End of Course Examination Program

Evaluation of US History and Constitution
Spring 2022 Test Data

Report provided to the Education Oversight Committee

Dr. Christine DiStefano

May 2023

End of Course Examination Program
Technical Evaluation of Spring 2022 Test Data
US History and Constitution

Table of Contents

Section 1

Statewide System of Standards and Assessment

1.1 South Carolina's End of Course Examination Program

As part of South Carolina's Accountability Program, students attending public schools take standardized assessments to gauge student progress and school performance. The End-of-Course Examination Program (EOCEP) is a statewide assessment program for high school students after completion of "gateway" courses in essential subject areas. The gateway courses were determined by South Carolina's State Board of Education and currently include the following courses: Algebra 1, Intermediate Algebra, Biology 1, English 1, English 2, and United States History and the Constitution (https://ed.sc.gov/tests/high/eocep/).

Scores from the EOCEP are used in a variety of ways, such as contributing to students' overall course grade, providing information reported on school report cards, and to provide accountability evidence to meet state and federal requirements. As listed in the South Carolina State Board of Education Regulation 43-262, the purposes and uses of the EOCEP tests are stated:

A. The examinations shall encourage instruction in the specific academic standards for the courses, encourage student achievement, and document the level of students' mastery of the academic standards.

B. The examinations shall serve as indicators of program, school, and school district effectiveness in the manner prescribed by the Education Oversight Committee in accordance with the provisions of the Education Accountability Act of 1998 (EAA).

C. The examinations shall be weighted 20 percent in the determination of students' final grades in the gateway courses.

1.1a. Review of the USHC End of Course Examination. As part of the requirements for receiving a high school diploma in South Carolina, students are required to pass a high school credit course in United States history. The United States History and Constitution (USHC) course and the related end-of-course test satisfy this requirement. In 2019, the South Carolina State Board of Education adopted the South Carolina Social Studies College- and Career-Ready Standards and the USHC EOCEP was revised to align to the new state standards. The USHC test administered during the 2021–2022 school year was based on the revised standards. Per the South Carolina Code of Laws, Section 59-18-320 notes the requirement of a technical review of an instrument prior to statewide adoption (https://www.scstatehouse.gov/code/title59.php). Given the change in USHC standards and creation of a new assessment, a technical review of the USHC EOCEP was conducted.

The Education Oversight Committee supported the current study as part of the responsibilities stated in the Education Accountability Act. This report evaluates the testing procedures, test construction process, and psychometric information to ensure that the EOCEP US History and Constitution produces reliable and valid information for use to evaluate student progress, school performance, and federal accountability. Review of the USHC EOCEP

information was conducted according to best practices educational measurement, as detailed by the *Standards for Educational and Psychological Testing* (*Standards*; AERA, APA, NCME, 2014).

This report is structured to provide information across multiple areas aligned with the U.S. Department of Education's Assessment Peer Review Guidelines (https://www2.ed.gov/admins/lead/account/saa/assessmentpeerreview.pdf). The objective is to conduct a review of the testing processes (e.g., test development, administration, scoring, reporting, etc.) related to the USHC assessment to ensure the quality of the instrument for operational administration as part of the South Carolina's end-of-course testing program. Data sources for the peer evaluation were provided by the South Carolina Department of Education (SCDE), the test contractor, Data Recognition Corporation (DRC), archival documents retrieved from the SCDE website (e.g., test blueprints, testing schedules, USHC revised standards, etc.) or from DRC/SCDE associates, and meetings/discussions with DRC and SCDE personnel. Datasets were provided by DRC, which included information about individual items and related psychometric indices (e.g., difficulty estimates, etc.). All parameters were calculated by the test contractor; no additional estimation of item or test parameters was conducted.

This report examines selected critical elements stated in Peer Review Guidelines; however, not all critical elements are necessary for the USHC evaluation. The *EOCEP 2021-22 Technical Report* provided by DRC includes a detailed alignment to specific *Standards* considered with the USHC assessment (DRC, 2022a). To focus discussion and attention on the review of the testing situation and evaluation of the Spring 2022 USHC data, individual *Standards* are not noted as these are included in the *EOCEP Technical Report*.

The dataset analyzed for this report is from the Spring 2022 administration of the USHC assessment as part of the EOCEP program. The USHC test administered during the 2021–2022 school year was a newly developed instrument, constructed in response to the state's adoption of revised US History and the Constitution Standards in 2019. As the instrument was new, the state's requirement to use the USHC scores in course grade calculations was waived and the test scores and item information examined here did not have any effect upon a student's course grade. Further, while most students have returned to in-person schooling, it is recognized that lingering effects of the COVID-19 health pandemic may have affected the scores. In light of these caveats, the USHC assessment results provide preliminary information concerning the appropriateness of the instrument to measure the standards currently in place across South Carolina. The results should be interpreted in the context of circumstances related to the COVID-19 pandemic, including school closures, nonstandard instruction delivery modes in the 2021–2022 school year, potential diminished opportunity to learn for students, and other unknown effects of the pandemic on teachers, students and their families.

1.2 The End-of-Course Program and the USHC EOCEP Assessment

The EOCEP USHC assessment is a required element by all South Carolina public-school students taking the US History and Constitution course as part of a credit bearing requirement to earn a high school diploma. The SCDE website provides information about the EOCEP.
for stakeholders to learn about the state's end of course examination program. For example, the website states test items are aligned to the South Carolina Academic Standards within each content area and the test items assess the stated content knowledge and skills. End-of-course examinations gateway subjects are offered three times a year (Fall/Winter, Spring, Summer) and tests may be taken in electronic or paper format.

Each test included in the EOCEP has a section to describe the test. Information regarding the USHC examination is provided in multiple areas of the SCDE website including the EOCEP link under the Testing and Assessment Tab (https://ed.sc.gov/tests/high/eocep/) and the Social Studies Instructional area of the SCDE website (https://ed.sc.gov/instruction/standards-learning/social-studies/). USHC blueprint information and standards covered by the examination are easily accessible. There is a separate information section for students and parents; while not directly related to test design, this information is helpful for stakeholders' understanding of the broader EOCEP. Translated documents for parents (e.g., assessment brochures, sample individual student reports) are provided in English, Large Print (English) and 10 additional languages.

1.3. EOCEP Population and USHC Examinees

All public middle school, high school, alternative school, virtual school, and adult education students enrolled in courses in which the academic standards corresponding to the EOCEP subjects are taught, regardless of course name or number, must take the appropriate end-of-course test. The population of students eligible for the EOCEP includes most high schoolers in South Carolina, including students with an Individual Education Plans (IEP) or 504 plans who are able to take the test with appropriate accommodations and supports (e.g., large print, Braille, read aloud administration, Sign Language Administration). This includes students as required by the Individuals with Disabilities Education Improvement Act (IDEA) and by Title 1 as noted by the Elementary and Secondary Education Act (ESSA).

In addition, the state testing policy includes suspended students, home school students who are registered through the district or local school board, homebound students, and home-based students as part of the EOCEP population. Also included are English as a Second Language/English Learning (ESL/EL) students, charter school students (including virtual charter schools), and students who are incarcerated. The *2021-22 EOCEP Technical Report* (DRC, 2022a) defines these groups as well as Special Groups of students including: Adult Education Students with Disabilities; Home School Students, Foreign Exchange Students, among others, which may be included in the EOCEP population of examinees. The population of EOCEP test takers does not include students who meet eligibility criteria for alternate assessments as determined by their IEP team. In addition, the course does not apply for students who are enrolled in a non-diploma course.

As the EOCEP does include students that can take the test with approved accommodations that are part of a student's IEP or 504 plan, the SCDE website details the definition of an accommodations and the purpose of such measures relative to test taking practices. Accommodation details are easily found under the Tests section of the SCDE website, within the EOCEP block of information (https://ed.sc.gov/tests/assessment-information/testing-swd/accommodations-and-customized-forms/).

The *Technical Report* provided by DRC details demographic characteristics of students who participated in any of the USHC EOCEP administrations during the 2021-22 academic year (Fall/Winter 2021, Spring 2022, and Summer 2022). As stated in the *Technical Report*, 53,055 students participated in the USHC assessments during this time period. Across the three examinations, the USHC population of test takers was roughly equally split by gender (49% male, 50% female), with students of White (50%), African American/Black (29%) or Hispanic (11%) made up the three most prominent racial/ethnic groups. Roughly 90% of the USHC sample were proficient in English (English Speaker II). Of the USHC 2021-22 population tested, roughly 8.5%

were on an Individualized Education Plan; 18% Gifted Learners (academically, artistically, or both), and 3% of examinees had a 504 educational plan.

Spring 2022 database information was used to estimate an average of 33,739 students taking the spring assessment (using information from across forms and USHC items responses). The Spring 2022 administration captured the majority of the USHC population, encompassing roughly 63.4% of the USHC test takers reported in the *2021-22 EOCEP Technical Report*. Data from the Spring 2022 database will be used for analyses and investigation of item parameters.

1.4. Summary: Statewide System of Standards and Assessment

The SCDE website provides information describing South Carolina's EOCEP. Information is clearly detailed for educators and other stakeholders. Relevant information about the purposes of the testing program, uses of the information, and areas tested with the EOCEP is easily accessible. Detailed information is presented on the SCDE website regarding the purpose of the test and uses of USHC results. The information is provided in multiple places within the SCDE website, allowing stakeholders to come across the same information from different search avenues. Information regarding the purpose of the USHC, information to be covered, and other information (e.g., sample items, data reviews of results from past USHC administrations, etc.) are readily and easily accessible on the SCDE website.

The database provided from DRC included responses from over 33,000 students who took the USHC end of course assessment in Spring 2022. The large number of students involved with the spring SHC test administration is acceptable to produce stable psychometric index estimates for use in the peer evaluation.

Section 2

Assessment Systems Operations Related to the USHC EOCEP

This section provides a review of the USHC examination to align with current recommendations for best practices of test development and test construction (e.g., Bandalos, 2018; Green, 2009; Mertler, 2016). The test specifications, test blueprint, test administration manual, and item development procedures are examined. Proper test development procedures support use of USHC results to assess student knowledge and provide accountability evidence.

Test specifications typically contain two components: a test description and a test blueprint. The test description specifies aspects of the test such as the test purpose, the target examinee population, and the overall test length. The test blueprint provides a listing of the major content areas and cognitive levels intended to be included on each test form. Testing administration procedures are reviewed to ensure alignment with best practices. This section was informed by the SCDE document "*United States History and the Constitution Standards and Assessment Crosswalk*" conducted in March 2023 to show how the USHC examination was updated to align with the 2019 South Carolina Social Studies College- and Career-Ready Standards, SCDE website documentation, and datasets information provided from DRC. A detailed evaluation is provided after each the review and discussion of each component; the summary section provides an overall reflection of the elements in Section 2.

2.1. USHC Assessment: Test Design and Test Development

The test design and test development components are essential to the validity process On the SCDE website (https://ed.sc.gov/tests/high/eocep/), the *Tests* area provides additional information about all EOCEP tests, a description of the purpose of the testing program, how scores are used in calculation of student grades and how EOCEP scores are used as part of federal accountability requirements. Additional important information such as: testing window dates, performance level descriptors, and data reviews of past test administrations are noted.

Each test in the EOCEP has a separate section. For the USHC assessment, links are displayed allowing educators and stakeholders easy access to standards and the test blueprint. Additional information includes a data review discussing results from past USHC assessments, including general observations of student skills and suggested instructional strategies to accommodate potential areas of lower performance. When preparing students, teachers can easily link to the revised Social Studies standards (https://ed.sc.gov/instruction/standards-learning/social-studies/) for more detail about the content areas, indicators, instructional resources and activity ideas.

2.1a. Test Blueprint. The Test Blueprint provides an overall description of the USHC administration and construction. The test description is included as a bulleted list and includes pertinent information of test length, projected time needed to take the assessment, test administration, and scoring information (Note: current example: https://ed.sc.gov/tests/tests-files/eocep-files/ushc-test-blueprint-2022-23/). Information in the tables and bulleted list is simple,

easy to read, and focuses the reader's attention on the most important aspects of the USHC test (e.g., number of items total and per area, item difficulty levels, item formats).

Test Blueprint: Coverage of Standards. The revised EOCEP USHC assessment measures five main content areas (standards) with six indicators per content area (30 indicators total). This is a reduction from the previous version of the USHC (Previous: 8 standards and 47 indicators). The reduction allowed the revised USHC assessment to focus on a more conceptual understanding of history, while allowing more opportunities for analysis and inquiry skills. All six areas are stated to have roughly equal weight to the overall test, with between 10 and 12 items per standard. The blueprint states that the USHC assessment consists of 55 total items. Table 1 provides a summary of the test blueprint information by test reporting/content area as included on the USHC EOCEP.

