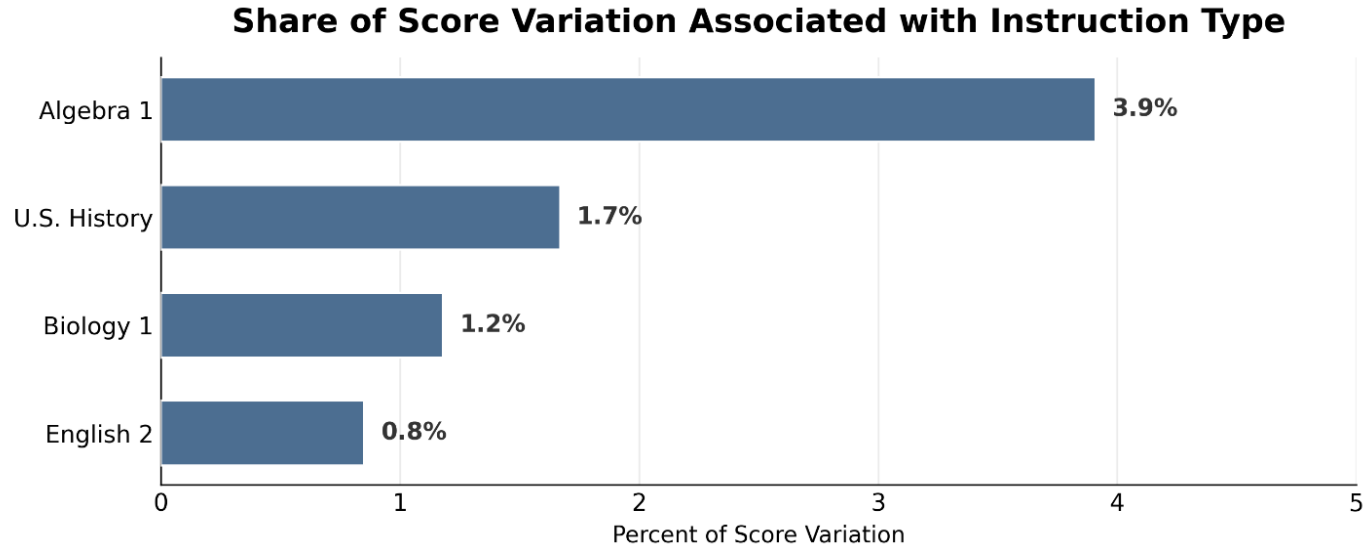
Note: Among tested students; n = tested students in the group. Formats with fewer than 20 tested students are not shown. Virtual SC results describe its tested minority (31.5 percent of enrollees).

Figure A13: Average scale score by delivery type, U.S. History.



Note: Among tested students; n = tested students in the group. Formats with fewer than 20 tested students are not shown. Virtual SC results describe its tested minority (31.5 percent of enrollees).

Figure A14: Share of score variation associated with delivery type by subject.



Note: Partial eta squared from Welch's ANOVA, among tested students. Association, not causation: groups are observational, not assigned. Even the largest share leaves over 96 percent of variation associated with other factors.

Appendix B: SC READY Tables and Charts

Table B1. SC READY participation by instructional delivery type

Instruction type	ELA students	Percent tested (ELA)	Math students	Percent tested (Math)
Face-to-Face	269,906	97.2	272,423	97.2
Unknown	36,163	97.0	33,763	97.3
District Virtual	5,297	92.5	5,401	92.8
Hybrid	2,161	93.2	2,039	92.8
VirtualSC	153	97.4	53	96.2
Distance Learning	46	73.9	47	66.0
Out-of-State Virtual	*	*	*	*

** fewer than 20 students, suppressed. Students counted within each subject's delivery assignment.*

Table B2. Mean ELA scale scores by grade and instructional delivery type (tested count in parentheses)

Instruction type	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Face-to-Face	481.3 (47,280)	535.6 (46,282)	553.7 (46,898)	570.2 (47,137)	588.6 (46,935)	581.1 (35,374)
Unknown	500.3 (1,820)	539.2 (1,863)	562.0 (1,704)	573.3 (10,110)	591.2 (10,032)	609.2 (10,634)
District Virtual	444.2 (523)	487.3 (562)	525.2 (675)	555.0 (993)	571.4 (1,244)	585.5 (1,300)
Hybrid	447.4 (220)	486.3 (235)	524.4 (223)	538.2 (413)	563.1 (532)	578.0 (538)
VirtualSC	508.3 (20)	*	*	*	588.7 (90)	562.1 (25)
Distance Learning	*	*	*	*	*	524.3 (33)
Out-of-State Virtual	*	*	*	*	*	*

** fewer than 20 tested students, suppressed.*

Table B3. Mean Math scale scores by grade and instructional delivery type (tested count in parentheses)

Instruction type	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Face-to-Face	463.0 (47,223)	499.8 (46,378)	540.6 (46,807)	530.7 (47,060)	550.4 (47,535)	552.8 (37,420)
Unknown	475.8 (1,886)	502.8 (1,741)	549.7 (1,789)	533.0 (10,209)	552.9 (9,622)	565.0 (8,516)
District Virtual	380.5 (529)	418.0 (585)	466.2 (677)	471.3 (998)	504.7 (1,246)	524.1 (1,366)
Hybrid	384.2 (225)	417.1 (238)	465.8 (227)	474.3 (384)	505.6 (402)	517.4 (563)
VirtualSC	*	*	*	*	525.7 (28)	*
Distance Learning	*	*	*	*	*	452.3 (33)
Out-of-State Virtual	*	*	*	*	*	*

** fewer than 20 tested students, suppressed.*

Table B4. Prior-achievement-adjusted differences, full model output (scale points versus Face-to-Face)

Subject	Model	Instruction type	Adjusted difference (points)	95% interval	Significant
ELA	ANCOVA	District Virtual	-10.8	-12.2 to -9.3	Yes
ELA	ANCOVA	Hybrid	-8.0	-10.3 to -5.6	Yes
ELA	ANCOVA	Distance Learning	+6.2	-16.6 to +29.0	No
ELA	ANCOVA	Unknown	+0.9	+0.3 to +1.5	Yes
ELA	ANCOVA	VirtualSC	-1.7	-10.2 to +6.8	No
ELA	ANCOVA + poverty	District Virtual	-10.5	-11.9 to -9.0	Yes
ELA	ANCOVA + poverty	Hybrid	-7.3	-9.7 to -4.9	Yes
ELA	ANCOVA + poverty	Distance Learning	+6.2	-16.9 to +29.3	No
ELA	ANCOVA + poverty	Unknown	+1.0	+0.4 to +1.6	Yes
ELA	ANCOVA + poverty	VirtualSC	-1.8	-10.3 to +6.7	No
Math	ANCOVA	District Virtual	-23.8	-25.4 to -22.2	Yes
Math	ANCOVA	Hybrid	-22.5	-24.9 to -20.0	Yes
Math	ANCOVA	Distance Learning	-28.5	-43.4 to -13.7	Yes
Math	ANCOVA	Unknown	-2.1	-2.8 to -1.4	Yes
Math	ANCOVA	VirtualSC	-0.4	-14.1 to +13.4	No
Math	ANCOVA + poverty	District Virtual	-24.0	-25.6 to -22.4	Yes
Math	ANCOVA + poverty	Hybrid	-22.4	-24.9 to -20.0	Yes
Math	ANCOVA + poverty	Distance Learning	-28.4	-43.2 to -13.6	Yes
Math	ANCOVA + poverty	Unknown	-1.9	-2.6 to -1.3	Yes
Math	ANCOVA + poverty	VirtualSC	-0.5	-15.0 to +14.0	No

Grades 4 to 8 with a prior-year score (240,601 ELA; 240,834 Math); the first model adjusts for prior score and grade, the second adds student poverty; robust intervals; groups with fewer than 20 students with a prior-year score excluded (Out-of-State Virtual had 1). Source of Table 4.

Table B5. Poverty composition by instructional delivery type, SC READY analysis population

Instruction type	ELA student assignments	ELA percent in poverty	Math student assignments	Math percent in poverty
Hybrid	2,161	71.9	2,039	69.6
Distance Learning	46	70.5	47	71.1
District Virtual	5,297	67.3	5,401	67.2
VirtualSC	153	65.3	53	60.8
Face-to-Face	269,906	62.8	272,423	62.7
Unknown	36,163	62.5	33,763	62.8

Poverty percentages are calculated among students with known poverty status.

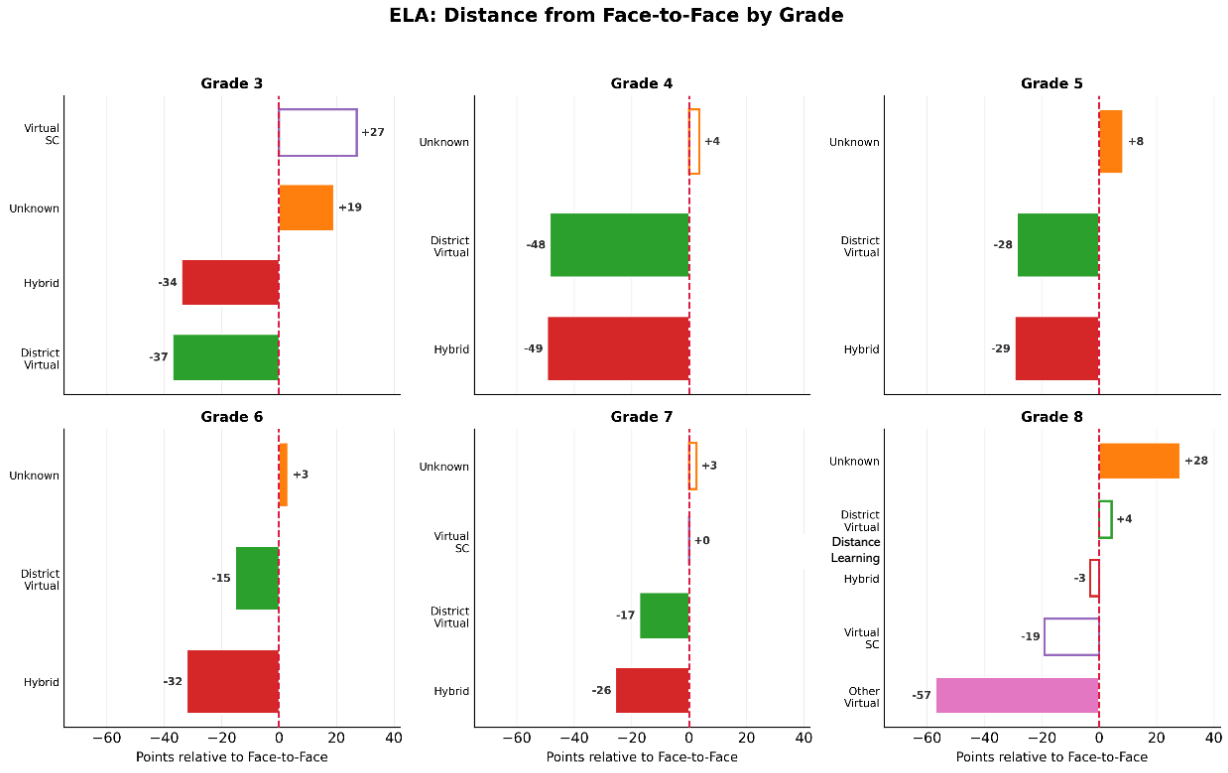
Table B6. Share of score variation associated with instructional delivery type, by grade and subject

Grade	Subject	Percent of variation
3	ELA	0.29
4	ELA	0.31
5	ELA	0.18
6	ELA	0.12
7	ELA	0.17
8	ELA	2.01
3	Math	0.75
4	Math	0.80
5	Math	0.85
6	Math	0.65
7	Math	0.52
8	Math	1.01

Welch's analysis of variance. All overall Welch tests were significant at p < .001. The grade 8 ELA value reflects the Unknown group (Finding 7).

