End of Course Examination Program

Technical Evaluation of Fall 2019 Field Test Data: English 2

Report provided to the Education Oversight Committee

Christine DiStefano

June 2020

End of Course Examination Program
Technical Evaluation of Fall 2019
English 2 Field Test Data

Table of Contents

Description of the South Carolina EOCEP and English 2 Assessment

Introduction

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 the State Board of Education in South Carolina and currently include seven named high school courses: Algebra 1, Intermediate Algebra, Biology 1, English 1, English 2, and United States History and the Constitution. Scores from the EOCEP are used in a variety of ways, such as: contributing to students' overall course grade, playing a role on school report cards, and providing accountability evidence to the United States Department of Education.

The English 1 end-of-course test scores have been used to provide accountability evidence; however, the English 1 tests are being phased out of this role and the English 2 end-of-course (EOCEP English 2) assessment will take its place. The EOCEP English 2 test is scheduled to be in operation at the start of the 2020-21 academic year. Per the South Carolina Code of Laws (https://www.scstatehouse.gov/code/title59.php), a technical evaluation of the EOCEP English 2 is required prior to its statewide adoption and administration:

> **SECTION 59-18-320.** Review of field test; general administration of test; accommodations for students with disabilities; adoption of new standards.
>
> (A) After the first statewide field test of the assessment program in each of the four academic areas, and after the field tests of the end of course assessments of high school credit courses, the Education Oversight Committee, established in Section 59-6-10, will review the state assessment program and the course assessments for alignment with the state standards, level of difficulty and validity, and for the ability to differentiate levels of achievement, and will make recommendations for needed changes, if any. The review will be provided to the State Board of Education, the State Department of Education, the Governor, the Senate Education Committee, and the House Education and Public Works Committee as soon as feasible after the field tests. The Department of Education will then report to the Education Oversight Committee no later than one month after receiving the reports on the changes made to the assessments to comply with the recommendations.

The Education Oversight Committee supported the current study as part of responsibilities as listed in the Education Accountability Act. This report evaluates psychometric information necessary to ensure that the EOCEP English 2 produces reliable and valid scores for use regarding student progress, school performance, and federal accountability. Information detailed in this reports on necessary factors such as alignment of the test content to English 2 standards, blueprint review, documenting test/item construction principles, and review of psychometric indices associated with items. Review of EOCEP English 2 materials was conducted according to best practices educational measurement, as detailed by the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 2014).

Data for the evaluation were provided by the South Carolina Department of Education (SCDE), the test contractor (Data Recognition Corporation, DRC), archival documents from the SCDE website (e.g., test blueprints, testing schedules, English 2 Standards, etc.), and meetings/discussions with Education Oversight Committee and SCDE associates. This report

used spring 2019 EOCEP English 2 field test administration data. Values in the dataset provided by DRC includes psychometric indices (e.g., difficulty values) and item information (e.g., information about item distractors) for the pool of items used on all Spring 2019 EOCEP English 2 field tests. Item parameter information was calculated by the test contractor, DRC, and relayed to the Education Oversight Committee through the SCDE.

This report is structured to provide information across multiple areas important for gaining trustworthy scores from the EOCEP English 2 examination. For each area, the report discusses (with a nontechnical focus) what is being measured, what criteria and/or guidelines were used to evaluate the information, and results and any recommendations for change.

EOCEP English 2 Test Population

The EOCEP English 2 assessment is a required element by all public school students who are taking English 2 as part of a credit bearing requirement for high school graduation. This group includes most of the high school students in South Carolina and contains students with an Individual Education Plans (IEP) or 504 plans who are able to take the test with appropriate accommodations and supports. This includes students as required by the federal Individuals with Disabilities Education Improvement Act (IDEA) and by Title 1 as noted by the Elementary and Secondary Education Act (ESSA). As noted by the SCDE memorandum (Jones, 2018):

> *With the exception of students who take alternate assessments, the English 2 field tests must be administered to:*
>
> - *Students who are enrolled in a credit bearing English 2 course (year-round or spring semester).*
> - *Students who are in their second year or above of high school, whose projected high school outcomes are non-diploma, and who are enrolled in an English 2 aligned course.*

The population of EOCEP English 2 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 who 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/).

Field Test Sample: EOCEP English 2 Test Takers, Spring 2019

English 2, and other gateway courses, are typically taken by students in high school; however, the year that the course is taken may vary according to an individual's high school selection of courses. For students following a traditional progression, the majority of students taking the EOCEP English 2 will be in grade 10.

Table 1 provides information for the population of spring 2019 EOCEP English 2 test takers, by grade level. Over 37,000 students participated in the assessment. As expected, the majority of test-takers were 10th grade students; very few 11th or 12th grade students took the EOCEP English 2. The number of students involved with the spring 2019 EOCEP English 2 field

test administration is acceptable to produce stable estimates of psychometric indices for evaluation.