Table 1. EOCEP US History and Constitution: Test Blueprint

Reporting Category (Key Concepts)	Number of Indicators	Number of Items per Reporting Category	Percentage of Assessment
Standard 1: Foundations of American Republicanism	6	10-12	18-22%
Standard 2: Expansion and Union	6	10-12	18-22%
Standard 3: Capitalism and Reform	6	10-12	18-22%
Standard 4: Modernism and Interventionism	6	10-12	18-22%
Standard 5: Legacy of the Cold War	6	10-12	18-22%

Evaluation: Test Blueprint. Examination of the Spring 2022 USHC test data showed that the number of items per standard aligned with the Blueprint. All six indicators were assessed from each standard (i.e., 100% of the standard was represented on the test). Across the five reporting categories (i.e., Standards), each area was equally weighted, accounting for 20% of the 55-item test. The blueprint information is well suited to inform stakeholders of what is expected on the EOCEP USHC assessment.

2.1b. Depth of Knowledge. The EOCEP USHC uses the Depth of Knowledge (DOK) classification system to categorize items into cognitive complexity levels. The DOK categorizes items into one of four categories, where higher numbers indicate higher levels of complexity, with levels defined as:

Level 1. Recall and Reproduction: Tasks at this level require recall of facts or rote application of simple procedures. The task does not require any cognitive effort beyond remembering.

Level 2. Skills and Concepts: This level requires some decision making. Tasks which include more than one mental step (e.g., comparing, predicting, organizing) are included.

Level 3. Strategic Thinking: Tasks at this level use planning skills and higher order thinking skills to solve more abstract tasks. Tasks with more than one correct answer or justifying a position are examples.

Level 4. Extended Thinking: At the most complex cognitive level, these tasks require synthesis of information from multiple sources or transfer of knowledge from one domain to another.

It is not typical for standardized tests to include items at DOK Level 4; however, the USHC assessment should have a mix of items across Levels 1 through 3. The test may be considered a "potentially high stakes" test as a sizable part of a student's grade (20%) is linked to the EOCEP test score and for some students, passing the US History and Constitution course may be dependent upon the end-of-course exam score. Test construction recommendations suggest that the test includes varied skills, including a mix of easier DOK (Level 1) and more complex DOK (Level 3) levels. The test blueprint should describe the total number of items to be included in each content area as well as the total number of items at each DOK level. This information assists teachers and students target time and content allocations for test preparation activities. As stated on the test blueprint, it is estimated that between 0% - 15% of the USHC test will be DOK Level 1 items, between 55% - 75% of items at Level 2, and between 25% and 45% at Level 3.

In addition, the revised 2019 Social Studies standards associate indicators with one of six skill levels based on disciplinary skills aligned to DOK levels. The redesign of the standards and integration of skill levels was conducted to encourage inquiry, higher order thinking skills, and meaningful learning of Social Studies content needed for college, career, and civic readiness. These changes inherently increased the rigor of the standards by requiring students to use the identified historical thinking skills to make broader connections between historical events and developments. The six skill levels are presented in Table 2.

Table 2. Skill Levels Associated with the 2019 Social Studies College- and Career-Readiness Standards

Skill Level	Definition
Comparison	Generate comparisons based on common or differing characteristics or contexts.
Causation	Analyze multiple causes and effects, to include distinguishing long-term and short-term examples.
Periodization	Organize a historical narrative into time periods using units of time (e.g., decades, half-centuries, centuries).
Context	Describe historical developments using specific references to time, place, and broader circumstances.
Continuities and Changes	Recognize patterns of historical continuities and changes and identify turning points in history.
Evidence	Identify source, and utilize different forms of evidence, including primary and secondary sources, used in an inquiry-based study of history.

Evaluation: Blueprint DOK. From the blueprint review of DOK levels, the test will be more heavily weighted at DOK Level 2 (Skills and Concepts), with between 55% and 75% of the items at this complexity level. Including most of the USHC items Level 2 is appropriate, given the purpose of the end of course examination. In addition, having the fewest percentage of items at DOK Level 1 is acceptable, as this positions the USHC assessment at (approximately) a medium to medium-hard level of complexity, with most items beyond basic recall of information.

This "hardness" level is appropriate to assess a student's comprehension of material presented after participation with US History and Constitution course content.

2.2 USHC Assessment: Item Development

This section discusses the item development. Items for the USHC EOCEP assessment were constructed to assess the content knowledge and skills described in the 2019 Social Studies standards revision. Information reported in this section on the development of items comes from SCDE website documentation, the *2021-22 EOCEP Technical Report* (DRC, 2022a), and discussions with SCDE and DRC personnel.

2.2a. Item Formats. A variety of item formats may be used on the EOCEP assessments. The majority of items are typically (closed response) test questions which require selection of the correct answer(s) to achieve full credit. Multiple choice, or Selected Response, is the most commonly encountered item format on standardized tests. This format largely consists of an item stem and options for the respondent to select the correct response(s) from a set of alternatives, or distractor choices. According to best practices for test construction (Green, 2009), the distractor options should be plausible responses and help to distinguish among examinees with varying levels of knowledge. Closed response questions can be machine scored, allowing many examinees to be tested in an efficient manner (Green, 2009). Typically, Selected Response items ask respondents to select the correct response from four possible alternatives, three of which are distractors and one correct (keyed) alternative.

Other objective response items per session may be present. Multiple Selection items prompt students to select a number of correct answers (e.g., "*Choose two answers*…"). The multi-select items may have 5 or 6 options to select from. In order to receive credit for a correct response, students must select all of the correct answer choices. Evidence Based items are two-part items. Students read a piece of text or passage and choose the best answer from the answer choices. Students will then be asked to support their response with evidence from the text—for example, to select multiple evidence statements, place multiple dates or steps in correct sequence, etc. In order to receive a correct response, students must answer both parts of the item correctly. Technology Enhanced items (for online test takers) ask students to interact with an item by using technology to provide their response, such as "drag and drop" where elements are moved into different positions, highlighting text, or clicking on images. (If needed, comparable selected response items are used as a replacement for the technology enhanced items on paper/pencil test administrations).

The SCDE website, EOCEP tab, provides online training tools for teachers and students interested in practicing specifically with online test-taking and/or technology enhanced items (https://wbte.drcedirect.com/SC/portals/sc). Teachers may also use released items to help students practice with types of items and DOK levels to be encountered on the USHC assessment.

2.2b. Item Pool Construction. A large pool of items was constructed for the USHC forms, where items were vetted by multiple committees. As noted in the *2021-22 EOCEP Technical Report* (DRC, 2022a) *"Newly developed items were reviewed by committees of South Carolina educators for content area and bias and sensitivity issues; items approved by these committees and the SCDE were field-tested among South Carolina students. Items demonstrating satisfactory*

performance on field tests became eligible for inclusion in operational forms during the subsequent administration" (p.22).

For EOCEP assessments, all items in the item pool were evaluated by item development committees using the following criteria:

- **Content alignment**— match of items to a standard and indicator to ensure alignment,
- **Rigor-level alignment**—evaluation of cognitive complexity and appropriateness to the level of rigor required,
- **Technical design**—item is current and accurate and stem, stimuli, distractors, and answer options are clear and concise, appropriate for the grade level, and considerate of students with special needs,
- **Universal design**—item provides for an accessible assessment for all students, focusing on language demand, format/complexity, and graphics/visuals, and
- **Fairness in testing**—item generates valid test scores for all groups of test takers through avoiding bias in test items and/or content area and avoiding language that unduly distracts students or disrupts their performance.

Activities for reviewing newly constructed USHC items were conducted by DRC in collaborating with SCDE staff, Content Specialists, and Bias and Sensitivity review committees. The members of the review committees provided feedback for each item, and committee facilitators recorded the committee decisions. Items accepted for use on the EOCEP assessments constituted the pools of items from which subsequent test forms for future Spring administrations may be created. As stated in the *2021-2022 EOCEP Technical Report*, a total of 160 items were developed for the USHC item pool (DRC, 2022a). The number of items by item format is summarized in Table 3.

Table 3. Item Formats, USHC Item Pool

Item Format	Number of Items	Percentage of Item Pool
Multiple Choice	140	88%
Evidence Based Selected Response	5	3%
Multiple Selection	9	6%
Technology Enhanced	6	4%
Total	160	

Note: Percentages may not total 100% due to rounding

Evaluation: USHC Item Formats and Item Pool. In summary, The SCDE website describes the item formats which students may encounter and provides support for teachers to practice these skills with students. The variety of formats helps to ensure that students are being assessed with best practices. The USHC item pool includes a majority of multiple-choice type items (roughly 85%), which is not unusual for standardized tests and the test blueprint notes that a variety of item formats may be encountered. Materials are provided on the SCDE website for teachers and students to practice with released items and with the online test taking platform.

Materials detailing construction of the EOCEP US History and Constitution item pool are described in the *Technical Report* provided by DRC. Items for the USHC have undergone an extensive review by multiple committees to ensure they are appropriate for all learners, at the appropriate level of rigor, and aligned with the content. The item pool developed by South Carolina educators is sufficiently large to construct a variety of USHC EOCEP alternative forms, while examining field test and other statistics to ensure psychometric quality of the content.

2.2c. USHC Form Construction. For test security, more than one USHC form is constructed; however, specific guidelines need to be followed to ensure forms are equivalent in content and rigor, psychometric quality of items and coverage of the standards. The *2021-22 EOCEP Technical Report* (DRC, 2022a) describes in detail the procedures used to construct forms for the USHC testing and the criteria used to evaluate items and item content. Forms were created for each testing window as well as paper- and custom forms. The majority of USHC students take the end-of-course assessment at the Spring testing using an online platform (DRC INSIGHT). These forms included eight field test items (total of 63 test items); other forms included 55 items. Regardless of form or testing window, 55 items are scored for the USHC EOCEP.

The 20-forms available for Spring testing allow rotation of forms within and across time points, enhancing test security. While same items were used across forms, the items were placed in similar, but different positions across forms (e.g., an item with position between 50 to 54 across the 20 different test forms). As common scored items were used across the forms, the DOK and psychometric levels of the forms are equivalent.

Evaluation: USHC Form Construction. Forms created for the USHC EOCEP were thoughtfully constructed according to best practices, with reviews and examination by numerous committee members of educators, SCDE, and DRC personnel. Each step was detailed in the *Technical Report* to provide a clear understanding of what procedures were followed. A total of 20 different forms were created for the USHC EOCEP Spring testing opportunities, each with 63 items. The 55-scored items comprise 87% of a given form and field test items only 13%. The relatively low percentage of field test items is sufficient to collect information about item performance without overly burdening or distracting students. The form creation process, number of forms created, and varied item placement across forms provides an opportunity for evaluation of new (field test) items and enhances test security.

2.3 USHC Assessment: Test Administration Procedures

As part of the state-wide standardized testing program, the EOCEP USHC assessment follows state and district regulations related to testing procedures such as: adherence to test security, regulations for distribution of materials, confidentiality mandates, and reporting of test violations. As with other standardized tests administered in South Carolina, District Test Coordinators and School Test Coordinators oversee test security and appropriate testing practices for the USHC assessment. This analysis includes a review of test administration procedures, instructions provided for those administering the assessment, instructions provided for students, accommodations, and test security procedures.

As part of the EOCEP, the USHC assessment is largely delivered online through the test contractor's online platform, DRC INSIGHT. This platform is responsible for delivering the assessment, storing responses, scoring the test, and providing test reports to students, districts, and schools. Paper-and-pencil test administrations are available if required as part of a student's educational plan due to disability. Tests may be administered to examinees during the academic

year's testing windows. The testing windows for all tests are prominently displayed on the SCDE website under the Assessment tab (https://ed.sc.gov/tests/assessment-information/). Detailed instructions for test administration are stated for district test coordinators and school test coordinators in a detailed Test Administration Manual (TAM). The TAM is easy to find on the SCDE website, EOCEP tab (e.g., https://ed.sc.gov/tests/tests-files/eocep-files/spring-2023-tam/).

Instructions for students are read aloud by the Test Administrator. The instructions follow a script, helping to ensure fidelity of test administration as all students in the state will receive the same instruction. Instructions are short, direct sentences with clear, easy to understand language. The TAM includes a section on appropriate accommodations for students and documentation regarding how approval for use of accommodations is determined.

Evaluation: USHC Test Administration Procedures. The TAM clearly describes testing instructions, including a listing of steps to be taken before testing, during testing, and after testing. Test security procedures are clearly detailed in the TAM and the TAM Appendix includes the confidentiality forms to be completed by school/district testing personnel. Links to report test violations are included in the TAM and on the SCDE website. The SCDE website provides easy to find information about test security regulations that must be followed during testing (https://ed.sc.gov/tests/assessment-information/test-security).The test administration procedures are clear and complete. The document provides clear instructions for district/school testing personnel to follow. In addition, the TAM provides advice on scenarios which may arise (e.g., student getting sick during testing, disruptive students, suspected cheating) and recommendations for handling the situation.