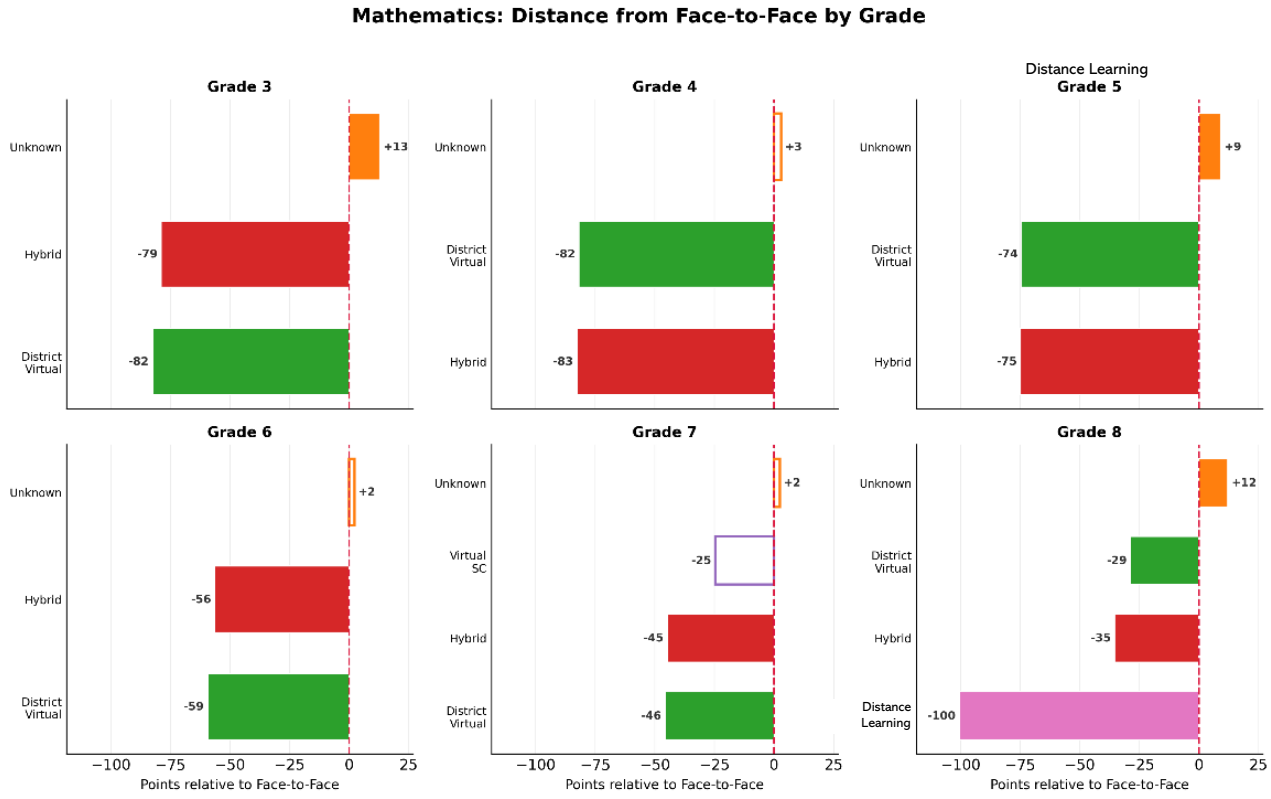
Appendix B figures

Figure B1. ELA score differences versus Face-to-Face by grade.



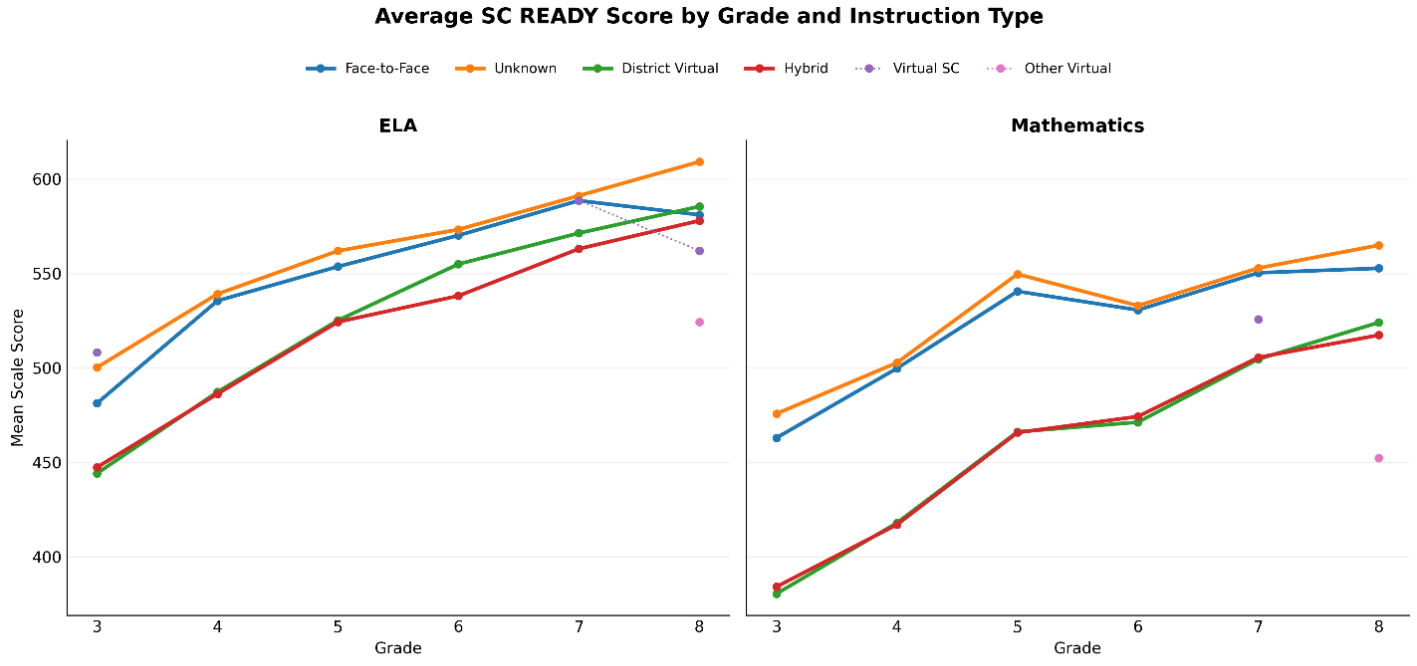
Note: Among tested students; formats with fewer than 20 tested students in a grade are not shown. Solid bars differ significantly from Face-to-Face (Games-Howell); open bars are not statistically distinguishable.

Figure B2. Math score differences versus Face-to-Face by grade.



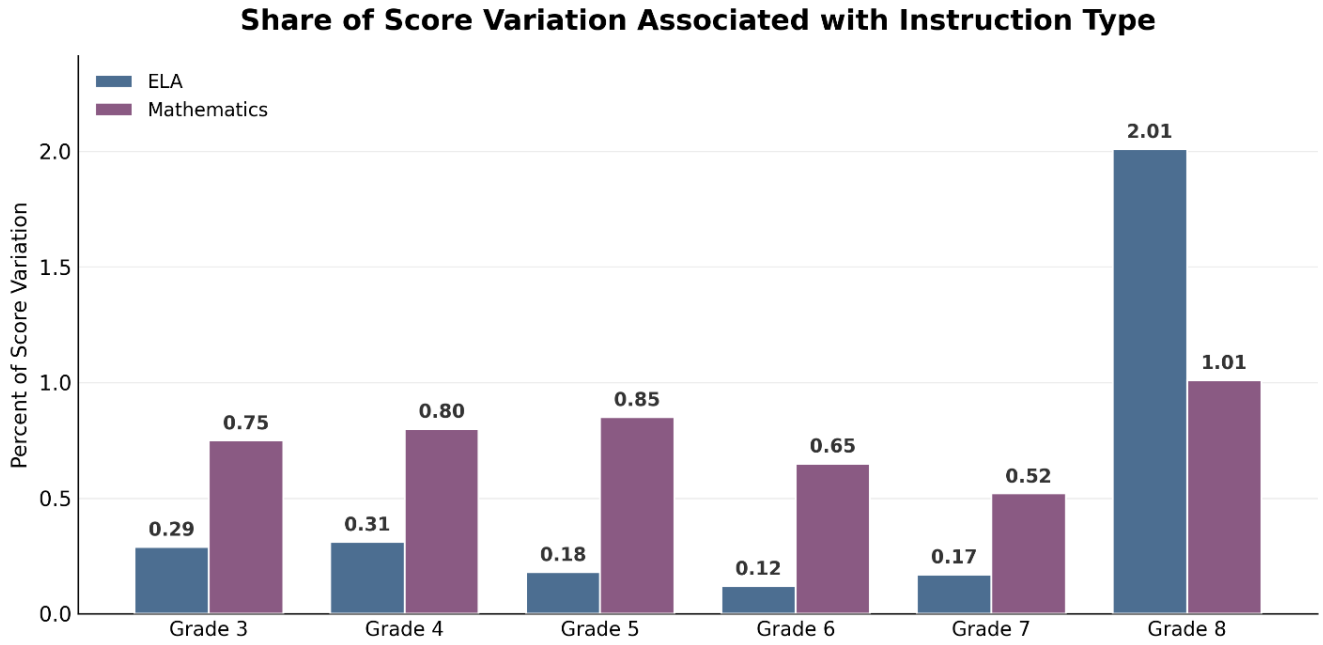
Note: Among tested students; formats with fewer than 20 tested students in a grade are not shown. Solid bars differ significantly from Face-to-Face (Games-Howell); open bars are not statistically distinguishable.

Figure B3. Mean SC READY scores by grade and instructional delivery type.



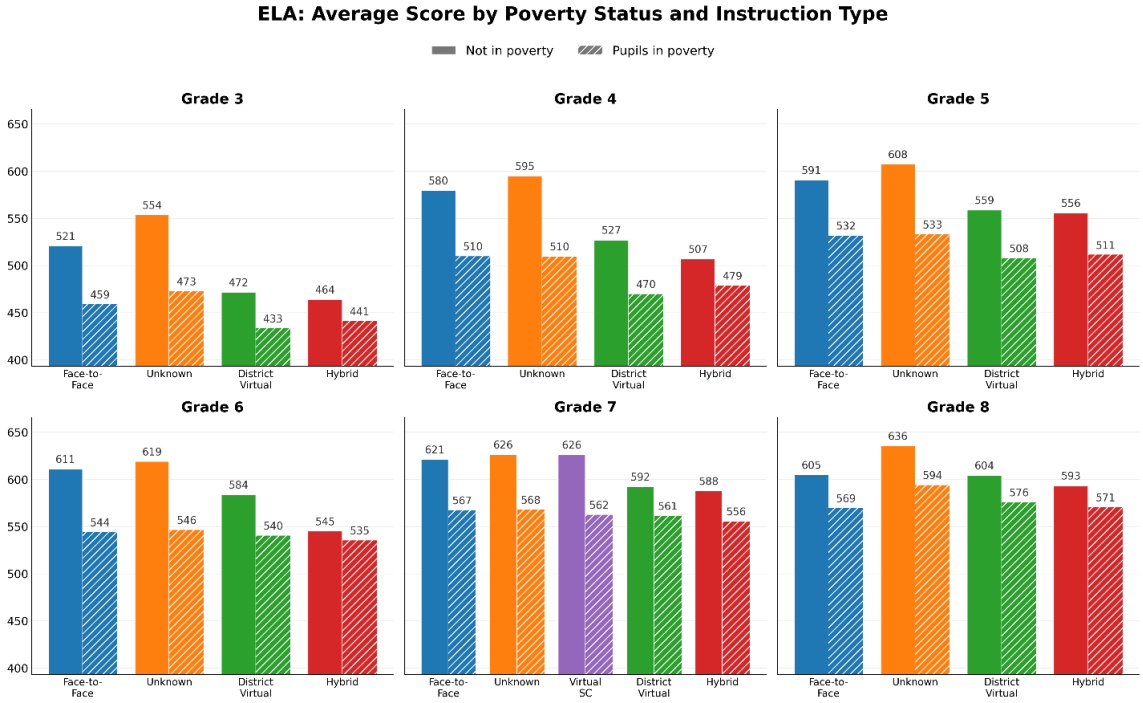
Note: Among tested students. A line is broken in grades where the format has fewer than 20 tested students. Thin dotted lines: small formats.

Figure B4. Share of score variation associated with instructional delivery type, by grade and subject.



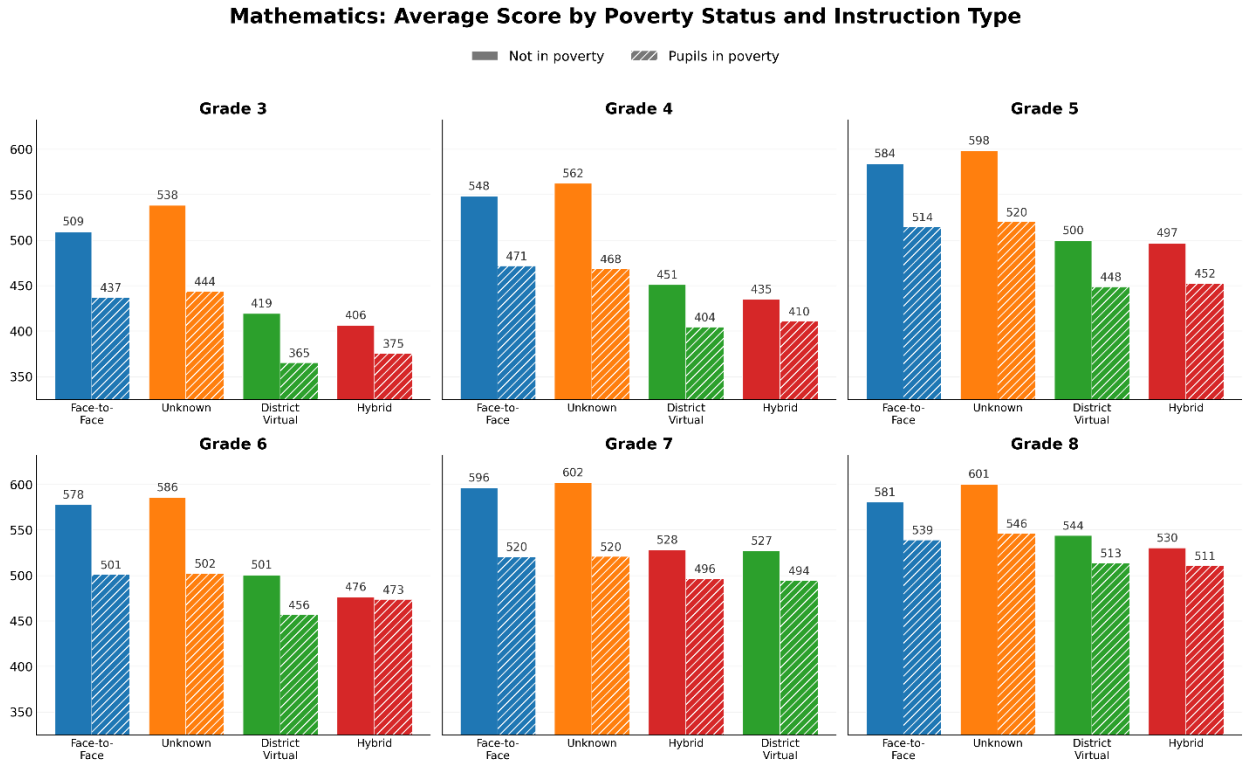
Note: Partial eta squared from Welch's ANOVA, among tested students. Association, not causation: groups are observational, not assigned. The grade 8 ELA bar reflects the Unknown group, not delivery type (see report text).