As part of the EOCEP program, the USHC testing adheres to the same procedures as other standardized test administrations. Standardized information as detailed above (i.e., use of TAM, test coordinators, etc.) helps to ensure that all USHC test takers receive the same procedures and conditions, regardless of test format or test window in which the USHC is taken. These administration procedures provide clear directives to deliver the USHC end of course assessment properly and with fidelity. Clear, objective information that is followed by all district/school testing personnel helps to ensure uniform testing procedures are delivered to all USHC examinees across the state. Easily accessible information helps ensure that all testing coordinators are well-informed, have appropriate training, and follow relevant security procedures. Access to uniform testing procedures can help ensure validity associated with the EOCEP and USHC scores for use with accountability and decision making.

2.4 Summary: Assessment Systems Operations Related to the USHC EOCEP

The assessment systems operations section evaluates the procedures used to develop and administer an assessment such as the test specifications, test blueprint, item development procedures and administration procedures. The SCDE provides clear, easy to understand test specifications prominently on their website; the test specifications are provided as a bulleted list, along with the test blueprint. The USHC EOCEP test blueprint includes a listing of the content areas and cognitive levels to expect; these levels are aligned with the data analyzed from the Spring 2022 USHC test administration. Other information, such as sample items, past data reviews, and suggestions for teaching/activities are readily available. The materials help students and teachers understand what is to be included and how to prepare for the assessment. Test administration procedures are clearly documented and defined for testing administrators in the TAM. The detailed instructions support the standardization procedures; uniform directions and

common materials are provided for all test takers across the state and across time points. In summary, the assessment operation procedures for the USHC EOCEP examination align with current recommendations for best practices of test development, construction and administration (e.g., Bandalos, 2018; Green, 2009; Mertler, 2016).

Section 3

Technical Quality – Validity

As stated in the *Standards*, validity is as defined the degree to which evidence and theory support the interpretations of test scores their intended uses. "*Validity is, therefore, the most fundamental consideration in developing tests and evaluating tests*" (AERA, APA, & NCME, 2014, p. 11). Test score validation is the process by which the interpretations associated with test scores hold meaning, providing trustworthy information for decision-making events. With the EOCEP, these events may include representing an examinee's level of USHC knowledge, evaluating school performance, or comparing relative performance across districts. Validity is an ongoing process, including all aspects of the testing process including design, content area specifications, item development, psychometric quality, scoring, and inferences made from the results.

Section 3 investigates the technical quality associated with the USHC EOCEP examination including evaluation of content, cognitive processes, internal structure, relations to other variables, and an assessment of overall validity. Information for this section comes from evaluation of SCDE documents and Spring 2022 USHC EOCEP database provided by DRC.

3.1 Overall Validity, Including Validity Based on Content

Content validity entails careful assessment of the items and domains included on an examination (Bandalos, 2018). The information helps to ensure that the material included on the test is representative of the target domain (i.e., USHC course content). Careful specification of content and review of the items representing the target domain is needed to ensure that the information obtained from administering the USHC is relevant (i.e., construct-irrelevant variance minimized), with the full range of the construct(s) measured (i.e., construct underrepresentation minimized). As noted in the *EOCEP Technical Report* (DRC, 2022a), the attention paid to the USHC test development process and the involvement of South Carolina educators contributes to the validity of the assessment. As an initial content review, alignment of the end of course test content was compared with the USHC Standards to review the accuracy of the included test content to the guidance provided on the test blueprint materials.

3.1a. Domain Alignment to Test Blueprint. Item alignment to USHC test blueprint was conducted by reviewing the standard descriptions from the Spring 2022 examination and matching these to the stated USHC Domain. Item descriptors appeared aligned with content areas; no mismatch between descriptor and domain was noted. Items statistics were reviewed to determine that the number of items stated, percentage of items to the total test, and standards included were in line with the guidance reported in the test blueprint.

All USHC categories were in line with information reported in the test blueprint in terms of domain coverage and percent of total test. Each reporting area assessed all six indicators included in the area (100% coverage). In addition, the number of items on an assessment matched the number stated on the blueprint. Each Reporting Category contained 11 items, in accordance with the number stated in the test blueprint. Each of the five categories category contributed an equal amount of weight to the overall USHC examination, each contributing 20%

of the total test content. Each indicator was given one or two items, 11 items per each standard. Table 4 summarizes domain coverage of the USHC examination.

Table 4. EOCEP US History and Constitution: Domain Coverage

Reporting Category (Key Concepts)	Number of Indicators	Percent of Domain Coverage	Number of Items From Blueprint	Number of Items per Reporting Category	Percentage of Assessment
Standard 1:Foundations of American Republicanism	6	100%	10-12	11	20%
Standard 2: Expansion and Union	6	100%	10-12	11	20%
Standard 3: Capitalism and Reform	6	100%	10-12	11	20%
Standard 4: Modernism and Interventionism	6	100%	10-12	11	20%
Standard 5: Legacy of the Cold War	6	100%	10-12	11	20%

3.1b. Item Formats Included on the USHC Examination. Item formats were examined for the different forms administered in Spring 2022. While there are 55 scored items, Spring testing with the USHC examination included 63 items for the "typical" testing scenario (e.g., online), where 55 items were operational and eight additional (unscored) items were used to collect field test data. Other forms (e.g., paper/pencil, "emergency" forms for security compromises, etc.) include 55 items. The test blueprint notes that a variety of item formats may be used. Table 5 examines item formats from the USHC Spring 2022 EOCEP.

The majority of items included on the USHC examination were multiple choice format questions, comprising 94% of the spring assessments, for both typical and "other" forms. Depending on the time of year that the test was taken, between 4% and 6% of the items were technology enhanced. Evidence based selected response, multiple selection, and drag-and-drop items were included under the Technology Enhanced heading. The percentage of the different item formats percentages made up between 2% and 5% of a USHC test; for test security, the percentage of items for various form administrations is not broken down further.

Table 5. Item Formats Included on USHC Forms

Item Format	Other USHC Forms	Percentage of Assessment	Typical USHC Forms	Percentage of Assessment
Multiple Choice	53	96%	59	94%
Technology Enhanced	2	4%	4	6%
Total	**55**	**100%**	**63**	**100%**

Best practices of test construction state that the correct answer for items should be varied across options (e.g., all keyed responses are not A) and should not follow a pattern (e.g., A, B, C, A, B, C, etc.) (Green, 2008). To ensure that best practices of test construction were followed, the

correct keyed response for the USHC EOCEP items was investigated for the set of 55-operational items used in Spring 2022. For the 55-items, items were varied with each letter (e.g., A) being the correct option an equal number of times (less the technology enhanced items). For test security, the number of items by response is not revealed, but best practices of test construction were followed in construction of the USHC correct responses.

Evaluation: USHC Domain and Item Format Alignment to Test Blueprint. In sum, USHC items align with the Key Reporting Areas noted in the Test Blueprint. The number of items by standard was in concordance with the percentage of items to be expected by content domain. Each Key Reporting area was given equal weighting to the overall assessment. The correct answer was also equally distributed across response options (e.g., A) and varied for the keyed response. The test blueprint notes that different item formats may be encountered on the USHC end-of-course examination. While the majority of the test is multiple choice, other item types were present; the percentages reflect the percentages in the available item pool for the USHC EOCEP assessment.

3.2 Validity Based on Cognitive Processes

As noted from the *USHC Standards and Assessment Crosswalk* (SCDE, 2023) which outlines the similarities/differences between the 2011 and 2019 South Carolina Social Studies standards and assessment, the changes made to the revised standards allowed greater opportunities for deeper analysis and inquiry. These changes increased the rigor of the standards by requiring students to use the six noted historical thinking skills to make broader connections between historical events and developments.

3.2a. USHC Historical Skills and DOK Levels. The historical thinking skills included in the 2019 revised standards ask students to interact with social studies content to make comparisons, analyze evidence, or determine patterns of continuity and change. To examine the cognitive processes included on the USHC EOCEP assessment, items from the Spring 2022 test were examined by skill level. As shown in Table 6, the USHC EOCEP assessment includes all six skill areas, with roughly equal weight given to each area. The number of items per area ranged between 8 to 10 items, accounting for between 15% to 18% of the USHC test.

Table 6. Historical Skill Areas Included on the USHC EOCEP, Spring 2022

Skill Area	Number of items	Percent
Comparison	9	16%
Causation	10	18%
Periodization	9	16%
Context	8	15%
Continuities and Changes	10	18%
Evidence	9	16%
Total	**55**	

Note: Percentages may not total 100% due to rounding

The revised Social Studies standards and deeper cognitive levels demonstrated with the historical skills resulted in increased cognitive complexity of USHC test items. The testing blueprint noted this shift, stating between 0% - 15% of the USHC test items would be at the lowest DOK Level (Level 1), between 55% - 75% at Level 2, and between 25% and 45% of items at at

the highest DOK level. As noted in the *USHC Standards and Assessment Crosswalk* (SCDE, 2023), no DOK table was included with the 2011 USHC Test blueprint. Table 7 reports the DOK Levels of the items on the Spring 2022 USHC EOCEP assessment. The majority of the items are at higher DOK levels. There are only 4% of the items at the lowest level; 96% are at DOK Level 2 and Level 3 combined. The majority of USHC items are at DOK Level 2 (75% of items).

Table 7. Depth of Knowledge Levels, Spring 2022 USHC EOCEP Assessment

DOK Level	Number of items	Percentage of Assessment
1	2	4%
2	41	75%
3	12	22%
Total	**55**	

Note: Percentages may not total 100% due to rounding

Evaluation: USHC Historical Skills and DOK Levels. Overall, the USHC assessment test is more rigorous than the past assessment, which is aligned with the more challenging Social Studies standards adopted in 2019. The assessment includes all six historical areas, devoting roughly equal attention to each area. The skill levels for the items are at higher levels (DOK 2 and 3), challenging students to use skills. This is also noted with the *USHC Standards and Alignment Crosswalk* document (2023), which showed 2011 Social Studies standards and the related USHC EOCEP assessment included more skills at lower cognitive levels. The SCDE website includes materials for stakeholders to become familiar with the types of test questions at different DOK levels, so examinees gain experience with the types of questions posed and responses expected.

3.3 Validity Based on Internal Structure

Validity evidence of a measure's internal structure involves examination of how well test items function to measure the construct of interest. If the items function acceptably as observable representations of the construct, then the evidence supports using scores from the instrument. Analyses used to support validity focus on the individual items using data from the target population to investigate characteristics of individual items and their relation to the total construct. This section examines item analysis information for the Spring 2022 USHC EOCEP administration database provided by DRC and the *2021-22 EOCEP Technical Report* (DRC, 2022a). Item statistics were calculated using Classical Test Theory techniques and modern test theory techniques under the Rasch measurement model. All item parameters and statistics were calculated by DRC. Summaries of item statistics (e.g., mean difficulty values, standard deviations) were computed; no additional psychometric analyses were performed.

3.3a. CTT Based Item Analysis. Two Classical Test Theory (CTT) indices were included in the dataset: item difficulty and adjusted point-biserial correlation values. CTT-based item difficulty (p) is defined as the proportion of students out of the total number of examinees answering an item correctly. Higher p values indicate easier items (i.e., a greater number of students selected the correct answer) and low p-values indicate more difficult items. Items which are too difficult (or, conversely, too easy) do not differentiate between low performing and high performing students. A difficulty value of p = .5 provides the highest level of differentiation (Bandalos, 2018).

The adjusted point biserial correlation (or item-total test correlation) is a measure of association, illustrating how well an item discriminates between high performing and low performing examinees. Values are calculated as the correlation between an item's score (correct/incorrect) and the total score, with the item in question removed from the total test score. The normal range of point biserial scores for items is –1 to +1, with higher values indicating that the item discriminates well between high and low performing students (Bandalos, 2018). Values of the point biserial may be positive, meaning that the item is discriminating appropriately. Negative values indicate that the item is not discriminating as intended, illustrating the tendency for higher ability students to select an incorrect answer and more of lower ability students to select the correct answer. Values that are close to zero or negative may indicate a flawed item. A value of zero means that there is no discrimination between high and low ability test takers. A high point-biserial coefficient means that students selecting the correct response are students with higher total scores and students selecting incorrect responses to an item tend to have lower total scores. In genera adjusted point biserial correlation, values should be at a moderate to higher correlation value (e.g., roughly .3 to .5) (Bandalos, 2018). In general, items should not have a low discrimination value (e.g., < .20), as this indicates that the item cannot differentiate between examinees with high and low ability. The *2021-22 EOCEP Technical Report* states recommended psychometric guidelines for including items on a test form (DRC, 2022a). This includes a p-value between 0.30 and 0.85 and an adjusted point-biserial correlation greater than 0.20.

Evaluation: CTT Difficulty. The average CTT-base difficulty value across the 55-items administered in Spring 2022 was p = .53 (standard deviation = .12), meaning, on average, students answered 53% of the EOCEP USHC Spring 2022 items correctly. This is at a moderate difficulty level, with the p-value very close to the value which maximizes differentiation among examinees. Figure 1 provides a histogram of difficulty values. USHC items yielded a minimum difficulty of p= .31 (i.e., 31% of examinees answering the item correctly) to a maximum of p = .76 (i.e., 76% of examinees answering the items correctly). As shown in the figure, the set of Spring 2022 EOCEP USHC items included a mixture of "harder" and "easier" items.

Item difficulty values were reviewed to determine the number of USHC items that were challenging for students, where p <.50. There were 44% items on the assessment with p-values below .50, seen as more challenging items for students (24 of 55 items). Slightly more than half, 56%, of items that were easier, noted at or above a CTT-based difficulty level of p ≥.50 (31 of 55 items). The USHC assessment is roughly balanced in terms of item difficulty.

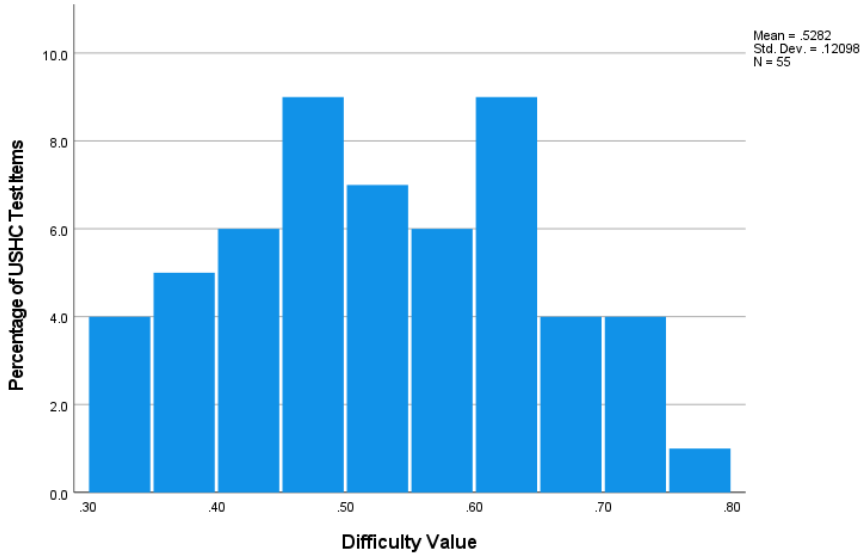

Figure 1. EOCEP USHC CTT-based Difficulty Values, Spring 2022

CTT difficulty values were examined across item format; however, there are relatively few technology-enhanced items on the Spring 2022 USHC EOCEP administration. Descriptive statistics are provided in Table 8. Technology-enhanced items reported a lower average p-value, showing that these items as a set were more difficult than the multiple-choice items, but with much larger fluctuation across the set. Overall, the different item formats were roughly equal in terms of average difficulty.

Table 8. Descriptive Statistics for USHC Spring 2022 Difficulty Values, By Item Format

Item Format	N	Mean	Standard Deviation	Minimum	Maximum
Multiple Choice	52	.53	.12	.31	.76
Technology Enhanced	3	.47	.22	.31	.73

Over the set of 55 EOCEP USHC items administered in Spring 2022, the item difficulty values appear to be acceptable, given the purpose of the test. Average values generally report a test of moderate difficulty, with many of the items approximating the midpoint, p=.50, level of difficulty.

Evaluation: CTT Discrimination. Across the 55-items USHC administered in Spring 2022, the average discrimination value was 0.38, illustrating that the set of test items are discriminating acceptably between examinees of different ability levels. Generally, USHC examinees with lower total test scores chose incorrect responses and higher ability students chose correct responses. Adjusted point biserial correlation values ranged from .17 to .54; however, most USHC EOCEP item discrimination values are between .35 and .45. The one item with a discrimination of .17 reported a moderate p-value (.52), but one of the item distractors reported a (low) positive value, highlighting some inconsistencies in examinee responses.

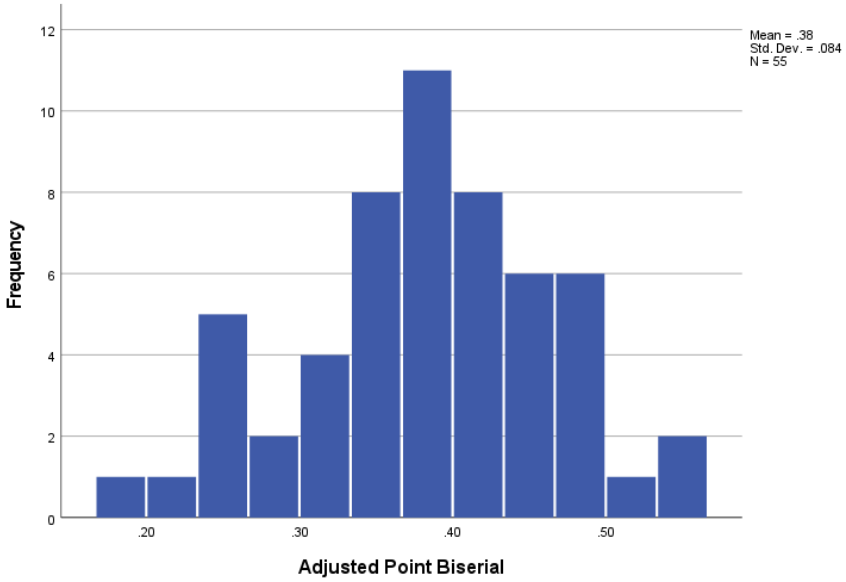

Figure 2. Distribution of USHC EOCEP Discrimination Values, Spring 2022

Considering item formats, mean adjusted point biserial values for the technology enhanced items were roughly equivalent to those for multiple choice items. Again, it is cautioned when interpreting values as few Technology Enhanced items included on any one USHC EOCEP form.

Table 9. Descriptive Statistics for USHC Spring 2022 Discrimination Values, By Item Format

Item Format	Number	Mean	Standard Deviation	Minimum	Maximum
Multiple Choice	52	.38	.09	.17	.54
Technology Enhanced	3	.39	.06	.34	.45

In summary, the USHC items are adequately discriminating between students with higher and lower skill levels overall and by item format. The discrimination levels are appropriate for the purpose of the assessment and values are in line with other state-wide examinations.

3.3b. Omit Rates and Distractor Analysis. A distractor analysis for selected response questions is an extension of item analysis. Here, we are no longer interested in how test takers select the correct answer, but how the distractors function to draw an examinee away from the correct answer. This is an important component to distinguish between examinees at varying levels of the latent domain. Distractors that are not effective are virtually useless and may provide a greater probability to select the correct answer by guessing.

Discrimination indices are calculated to determine if the distractor is selected by enough candidates for it to be an attractive alternative. Each distractor has a unique item discrimination adjusted point-biserial value used to analyze functioning and (if needed) to alert test developers that a distractor may need refined to increase effectiveness. However, instead of expecting a

positive, high point-biserial value, a negative correlation is of interest to illustrate students with lower ability select the option instead of the correct answer. Distractors which may be partially correct or appeal to higher ability students can be identified.

The omit rate discusses the number of USHC examinees who skipped an item and were examined to see if there were items which were "skipped" by many examinees. The *2021-22 EOCEP Technical Report* states that an omit rate five percent or lower (omit ≤ 5%) is used to select items for a form. Items which are skipped by many examinees may be problematic or confusing.

Evaluation: Omit Rates and Distractor Analysis. The omission rate for USHC items was not a concern as omitted counts were low across all 55 items included with the Spring 2022 administration. The average omission rate was .002% of USHC test examinees, well below DRC's stated criterion.

A distractor analysis was conducted for the Spring 2022 USHC multiple choice items. Item options were examined to see if the three incorrect options yielded negative discrimination, accompanied by a positive discrimination value for the correct option. Every one of the Spring 2022 USHC multiple choice items reported three negative point-biserial correlations for each of the incorrect options and a positive point-biserial correlation for the correct option. This information illustrates that the incorrect options were generally selected by USHC EOCEP examinees with lower ability levels, and the correct option was selected by generally selected examinees with higher ability levels. The USHC items and distractors are functioning according to recommendations from best practices of test construction.

3.3c. Rasch item fit statistics: Difficulty Values and Item Fit. DRC uses the Rasch measurement model to provide the US History and Constitution item parameter estimates. The Rasch model is a general name for a family of modern test theory models which compute the probability that an examinee will respond favorably to an item, given characteristics of the item. Characteristics are defined as the amount of the latent construct an individual possesses (i.e., ability in Rasch terminology) and the hardness of the item (i.e., item difficulty). The Rasch model produces scores for each person and each item on a common, interval-level scale (i.e., logit) scale. These common scores are called measures, and the process of putting both ability and item difficulty parameters on the same scale is termed calibration.

The Rasch measurement model relates person and item characteristics to the probability of choosing a correct response. This model-based approach is popular in the psychometrics field when dealing with standardized tests and is used to estimate item parameters, provide an estimate of the examinee's ability (which is then transformed from the raw scale to a scaled test score) and to investigate the psychometric properties of items and the test (Baker, 2001).

Rasch item parameters provide a model-based estimate of item difficulty. For dichotomously scored (i.e., objective response) items, difficulty is the location on the latent ability (termed Theta) variable where an examinee has a 50% chance of answering the item correctly. A characteristic of the Rasch model is that all items are thought to have the same item discrimination, but varying levels of item difficulty. The difficulty parameter is defined as the point on the ability scale (i.e., location on the latent scale, Theta) at which the probability of providing a correct response to an item is .5 (or 50%). Difficulty values are typically within the range –3 < = *difficulty* < = +3. (Baker, 2001; Smith & Smith, 2004). Item difficulty parameters can be interpreted relative to ability level. As stated in Baker (2001, p. 34-35) "an item whose difficulty is –1 function

better among lower ability examinees while an item with a difficulty value of +1 does best to distinguish between examinees functioning at higher ability levels."

Infit and Outfit are Rasch-based fit statistics which may be used to assess USHC items fit to the Rasch model. The fit measures are obtained through the calibration process. These values are useful to examine for USHC items, as items which do not fit the Rasch model (i.e., misfitting items) do not produce trustworthy parameter estimates. As stated in the Winsteps user's manual (Linacre, 2006, http://www.winsteps.com/winman/diagnosingmisfit.htm**):**

> Outfit measures are more sensitive to unexpected observations by persons on items that are relatively very easy or very hard for them (and vice-versa). Infit measures are more sensitive to unexpected patterns of observations by persons on items that are roughly targeted on them (and vice-versa).

Infit and outfit values can be reported as unstandardized values, standardized values, or mean square values; however, generally mean square values are recommended for interpretation (Linacre, 2006). Expected values for the mean squares should approximate 1.0. Values greater than 1.0 (underfit) indicate unmodeled noise or other sources of variance in the data and may degrade measurement. Mean square values less than 1.0 (overfit) indicate that the model predicts the data too well and may cause summary statistics to report inflated values. The *2021-22 EOCEP Technical Report* (DRC, 2022a) notes that the Infit and Outfit mean square values for all items on the USHC should be within the acceptable range of 0.7 to 1.3. Items which fall outside of targeted bounds are flagged for review by DRC psychometric staff.

Evaluation: Rasch Based Difficulty Indices. Difficulty (i.e., location) values for the Spring 2022 USHC EOCEP objective response items (multiple choice and technology enhanced) were evaluated. For the set of USHC items, the mean Rasch difficulty value was .05, meaning the set of items was targeted at the average on the latent ability distribution. As shown in Figure 3, the difficulty values cover a smaller range of ability levels within ± 1.5, ranging from a minimum ability value of -1.37 to a maximum value of 1.45. The majority of items on the EOCEP USHC are within a range of ± 0.5, showing that the test items are largely targeted to an average level of ability and are not overly difficult for the set of examinees.

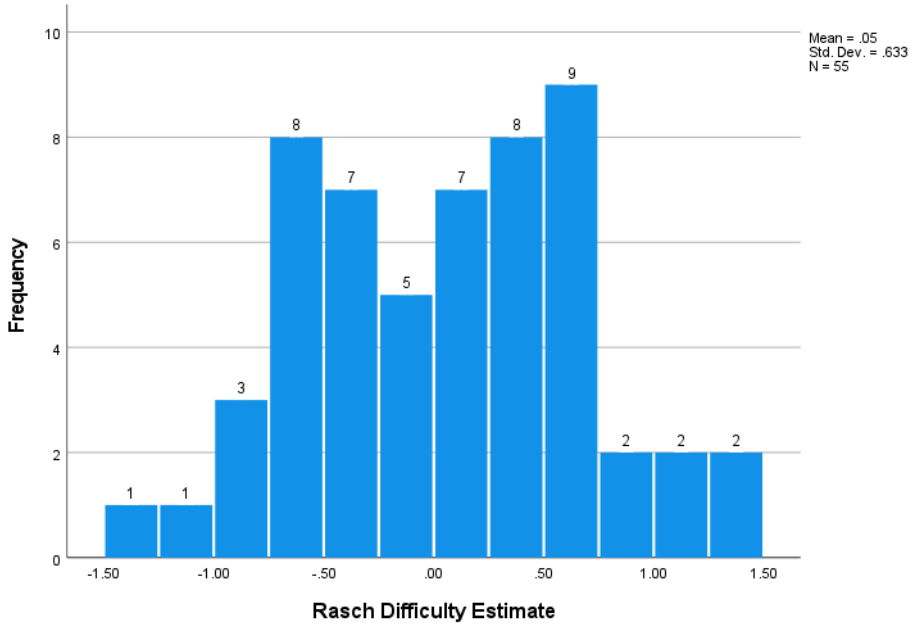

Figure 3. Rasch Difficulty Measures Estimates, USHC EOCEP Spring 2022

Examining the distribution of Rasch-based difficulty estimates by half standard deviations, roughly 50% of the USHC Spring 2022 EOCEP assessment is targeted to a difficulty level between -0.5 to +0.5. There are four items on the test (roughly 7%) targeted to examinees above an ability estimate of 1.0. This means that the majority of test items are appropriate for students with lower to slightly above average ability in USHC. Table 10 provides a frequency chart, in categories of width 0.5, of item location (difficulty) values for the set of 55 objective response items on the Spring 2022 USHC assessment.