Figure B5. ELA scores by poverty status and instructional delivery type, by grade.



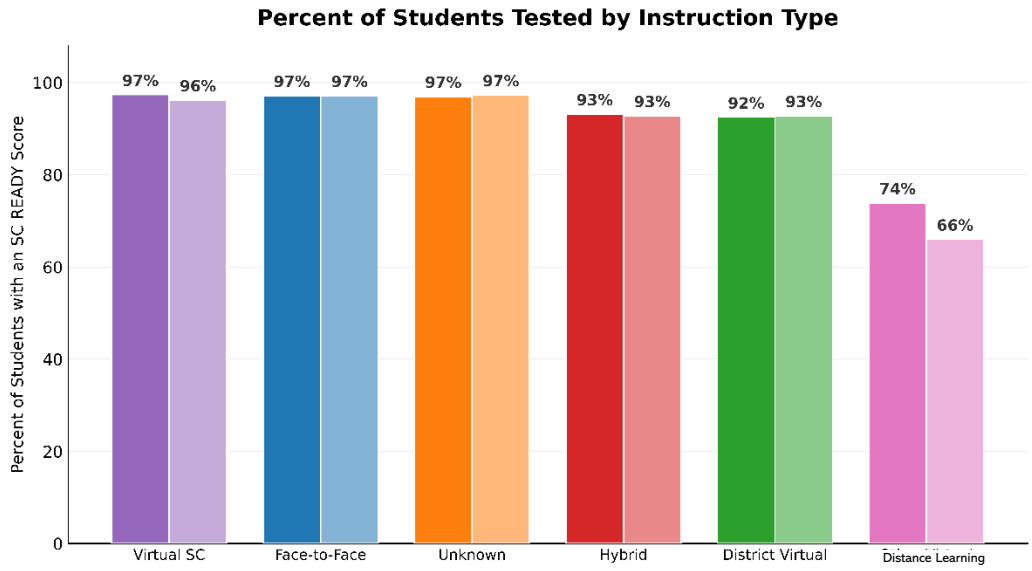
Note: Among tested students with known poverty status. Solid: not in poverty. Hatched: pupils in poverty. Types with fewer than 20 tested students in either group are not shown. All panels share one score scale; scores remain grade-specific tests.

Figure B6. Math scores by poverty status and instructional delivery type, by grade.



Note: Among tested students with known poverty status. Solid: not in poverty. Hatched: pupils in poverty. Types with fewer than 20 tested students in either group are not shown. All panels share one score scale; scores remain grade-specific tests.

Figure B7. SC READY test participation by instructional delivery type.



Note: Of students whose subject was delivered by each format, the share with a score. Solid: ELA. Lighter: Math. Formats under 20 students not shown.

Appendix C: Data Preparation, Methods, and Sensitivity Checks

Units of analysis. SC READY: one record per student per subject (313,727 students, grades 3 to 8). EOCEP: one record per student per subject (255,599 records for 183,612 students from 301,744 eligible enrollments, grades 7 to 12; 242,174 records carry a score). Each student's coursework is classified into one delivery type per subject: among coded eligible courses, the type with the most instructional days; ties resolve to the non-Face-to-Face type; if no coded subject course is available, the assignment is Unknown. In SC READY, Unknown also includes students for whom no course for that subject is found in the enrollment files.

Assignment correction. An earlier specification allowed 6,800 EOCEP student-subject records, 2.7 percent, to appear in more than one delivery group with the same score. After the unit was corrected, headline pass rates moved by one to three points and VirtualSC participation moved from 35.9 to 31.5 percent. The earlier EOCEP models also omitted the documented test administration window control. Correcting the unit and adding that control produced somewhat larger adjusted EOCEP differences, including Algebra 1 District Virtual moving from about minus 25 to minus 32 percentage points. The SC READY correction changed the adjusted results very little.

Eligibility sensitivity. Restricting EOCEP eligibility to a narrow core-course list produced substantively similar participation and pass rates. The narrower definition retained 93.6 percent of eligible EOCEP enrollment rows; VirtualSC participation was 32.6 percent under the restriction.

Models. EOCEP: linear probability models for passing and score models in standard deviation units; school, grade, and test administration window fixed effects; standard errors clustered by school; estimates reported only for groups with at least 10 schools and 20 tested students within a subject. SC READY: analysis of covariance on the SY 2024-25 score with prior-year score and grade as covariates, with and without a student poverty indicator; heteroskedasticity-robust intervals; grade 3 excluded (no prior test).

Data quality rules. Duplicate EOCEP tests (131 students) resolved by keeping the highest score; 537 EOCEP tester records and 59 prior-year records with invalid identifiers removed. Poverty status from 2025 Report Card files covered 98.3 percent of the EOCEP analysis population and 97.7 percent of the SC READY analysis population; students with conflicting records were classified as pupils in poverty if any record said so. Results for any group with fewer than 20 tested students are suppressed throughout.

1205 Pendleton Street
Room 502 Brown Building
Columbia, SC 29201
www.eoc.sc.gov

SC EDUCATION
OVERSIGHT COMMITTEE
Reporting facts. Measuring change. Promoting progress.

The South Carolina Education Oversight Committee (EOC) is an independent, nonpartisan group of 18 educators, business people, and elected officials appointed by the legislature and governor. The EOC enacts the South Carolina Education Accountability Act of 1998, which sets standards for improving the state's K-12 educational system. The EOC reviews the state's education improvement process, assesses how schools are doing, and evaluates the standards schools must meet to build the education system needed to compete in this century.

EDUCATION OVERSIGHT COMMITTEE

DATE: September 21, 2026

SUBCOMMITTEE:
Academic Standards & Assessments Subcommittee

ACTION ITEM:
Evaluation of WIN Assessment

PURPOSE/AUTHORITY
§SECTION 59-18-320 Review of field test; general administration of test; accommodations for students with disabilities; adoption of new standards.

(D) Any new standards and assessments required to be developed and adopted by the State Board of Education, through the Department of Education for use as an accountability measure, must be developed and adopted upon the advice and consent of the Education Oversight Committee.

CRITICAL FACTS
Changes in the new WIN assessment prompted a review of both the process that occurred to change the standards and a psychometric review of the assessment pursuant to Section 59-18-320 of the SC Code of Laws.

TIMELINE/REVIEW PROCESS

- **September 21, 2026:** ASA subcommittee review of UGA review of WIN assessment
- **October 26, 2026:** anticipated EOC review of UGA analysis of WIN assessment; review to be provided to the State Board of Education, the State Department of Education, the Governor, the Senate Education Committee, and the House Education and Public Works Committee

ECONOMIC IMPACT FOR EOC

Cost: $29,902.98 assessment conducted by University of GA K-12 Assessment Solutions)

ACTION REQUEST

☒ **For approval** ☐ **For information**

ACTION TAKEN

☐ **Approved** ☐ **Amended**
☐ **Not Approved** ☐ **Action deferred (explain)**

Mary Frances Early College of Education
K–12 Assessment Solutions
UNIVERSITY OF GEORGIA

WIN Work Ready Assessments

Pre-Operational Assessment and Equating Processes

UGA Gwinnett Campus:
2530 Sever Road
Suite 149
Lawrenceville, GA 30043

Contact Us:
Toll free: (888) 392-8977
Web: https://coe.uga.edu/directory/k-12-assessment-solutions

Table of Contents

1. Purpose and Scope of Review 3
1.1 Assessment Purpose 3
2. Test Design and Blueprint 4
2.1 Development of the Test Blueprint 4
2.2 Item Types 4
2.3 Universal Design Principles 4
2.4 Evaluation and Recommendations 5
3. Item Development and Review 5
3.1 Item Development 6
3.2 Item Review 6
3.3 Evaluation and Recommendations 6
4. Field Testing and Psychometric Analysis 7
4.1 Field Testing 7
4.2 Item Selection and Form Construction 8
4.3 Alignment Study 8
4.4 Other Testing Activities 9
4.5 Evaluation and Recommendations 9
5. Standard Setting and Cut Scores 10
5.1 Standard Setting Methodology 10
5.2 Panelists and Training 11
5.3 Rounds of Judgment and Use of Data 12
5.4 Evaluation and Recommendations 12
6. Scale Scores, Achievement Levels, and Reporting 13
6.1 Scoring and Scale Scores 13
6.2 Achievement Levels 13
6.3 Reporting and Intended Uses of Results 14
6.4 Evaluation and Recommendations 15
7. Conclusion 15
References 17

1. Purpose and Scope of Review

This review evaluates the technical and procedural quality of the pre-operational processes used for the WIN Work Ready Assessments from item and test development to operational use. The purpose of this review is to determine whether assessment development activities were conducted in accordance with generally accepted assessment best practices and whether sufficient evidence exists to support valid and defensible interpretations of scores.

Subject matter experts from the University of Georgia reviewed available documentation and evidence related to assessment design, item development, field testing, psychometric evaluation, form construction, alignment, standard setting, and readiness for operational implementation.

For the purposes of this evaluation, the University of Georgia reviewed the following primary documents provided by the WIN Team:

- WIN Work Ready Assessments Technical Report
- WIN ASA_Alignment Report 2025
- WIN ASA_Field Testing Summary
- WIN ASA_Standard Setting Report 2025

In addition to these reports, the WIN Team also provided ancillary materials and artifacts supporting test development activities, such as administration materials, panel training materials, and rating sheets.

1.1 Assessment Purpose

The WIN Work Ready Assessments were designed to measure foundational academic skills required for success in job-specific training programs and workforce preparation activities. Assessment development was grounded in the U.S. Department of Labor Occupational Information Network (O*NET, released in 1997), an online database that provides profiles for hundreds of occupations including required skills, daily job duties, experience requirements, wages and employment trends, and job zones. Assessment items are aligned to learning objectives linked to O*NET Job Zones and associated workplace skill expectations. The assessment framework was revised in 2024-2025 following a re-review of O*NET data that identified additional workplace readiness skills requiring additional learning objectives (WIN, 2026a, 2026b, 2026d).

The intended interpretation of assessment results is that examinees are credentialed and possess the foundational math, reading, and data literacy skills necessary to successfully engage in training at a particular job level. The results are not intended to indicate readiness for a specific job, but rather readiness for job specific training and positioned for successful progress.

The development of the current assessment edition involved a multi-year process starting with research and framework development, followed by a 2024 update that incorporated new high-demand workplace skills. Key activities to support these updates included refreshing the item pool, developing and field testing new items, conducting alignment studies to ensure the match between learning objectives and operational forms, and establishing new performance standards

and cut scores using a Modified Angoff methodology (Cizek & Bunch, 2007; WIN, 2026b, 2026c, 2026d).

2. Test Design and Blueprint

This section summarizes how the WIN Team translates the O*NET Content Model into test blueprints that define what content is measured on the WIN Work Ready Assessments and how learning objectives complexity progresses over five achievement levels. Consistent with the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014), the WIN Team documents what is being measured, the intended examinee population, and the intended uses of test scores.

2.1 Development of the Test Blueprint

The ***test blueprint*** serves as the primary mechanism for ensuring content alignment and supporting valid interpretations of test scores. Specifically, it describes which learning objectives are assessed, how often they appear, and how much each objective contributes to the overall score (Nunally, 1978; Sireci, 1998).

Assessment development for the WIN Work Ready Assessments was guided by test blueprints organized around learning objectives, achievement levels, and content strands. Work Ready Math, Work Ready Reading, and Work Ready Data each include five achievement levels representing increasing complexity and workplace readiness demands. Learning objectives, discrete skills, and strands were used to define the content domain and guide item development activities (WIN, 2026a, 2026b). Item development, review, and selection processes are governed by the test blueprint.

2.2 Item Types

All assessment items are traditional selected-response (multiple-choice) items, including a stem (question) and requiring examinees to select one correct answer from four options. Available documentation describes item development and blueprinting procedures based on selected-response formats and does not indicate that alternative item types were considered during the development of test specifications or test blueprints.

2.3 Universal Design Principles

Universal design principles are applied from the outset to make tests accessible, fair, and appropriate for the widest range of students (Laitusis & Karvonen, 2025). Rather than addressing accessibility only after tests are developed, the guiding design specifications for the WIN Work Ready Assessments consider accessibility across all phases of development to support valid measurement for all learners.

Universal design principles guide decisions related to language demand, visual presentation, response formats, and interaction design. Items are reviewed for clarity, readability, and potential sources of bias, and accessibility features. These practices are intended to reduce barriers to access by addressing accessibility considerations during assessment development.

The technical report indicates that each item is reviewed at the concept stage and again throughout the development process for bias and sensitivity. Checklists were used to flag stereotyping, gender, ethnicity, religion, or class biases, and emotionally charged topics (WIN, 2026a).