Table 10. Frequency Table of Rasch-Based Difficulty Estimates, Spring 2022 USHC EOCEP Items

Item Location	Frequency	Percent	Cumulative Percent
-1.5 to -1.0	2	3.6	3.6
-1.0 to -0.5	11	20.0	23.6
-0.5 to 0.0	12	21.8	45.5
0.0 to 0.5	15	27.3	72.7
0.5 to 1.0	11	20.0	92.7
1.0 to 1.5	4	7.3	100.0

Note: Categories for the frequency table are inclusive of the lower bound.

Evaluation: Rasch Based Fit Indices. Table 11 provides the descriptive summary of the Rasch Infit and Outfit measures for items on the Spring 2022 data. All items included on the USHC EOCEP Spring 22 administration yielded average Infit and Outfit values close to the expected value of 1. All item values were within recommended bounds for the Infit statistic, even for the minimum and maximum values. One item demonstrated an Outfit value of 1.60, just slightly outside of the recommended cutoff. Examination of this item shows that it is one of the hardest on the USHC Spring 2022 assessment (p = .325, Rasch ability measure = 1.45). All other Outfit values were within the recommended bounds set by DRC. The information indicates that the Rasch model provides an acceptable fit to items included on the EOCEP USHC Spring 2022 assessment.

Table 11. Average Standardized Rasch Fit Indices, USHC EOCEP Assessment Spring 2022

Rasch Fit Index	N	Mean	Standard Deviation	Minimum	Maximum
Infit	55	1.01	.10	.84	1.33
Outfit	55	1.02	.16	.76	1.60

3.4 Validity based on relations to other variables.

To support validity associated with the USHC EOCEP test score inferences, the test scores should associate in meaningful and expected ways to other variables (Bandalos, 2018). Different constructs measuring different areas may be related, but should not be too highly related to each other, to show that the constructs are distinct (i.e., divergent validity). The relationship between the scores from tests measuring different constructs can be assessed through evaluation

of correlations among observed scores. Evaluation of correlation values among measures of different constructs (i.e., other EOCEP scores) provides divergent validity evidence.

3.4a Intercorrelations with EOCEP scales. Correlations between USHC EOCEP test scores and other EOCEP content area scores were examined to provide evidence of external validity. Intercorrelations between the USHC EOCEP scores and other content areas were obtained from the *2021-22 EOCEP Technical Report* (DRC, 2022a). Correlation values were computed using the combined population of EOCEP examinees, across Fall/Winter, Spring, and Summer administrations. Scores needed to be present on both examinations; therefore, the number of examinees included in the calculation varies from roughly 1,300 to just over 4,073 pairs of end-of-course examination scores.

Table 12 reports inter-correlations among content areas for examinees taking the USHC along with another EOCEP assessment during the 2021-22 academic year. For the available set of examinees, the correlation coefficients were in a moderate to high range, showing a relationship among scores for a given examinee across content areas. However, correlation values are not excessively high (e.g., .90 or greater), suggesting that the EOCEP assessments are measuring different content areas. As noted in the *2021-22 EOCEP Technical Report*, EOCEP test scores do share a substantial amount of variance, suggesting that a similar trait may be measured for examinees, such as general ability (e.g., students scoring high on the USHC tended to score similarly in other EOCEP areas) (DRC, 2022a). Overall, the values are sufficient to suggest divergent validity of USHC content with other tested areas.

Table 12. Correlations Between USHC and EOCEP Scores, 2021-22 Academic Year

Algebra 1 (N=1,326)	**Biology 1** (N = 4,073)	**English 2** (N = 3,729)
0.54	0.74	0.74

3.4b. Consequential Validity. As test scores are used to make judgements about students' level of content knowledge, a comprehensive view of validity includes an assessment about the consequences (intended and unintended) related to the uses of the test scores. When evaluating validity evidence, the current viewpoint suggests that test users and test developers consider consequential validity (AERA, APA, & NCME, 2014; Messick, 1989). However, unlike other indicators of validity, consequential validity has less to do with data analysis and more to do with making inferences that scores are appropriately interpreted and used.

To address the intended consequences of the USHC assessment, the purposes of the assessments must be clearly specified, helping to ensure that the uses of the USHC scores are aligned with the intent of the end of course testing program. From the SCDE website, the intended purposes for USHC scores are clearly stated, showing how the scores should be used and the potential impact on various groups of stakeholders.

Table 13. Uses of USHC EOCEP Results, by Users

User	Uses of USHC EOCEP Results
Students	USHC scores contribute to the requirement of passing a high school credit course in United States history (20% of course grade). Passing this gateway course is required to receive a South Carolina high school diploma.
Schools and Districts	USHC results are used in the calculation of middle school and high school Absolute Ratings and Growth Ratings.

In addition, information regarding how to interpret 2021-22 USHC EOCEP scores are provided on the SCDE website. The SCDE provides a very detailed EOCEP User's Guide (https://ed.sc.gov/tests/tests-files/eocep-files/2021-2022-score-report-users-guide/), which explains all components included on Individual, School, and District reports. Sample Individual Score reports are provided for stakeholders (e.g., parents, students, teachers) to review (https://ed.sc.gov/tests/tests-files/eocep-files/2021-2022-sample-individual-student-report-english/) prior to receiving actual reports. For educators interested in additional information, professional development opportunities are provided for stakeholders to learn specifically about the USHC EOCEP assessment, including how to view, interpret, and use assessment data (https://ed.sc.gov/instruction/standards-learning/social-studies/professional-learning-opportunities/).

To reduce unintended consequences as well as improper use of scores, score reports include a caveat:

- *Consider how conditions for learning, disrupted by the pandemic, may have impacted student performance. As a reminder, a single score does not provide a complete or precise measure of student achievement. When interpreting results, please take into consideration other measures of achievement.*

In summary, the information included on the SCDE website is easily accessible and helps to support that the USHC EOCEP scores are appropriately interpreted.

3.5 Summary: Technical Quality – Validity

Validity is an essential characteristic of a testing program and is necessary to support the meaning and interpretation of scores. The USHC EOCEP provided validity evidence in multiple areas. EOCEP USHC test items, blueprint alignment, and adherence to best practices of item construction appear sound. The Test Blueprint accurately represented the percentage of items to be expected by content domain, historical thinking skills, and DOK levels. The USHC assessment includes 100% domain coverage of each of the five standards, with an equal number of items (i.e., percentage of the assessment associated with each standard).

Investigation of psychometric descriptive information showed that the USHC EOCEP was moderately difficult (average p = .53), targeted at an average examinee ability level (average ability (theta) measure = .05), and able to discriminate between higher and lower ability examinees (average adjusted point-biserial correlation = .38). Distractors for the multiple-choice items are functioning as intended to discriminate among students with different levels of USHC content knowledge. Practically every USHC item on the Spring 2022 administration met psychometric criteria to demonstrate good fit using both classical and modern test theory methodology. One USHC item yielded a Rasch-based Outfit estimate of 1.60, and another item with a lower discrimination (.17) were the only two items out of the 55 USHC items with values slightly outside of recommended bounds. The USHC assessment illustrates acceptable divergent validity with other EOCEP forms. The SCDE website provides a wealth of information and materials to help stakeholders understand how to use and interpret USHC EOCEP scores, thus promoting consequential validity.

Finally, information regarding testing procedures in the TAM is clear, illustrating detailed instructions for conducting the USHC EOCEP assessment from start to finish. In summary, evidence of validity supported by many activities USHC EOCEP Spring 2022 assessment is thoughtfully constructed and psychometrically sound.

Section 4

Technical Quality – Other

The U.S. Peer Review Critical Elements require review of additional technical aspects which support the use of test scores, including examination of reliability evidence, fairness and accessibility evidence, and investigation of the full performance continuum, scoring, and use of multiple forms. This section provides a review of the critical element areas in the Technical Quality – Other category as related to the USHC EOCEP assessment. Information for this section comes from evaluation the of Spring 2022 USHC test database, the *2021-22 EOCEP Technical Report* (DRC, 2022a) and the *USHC EOCEP Standard Setting Report* (DRC, 2022b).

4.1. Reliability

Reliability is defined as the degree to which similar results would be obtained if the testing was repeated (Bandalos, 2018). In other words, reliability provides a measure of the consistency of test scores if the test were re-administered under similar conditions. A reliability estimate provides a measure of the amount of test variance that can be attributed to true score differences; the remaining variance is considered to be due to measurement error. Therefore, lowering measurement error can contribute to greater accuracy, or higher reliability.

There are different forms of reliability and estimates may be calculated with modern or classical test theory methods. The *EOCEP Technical Report* (DRC, 2022a) describes a variety of methods that were used to estimate reliability and error of the Spring 2022 USHC assessment, using both classical and modern techniques. In this section, the classical indices of the reliability of raw scores and standard error of measurement (SEM) classical indices are discussed. Values discussed here were taken from the *EOCEP Technical Report (*DRC, 2022a); however, the technical manual describes additional procedures and supports.

4.1a. Classical Test Theory Reliability Estimates. As the USHC EOCEP is given to one group of examinees at one point in time on one occasion, an internal consistency estimate is appropriate. Under the CTT framework, internal consistency provides an estimate of how consistently examinees perform across the set of test items under a single test administration. DRC estimated internal consistency reliability using Cronbach's alpha. Estimates can range from 0 to 1, with higher levels representing greater levels of consistency. For higher stakes assessments, Cronbach's alpha estimates at or above 0.85 are desirable (Bandalos, 2018). The standard error of measurement (SEM) is often estimated along with reliability to provide a measure of precision. The values of SEM are a function of the amount of measurement error in the testing situation (i.e., 1 – reliability), as well as the amount of variability in the (observed) test scores. Smaller SEM values indicate more accurate estimation of an examinee's ability in the raw score metric of the test.

Evaluation: Classical Test Theory Reliability Estimates. USHC EOCEP estimates of Cronbach's alpha and SEM were computed for the Fall/Winter 2021 and Spring 2022 test online administrations. Values for both alpha and SEM were estimated for the entire group of test takers and subgroups of students by gender, race, students with disabilities, and English Learners. Using a .85 alpha value as a cutoff, all but one was above .85. The majority of reliability estimates were between .90 - .91 for the USHC and by groups. In the Spring 2022 administration, the alpha value

estimated for the English Language Learners subgroup was under the stated cutoff, but at a value of .83, was very close to the boundary.

Standard error of measurement values were at roughly the same level for both USHC administrations (Fall/Winter and Spring) and across the subgroup reported. SEM values were within a narrow range, from 3.15 to 3.38. The close proximity of SEM values shows that the USHC scores have roughly the same level of precision across all groupings. The values are low (roughly 3 points out of 55), demonstrating accuracy (i.e., low levels of little measurement error) associated with USHC scores.

4.2. Fairness and Accessibility

Fairness and accessibility in testing imply that all eligible students are provided with a fair test and provided an equal opportunity to participate in assessment. Typically, tests are investigated to ensure that the measure performs similarly for different groups of test takers, despite differences in personal characteristics. Examinees may be grouped according to personal characteristics (e.g. gender) to ensure that the USHC does not give any one group an unfair advantage. Here, fairness is examined using differential item functioning (DIF), which is discussed in general terms; interested readers can refer to item response theory textbooks for more technical information about calculating DIF indices (e.g., Baker, 2001). Accessibility refers to providing an equitable opportunity to participate in the assessment process. This may refer to areas such as accommodations, number of assessment periods, and standardized procedures. Both areas are discussed in this section.

4.2a. Accessibility. Many actions related to accessibility have been discussed as part of the test development and design process. The design of the test using accessibility supports (i.e., universal design principals, standardization and testing administration processes, custom forms, allowing accommodations, etc.) have been discussed earlier as part of Section 1 (Statewide System of Standards & Assessment) and Section 2 (Assessment Systems Operations). For example, reviews of item content for bias and alignment, use of a test blueprint, clear definitions of standards, and uniform procedures help to ensure that all USHC examinees have the same access to and experience with the test as part of the EOCEP assessment. All students with documented disabilities with IEPs or 504 Plans must have necessary accommodations documented (e.g., extra time) prior to the time of testing. As noted in Section 1, information regarding test accommodations is clearly defined as part of South Carolina's statewide system of assessment.