2.4 Evaluation and Recommendations

Evaluation

The WIN documentation indicates that the test blueprint guides item development, form construction, and score reporting, supporting representative coverage of the intended content domain. The WIN test blueprint provides a structured framework linking learning objectives, content strands, and achievement levels to the intended interpretations of scores. The assessment design incorporates procedures intended to promote accessibility, fairness, and content alignment through the application of universal design principles and multiple stages of item review. These practices reduce the need for accommodations by addressing access at the design stage rather than retrofitting accessibility after development.

The selected-response item type supports efficient administration and scoring while providing broad coverage of the targeted content domain. However, some learning objectives may involve skills that could also be measured through performance-based or constructed-response tasks such as “Measure the length, width, and height of products and materials” to fully assess the intent of the skill.

Overall, the documented blueprint and design processes are consistent with generally accepted assessment development practices and provide an appropriate foundation for subsequent development activities.

Recommendations

The WIN Team could expand technical documentation describing the development, review, and approval of test blueprints, including the roles of subject matter experts and other stakeholders involved in decision-making. Additional documentation regarding blueprint governance would strengthen evidence supporting content validity and assessment oversight.

Additionally, future reviews could consider evaluating the alignment between selected-response item types and the skills being measured, particularly for learning objectives that emphasize complex performance or application, to ensure that assessment methods continue to provide the strongest evidence of intended knowledge and skills.

3. Item Development and Review

Item development and review processes are critical to assessment quality because test score validity depends on the quality, alignment, fairness, and accessibility of individual test items. Effective assessment programs use well-established item development procedures, multiple

stages of review, and quality control processes to ensure items accurately measure intended knowledge and skills while minimizing irrelevant barriers to performance.

The WIN documentation describes a structured item development process involving subject matter experts with relevant domain knowledge and experience and well versed in academic and workforce standards, assessment design principles, accessibility, and bias and sensitivity review process. Items were developed using established specifications and underwent multiple review stages before field testing, including content review, editorial review, bias and sensitivity review, accessibility review, and quality assurance review (WIN, 2026a). Item development procedures emphasized alignment with learning objectives, clarity of language, fairness, and adherence to universal design principles. This is in accordance with the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014).

3.1 Item Development

Item development is a multi-step process designed to ensure that test items are accurate and aligned to the learning objectives. This process is typically guided by item specifications and style guidelines that define expectations for content, format, and rigor. The technical report describes the criteria for developing the parts of an item (WIN, 2026a).

Every item is reviewed at the "concept stage" and throughout the development process, particularly for bias and sensitivity issues. Checklists are used to determine no stereotyping is present and there is no bias based on gender, ethnicity, religion, or socioeconomic status. Following stereotyping and bias reviews, items were reviewed for potentially emotionally charged topics.

The item development procedures emphasize alignment to learning objectives, multiple review activities, and consideration of fairness and accessibility throughout development.

3.2 Item Review

Professional editors review items for length, text density, word usage, sentence structure, and grammar to improve readability and clarity. Following this, the executive content development team conducts a final quality assurance review to verify that items align with learning objectives, contain a single correct answer, feature parallel options, remain free of bias or sensitivity issues, and utilize clear language and graphics (WIN, 2026a).

3.3 Evaluation and Recommendations

Evaluation

The documentation indicates that WIN implemented a structured and multi-stage item development and review process designed to support item quality, alignment, fairness, and accessibility. Item development activities were guided by learning objectives and included multiple review points, checking for content accuracy and alignment, bias and sensitivity, editorial quality, and overall quality assurance. This is evidence that multiple safeguards were incorporated to support the quality of items before field testing.

Overall, the documented procedures reflect generally accepted assessment development practices by emphasizing alignment to learning objectives, systematic review, and attention to fairness and accessibility throughout the development process. There is reasonable evidence that item development and review activities were conducted in an industry-accepted manner.

Recommendations

No significant concerns were identified regarding item development and review procedures. Future technical documentation could provide additional information regarding any training that is provided to item developers and reviewer guidelines and process for agreement. Such information would provide additional transparency regarding item quality control and development governance.

4. Field Testing and Psychometric Analysis

Before new test items are used to calculate student scores, they are first administered through a structured field testing process. After test items have been reviewed and approved by content experts for accuracy and by reviewers for potential bias, the items are embedded in operational test forms as field test items. Responses to field test items do not affect scores; rather, they provide data to evaluate how items function.

4.1 Field Testing

Newly developed items were field tested between January and April 2025 using a matrix-sampling design in which unscored field-test items were embedded within operational assessment administrations (WIN, 2026c). Examinees were presented with five to eight embedded field-test items under operational testing conditions. This approach allowed field-test responses to be collected under conditions that closely resembled operational administration while minimizing burden on individual examinees.

Field-test sample sizes were substantial, including approximately 58,427 Work Ready Math examinees, 56,016 Work Ready Reading examinees, and 56,128 Work Ready Data examinees (WIN, 2026c). Approximately half of the sample consisted of South Carolina high school students participating with prior approval from the South Carolina Department of Education.

After field testing is complete, a series of statistical analyses are conducted to evaluate how each item performed. The analyses examine multiple aspects of item quality, including:

- **Difficulty** – whether the item is appropriately challenging for students.
- **Discrimination** – how well the item distinguishes between students with stronger and weaker overall performance.
- **Distractor performance (for selected-response items)** – how students respond to the incorrect answer choices and whether those options function as intended.

These analyses were used to evaluate item performance and identify items that warranted additional review or revision. Items violating one or more criteria were flagged for review and possible revision. Modified items were not immediately transitioned into operational use but instead remained eligible for future field testing to establish new psychometric evidence.

The field testing process generated empirical evidence regarding item functioning under operational testing conditions and informed subsequent psychometric review activities. The large sample size provides a strong basis for evaluating item functioning and supporting psychometric decision-making.

4.2 Item Selection and Form Construction

Form construction is the process of selecting items for operational use in a manner that maintains content representation, psychometric quality, and comparability across test forms. Effective form construction helps ensure that score interpretations remain consistent regardless of which operational form an examinee receives.

WIN documentation indicates that forms were developed collaboratively by the content development team and psychometrician following the results from field testing. Item selection was informed by assessment blueprints and psychometric analyses. The assessments use fixed-form designs, and two parallel forms were developed for each Work Ready assessment. Operational forms were equated using common-item psychometric methodologies to support score comparability across administrations (WIN, 2026a).

4.3 Alignment Study

The WIN Team conducted an alignment study in August 2025 as part of the assessment revision process (WIN, 2026b). The study was designed to evaluate whether assessment items appropriately reflected the intended learning objectives and supported valid score interpretations.

The alignment study was conducted following substantial revisions to Work Ready Math, Work Ready Reading, Work Ready Data, including the addition of new learning objectives identified through a review of updated O*NET workforce-readiness data. Eight subject matter experts participated in the standard setting process. Panelists represented across baccalaureate, graduate, and post-graduate levels and had expertise in K-12 and adult education, workforce experience, and corrections experience. They represented four states in the South; one participant was from South Carolina. Panelists were selected through a structured recruitment and interview process intended to ensure relevant expertise and diverse perspectives (WIN, 2026b).

Prior to conducting the review, panelists received an orientation and training session led by the WIN psychometrician. Reviewers were provided with assessment forms, learning objectives, examples of the skills associated with each objective, and standardized scoring materials. Panelists independently reviewed items and evaluated whether each item appropriately measured its intended learning objective.

The review process required panelists to compare the content and cognitive demands of each item to the specified learning objectives. Because the learning objectives incorporated both knowledge and skills and level of complexity, alignment judgments simultaneously evaluated content representation and degree of rigor. Individual judgments were collected, and agreement rates were calculated across reviewers. Figure 1 shows a sample score sheet that documents individual reviewer decisions for each item.

Figure 1. Panelist Score Sheet [Sample]

	ALIGNMENT			ROUND 1					ROUND 2				
Question #	FORM 1	FORM 2		Question #	level 2	level 3	level 4	level 5	Question #	level 2	level 3	level 4	level 5
1	Yes	Yes		1	Yes	Yes	Yes	Yes	1	Yes	Yes	Yes	Yes
2	Yes	Yes		2	Yes	Yes	Yes	Yes	2	Yes	Yes	Yes	Yes
3	Yes	Yes		3	Yes	Yes	Yes	Yes	3	Yes	Yes	Yes	Yes
4	Yes	Yes		4	Yes	Yes	Yes	Yes	4	Yes	Yes	Yes	Yes
5	Yes	Yes		5	Yes	Yes	Yes	Yes	5	Yes	Yes	Yes	Yes
6	Yes	Yes		6	Yes	Yes	Yes	Yes	6	Yes	Yes	Yes	Yes
7	Yes	Yes		7	Yes	Yes	Yes	Yes	7	Yes	Yes	Yes	Yes
8	Yes	Yes		8	Yes	Yes	Yes	Yes	8	Yes	Yes	Yes	Yes
9	Yes	Yes		9	Yes	Yes	Yes	Yes	9	Yes	Yes	Yes	Yes
10	Yes	Yes		10	Yes	Yes	Yes	Yes	10	Yes	Yes	Yes	Yes

Results demonstrated strong alignment for Work Ready Math and Work Ready Reading and moderate-to-strong alignment for Work Ready Data (WIN, 2026b). While the Data assessment showed a higher proportion of moderately aligned items than the other content areas, no Data items were classified as weakly aligned.

4.4 Other Testing Activities

In addition to development activities, assessment programs establish procedures to support secure and standardized administration, accessibility, and score integrity. The assessments are typically delivered online through a web-based proprietary system. Online versions are offered in in English, Spanish, and American Sign Language. Paper versions including standard booklet, large print, and Braille are also available. The technical report itself did not provide details regarding administration procedures, test security requirements, accommodation policies, or administrator training presentations. This may be because WIN provides customized test administration materials for partners like South Carolina districts and schools that are posted in the state's online resource center.

4.5 Evaluation and Recommendations

Evaluation

The documentation indicates that WIN implemented field-testing and psychometric evaluation processes that support developing quality items and operational forms. Field-test items were administered under conditions closely resembling operational testing. The documented field testing process reflects accepted assessment development practices by collecting item performance data under operational testing conditions. Further, sample sizes exceed those typically required for stable estimation of classical and IRT item statistics. These analyses provide evidence that test items are technically sound, measure student knowledge accurately, and operate fairly across student groups.

The documented form construction procedures demonstrate appropriate consideration of both content and psychometric characteristics. The use of blueprint-driven item selection, psychometric review, parallel forms, and equating procedures provides evidence that WIN took reasonable steps to support consistent score interpretation across forms.

The alignment study provides additional validity evidence supporting the relationship between assessment content and the learning objectives. WIN's documentation describing panelist expertise, training, independent evaluations, agreement rates, and process for flagged items supports the conclusion that operational forms adequately represent the intended workplace-

readiness learning objectives and contribute to defensible interpretations of scores. The documented procedures provide evidence that WIN evaluated and improved content alignment through a systematic review process.

Overall, the field testing, psychometric analysis, form-construction, and alignment-review activities are consistent with generally accepted assessment-development practices and provide substantial evidence supporting the technical quality of the operational assessments.

Recommendations

Future technical documentation could provide additional details regarding item selection criteria and form construction decision rules to increase transparency regarding development of operational forms; however, the information reviewed provides sufficient evidence that accepted form development practices were followed.

Continued review of moderately aligned Work Ready Data items may further strengthen alignment evidence in future assessment revisions and support continued monitoring of content representation. No significant concerns were identified regarding the alignment study methodology.

Future technical documentation could include descriptions of WIN-recommended administration procedures, test security protocols, accommodation policies, administrator training, and test security measures to provide standardized guidance and evidence supporting secure and consistent implementation. Although supporting state administration materials were reviewed, incorporating recommended best practices into technical documentation would provide additional evidence supporting standardized implementation across testing settings.

5. Standard Setting and Cut Scores

Standard setting is the process through which performance standards and associated cut scores are established to support meaningful interpretation of assessment results. Because the WIN Work Ready Assessments underwent substantial revision in August 2025, including the addition of new learning objectives derived from updated O*NET analyses and lengthening the test forms, WIN conducted a new standard setting study prior to the operational release of the revised assessments in December 2025.

The standard setting process was designed to establish the minimum score required for each of the five achievement levels used by the Work Ready Math, Work Ready Reading, and Work Ready Data assessments. These achievement levels are linked to progressively more complex foundational workforce-readiness skills and corresponding O*NET Job Zones. The resulting cut scores provide the basis for classifying examinees into achievement levels and support interpretation of assessment results in relation to workforce preparation and career readiness.

5.1 Standard Setting Methodology

WIN employed a Modified Angoff (Yes/No) method to establish cut scores for each assessment (WIN, 2026d). The study was conducted virtually and facilitated by a WIN psychometrician. The Modified Angoff approach is a widely recognized standard setting methodology used in

educational and credentialing assessments because it relies on subject matter expert judgments regarding the performance expected of minimally competent candidates at each performance level.

Under the Modified Angoff approach, panelists reviewed every item and determined whether a borderline examinee at a specified achievement level would be expected to answer the item correctly. This approach simplifies judgments to a binary decision, thereby reducing cognitive burden and promoting consistency of judgments across panelists. The methodology focuses on the knowledge and skills represented in the performance/achievement level descriptors and the characteristics of examinees functioning at the threshold between adjacent achievement levels. Panelists discussed the skills that distinguish adjacent achievement levels and used those distinctions to guide judgments regarding expected item performance. This process ensures that cut score recommendations are rooted in the intended meaning of the achievement levels rather than item difficulty alone.

5.2 Panelists and Training

Eight subject matter experts participated in the standard setting process. These panelists were the same committee from the August 2025 Alignment Study (WIN, 2026b, 2026d). Panelists represented across higher education levels and had expertise in K-12 and adult education, workforce experience, and corrections experience. They represented different states; one participant was from South Carolina.

Prior to conducting the review, panelists received an orientation and training session led by the WIN psychometrician. Training included an overview of the purpose of standard setting, the role of subject matter experts, the definition and interpretation of the performance level descriptors, security requirements related to assessment materials, and instruction on implementing the Modified Angoff methodology. Sample items were reviewed, and panelists practiced making judgments using the established procedures before beginning. Figure 2 shows a sample score sheet provided during training that illustrates the process for individual panelist decisions for each item.

Figure 2. Panelist Judgement Score Sheet [Sample from Training Materials]

	ROUND 1			
Question #	level 2	level 3	level 4	level 5
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				

The training was intended to establish a common understanding of achievement level expectations and judgment procedures.

5.3 Rounds of Judgment and Use of Data

The standard setting process included two rounds of independent judgments (WIN, 2026d). During the first round, panelists independently reviewed items and recorded Yes/No judgments regarding whether a borderline examinee at each achievement level would answer the item correctly. Panelists made these judgments using the performance level descriptors, assessment forms, and scoring materials provided during training. Individual judgments were collected, and agreement rates were calculated across panelists.

Following completion of the first round, panelists participated in facilitated discussions focused on areas of disagreement and items that generated the greatest variability in ratings. To support these discussions, panelists were provided with summary information showing the distribution of cut score recommendations, average recommendations, and measures of variability across panelists. Individual panelist identities were protected through the use of numeric identifiers. Discussions focused on the interpretation of performance level descriptors, the characteristics of borderline examinees, and the rationale underlying differing judgments. Importantly, panelists were not encouraged to seek consensus regarding cut scores; rather, discussions were intended to improve shared understanding of the achievement level expectations and judgment task.

Following discussion, panelists completed a second round of independent judgments. After Round 2, summary statistics and panelist distributions were again reviewed and discussed. The documentation notes that panelists did not express concerns regarding the resulting recommendations after the second round and that the degree of change between rounds was relatively small. Based on the stability of results and supporting research indicating that two rounds frequently produce defensible results, the decision was made not to conduct a third round of judgments.

The documentation describes how panel recommendations were reviewed and transformed into operational cut scores that were subsequently adopted for operational use. The process included trained subject matter experts, defined performance level descriptors, training methods with borderline examinees, multiple rounds of judgment, discussion, and evaluation of panelist convergence. Documentation additionally describes how resulting panel recommendations were transformed into operational cut scores and adopted without modification (WIN, 2026d).

5.4 Evaluation and Recommendations

Evaluation

The documentation reviewed provides evidence that the standard setting process was a formal testing event, facilitated by a psychometrician, informed by established performance/ achievement level descriptors, and supported by educator and workforce expert judgments. The standard setting report provides evidence of panelist convergence through documentation of individual ratings, means, and standard deviations across both rounds. The inclusion of multiple rounds, structured discussion, performance level descriptor review, and quantitative feedback are consistent with accepted practices intended to improve the quality and consistency of expert judgments and the resulting cut score recommendations. Although the documentation acknowledges limitations associated with panel size and the absence of impact data at the time of the standard setting, the overall evidence indicates that the panel recommendations were

translated into operational cut scores through a systematic, documented, and professionally accepted process suitable for operational use.

The standard setting process is consistent with generally accepted practices for criterion-referenced assessments and provides sufficient evidence supporting the defensibility of achievement level classifications and associated score interpretations. No significant concerns were identified regarding the standard setting methodology or implementation procedures.

Recommendations

All panelists were drawn from southeastern states (South Carolina, Kentucky, Tennessee, and Florida). Future standard setting studies may benefit from including subject matter experts from additional geographic regions to further broaden the range of perspectives informing achievement level expectations.

In addition, future reviews could incorporate post-administration impact data and stakeholder feedback to provide supplementary evidence regarding the appropriateness and consequences of the achievement level classifications after operational implementation.

6. Scale Scores, Achievement Levels, and Reporting

Following administration and scoring, examinee responses are transformed into achievement-level classifications and reported in a manner intended to support interpretation of foundational workforce-readiness skills. The intended interpretation of assessment results is that examinees possess varying levels of foundational math, reading, and data literacy skills associated with workplace preparation and success in occupational training.

6.1 Scoring and Scale Scores

Items are scored dichotomously, with responses marked either correct or incorrect.

Once responses are scored, raw scores are transformed to a reporting scale. The technical report indicates that item response theory methodologies are used to support score scaling and comparability across operational forms, while the standard-setting report documents the conversion of raw-score cut scores to the operational reporting scale through test characteristic curve and theta-based transformation procedures. Reliability estimates were strong, ranging from 0.84 to 0.91, and classification accuracy estimates ranged from approximately 0.75 to 0.78 (WIN, 2026a), providing evidence that scores are sufficiently consistent for their intended uses.

6.2 Achievement Levels

The WIN Work Ready Academic Skills Assessments include five achievement levels for each content area: Level 1, Level 2, Level 3, Level 4, and Level 5. The standard setting event defined the characteristics of borderline examinees and established cut scores, and raw cut scores were established for each achievement level across the three assessments.

Table 1 shows a sample set of achievement levels and descriptors for math.

Table 1. Math Achievement Levels and Anchors [Sample from Standard Setting]

Achievement Level	Achievement Level Descriptor (Anchors)
1	A participant who has earned an Achievement Level 1 in Work Ready Math has the foundational math skills to succeed in preparatory programs for O*NET Job Zone 1 career pathways. Participants can apply basic problem solving strategies and mathematical understandings to solve 1-step mathematical problems in the context of Job Zone 1 occupations.
2	A participant who has earned an Achievement Level 2 in Work Ready Math has the foundational math skills to succeed in preparatory programs for O*NET Job Zone 2 career pathways. Participants can apply problem solving strategies and mathematical understandings to solve mathematical problems in the context of Job Zone 2 occupations.
3	A participant who has earned an Achievement Level 3 in Work Ready Math has the foundational math skills to succeed in preparatory programs for O*NET Job Zone 3 career pathways. Participants can apply problem solving strategies and mathematical understandings to solve mathematical problems in the context of Job Zone 3 occupations.
4	A participant who has earned an Achievement Level 4 in Work Ready Math has the foundational math skills to succeed in preparatory programs for O*NET Job Zone 4 career pathways. Participants can apply problem solving strategies and advanced mathematical understandings to solve mathematical problems in the context of Job Zone 4 occupations.
5	A participant who has earned an Achievement Level 5 in Work Ready Math has the foundational math skills to succeed in preparatory programs for O*NET Job Zone 5 career pathways. Participants can apply advanced, complex mathematical reasoning to solve problems in the context of Job Zone 5 occupations.

The achievement levels are directly linked to learning objectives, performance (achievement) level descriptors, and O*NET job zone expectations.

6.3 Reporting and Intended Uses of Results

Assessment results are intended to communicate examinees' levels of foundational workforce-readiness skills in mathematics, reading, and data literacy. They are designed to determine whether examinees possess the foundational academic skills necessary to succeed in training associated with various occupational pathways.

The WIN documentation reviewed primarily focused on assessment development and technical characteristics. Limited documentation was available regarding operational score reports, stakeholder-facing interpretive guides, or other reporting resources intended to support score use and interpretation.

6.4 Evaluation and Recommendations

Evaluation

The documentation reviewed indicates that WIN implemented scoring, scale score reporting, and achievement level classification procedures that are consistent with generally accepted psychometric practices. Assessment forms were developed and equated using common-item methodologies and psychometric analyses designed to maintain consistency of score interpretations across administrations.

Reported reliability and classification accuracy estimates provide evidence that assessment results are sufficiently consistent for their intended purposes. Additional psychometric analyses, including dimensionality analyses and differential item functioning, further support the appropriateness of the scoring model and resulting score interpretations.

The documented relationship among learning objectives, achievement level descriptors, standard setting judgments, and cut scores provides evidence supporting meaningful interpretation of achievement classifications. The achievement level framework appears aligned with the intended use of the assessments as measures of foundational workforce-readiness skills rather than preparation for specific occupations (WIN 2026a, 2026d).

Overall, the scoring, scaling, and achievement-level classification procedures provide reasonable evidence supporting the consistency, interpretability, and intended uses of assessment results.

Recommendations

The WIN documentation describes the intended uses and interpretations of assessment results; however, limited evidence was available regarding operational score reports, user guides, or stakeholder-facing materials intended to support score interpretation.

Future technical documentation and reviews could include sample score reports, interpretative guides, and related reporting resources to provide additional evidence that score meanings, limitations, and appropriate uses are communicated clearly to students, educators, workforce partners, and other stakeholders.

7. Conclusion

Based on the documentation reviewed, the WIN Team implemented a comprehensive assessment development process that included blueprint development, item development and review, field testing, psychometric analysis, alignment studies, standard setting, and score scaling activities. The procedures described are generally consistent with accepted assessment development practices and provide evidence supporting the validity and defensibility of score interpretations for the intended purposes of the WIN Work Ready Assessments.

Noted strengths of the assessment program include the structured item review process, extensive field testing sample sizes, documented psychometric analyses, formal alignment study, and well-accepted standard setting methodology. These activities collectively contribute evidence supporting content validity, score reliability, and achievement level interpretations.

Recommendations provided throughout this report focus primarily on expanding technical documentation related to blueprint governance, form construction procedures, score reporting resources, and recommended administration practices. These recommendations are intended to enhance transparency and documentation rather than address substantive concerns regarding assessment quality. Specific opportunities include providing additional information about blueprint approval processes, form construction procedures, and reporting materials that support appropriate score interpretation and use. Although administration and test security materials and training are customized for each state, WIN may also benefit from documenting a core set of recommended practices that apply across states.

Overall, the evidence reviewed supports the conclusion that the WIN Work Ready Assessments were developed through a process consistent with professional assessment practices and are appropriately prepared for operational use within the intended context.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Cizek, G. J., & Bunch, M. B. (2007). *Standard setting*. SAGE Publications, Inc., https://doi.org/10.4135/9781412985918.

Laitusis, C. C., & Karvonen, M. (2025). A case for reimagining universal design of assessment systems. *Educational Measurement: Issues and Practice, 44,* 5-14. https://doi.org/10.1111/emip.12674.

Nunally, J. C. (1978). *Psychometric theory*. New York: McGraw Hill.

Sireci, S. G. (1998). The construct of content validity. *Social Indicators Research*, 45, 83–117.

Worldwide Interactive Network, Inc. (WIN Learning). (2026a). WIN Work Ready Assessments technical report (Version 6.2026) [Confidential technical report].

Worldwide Interactive Network, Inc. (WIN Learning). (2026b). WIN Academic Skills Assessments alignment report 2025. (Updated May 2026) [Confidential technical report].

Worldwide Interactive Network, Inc. (WIN Learning). (2026c). WIN Academic Skills Assessments field testing summary [Confidential technical report].

Worldwide Interactive Network, Inc. (WIN Learning). (2026d). WIN Academic Skills Assessments standard setting report 2025 (Updated May 2026) [Confidential technical report].

Mary Frances Early College of Education
K–12 Assessment Solutions
UNIVERSITY OF GEORGIA

WIN Work Ready Assessments

Evaluation of Math, Reading, Data

2026

Gwinnett Campus:
2530 Sever Road
Suite 149
Lawrenceville, GA 30043

Contact Us:
Toll free: (888) 392-8977
Web: https://coe.uga.edu/directory/k-12-assessment-solutions

Table of Contents

Summary 3
1. WIN Work Ready Assessments 5
1.1 South Carolina Career Ready Test Overview 5
1.2 Review Process 6
2. Test Blueprint Review 7
2.1 Coverage of Learning Objectives 7
2.2 Depth of Knowledge Distribution 8
3. Evaluation of Overall Validity 9
3.1 Coverage by Achievement Level 9
3.2 Alignment to Learning Objectives 11
3.3 Depth of Knowledge 12
3.3.1 Distribution of DOK Levels 12
3.3.2 Review Items for DOK Level 14
3.4 Internal Structure 21
3.4.1 CTT Item Statistics 21
3.4.2 1-PL Item Statistics 23
3.4.3 Dimensionality 24
3.4.4 Reliability 25
3.4.5 Measurement Invariance 25
3.4.6 Equating Multiple Forms 26
4. Achievement Level Classifications 27
References 28

Summary

This document reviews the WIN Work Ready Math, Reading, and Data Assessments and evaluates the extent to which these assessments provide valid and reliable evidence for measuring foundational employability skills associated with career readiness. The review examines the alignment of the assessments of their intended learning objectives and achievement levels, evaluates evidence related to content validity and internal structure, and considers the reliability and appropriateness of achievement level classifications used to support credential attainment and reporting.

The review of the WIN Work Ready Assessments yielded several key findings. Overall, the test blueprints appropriately define the knowledge and skills measured by each assessment and reflect the progression of achievement levels aligned to O*NET Job Zones. The operational forms were consistent with blueprint specifications, and all learning objectives identified on the blueprints were represented across the reviewed forms. Item reviews indicated that the assessments are generally well aligned to the intended learning objectives, with only a small number of items on the WIN Work Ready Data recommended for content review or objective re-alignment. Overall, the alignment between assessment items and the intended learning objectives was strong and supports the content validity of the assessments.