Evaluation: Accessibility The testing windows are clearly stated on the SCDE website (https://ed.sc.gov/tests/tests-files/assessment-information/2021-22-assessment-schedule/), allowing students multiple time points to take the assessment. Students have access to the assessment during any semester of the academic year when the US History and Constitution course is taken. Teachers and school test coordinators have access to the Testing Windows and test schedule, as these are posted on the SCDE website prior to the start of an academic year to assist with planning and preparation activities. Multiple time points for taking the assessment and testing windows posted well in advance may be considered accessibility measures, as these allow access for students to progress to their degree objectives and take the USHC EOCEP assessment in a timely manner.

4.2b. Differential Item Functioning. A DIF study examines the actual test performance of examinees in different demographic groups for examinees at the same ability level. If

examinees with the same ability, but from different groups perform differently, a characteristic about the question could be unfairly disadvantaging one group, causing a difference to appear. With a DIF analysis, focal and reference subgroups within a category are compared, where examinees typically considered as disadvantaged are categorized as the focal group (e.g., female, African Americans), and the advantaged examinees are categorized as the reference group (males, Caucasian students).

There are multiple tests and indices for DIF reported in the *EOCEP Technical Report* (DRC, 2022a); however the Mantel-Hanzel test is reported here as it was the index included for review in the Spring 2022 USHC database. This index is a standard in the psychometric industry for examining DIF (see https://www.winsteps.com/winman/mantel_and_mantel-haenszel_dif.htm for more information about how the statistic is calculated in WINSTEPS). As is typical in test construction, questions are classified into three categories: A, B, or C, which are termed the Educational Testing Service standards. These categories are defined as:

- Category A contains the questions with little or no difference between the two matched groups. DIF is negligible.
- Category B contains questions with small to moderate differences, and
- Category C contains the questions with the greatest differences (i.e., moderate to large DIF).

DIF analyses typically include a + or – sign to denote how DIF is exhibited. A positive sign (e.g., C+, B+) illustrates the presence of DIF favors the focal group (disadvantages reference group) where a negative sign (e.g., C-, B-) gives advantage to the reference group (disadvantages the focal group). In other words, positive DIF values mean that the item is more difficult for members of the reference group than for those examinees in the focal group, for examinees with the same level of ability. An assessment will ideally be comprised of category A items if the test pool is sufficient. Category B questions may be used, with preference for items with smaller DIF values (all other aspects, including content coverage, etc. equal). Items exhibiting category C level DIF should not be used, if possible.

Evaluation: Differential Item Functioning. For the USHC, DIF was investigated for the 55 items included in the Spring 2022 administration using following demographic groups. The reference group and focal groups are taken from the *EOCEP Technical Report* (DRC, 2022a).

- ***Gender:*** Two groups are included. The focal group is females; the reference group is males.
- ***Race/Ethnicity:*** Six groups are included. The focal groups are students whose race/ethnicity is reported as Black, Hispanic, Two or More races, Asian/Pacific Islander, American Indian, or Other; the reference group is students whose race/ethnicity is reported as White.
- ***Disability Status:*** Two groups are included. The focal group is students identified with a disability; the reference group is all others.
- ***English Proficiency Status:*** Two groups are included. The focal group is students identified as multilingual or LEP learners; the reference group is all others.

The *Technical Report* noted that a minimum number of cases was set for both the focal group (n = 200) and the reference group (400) to ensure sufficient power to detect differences among groups. A total of seven focal-reference group tests were conducted by: a) Race/ethnic (groups of Black, Hispanic, Two or More races, Asian/Pacific Islander vs. White), b) Gender, c) English Language Learners, and d) Disability status. DIF tests among the remaining groups were not computed due to low sample sizes.

The 55 items from the Spring 2022 testing were investigated for DIF across groups. For the set of 385 DIF tests (55 USHC items x 7 DIF pairs) no items demonstrated C level DIF and only two items demonstrated B level (moderate DIF). Both DIF tests showing moderate DIF on the USHC were found in comparison of students with limited English proficiency and English proficient examinees. Table 15 provides a summary of the DIF tests. Roughly 99.5% of the DIF tests conducted demonstrated negligible DIF. Considering item format type, no DIF was observed for the three constructed response items; any DIF identified was constrained to multiple choice items. In summary, USHC EOCEP items did not demonstrate excessive levels of DIF for the Spring 2022 items reviewed, with most items showing little DIF.

Table 15. DIF Investigations USHC Test Items, Spring 2022

DIF Classification	Frequency	Percent	Cumulative Percent
A-	85	22.1	22.1
A+	298	77.4	99.5
B-	1	0.3	0.3
B+	1	0.3	0.3
	385	**100.0**	100.0

Note: Percentages may not total 100% due to rounding

4.3 Full Performance Continuum

To judge impact, the assessment should be able to categorize students into different ability levels along the performance continuum, where scores report amount of USHC content knowledge examinees possess. These performance levels can be used for a variety of purposes, including accountability reporting. DRC and SCDE personnel held a workshop in June 2022 to recommend performance standards for the revised USHC EOCEP assessment (DRC, 2022). The summer workshop involved 14 educators and stakeholders from across South Carolina. The purpose of the meeting was to develop cut scores to divide students into four achievement levels: Does Not Meet Expectations, Minimally Meets Expectations, Meets Expectations, and Exceeds Expectations. Data evaluated in this section was taken directly from the *SC EOCEP USHC 2022 Standard Setting Technical Report* (DRC, 2022b) provided by DRC. The performance level descriptors and grade associated grade level(s) are reported in Table 16.

The performance levels are related to a student's ability, which is estimated by the Rasch person measure. Considering that the population of examinee's ability scores represent a normal distribution, this distribution is centered at 0, with lower (negative numbers) representing lower than average ability, positive numbers representing higher ability. The larger the number, the higher (or lower) the ability estimate. Using the Rasch-calibrated estimates, these raw scores (on the Theta metric) may be transformed and categorized for accountability reporting. As the ability score is used to create a student's EOCEP USHC different cut scores produce different letter grades. Impact data illustrates the effect of using the "cuts" to determine the percentage of EOCEP examinees that would receive a given letter grade. The discussions outlined in the *Standard Setting Technical Report* (DRC, 2022b) detail the procedures used to arrive at the final cut scores.

Table 16. Description of USHC EOCEP Performance Level Descriptors, Summer 2022

PLD	Description of USHC EOCEP Performance Level Descriptor (PLD)	Grade Level(s)
Does Not Meet Expectations	The student Does Not Meet Expectations as defined by the course content standards. The student needs substantial academic support to be prepared for and to be on track for college and career readiness.	F
Minimally Meets Expectations	The student Minimally Meets Expectations as defined by the course content standards. The student needs additional academic support to be on track for college and career readiness	D
Meets Expectations	The student Meets Expectations as defined by the course content standards. The student is on track for college and career readiness.	C & B
Exceeds Expectations	The student Exceeds Expectations as defined by the course content standards. The student is well prepared for college and career readiness.	A

Evaluation: Full Performance Continuum. Detailed information about the cut-score process used (i.e., Bookmark Procedure), materials evaluated (e.g., Ordered Item booklets), and other information (e.g., discussion rounds, workshop evaluations, etc.) are provided in the *Standard Setting Technical Report* (DRC, 2022b). The process resulted in four cut scores needed divide the latent (USHC) ability distribution into letter grades. Using information from discussions over three rounds of the Bookmarking procedure, educators constructed cut-scores for the ability distribution of USHC EOCEP examinees. As five "grades" are needed (A, B, C, D, F), four cut-points (i.e., cut-scores) in the ability distribution were required.

Table 17 provides the cut-score estimates. Ability estimates range from negative infinity to positive infinity, thus no minimum for a grade of "F" is needed. As expected, the higher the performance level, the higher the students' estimated ability. Ability estimates were lower than average (i.e., ability = 0) only for the lowest performance levels (F and D). Ability estimates higher than average are needed for B and A "grades", with a grade of C close to the average level ability of 0. Overall, the USHC EOCEP ability estimates appear to be acceptable to distinguish between USHC examinees at different ability levels.

Table 17. Cut Scores on the Ability Scale Associated Grade, USHC Spring 2022

USHC EOCEP Ability Distribution Cut-Scores			
F/D	**D/C**	**C/B**	**B/A**
-0.1584	0.2286	0.8355	1.3325

Note: cut-scores based on the (unstandardized) Rasch Person-measure metric

To examine impact, the percentage of USHC examinees falling into the Meets + Exceeds level (i.e., grade of C or higher) was examined. Table 18 provides the percentages of USHC EOCEP examinees in each category. While test scores fall along the performance continuum, only 40% of USHC examinees reached the Meet + Exceed level; roughly 60% of examinees at a D or an F level. As the USHC EOCEP Spring 2022 test scores were not included with a student's end-of-course grade, some students may not have expended as much effort with the testing situation as they would have done if the test score contributed 20% to the final course grade.

Table 18. Impact Data for South Carolina USHC EOCEP, Spring 2022

PLD	Does not Meet	Minimally	Meets		Exceeds	Meets + Exceeds	
Letter Grade	F	D	C	B	A	Percentage C or Higher	
Percentage	40.6%	18.8%	18.4%	19.0%	15.0%	40.7%	

After review by the SCDE and approval by the Superintendent of Education, the final cut scores providing the percentage of students per category were recommended for use by the SCDE starting with the 2022-23 administration of the USHC examination. These cut scores appear to be appropriate for distinguishing among USHC examinees. However, re-examination may be useful once the USHC EOCEP assessment is included as part of the overall course grade.

4.4 Scoring

The Rasch measurement model is used to estimate an examinee's placement on the ability continuum; however, these values may not be easily interpreted by stakeholders. For example, negative ability values and/or values that appear small may be misinterpreted. To produce EOCEP scores which were more meaningful to stakeholders, the ability estimates are transformed to scale scores. The scale was chosen so that it was not tied to a particular assessment and allowed comparison across tests within the state's EOCEP.

The score metric used in the EOCEP was determined by the SCDE. To facilitate interpretation, the range of scale scores was set to have a minimum score of 0 and maximum score of 100. Additionally, the scale is constructed so that each standard letter grade of A, B, C, D, and F corresponds to the South Carolina grading scale with scale score values of 90, 80, 70, 60, and 50 for letter grades of A through F, respectively.

In addition to the total test score, students receive information on their performance in each EOCEP Reporting Category. For the USHC, these are the five Social Studies standards identified on the USHC test blueprint as Key Reporting Areas. An examinee's performance level is reported for each area in terms of *Low*, *Middle*, or *High* performance; these levels are based on an examinee's performance on the subset of items that assess the standard.

4.4 Evaluation: Scoring. The scoring metric used for EOCEP scale scores ranging from 0 to 100 aligns well with "traditional" expectations of testing. Documentation regarding the USHC assessment clearly states how the scale scores should be interpreted using the performance level descriptors, letter scores, and numerical scores (https://ed.sc.gov/tests/tests-files/eocep-files/pld-user-guide-ushc/). The information is presented in a clear and easy to understand format:

- Does Not Meet corresponds to a scale score in the range of 0-59 (F).
- Minimally Meets corresponds to a scale score in the range of 60-69 (D).
- Meets correspond to a scale score in the range of 70-89 (B/C).
- Exceeds corresponds to a scale score in the range of 90-100 (A)

An example of how to interpret student performance in the Key Reporting areas is provided. In sum, the scoring information presented in the *Standard Setting Technical Report* (DRC, 2022b) is clear for stakeholders to understand the relationship between the Rasch scores, how these are transformed to scale scores, and the meaning of the scores in multiple forms.

The *EOCEP Technical Report* (DRC, 2022a) provides a summary of the total test scale scores across the three administrations of the USHC examination held during the 2021-22 academic year. The distribution of USHC scores in scaled format is shown in Table 19. As shown, these align with other information presented earlier, showing an average score of 65 (Minimally Meets/D range) and a distribution of USHC scale scores skewed toward the lower end of the score distribution.

Table 19. Distribution of Scale Scores, USHC EOCEP 2021-22 academic year

				Percentile				
	Examinees	**Mean Scale Score**	**Std. Deviation Scale Score**	**10th**	**25th**	**50th**	**75th**	**99th**
USHC	53,055	65.08	19.67	40	50	63	81	100

Overall, the information regarding the scoring was acceptable for the USHC EOCEP. The ability level raw scores are thoughtfully transformed to align with stakeholders' expectations and information relating scaled scores to performance level descriptors is useful for interpretation of skills. The letter grades and numerical scores are helpful to understand examinee performance with a scale aligned to the South Carolina grading scale. The USHC scoring information is found to be reasonable for the EOCEP assessments and administrations. Scoring information converts students' scores to multiple formats, including scaled scores, letter scores, PLDs, and Reporting Area categorizations. These different formats are useful for a variety of purposes and may be interpreted by many different stakeholder groups.