With respect to cognitive complexity, the WIN Work Ready Assessments incorporate a range of Depth of Knowledge (DOK) levels intended to measure increasingly complex applications of skills and knowledge. The DOK distributions identified in the *Test Maps* appear reasonable and generally consistent with the intended purpose of workforce-readiness assessments. However, item-level reviews indicated that a number of items classified as DOK 3, particularly within the Reading and Data assessments, function operationally at a DOK 2 level because the correct responses can often be identified through straightforward interpretation or application of readily available information. While these findings do not raise concerns regarding overall assessment quality, a broader review of DOK classifications and distributions is recommended to ensure that the cognitive demands of operational forms reflect the intended blueprint specifications.

Evidence based on the internal structure of the assessments indicated that the WIN Work Ready Assessments generally function as intended. Classical Test Theory analyses demonstrated appropriate ranges of item difficulty and strong item discrimination across the Math, Reading, and Data assessments. Item response theory analyses showed that item difficulties span the ability continuum, supporting the validity of score interpretations and classification decisions. Dimensionality analyses provided evidence that each assessment measures a single dominant construct, and reliability estimates were strong. The multiple forms equating process was sound, supporting the use of multiple forms for each assessment. Due to insufficient sample sizes, DIF and subgroup analyses are incomplete, but are strongly recommended for additional evidence regarding fairness and validity. Additionally, model fit evaluations and post-equating checks are recommended for continued assurance of valid score comparisons.

The evaluation also examined achievement level classifications for the WIN Work Ready Assessments. Achievement level cut scores were established using a recognized standard-setting methodology and applied to 1-PL-based ability estimates to classify students into five

achievement levels corresponding to established O*NET Job Zones. Classification quality indices indicated acceptable classification accuracy and consistency. When feasible given blueprint and content constraints, we recommend adding more difficult items to better scale and classify higher ability examinees.

Overall, the WIN Work Ready Assessments demonstrate sound technical quality and provide valid and reliable evidence regarding the prerequisite career readiness skills required for success in occupations aligned to the corresponding O*NET Job Zones. The assessments are generally aligned to their intended learning objectives, exhibit appropriate psychometric characteristics, and support meaningful achievement level classifications linked to career readiness outcomes. Although the assessments are functioning as intended, targeted refinements, including reviewing flagged content and DOK alignment issues, conducting differential item functioning analyses when feasible, and expanding item coverage at higher difficulty levels, would further strengthen the validity, interpretability, and utility of assessment results.

1. WIN Work Ready Assessments

This report reviews the WIN Work Ready Assessments, a suite of career-readiness assessments covering academic skills measuring Math, Reading, and Data. This review evaluates the three component subtests, or employable skills, of the WIN Work Ready Assessments. These subtests are utilized for the *South Carolina Career Ready Test* and are conducted on behalf of the South Carolina Oversight Committee in collaboration with the South Carolina Department of Education.

1.1 South Carolina Career Ready Test Overview

The *South Carolina Career Ready Test* is a statewide assessment administered to all eleventh-grade students to measure foundational workplace skills identified as essential for career readiness. The assessment evaluates students in four key areas: soft skills, math, reading, and data analysis.

The South Carolina Career Ready Test incorporates the WIN Work Ready Assessments to measure students' proficiency in employability skills. South Carolina students who pass the WIN Work Ready Assessments earn the South Carolina Work Ready Credential, demonstrating mastery of foundational math, reading, and data analysis skills. Students who also pass the soft skills component earn the South Carolina Work Ready Essential Soft Skills Credential, which validates competencies in effective communication, professionalism, teamwork and collaboration, critical thinking, and problem solving. Both credentials are endorsed by the National Work Readiness Council and are designed to help educators, workforce development professionals, and employers determine students' readiness for post-secondary education, technical training, apprenticeships, on-the-job training, and employment opportunities.

The mathematics, reading, and data assessments are assigned a scale score and an achievement level score from 1 to 5 based on the U.S. Department of Labor O*NET Job Zones, which categorize occupations based on the education, experience, and training typically required for successful job performance. Students who achieve Levels 2 through 5 earn a corresponding Work Ready Credential, reflecting readiness for occupations within those job zones.

The WIN Work Ready Assessments are organized around learning objectives aligned to achievement levels and O*NET Job Zones. Within each assessment, related skills are grouped into strands that represent broad domains of workplace competencies and help define the knowledge and skills measured by each assessment. Together, these components guide the development of authentic assessment items that measure students' ability to apply career readiness skills in workplace-relevant contexts.

1.2 Review Process

With the support of the Education Oversight Committee, experts from the University of Georgia evaluated the WIN Work Ready Assessments for reliability and validity in assessing the prerequisite career readiness skills required to successfully complete preparatory programs for jobs categorized by the corresponding achievement level / job zone. The evaluation process follows best practices in educational measurement, as detailed by the Standards for Educational and Psychological Testing (AERA, APA, NCME, 2014).

The following materials were provided by the WIN Team for evaluation:

- 100A and 200B Test Forms
- For each Work Ready assessment:
    - Blueprint and Test Maps
- WIN Work Ready Assessments Technical Report

The Test Maps included metadata about individual items and psychometric indices. The test contractor calculated all parameters; no additional estimation of item or test parameters was conducted. The items reviewed for content validity were presented in the new/current edition of all three-component assessment, and the psychometric review is based on the summaries from the Technical Report which reported on the new editions released December 2025. The n-counts are relatively small and will be updated as more data is collected.

This report was prepared by the University of Georgia and examines critical elements of the WIN Work Ready Assessments test designs and summarizes findings and recommendations for each.

2. Test Blueprint Review

A test blueprint review is crucial for an assessment's validity, fairness, and reliability. It aligns assessment items to learning objectives and cognitive complexity levels such as Depth of Knowledge (DOK). Traditionally, such as in state assessments, test blueprints are content based maps and used to define how many items are dedicated to each content domain. The WIN Work Ready Assessments use a progression of skills and cognitive demand. They are designed with Achievement Levels 1-5 aligned to O*NET Job Zones 1-5 with roughly equal items aligned to each achievement level.

2.1 Coverage of Learning Objectives

The learning objectives stated in the test blueprints encompass all the learning objectives that underpin both the curriculum and the assessments. The learning objectives reflect the most common (most in-demand) foundational work readiness skills required across the nearly 1,000 occupations classified by O*NET. The learning objectives are comprehensive for the curriculum and the assessments but are not occupation specific and therefore do not reflect all O*NET defined discrete skills.

Because all learning objectives are represented on the test blueprints, the Subject Matter Experts (SMEs) did not need to evaluate and ensure balanced weighting across content domains (reporting categories). Rather, each achievement level comprises approximately 20% of each assessment. Table 1 through Table 3 summarize the number of learning objectives and items aligned with each achievement level as reflected in the provided *New Blueprint and Test Form Composition*.

Table 1. Test Blueprint for Work Ready Math

Achievement Level	Test Blueprint	
	Number of Objectives	Number of Items per Achievement Level
1	7	8
2	7	8
3	7	8
4	6	8
5	4	8
Total Number of Items		**40**

Table 2. Test Blueprint for Work Ready Reading

Achievement Level	Test Blueprint	
	Number of Objectives	Number of Items per Achievement Level
1	5	8
2	6	8
3	5	8
4	5	8
5	4	8
Total Number of Items		**40**

Table 3. Test Blueprint for Work Ready Data

Achievement Level	Test Blueprint	
	Number of Objectives	Number of Items per Achievement Level
1	4	6
2	5	6
3	6	7
4	7	7
5	3	6
Total Number of Items		**32**

Recommendation: N/A

2.2 Depth of Knowledge Distribution

The WIN Work Ready Assessments use a progression of Achievement Levels 1-5 aligned to O*NET Job Zones 1-5 (National Center for O*NET Development, 2026), which are the foundations for scoring, standard setting, credential attainment, and reporting. The framework for each of the assessments has increasing cognitive complexity built into the learning objectives. Each achievement level builds on the prior level, with the application of skills increasing with each level.

The assessments do not specifically employ Webb's Depth of Knowledge (DOK) as an operational assessment framework. The test blueprints and score reports do not classify items by DOK level or provide DOK distributions. However, in the Test Maps, all items are assigned to a DOK level, indicating that DOK is actively used during item development and blueprinting.

Recommendation: N/A

3. Evaluation of Overall Validity

Content validity is essential for ensuring an assessment accurately measures the intended knowledge and skills (Bandalos, 2018). It involves a thorough evaluation of the items and domains to ensure they represent the target domain. This review ensures that the information gathered from administering the assessment is relevant and minimizes construct-irrelevant variance. Furthermore, content specification and item review help to ensure that the full range of the construct(s) is measured, minimizing construct underrepresentation.

To verify content validity, subject matter experts (SMEs) compared the WIN Work Ready Assessments with the provided *New Blueprint and Test Form Composition* tables (December 2025). The assessments were reviewed for item alignment to learning objectives and Depth of Knowledge (DOK) levels. Two SMEs independently reviewed each item for objective alignment. A third SME resolved any discrepancies. The panel of SMEs then held a consensus meeting to finalize alignment recommendations. The internal structure of the assessments was reviewed by an educational measurement expert at the University of Georgia.

3.1 Coverage by Achievement Level

Items on the WIN Work Ready Assessments were reviewed for alignment to achievement level and evaluated against each blueprint. First, the provided *Test Maps* were evaluated against each *New Blueprint and Test Form Composition* tables. Then, all items were reviewed for alignment to each achievement level.

Table 4 through Table 6 summarize the number of items aligned with each achievement level as reflected in the provided *Test Map*s.

Table 4. Coverage by Achievement Level for Work Ready Math

Achievement Level	Test Blueprint		Test Map	
	Number of Objectives	Number of Items per Achievement Level	Number of Items per Achievement Level	% of Assessment
1	7	8	8	20%
2	7	8	8	20%
3	7	8	8	20%
4	6	8	8	20%
5	4	8	8	20%
Total Number of Items			**40**	**100%**

Table 5. Coverage by Achievement Level for Work Ready Reading

Achievement Level	Test Blueprint		Test Map	
	Number of Objectives	Number of Items per Achievement Level	Number of Items per Achievement Level	% of Assessment
1	5	8	8	20%
2	6	8	8	20%
3	5	8	8	20%
4	5	8	8	20%
5	4	8	8	20%
Total Number of Items			**40**	**100%**

Table 6. Coverage by Achievement Level for Work Ready Data

Achievement Level	Test Blueprint		Test Map	
	Number of Objectives	Number of Items per Achievement Level	Number of Items per Achievement Level	% of Assessment
1	4	6	6	19%
2	5	6	6	19%
3	6	7	7	22%
4	7	7	7	22%
5	3	6	6	19%
Total Number of Items			**32**	**100%***

**due to rounding at each level, total does not equal 100%*

The number of items on the WIN Work Ready Assessments by achievement level align with the number stated on the test blueprints.

Math: All learning objectives on the test blueprint are represented on both forms of the math assessment.

Reading: All learning objectives on the test blueprint are represented on both forms of the reading assessment.

Data: All learning objectives on the test blueprint are represented on both forms of the data assessment.

Recommendation: N/A

3.2 Alignment to Learning Objectives

All items on the WIN Work Ready Assessments were reviewed for alignment to learning objectives. Table 7 highlights items flagged during review. There were no items flagged for learning objective alignment in the Work Ready Math and Work Ready Reading assessments.

Table 7. Alignment to Learning Objectives for Work Ready Data

Form	Item #	Objective on Test Map	Suggested Re-alignment	Notes
100A	20	DGI4.1	DAD4.5	Item does not align to the sources stated in DGI4.1: interviews, surveys, and focus groups. Recommend alignment to DAD4.5: trends and relationships between variables.
200B	11	NA	NA	Item does not have a correct answer, review content. The key is stated as D but there is no error; Katie Kite's rental fee is correct. The rental is 8 days, so 1 week + 1 day = $700.
200B	20	DGI4.1	DAD4.5	Same item has Form A, Item 20.
200B	21	NA	NA	The purpose and structure of the graphic is unclear. It appears to function as a range map or spatial display rather than a traditional chart or graph (such as coordinate grid), but key elements of the display are difficult to interpret. While Answer A appears to be the best choice because changing the order of the ranges may improve readability, the item may benefit from clearer visual design or a more explicit indication of the graphic's purpose.
200B	29	NA	NA	The correct answer is C, but choice B may also be plausible because the trend for admitted students could reasonably be extrapolated to approximately 5,100 in Year 6. The graph may support more than one defensible prediction. Consider revising the values, answer choices, or

				graph scale to ensure a single best answer is clearly supported by the data.

Recommendation: Overall, the items on the WIN Work Ready Assessments are well aligned with the stated learning objectives. Review alignment and content flag for the Data assessments.

3.3 Depth of Knowledge

All items are assigned to a DOK level in the Test Maps only, suggesting that WIN uses DOK as an internal measure of cognitive complexity during item development and blueprinting, while achievement levels are used for external reporting and credential decision-making.

Items on the Test Map are classified as DOK 1, DOK 2, or DOK 3. This evaluation uses the following descriptions as guidelines for DOK. As DOK levels increase, the cognitive demand on students also increases; higher DOK levels require more than just recalling facts; they require deeper understanding, application, analysis, and synthesis.