4.5 Multiple versions of an assessment

To adhere to test security directives, multiple forms of the USHC EOCEP are administered during a testing situation. As noted earlier, DRC uses the Rasch measurement model to calibrate ability and item difficulty parameters on the same scale is termed calibration. Use of the Rasch model for calibration has many advantages, when assumptions behind the method are met. These include aspects such as: mapping persons and items onto the same scale, one-to-one mapping of raw number correct scores to Rasch estimates of ability, the ability to handle missing items, and availability of diagnostic statistics to evaluate the model and data fit (Bond & Fox, 2007; Wright & Stone, 1979). The Rasch model is often used for large scale standardized test programs, such as the EOCEP.

After Rasch calibration, scores on the different USHC forms can be linked and equated. Linking and equating are related, but different, processes. Equating is the process of adjusting scores on forms so forms can be used interchangeably (Kolen & Brennan, 2004). Linking is the mechanism that establishes the comparability between tests. All equated scores can be placed on one scale.

Beyond test security, providing multiple versions of an assessment provides an opportunity for field testing new items. For the 2021-2022 administrations of the USHC examination, field test items were added to the Spring 2022 assessments. Multiple forms of the

USHC were administered in Spring 2022, with additional field test items of many different item formats (e.g., multiple choice, drag-and-drop, etc.) tested.

Evaluation: Multiple Assessment Forms. The EOCEP *Technical Report* (DRC, 2022a) provides a detailed check of assumptions underlying the Rasch model. Examination of item parameters for Infit, Outfit, Dimensionality (to ensure that one primary dimension is assessed), and Local Independence using an analysis of residual correlations (to ensure that no remaining variance is left to explain after extracting the primary dimension) are detailed. Checks on assumptions are necessary to provide assurance that the Rasch model fits the USHC data/persons acceptably and that information generated from the Rasch model is trustworthy for interpretation and use in decision-making.

After providing evidence that the underlying assumptions of the Rasch model were met, the *EOCEP Technical Report* (DRC, 2022a) details the multiple steps used for linking and equating across EOCEP test forms using a smaller set of linking items. The EOCEP equating design used a network of loops (Wright & Stone, 1979) to connect multiple forms through sets of common items. This design allows for verification of link coherence, meaning that the linking parameter used provides stable estimates. Steps to conduct the equating procedures are presented in a series of 12 statements which outline the decision-making process and provide specific guidance if steps in the process are not met (e.g., determine robust *Z* statistics if needed). The steps in the equating process are broken into small pieces, providing clear instructions in the *EOCEP Technical Report* (DRC, 2022a) to show stakeholders how the test equating procedures are conducted by DRC. Concerning the USHC, the *EOCEP Technical Report* (DRC, 2022a) notes that standard setting was to be conducted after the Spring 2022 administration and that later administrations of the USHC will conduct post-equating checks to ensure adequacy of the process.

The USHC EOCEP Spring 2022 assessment included 20 different forms each with 63 items (total of 1,260 items administered). From these forms eight field test items were included along with the 55 operational USHC items. Of the 1,260 USHC items administered across the different forms seen by examinees, the majority of items (used with 1,180 items or roughly 94% of items administered). Technology enhanced formats accounted for a smaller amount of the total at 80 items (roughly 6% of USHC items administered). Technology-enhanced formats included 46 Drag and Drop (DRD) items (3.7% of total items across forms), 29 Multiple Selection (MS) items (2.3% of total items across forms), and five Evidence Based Selected Response (EBSR) items (0.4% of total items across forms). Figure 4 provides a breakdown of all the items administered across the 20 different USHC forms used in Spring 2022. It is noted that 55 items across the forms are duplicated (1,110 items); however, further breakdowns were not conducted to help promote test security.

4.6 Summary: Technical Quality -Other

Other technical aspects provide additional evidence to support the usefulness and meaningfulness of test scores. The information provided in this section showed that the USHC EOCEP Spring 2022 assessment provides consistent scores with an acceptable level of accuracy. Accessibility and fairness are apparent through many different sources of evidence, such as universal design procedures for constructing items, bias and sensitivity reviews of content, availability of custom formats and accommodations. There is minimal differential item

functioning present between subgroups, with only two items exhibiting moderate DIF. These two items relate to English Language learner and native English speaker differences. Item wording

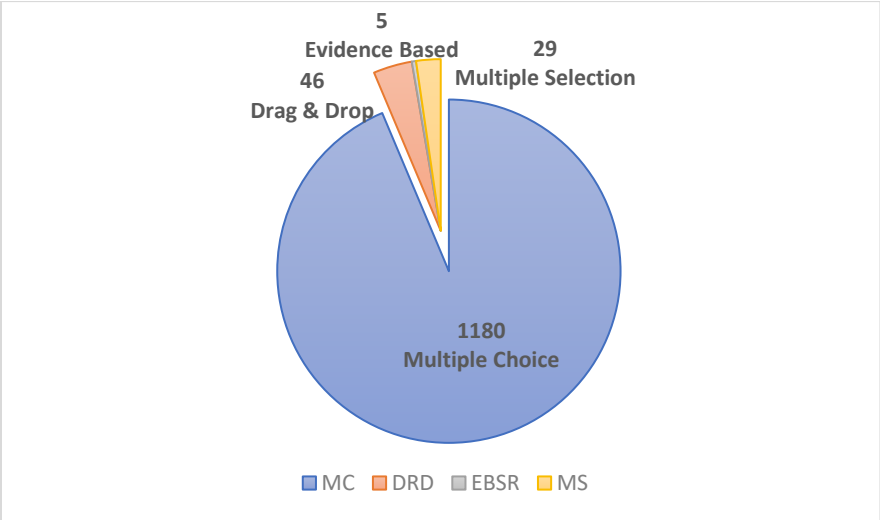

Figure 4. Item Formats included on the Spring 2022 Multiple forms, by Item Type

for all components (e.g., stem and distractors) can be examined to ensure that no bias or confusing wording is present. Multiple USHC EOCEP test forms were provided in Spring 2022 to adhere to test security, where most items utilize a multiple-choice format.

The *Standard Setting Technical Report* (DRC, 2022b) provides detailed information regarding the workshops used to construct cut-points for performance levels. The cut-points relate in a meaningful way to the letter grades. While USHC examinees fall along the performance continuum, a majority of the Spring 2022 USHC EOCEP students did not meet or only minimally met USHC standards. This may be due to the Spring 2022 test not contributing to the overall grade with the first administration of the new USHC EOCEP examination. The *Standard Setting Technical Report* clearly details the decision-making steps and processes conducted during the standard setting process. Overall, the information provides additional technical support to enhance validity associated with the USHC scores.

Section 5

Inclusion of all students

As detailed earlier, all students, including those with a current IEP or 504 Accommodation Plan, enrolled in the U.S. History and the Constitution course must participate in the USHC EOCEP. The *EOCEP Technical Report* (DRC, 2022a) detailed many different groups of students that participated in the 2021-22 academic year's three assessment periods (Fall/Winter, Spring, or Summer administrations). Over 53,000 students took the assessment, encompassing a diverse assortment of students from a wide variety of ethnic/racial groups, IEP status, gifted learner status, 504 plans, and level of English proficiency.

Previous sections of the report have detailed many efforts put forth by the SCDE and DRC to include all students in the EOCEP testing. Careful attention was used when constructing the USCH, where test developers were attentive to aspects of inclusion throughout the processes of test design, test construction and item writing. In addition, after administration of the USHC, data were reviewed by DRC and the SCDE to examine items for differential item functioning and lack of fit to the Rasch model. These activities help to construct a USHC EOCEP assessment that is inclusive of all students. In this section a few additional inclusion activities are detailed. Data for these analyses came from archival sources posted on the SCDE website and the EOCEP *Technical Report* provided by DRC (2022a).

5.1 Including Students with Disabilities

Students who are not able to participate in the same manner as other students or with accommodations, may be eligible for the alternative assessment. IEP teams are provided guidance regarding student eligibility for the EOCEP (https://ed.sc.gov/tests/assessment-information/testing-swd/). The information available on the SCDE website details procedures for testing students with IEPs including allowable accommodations, training information for test administrators, information regarding testing materials, and access to a frequently asked questions page. The *South Carolina Accessibility Support Document* is also provided on the SCDE website (https://ed.sc.gov/tests/tests-files/students-with-disabilities/accessibility-support-document/) to assist educators when testing students with disabilities.

5.2. Procedures for Including English Language Learners

As with other EOCEP assessments, the USHC assessment is not available in languages other than English. While the test must be taken in English, appropriate accommodations for English Language Learners are available, where a student's need and eligibility for testing accommodations is based on multiple sources of evidence (e.g., English fluency level, teacher judgment, other accommodations used in the classroom). The SCDE website provides documentation for stakeholders to examine the means for determining student eligibility for accommodations and guidance on selection of appropriate accommodations for English Language Learners, including guidance on oral administration (https://ed.sc.gov/policy/federal-education-programs/esea-title-iii/).

5.3. Customized Materials and Formats

To be inclusive to all students, the EOCEP assessments are available in a variety of materials formats. These include customized materials, such as Braille and Large Print materials. Accommodations recommended by a student's IEP or 504 plan are also available during testing. As noted in the *EOCEP Technical Report* (DRC, 2022a), a variety of accommodations were used by USHC students over the 2021-22 academic year testing timepoints. Information from the *EOCEP Technical Report* detailing the Customized Materials and Accommodations used with the USHC EOCEP during the 2021-22 testings is detailed in Table 19.

Table 19. Customized Formats and Accommodations Used, 2021-22 USHC EOCEP Administrations (N = 53,055)

Custom Format	N	Percentage	Accommodations	N	Percentage
Braille	1	0.00	Setting	1,411	2.66
Sign Language signed administration	7	0.01		106	0.20
			Timing		
Large print	8	0.02	Scheduling	16	0.03
Oral administration	1,342	2.53	Response Options	5	0.01
			Presentation	32	0.06
			Supplemental Materials	16	0.03

Note: Number of test forms for Accommodations estimated from percentage reported in the 2021-2022 Technical Report.

Evaluation: Inclusion of All Students. Considering the areas described above in 5.1, 5.2, and 5.3, the USHC EOCEP strives to include all eligible students in the assessment process. Information presented on the SCDE website is easy to find and clearly states information needed to assist educators and IEP team members identify which students are eligible for the testing and what accommodations are allowed. Custom formats and accommodations provided by DRC were used during the 2021-22 academic year, showing that these methods are needed by some USHC EOCEP test takers for inclusion in the testing program.

5.4 Summary: Inclusion of Students

The procedures used to create the USHC EOCEP and documentation to assist educators with understanding accommodations and student eligibility for the assessment are thoughtfully constructed. The process was designed to be sensitive to and recognize all students' needs and be inclusive of all students with the USHC EOCEP assessment.

Section 6

Achievement Standards and Reporting

Standard setting is the process used to construct cutoff scores for an assessment (Cizeck & Bunch, 2006). For the USHC EOCEP, standard setting refers to the process to develop the scores aligned with the performance level descriptors (PLDs) which categorize students into ordered groups according to the amount of content knowledge possessed. This section reviews the standard setting procedures used to develop the cut-scores for the USHC EOCEP. Some information regarding standard setting was presented earlier in the discussion of impact (Section 4). Data for this section come from the *SC USHC EOCEP 2022 Standard Setting Technical Report* provided by DRC (2022b).

6.1. Standard Setting for the USHC EOCEP

Given the 2019 adoption of revised Social Studies standards and the subsequent revision of the USHC EOCEP assessment, new cut scores were needed to categorize examinees according to their amount of content knowledge. DRC and the SCDE collaborated on the USHC standard setting process. In June 2022, a two-day Standard Setting workshop was held; attendees included educators from around the state, DRC personnel, and SCDE staff. Fourteen teachers from across South Carolina participated in the workshop.

As stated in the *Standard Setting Technical Report* (DRC, 2022b), the objective for the workshop was to use the revised USHC materials and create cut-scores which would categorize examinees into performance levels aligned with the Profile of the South Carolina graduate. These performance levels descriptors are based on the amount of USHC content knowledge possessed by an examinee; the PLDs are stated below:

- *Does Not Meet Expectations.* The student Does Not Meet Expectations as defined by the course content standards. The student needs substantial academic support to be prepared for and to be on track for college and career readiness.
- *Minimally Meets Expectations.* The student Minimally Meets Expectations as defined by the course content standards. The student needs additional academic support to be on track for college and career readiness.
- *Meets Expectations.* The student Meets Expectations as defined by the course content standards. The student is on track for college and career readiness.
- *Exceeds Expectations.* The student Exceeds Expectations as defined by the course content standards. The student is well prepared for college and career readiness.

The Bookmark Procedure (Lewis, Mitzel, & Green, 1996) was used to conduct the Standard Setting. To create cut-points, workshop participants became familiarized with the USHC standards, the PLDs, and the skills that students with a certain level of competency should demonstrate at each performance level. Using an ordered item booklet (i.e., book of USHC test items ordered by item difficulty), participants placed a “bookmark” at the place that separated

students with different levels of competency according to the PLDs. Cut scores are created on the ability scale provided by the Rasch model; these cut points have an associated level of precision (i.e., standard error or measurement) associated with the ability value. As detailed in Section 3, under the Rasch measurement model, items are targeted to various levels on the ability scale, with some items more (or less) difficult for students at different ability levels. The probability of a correct response on an item can be plotted as a function of the ability of persons (e.g., item characteristic curve) given the item parameters. The first derivative of an item characteristics curve produces an item information curve. Item information curves peak at the item difficulty value, with less information provided by the item for those ability levels farther from the item difficulty value. For example, a very difficult item will provide little information for examinees with low ability because the item is already too hard and most examinees with low ability will get the item incorrect.

Item information functions may be summed across all the test to provide a measure of test information. Like item information, test information function shows which ability levels the test is targeted toward. The inverse of the test information function is known as the conditional standard error of measurement (CSEM). Like SEM, the CSEM value provides an estimate of the amount of measurement error. However, CSEM estimates can vary along the ability continuum, as some ability levels may be estimated with more precision (i.e. more information) than others. CSEM values are lower (i.e., more precision) when more information is present.

After the "bookmark" was placed, participants discussed the procedures and decisions leading to the cut-score placement. Three rounds of bookmarking were conducted; after each round, DRC staff used Spring 2022 USCH data to present impact findings and CSEMs for discussion. The SCDE also considered results of other assessments and policy implications before editing the final cut scores. Information from the discussions and data were used to adjust cut-points, as needed. The final cut scores created at the June 2022 Standard Setting workshop are noted in Table 20.

Table 20. PLD Cut Scores and CSEM Values, USHC EOCEP June 2022

	Minimally Meets	Meets	Exceeds
Ability	0.011	0.451	1.333
CSEM	0.282	0.286	0.328

The standards set by the June 2022 committee were also transformed to the "letter" grades associated with the South Carolina letter grading scale. Table 21 provides the cuts in the theta distribution as noted in the DRC (2022b) *Standard Setting Technical Report*. From the table, the A level remained similar to the Exceeds cut score and the C level was similar to the Meets cut score noted in Table 20; however, additional detail was added to create a cut on the ability curve at the F/D threshold and the C/B threshold.

Table 21. USHC EOCEP Letter Grading Scale Cut cores and Impact Data, June 2022

	Cut Scores				Impact Data				
	F/D	**D/C**	**C/B**	**B/A**	**F**	**D**	**C**	**B**	**A**
Ability	-0.1584	0.3386	0.8355	1.3325	40.56	18.78	14.77	10.97	14.62

Evaluation: Standard Setting. The *USHC EOCEP Standard Setting Technical Report* provided by DRC (2022b) provides a clear description of the standard setting processes, including a description of how to use the Bookmark method, description of CSEMs, and a discussion of the processes used by the participants. The Bookmark procedure was used to create cut-scores; this method is the most common method used and is widely accepted as representing best practice when conducting standard setting procedures. Cut-scores were created carefully, with multiple rounds of discussion and investigation of Spring 2022 USHC assessment data to examine the effect of the cut-point, investigate precision associated, and allow for fine tuning of the cut score placement. The procedures are well-documented and clear to understand the step-by-step procedures used by the SCDE and DRC.

The cut score values created in the standard setting workshop appear to be appropriate given the purpose of the USHC EOCEP. Values of the cut scores are not excessively high (nor low) on the ability distribution, with the Minimally Meets level set around the average of the ability distribution and Meets less than an ability level of 0.5. These levels are acceptable for the purpose of the USHC. The impact data shows the effect of the cut scores with the Spring 2022 USHC EOCEP assessment. As noted previously, the large percentage of "F" scores may be due to other factors (e.g., test not counting, lag from the pandemic) as well some lower scores appearing as the revised standards have also precipitated a change for teachers. In summary, the Standard Setting procedures produced acceptable scores to categorize USHC examinees into performance levels based on the level of content knowledge displayed.

6.2 Reporting

Score reports communicate the meaning of the test scores to various groups of users (e.g., educators, teachers, students and parents). The data from USHC EOCEP is used for a variety of purposes and by a variety of users; each stakeholder group needs to be able to clearly understand and interpret the information provided by assessment. A clear score report is essential to relay this information.

In terms of expecting the score reports, the 2022 EOCEP TAM provided a timeline for receipt of the EOCEP assessment Score Reports. The Assessment Schedule provides the date of delivery of data and paper reports to schools. Both documents are available on the SCDE website. To assist in interpretation of scores, the SCDE (2021) provides the *EOCEP Score Report User's Guide* (https://ed.sc.gov/tests/tests-files/eocep-files/2021-2022-score-report-users-guide/) which details information included in the various types of reports available for Individual Students, School Level, and District Level as well as which are delivered in Paper Reports or Electronic Score format.

The reports presented in the *User's Guide* include clear, detailed explanations (SCDE, 2021), providing information to assist with interpreting components of the report such as: 1) Scale scores (from 0 to 100), 2) Letter grade and the associated student performance level (with both letter and PLDs), and Student Performance on Reporting Areas (categorized as *Low*, *Middle*, or *High,* based on the subset of items that assess the standard). Sample reports are provided for each score report, with statistics and essential report elements numbered and explained. Where appropriate, descriptive statistics (e.g., Standard Deviation, Mean, Median, and Highest/Lowest Scale Score) are defined and an example is included to aid in interpretation.

Evaluation: Achievement Standards and Reporting. The SCDE website includes sufficient information to let USHC test users know when reports will be expected and provides a variety of reports to assist users with understanding and interpreting the information. The *EOCEP*

2022 Test Administrator's Manual (SCDE, 2022) details when preliminary scores can be expected; however, as the USHC test was not used in course grades, this assessment was not included in the 2021-22 Testing Schedule. It is assumed that USHC information will be included in the state testing schedule beginning in 2023-34 when the USHC EOCEP assessment is included as 20% of a student's grade.

The sample score reports included in the *EOCEP Score Report User's Guide* are very detailed and very easy to read and understand. The sample reports show stakeholder groups what to expect, definitions, and where to find the important components that are associated with the different types of reports generated. The *User's* Guide provides clear instructions on how to read the reports and where to find relevant information and are even documented with examples to help with interpretation of the information in context. All EOCEP reports templates are clear to understand, are colorful and engaging to read, with adequate spacing, and explanations in clear language. The *User's Guide* (SCDE, 2022) provides the information necessary for stakeholders to familiarize themselves with the USHC EOCEP and to understand how read the EOCEP reports and to interpret the information. This information is easy to find and to access from the SCDE website.

6.3 Summary: Achievement Standards and Reporting

The overall purpose of reporting test results is to communicate information about student performance to stakeholders. For the USHC EOCEP assessment, the achievement standards were created using a widely used procedure (i.e., Bookmark Procedure) with direction from a set of target stakeholders (i.e., South Carolina educators familiar with USHC standards and population). The *Standard Setting Technical Report* (DRC, 2022b) is very clear and easy to read, expressing discussions and details from the workshop. The achievement standards (PLDs) created make sense given the purpose of the USHC assessment. Final cut-scores from the USHC EOCEP will go into practice for the 2023-24 academic year. It is hoped that a reexamination of data will occur when the USHC EOCEP is included as 20% of a student's classroom grade, that the impact data will have fewer ratings at the low end of the PLD/letter grade scale.

The score reports provided on the SCDE website are useful to aid the user in understanding the meaning of the test scores. The reports and supplementary information developed by DRC are in alignment with best practices of the testing industry. The score reports are detailed, informative, yet also easy to read and comprehend. The information presented supports the use of the achievement standards and the score reports to assist test users and stakeholders.

Summary and Recommendations

This report summarized the results from the Spring 2022 operational testing of the South Carolina End of Course Educational Program, US History and Constitution examination (EOCEP SCDE). The EOCEP US History and Constitution course is a requirement for students seeking a high school diploma from South Carolina. The USHC EOCEP test scores serve multiple uses: contributing a sizable (20%) part of of a student's course grade, is used for school report card presentations, and for local and federal accountability purposes. This evaluation of the USHC EOCEP followed the U.S. Peer Review list of critical elements to review the processes associated with the USHC testing situation, from its start with the policy documentation to the score reports provided to end users. Overall, the USHC EOCEP is well constructed; any suggestions provided below are minor. Based on the evaluation, the following recommendations are provided.

1. Statewide System of Standards and Assessment

The SCDE website provides detailed information about the EOCEP and the USHC as part of this testing program. Information and resources about the purpose and uses of the testing program and the USHC are readily and easily accessible on the SCDE website.

2. Assessment Systems Operations Related to the USHC EOCEP

Information regarding the USHC test specifications is clear, easy to understand, and easy to access assessment prominently. USHC standards to be assessed, test blueprint, domain coverage, and skill levels as well as resources (e.g., sample items, past data reviews, and suggestions for teaching/activities) are readily available to assist stakeholders with test preparation. The Test Administrator's Manual provides detailed instructions to support test security and standardization.

3. Technical Quality – Validity

The USHC includes test items that are constructed through adherence to industry best practices. Items used in the Spring 2022 testing program met psychometric criteria to demonstrate good fit using both classical and modern test theory methodology. Consequential validity is addressed through providing information and materials to help stakeholders understand how to correctly interpret USHC EOCEP scores and how scores may be used.

Recommendation: The two USHC EOCEP items outside of recommended bounds (e.g., one item with an Outfit greater than 1.3 and item with a discrimination value under .20) may be examined in future administrations.

4. Summary: Technical Quality – Other

The USHC EOCEP Spring 2022 assessment provided consistent scores which demonstrated acceptable precision. Attention toward accessibility and fairness are apparent through many different sources of evidence, such as universal design procedures for constructing items, bias and sensitivity reviews of content, availability of custom formats and accommodations, and minimal differential item functioning across examinee subgroups. The standard setting

procedures and cut points relate ability scores in a meaningful way to the letter grades and performance level descriptors.

5. Inclusion of Students

The procedures used to create the USHC EOCEP, customized forms were developed to be sensitive to and recognize all students' needs and be inclusive of all students. Related procedures are clearly documented to assist with questions regarding accommodations and needs of specific student populations.

6. Achievement Standards and Reporting

The achievement standards and related cut-scores created from the standard setting workshop are appropriate given the purpose of the USHC EOCEP assessment. The process used to create cut-scores aligned with best-practices and documentation of the process showed how the standards were set. Score reports and supplementary information is readily available for stakeholders to gain additional information about the different types of score reports and score interpretations with materials that are detailed, informative, yet also easy to read and comprehend. Impact data from the Spring 2022 assessment showed that 60% of USHC EOCEP examinees did not meet or minimally met course standards; however, some of this discrepancy may be due to the uniqueness of the testing situation (i.e., new instrument, change to the revised standards, waiver of requirement that the USHC scores count 20% of the course grade)

Recommendation: Conduct a reexamination of the ability levels associated with cut-scores and impact data when the USHC EOCEP is included as 20% of a student's course grade.

Overall, the EOCEP US History and Constitution Spring 2022 resources evaluated showed the test to be appropriate, demonstrates psychometric soundness, and includes a variety of validity evidence to support for use of scores for decision-making and accountability purposes. Minor recommendations are provided to enhance the performance of the test for use with the South Carolina End of Course Examination Program.

Reference List

Baker, F. (2001). *The Basics of Item Response Theory.* ERIC Clearinghouse on Assessment and Evaluation, University of Maryland, College Park, MD.

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

Bond, T. G., and Fox, C. M. (2001). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Lawrence Erlbaum Associates: Mahwah, NJ.

Cizek, G.J. & Bunch, M. B. (2006). *Standard Setting: A Guide to Establishing and Evaluating Performance on Tests*. Sage Publications: Thousand Oaks, CA.

Data Recognition Corporation (2022a). 2021-22 *South Carolina End-of-Course Examination Program Technical Report.* Retrieved from DRC associates.

Data Recognition Corporation (2022b). *South Carolina United States History and Constitution Standard Setting Technical Report.* Retrieved from DRC associates.

Green, S. K. (2009). *Assessment is essential*. McGraw-Hill Higher Education: New York.

Kolen, M. J., & Brennan, R. L. (2004). Test Equating, Scaling, and Linking: Methods and Practices. Springer: New York.

Linacre, J. M. (2006) WINSTEPS: Rasch measurement computer program. Chicago: Winsteps.com. Information retrieved from: http://www.winsteps.com.

Mertler, C. A. (2016). *Classroom assessment: A practical guide for educators*. Routledge: Newy York.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13–103). Macmillan Publishing Co, Inc; American Council on Education.

Smith, E. V., & Smith, R. M. (2004). *Introduction to Rasch measurement: Theory, models and applications*. JAM press: Maple Grove, MN.

South Carolina Department of Education (2023). *USHC Standards and Assessment Crosswalk.* Information retrieved from SCDE associates.

Wright, B. D., & Stone, M. H. (1979). Best test design. University of Chicago Press: Chicago, IL.