- Level 1. Recall and Reproduction: This level requires students to recall basic facts, information, definitions, terms, or perform simple, routine procedures.
- Level 2. Skills and Concepts: This level requires engaging in mental processing beyond simple recall. Students need to apply concepts, use skills, and make decisions. It requires understanding and using knowledge.
- Level 3. Strategic Thinking: This level requires deep understanding, planning, using evidence, and more complex reasoning. Students must analyze, evaluate, and draw conclusions. The cognitive demands are more abstract and require justification.
- Level 4. Extended Thinking: This level requires students to make connections, relate ideas within or among content areas, and select or devise an approach to solve a problem. It often involves extended time and requires synthesis and in-depth analysis.

Items on the WIN Work Ready Assessments were reviewed for DOK distribution and evaluated against each Test Map.

3.3.1 Distribution of DOK Levels

Table 8 through Table 10 summarize the number and percentage of items aligned with each DOK level as reflected in the provided Test Map.

Table 8. DOK Distribution for Work Ready Math (Test Map)

DOK	Form 1		Form 2	
	Number of Items	% of Assessment	Number of Items	% of Assessment
1	7	17.5%	6	15%
2	27	67.5%	28	70%
3	6	15%	6	15%
Total	**40**	**100%**	**40**	**100%**

Table 9. DOK Distribution for Work Ready Reading (Test Map)

DOK	Form 1		Form 2	
	Number of Items	% of Assessment	Number of Items	% of Assessment
1	11	27.5%	13	32.5%
2	13	32.5%	13	32.5%
3	16	40%	14	35%
Total	**40**	**100%**	**40**	**100%**

Table 10. DOK Distribution for Work Ready Data (Test Map)

DOK	Form 1		Form 2	
	Number of Items	% of Assessment	Number of Items	% of Assessment
1	4	12.5%	2	6.25%
2	16	50%	18	56.25%
3	12	37.5%	12	37.5%
Total	**32**	**100%**	**32**	**100%**

Recommendation:

The DOK distributions for the WIN Work Ready Assessments appear reasonable and consistent with the intended purpose of a workforce-readiness assessment. While the intended proportion of DOK 3 items is substantial in each section, item-level reviews flag many items that function operationally as DOK 2. As a result, a full form-level analysis is recommended to determine whether the delivered assessments adequately reflect the cognitive complexity is intended.

> Math: The Math assessments demonstrate balanced DOK distribution with strong emphasis on application of math concepts and procedural problem-solving (DOK 2). The assessments focus on applying formulas, proportional reasoning, interpretation of information, data, graphs, and decision-making.

Reading: The Reading assessments suggest rigorous assessments with a substantial emphasis on DOK 3 reasoning. The item review in the following section suggests that many designated DOK 3 items function operationally as DOK 2. As a result, the operational forms may not fully achieve the intended distribution of higher complexity.

Data: The Data assessments place substantial emphasis on DOK 3 reasoning and evidence-based interpretation. However, similar to Reading, several items classified as DOK 3 appeared to function operationally as DOK 2. Consequently, the implemented forms may contain a lower proportion of DOK 3 items than intended.

3.3.2 Review Items for DOK Level

All items on the WIN Work Ready Assessments were reviewed for DOK level.

Evaluation: Table 11 through Table 13 highlights items flagged during review.

Table 11. Alignment to DOK for Work Ready Math

Form	Item #	DOK on Test Map	Suggested DOK	Notes
100A	22	1	2	Item asks students to find the volume of cylinder, then compute costs. Students apply knowledge and calculate using multiple steps. Item is similar to #21, classified as DOK 2.
200B	3	1	2	Item asks students to analyze multiple pieces of information, verify totals, identify discrepancies, and evaluate a business decision. The item requires more than recall or simple computation. Item is similar to Form 100A #8, classified as DOK 2.

Table 12. Alignment to DOK for Work Ready Reading

Form	Item #	DOK on Test Map	Suggested DOK	Notes
100A	20	3	2	Item asks students to analyze how word choice affects readers, but the distractors are relatively weak and only one interpretation is supported. Students are not required to evaluate multiple reasonable interpretations.

Form	Item #	DOK on Test Map	Suggested DOK	Notes
100A	22	3	2	Item asks students to analyze purpose and audience to identify where the passage would most likely appear, but there is only one clear fit. The distractors are implausible, limiting strategic reasoning.
100A	25	3	2	Item asks students to determine the purpose of a company letter. The item approaches DOK 3, but students are not required to justify how specific details or sections of the letter support the purpose.
100A	26	3	2	Item asks students to analyze how a word contributes to the meaning of the letter. The item stem targets abstract vocabulary analysis, but only one interpretation aligns with the context, limiting the need for deeper evaluation.
100A	29	3	2	Item asks students to identify words contributing to an optimistic tone. To increase the rigor, item could require students to justify their conclusion or compare multiple plausible tone interpretations and determine which is better supported by evidence.
100A	30	3	2	Item asks students to determine how the author supports a claim. Item approaches DOK 3 because students evaluate evidence; however, the distractors are easily eliminated and do not require students to evaluate the relative strength of multiple plausible pieces of evidence.
100A	33	3	2	Item asks students to identify evidence supporting a statement/idea. The item stem resembles evidence-based reasoning, but students are only required to match the statement to a supporting detail. Students do not evaluate multiple or competing pieces of evidence.
100A	34	3	2	Item asks students to determine which visual would add to the meaning of the

Form	Item #	DOK on Test Map	Suggested DOK	Notes
				passage. Item stem approaches DOK 3 reasoning, but only one choice clearly aligns with the passage's focus, reducing overall complexity.
100A	35	3	2	Item asks students to determine the format of a passage and explain its impact on readers. The second part of the item stem ("effect") suggests DOK 3 reasoning, but the choices combine structure and effect, allowing students to identify the correct choice without deeper reasoning.
100A	37	3	2	Item asks students to determine how an expert quote supports the author's purpose. Item stem approaches DOK 3 analytical thinking, but students are not required to explain why the quote supports the purpose better than alternative evidence.
200B	10	2	1	Item asks students to recall the meaning of "Open Enrollment" using information stated directly in the memorandum. Students locate the definition explicitly provided in the passage.
200B	20	3	2	Item asks students to analyze how word choice affects the tone, but the distractors are relatively weak and only one interpretation is supported. Students are not required to evaluate multiple reasonable interpretations or justify conclusions.
200B	22	3	2	Item asks students to identify the tone of passage. The item stem is attempting higher-level tone analysis, but the correct response is supported by explicit features of the passage, and the distractors are comparatively weak.

Form	Item #	DOK on Test Map	Suggested DOK	Notes
200B	25	3	2	Item asks students to identify the intended audience based on phrases. Students use contextual clues and document purpose to make an inference. The item does not require students to justify how multiple details support the audience determination or evaluate alternative audiences.
200B	28	3	2	Item asks students to determine the main purpose of the listed activities. Students synthesize information from a section of the passage, but the answer is a straightforward interpretation of the activities presented. The item does not require evaluation of competing purposes or supporting evidence.
200B	30	3	2	Item asks students to determine which activity would not contribute to the requirements. Students compare an activity against explicitly stated criteria and identify the mismatch. The item does not require strategic reasoning or evaluation.
200B	34	3	2	Item asks students to identify the effect of formatting. The item stem targets analysis of text structure; however, one answer clearly aligns with the purpose of the formatting while the distractors are relatively implausible.
200B	38	3	2	Item asks students to identify evidence supporting the author's claim. The item stem approaches evidence-based reasoning, but students are only required to match a claim to a supporting detail explicitly stated in the passage.

Table 13. Alignment to DOK for Work Ready Data

Form	Item #	DOK on Test Map	Suggested DOK	Notes
100A	1	2	1	Item asks students to identify the purpose of research questions. Students recall a basic concept about information gathering.
100A	19	3	2	Item asks students to interpret a graph and identify the occupation with the strongest employment opportunity. Students compare several data categories, but the item does not require them to evaluate multiple indicators or justify why one occupation represents the best opportunity.
100A	23	3	2	Item asks students to analyze a graph containing multiple variables and identify an industry that meets specific criteria. The item approaches DOK 3 because students must consider multiple data relationships simultaneously; however, students are only required to locate the correct industry rather than explain how the variables interact to support the conclusion.
100A	24	3	2	Item asks students to determine the relationship between advertising and sales. Students interpret the trends and relationships in a graph, but the item does not require DOK 3 level reasoning such as evaluating alternative explanations, justifying the strength of the relationship, or determining limitations in the data. Item is similar to #10 and #29.
100A	26	3	2	Item asks students to determine which conclusion is best supported by a graph. The item approaches DOK 3 because students evaluate data and compare conclusions; however, the distractors are relatively easy to eliminate and do not require students to weigh the relative

				strength of multiple plausible interpretations.
100A	31	3	2	Item asks students to identify an implication supported by survey data. The item stem targets higher-order reasoning by requiring students to move beyond direct observation; however, only one conclusion is plausible, limiting the need for deeper evaluation.
100A	32	3	2	Item asks students to compare data from two time periods and determine the most likely implication. The item approaches DOK 3 because students identify patterns and draw conclusions; however, students are not required to evaluate alternative interpretations of the data or explain the evidence supporting their conclusion.
200B	16	2	3	Item asks students to determine which survey response is related to a research question but not relevant to answering it. Students must evaluate the relationship between the research question and multiple pieces of information and distinguish between related and relevant evidence, aligned with DOK 3 reasoning.
200B	19	3	2	Item asks students to draw a conclusion supported by a graph showing job openings and unemployment. Students interpret trends, but the item does not require them to evaluate the strength of multiple plausible conclusions.
200B	23	3	2	Item asks students to determine the relationship between fuel economy and various driving speeds. Students interpret the overall pattern of the graph, but the item does not require DOK 3 level reasoning such as explaining how the variables interact or justifying one interpretation over competing choices.
200B	24	3	2	Item asks students to determine the relationship shown in a graph comparing age ranges across employment

				arrangements. Item does not require deeper evaluation beyond the visible relationships.
200B	26	3	2	Item asks students to determine the trends shown in a graph of the number of movers and mover rates over time. The answer is straightforward and does not require reasoning beyond the visible relationships.
200B	31	3	2	Item asks students to identify the most likely implication supported by data across multiple graphs. Students move beyond direct observation and make a modest inference from the graphs. However, students are not required to evaluate competing explanations or justify why one implication is better supported than another.

Recommendation:

Overall, the WIN Work Ready Assessments items varied in their alignment with the stated DOK levels. Most flags involved items designated as DOK 3 that appeared to function operationally as DOK 2 because students could arrive at the correct answer through straightforward or concrete interpretation, application of stated criteria, or identification of a clearly supported answer. A recurring pattern was that items intended to measure strategic thinking but did not require students to engage in the evaluation, justification, comparison, or evidence-based reasoning typically associated with DOK 3 items.

The most DOK alignment flags occurred in the Reading and Data assessments.

Math: Review the two flagged items and either update the associated metadata or revise the items to more accurately reflect the intended DOK classification. The Work Ready Math assessment generally demonstrates appropriate alignment between the intended and stated DOK levels. Overall, no substantial DOK alignment concerns were identified.

Reading: Although the test map identifies several Reading items as DOK 3, many function operationally as DOK 2. The item stems often attempt to target strategic thinking or abstract reasoning; however, the answer choices limit the cognitive demand of the task because the key is clearly supported by the text while the distractors are comparatively weak. As a result, students are required to identify and interpret parts of the text rather than evaluate, justify, and compare competing ideas that are typically associated with DOK 3 items.

Review the flagged items and consider revisions that require students to engage in strategic thinking or abstract reasoning. For example, items addressing author purpose, tone, and word choice could be strengthened by asking students to evaluate multiple plausible interpretations and determine which is best supported by the text. These revisions would more closely align item complexity with DOK 3 expectations for strategic thinking, evaluation, and evidence-based reasoning.

Data: Although the test map identifies several Data items as DOK 3, many function operationally as DOK 2. The item stems often attempt to target strategic thinking through data interpretation, trend analysis, prediction, and evidence-based conclusions; however, the answer choices frequently reduce the cognitive demand because the correct response can be identified from a straightforward reading of the graph, chart, or dataset while the distractors are comparatively weak. As a result, students are required to identify patterns, interpret information, and select supported conclusions rather than evaluate, justify, or analyze the relative strength of multiple conclusions/ explanations that are typically associated with DOK 3 items.

Review the flagged items and consider revisions that require students to engage in stronger evidence-based reasoning. For example, items involving trend analysis, predictions, implications, and conclusions could be strengthened by asking students to evaluate multiple plausible interpretations, determine which conclusion is best supported by the available data, compare competing explanations, or analyze how multiple variables interact to influence outcomes. These revisions would more closely align item complexity with DOK 3 expectations for strategic thinking, evaluation, and data-driven reasoning.

In addition to the item-level DOK review, a comprehensive, holistic review is recommended to determine whether each assessment adequately reflects the cognitive demands and distribution parameters specified in the test blueprint. Consider the cumulative impact of any proposed DOK reclassifications, as shifts across multiple items may affect the overall balance of cognitive complexity represented within and across forms.

3.4 Internal Structure

According to the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 2014), validity evidence based on internal structure relates to how test items, individually and collectively, align with the construct(s) being measured. To that end, this evaluation included summaries of classical test theory and Rasch model item analyses, dimensionality, reliability, and measurement invariance. No additional analysis was conducted. Rather, materials provided by SCDE and WIN, including the technical report, field testing summary, and psychometric statistics files, are summarized and interpreted.

3.4.1 CTT Item Statistics

A classical test theory (CTT) item analysis is conducted as a baseline check for the appropriateness of items. To support psychometric analyses and ability estimation, items need to

be sufficiently difficult for the population and function properly. CTT item difficulty is reported as the proportion of examinees who answered an item correctly. Lower values indicate harder items (e.g., 0.15 (15%) of the examinees answered the item correctly), and higher values indicate easier items (e.g., 0.85 (85%) of examinees answered the item correctly). For these assessments, a wide range of difficulty values should be observed to indicate utility across the breadth of examinee ability levels. Additionally, values near 0.50 provide strong information (Bandalos, 2018). One measure of CTT item quality is the item-total score (also termed point-biserial) correlation. The item-total correlation is a quantification of the degree to which individual items separate examinees with low and high scores. Values greater than 0.30 are considered satisfactory, while values less than 0.20 indicate low discrimination and suggest revision of the item (Nunnally & Bernstein, 1994). Negative values indicate that higher scoring examinees tend to get the item incorrect more often than low scoring examinees, which is a flag for immediate item review.

Evaluation: Results from CTT analysis are displayed in Table 14.

Overall, each assessment has an appropriate range of item difficulties to properly assess students across the span of ability levels. Additionally, most item-total correlations are above 0.30, indicating that items discriminate adequately. The field test summary document indicates stringent and appropriate criteria for item flagging, review, and revision, including flags for items that are too easy for the sample, too difficult for the sample, and items with low discriminatory power. Altogether, the difficulty and item-total correlation estimates provide preliminary evidence that the items are functioning properly for the math, reading, and data assessments.

Recommendation: N/A

Table 14. Classical Test Theory Item Statistics

Assessment	Number of Items	Mean *p*-value (range)	Mean Item-Total Correlation (range)	Percentage of Items Flagged
Work Ready Math (Field Test)	75	0.48 (0.08 - 0.96)	0.31 (-0.23 - 0.64)	31%
Work Ready Reading (Field Test)	215	0.61 (0.14 - 0.96)	0.37 (-0.06 - 0.59)	14%
Work Ready Data (Field Test)	67	0.53 (0.12 - 0.89)	0.31 (-0.05 - 0.57)	24%
Work Ready Math (Operational Form 1)	40	0.57 (0.22 - 0.96)	0.41 (0.23 - 0.55)	8%*
Work Ready Math (Operational Form 2)	40	0.57 (0.20 - 0.95)	0.43 (0.21 - 0.64)	8%*

Work Ready Reading (Operational Form 1)	40	0.65 (0.36 - 0.89)	0.43 (0.24 - 0.59)	0%
Work Ready Reading (Operational Form 2)	40	0.65 (0.38 - 0.94)	0.42 (0.22 - 0.57)	5%*
Work Ready Data (Operational Form 1)	32	0.61 (0.31 - 0.87)	0.38 (0.17 - 0.57)	3%
Work Ready Data (Operational Form 2)	32	0.59 (0.34 - 0.84)	0.38 (0.15 - 0.58)	3%

Note. Blue shaded rows are field tests. Green shaded rows are operational. *Some flagged (easy) items were included to cover learning objectives.

3.4.2 1-PL Item Statistics

The one-parameter logistic (1-PL) item response theory model was used for calibration and scaling. While it makes strong assumptions about items, the 1-PL model has preferrable measurement properties including sum-score sufficiency, invariant item ordering, and a common item-to-ability scale (Engelhard, 2013). These properties are useful for interpretation of items, ability estimates, and performance level classifications. The 1-PL model includes ability estimates and item parameter estimates. The 1-PL model assumes that each item discriminates equally and estimates a difficulty parameter for each item. This difficulty parameter is a location parameter, indicating the point on the ability scale where an examinee has a 50% chance of answering the item correctly. For the WIN assessment purposes, difficulty parameters should span the ability distribution (e.g., -3 to 3) to reliably locate all examinees.

When applying the 1-PL model, it is important to assess the degree to which the data fits the model. 1-PL model fit statistics, including infit and outfit, residual-based item fit statistics, and chi-square tests can quantify the differences between observed responses and model-predicted responses.

Evaluation: Results from 1-PL model psychometric analyses are displayed in Table 15. Overall, each assessment has a sufficient range of item difficulties to properly assess students across the span of ability levels.

Table 15. 1-PL IRT Model Item Parameters

Assessment	Number of Items	Mean 1-PL Difficulty (Range)
Work Ready Math (Field Test)	75	0.23 (-3.54 - 3.62)
Work Ready Reading (Field Test)	215	0.00 (-3.00 - 2.81)
Work Ready Data (Field Test)	67	0.32 (-2.09 - 3.04)
Work Ready Math (Operational Forms)	40	-0.83 (-3.98 - 1.28)
Work Ready Reading (Operational Forms)	40	-1.10 (-3.54 - 0.83)
Work Ready Data (Operational Forms)	32	-0.60 (-2.22 - 1.03)
Note. Blue shaded rows are field tests. Green shaded rows are operational.		

Recommendation: Each operational form's model and item fit statistics should be included for a complete evaluation. Additionally, if feasible within blueprint constraints, more difficult items (b > 1.2) could be included to more precisely scale and classify higher ability examinee.

3.4.3 Dimensionality

The item response theory models used to scale the WIN assessments have an underlying assumption of unidimensionality. That is, each test measures a single domain. If this is not the case, and more than one factor contributes to item responses, then the validity and interpretation of scale scores are called into question. To assess dimensionality, common approaches are exploratory factor analysis (EFA) and principal components analysis (PCA). Within both frameworks, when the first component accounts for substantially more variance than subsequent components, it is an indicator that the first factor is dominant and that the unidimensionality assumption has been satisfied (Zopluoglu & Davenport, 2017).

Evaluation: Results from EFA visualizations (scree plots) of the dimensionality of each assessment indicated that for each assessment, the first factor was dominant, and the unidimensionality assumption was met.

Recommendation: N/A

3.4.4 Reliability

Test score reliability is concerned with the consistency and precision of scores and is a function of the amount of measurement error (Wells & Wollack, 2003). Reliability is a necessary condition for validity because if scores are highly variable and error-ridden, they cannot be said to measure the construct(s) accurately. There are several ways to quantify reliability in the CTT framework, including Cronbach's alpha and the standard error of measurement. Cronbach's alpha ranges from 0 to 1 and quantifies the degree to which the items consistently measure the target domain. In high-stakes settings, alpha should be approaching or above 0.90 (Wells & Wollack, 2003). The standard error of measurement (SEM) provides an interval estimate around raw scores.

Evaluation: For the WIN assessments, Cronbach's alpha reliability estimates were strong, ranging from 0.84 - 0.91. SEM values ranged from 2.42 to 2.69. Raw score SEMs in this range are adequate considering there were 32 – 40 operational items. In summary, reliability is adequate for the WIN assessment scaling purposes.

Recommendation: Reliability estimates could be provided at the subgroup level for additional evidence of fairness and validity. Reliability estimates should also be provided at the achievement level cut-scores (e.g., conditional SEMs at cut scores).

3.4.5 Measurement Invariance

Test fairness is a fundamental aspect of conducting group comparisons and ensuring the validity of assessments, particularly when examining differences based on gender, ethnicities, culture, or treatment conditions. To achieve test fairness, it is essential to detect and prevent any form of unfairness throughout the entire testing process, including test design, development, administration, and scoring (Camilli, 2006). When a test is free of systematic bias, measurement invariance has been met.

Differential item functioning (DIF) analysis plays a crucial role in addressing test fairness by identifying potentially biased items in a test. DIF procedures assess whether examinees from different subgroups, who possess the same underlying ability or trait, have different probabilities of endorsing an item (Angoff, 1993). By identifying items that function differently across groups, DIF analysis helps to minimize the impact of factors unrelated to the construct being measured (Sireci & Rios, 2013). Biased items systemically advantage or disadvantage a specific subgroup because of factors irrelevant to the intended construct. By addressing DIF, the fairness and validity of the test can be enhanced, ensuring that an item is unbiased and measures the same construct across groups.

From a psychometric perspective, DIF is commonly analyzed using methods that compare item performance across groups. For the WIN assessments, subgroups of interest could be gender (male/female), racial/ethnic groups, disability status (no/yes), multilingual status (no/yes), and different form accommodations. To measure DIF, the Mantel-Haenszel (MH) delta statistic

quantifies the difference in item response distribution for two groups. Researchers at ETS developed thresholds to interpret MH delta values (Zwick, et al., 1999).

Evaluation: In the test development phase, item writers followed guidelines for fairness and sensitivity to minimize bias under Universal Design principles, and each item was reviewed by professional editors and the executive content team. Due to insufficient subgroup sizes, DIF analysis has not yet been completed.

Recommendation: Evaluation to be completed after DIF analysis is conducted.

3.4.6 Equating Multiple Forms

For each of the three WIN assessments, WIN administers two forms. This is done primarily for security purposes. When multiple forms of an assessment are administered, steps must be taken to ensure that all forms provide valid and reliable scores, are comparable, and do not bias against any individuals. Equating is the statistical process of adjusting scores from different forms so they can be used interchangeably (Kolen & Brennan, 2014). For example, if Form A is more difficult than Form B, then a score of 75 on Form A may correspond to a score of 84 on Form B. Equating is the statistical process of finding the equivalent score adjustments.

For the WIN assessments, all items are calibrated using equivalent groups and are therefore on the same scale. Using the 1-PL IRT model, each total score corresponds to an ability estimate and scale score. With the total score – ability correspondence for each form, the forms can be equated by mapping ability estimates on each form to the corresponding scores on each form.

Evaluation: The process outlined in the technical report is sound. For each assessment, forms were very similar in terms of difficulty, so equating adjustments were minimal. The true score equating process was applied appropriately using the 1-PL IRT model. Based on the process applied, different forms can be used interchangeably with no concerns with threats to validity or bias.

Recommendation: After future administrations, post-equating checks should be conducted to ensure that the 1-PL difficulty parameter estimates remain appropriate and support a valid equating process.

4. Achievement Level Classifications

For summative assessments, in addition to scale scores, it is often useful to provide achievement level classifications that are coupled with interpretable descriptors of skill and understanding development for each achievement level. In an item response theory framework, cut scores for each performance level must be determined and applied to classify examinees. For the WIN assessments, cut scores were determined in a standard setting process and applied to classify examinees into one of five achievement levels:

- **Job Zones 1-2:** occupations that need very little to some preparation
- **Job Zone 3:** occupations that need medium preparation
- **Job Zone 4:** occupations that need considerable preparation
- **Job Zone 5:** occupations that need extensive preparation

This evaluation concerns the degree to which these classifications are valid and reliable. Data for this section comes from the technical report (WIN, 2026).

Evaluation: A modified Angoff standard setting was used to determine the WIN assessment achievement level cut scores. This method is commonly applied in industry and accepted as a valid procedure for standard settings (Cizek & Bunch, 2007). The cut scores were then applied to the ability estimates derived from the 1-PL scaling process to classify students into the achievement levels. Classification accuracy estimates ranging from 0.75 - 0.78 and consistency indices ranging from 0.66 - 0.71 indicate acceptable consistency for classification into the five achievement levels.

Recommendation: As noted earlier, there were very few items located near Level 3/4 and Level 4/5 operational cut scores. More difficult items could be added to improve accuracy and consistency of Level 4 and Level 5 classifications.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). Standards for educational and psychological testing. American Educational Research Association.

Angoff, W. H. (1993). Perspectives on differential item functioning methodology. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 3-23). Lawrence Erlbaum Associates.

Bandalos, D. L. (2018). Measurement theory and applications for the social sciences. Guilford Publications: New York.

Camilli, G. (2006). Test fairness. *Educational measurement*, *4*, 221-256.

Cizek, G. J., & Bunch, M. B. (2007). *Standard setting*. SAGE Publications, Inc., https://doi.org/10.4135/9781412985918

Engelhard, G. (2013). *Invariant Measurement: Using Rasch Models in the Social, Behavioral, and Health Sciences*, 1st edition. Routledge/Taylor & Francis Group.

Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices,* 3rd ed. Springer Science + Business Media.

National Center for O*NET Development. O*NET OnLine Help: Job Zones. *O*NET OnLine*. Retrieved from https://www.onetonline.org/help/online/zones

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*, 3rd ed. McGraw-Hill.

Sireci, S. G., & Rios, J. A. (2013). Decisions that make a difference in detecting differential item functioning. *Educational Research and Evaluation, 19*(2-3), 170-187.

Webb, N. L. (2002). Depth-of-knowledge levels for four content areas. *Language Arts*, *28*(March), 1-9.

Wells, C. S., & Wollack, J. A. (2003). An Instructor's Guide to Understanding Test Reliability. Testing & Evaluation Services. University of Wisconsin.

Worldwide Interactive Network (2026). *WIN Work Ready Assessments: Technical Report, Version 6.2026*. Retrieved from WIN associates.

Zopluoglu, C., & Davenport, E. C. (2017). A note on using eigenvalues in dimensionality assessment. *Practical Assessment, Research, and Evaluation, 22*(7). Available online: http://pareonline.net/getvn.asp?v=22&n=7.

Zwick, R., Thayer, D. T., & Lewis, C. (1999). An empirical bayes approach to Mantel-Haenszel DIF analysis. *Journal of Educational Measurement, 36*(1), 1-28.