Table 1. Grade Level Distribution of EOCEP English 2 Examinees, Spring 2019 Field Test

Grade Level	Number of Examinees	Percentage of Examinees
9th	11,387	30.6%
10th	25,475	68.4%
11th	326	1.0%
12th	59	<.01
Total	37,247	100.0

Section A

EOCEP English 2: Test Regulations, Construction, and Performance

This section provides a review of the English 2 End of Course (EOCEP English 2) 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, blueprint, test administration, and scoring procedures are examined. Proper test development procedures support use of the EOCEP English 2 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, 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. The evaluation of test blueprint and construction materials largely used archival data from the SCDE website and information from conversations with SCDE personnel.

A.1 Regulations for Testing

The test description is a written document that provides background information about the examination. Elements such as the overall test length, the purpose of the testing, and the item types examinees may expect (e.g., multiple choice, open response) are typically stated. Test administration procedures, test-taking mode (e.g., paper-and-pencil or computer-based) and scoring procedures and scoring rubrics are also presented.

Evaluation: Test Description. On the SCDE website - Tests section, (https://ed.sc.gov/tests/high/eocep/), the Overview link 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: dates for fall/spring testing windows, webinars for assistance, scheduling for delivery of materials to schools, and report delivery schedules are noted for all gateway course testing.

Stakeholders can easily access EOCEP English 2 test description information on the SCDE website as part of the test blueprint (https://ed.sc.gov/tests/tests-files/eocep-files/2019-20-english-2-test-blueprint/). The test description is included as a bulleted list and includes pertinent information of test length, test administration, and scoring information. The bulleted list is simple, easy to read, and focuses the reader's attention on the most important aspects of the English 2 test (e.g., number of items, delivery over two Sessions, inclusion of an essay question, etc.).

A.2 Test Construction: Blueprint, Standards and DOK Levels

The content areas listed in the test blueprint provide information about the knowledge, skills, and abilities on an assessment. In addition to listing content areas, the test blueprint specifies the number (or proportion) of items to be included on each test form, by content area. These numbers/proportions reflect the relative importance of each content area (i.e., more items denote greater importance).

Standards. The EOCEP English 2 assesses four main content areas noted in the English 2 content standards; these areas are tested across two testing sessions: (1) Reading (including the areas of Reading Literary Text and Reading Informational Text) and (2) Writing (including the areas of Writing, Communication, Inquiry and the Text Dependent Analysis (TDA) component). The blueprint notes that Inquiry items from the English 2 standards contribute to the total EOCEP English 2 score, but not to the Reading or the Writing subscale scores.

The blueprint names the broader reporting area and individual indicators (i.e., smaller pieces which operationalize the standard in concrete learning objectives) included on the test. The document includes the larger domain and indicator/specific skills which may be included on the EOCEP English 2 (e.g., Reading Literacy Text, 5.1), along with a possible number of items.

DOK. The EOCEP English 2 uses the Depth of Knowledge (DOK) classification system to categorize items. The DOK categorizes the cognitive complexity of items into one of four categories, where higher numbers indicate higher levels of complexity. The DOK levels are 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 are 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 EOCEP English 2 exam should have a mix of items across Levels 1 through 3. The EOCEP English 2 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. For some students, passing English 2 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 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.

Evaluation of Test Blueprint – Standards Covered and DOK Levels Included. The EOCEP English 2 test blueprint is easy for stakeholders to find on the SCDE website

(https://ed.sc.gov/tests/tests-files/eocep-files/2019-20-english-2-test-blueprint/). This one document includes the test description as well as the standards covered and DOK levels to be expected.

Evaluation: Standards. Table 2 provides a summary of the test blueprint information by test reporting/content area as compared to the English 2 standards. The percent of the domain covered (as stated on the Test Blueprint) was computed by dividing the Number of standards on the test blueprint by the total number of English 2 standards in a given domain. The Text Dependent Analysis (TDA) is noted separately on the Test Blueprint; however, these skills are included as part of the Writing domain and were included in the computation of the domain coverage.

As noted, there are more English standards to be covered during the academic year than are included on the end-of-course assessment. This is understandable as the test provides a snapshot of learning at one time point and the English 2 standards provide the framework of skills to be practiced over the academic year.

The three content areas that comprise most of the test: Reading Literary Text, Reading Informational Text, and Writing include roughly 57% - 67% of the Standards within a given domain, providing acceptable coverage of the intended skills. The two areas that assess a lower percentage of their respective domains, Communication and Inquiry, include indicators that are not easily applicable to a standardized testing situation (e.g., English 2 Communication Standard 3.2- *Create visual and/or multimedia presentations, using a variety of media forms to enhance understanding of findings, reasoning, and evidence for diverse audiences*). To assist teachers and students, the Assessment Boundaries document (https://ed.sc.gov/tests/tests-files/eocep-files/english-2-assessment-boundaries/) provides a focused list of standards and indicators which are included on the EOCEP English 2 assessment.

Table 2. EOCEP English 2 Domains Coverage Noted by Test Blueprint

Domain	Number of English 2 Indicators	Number of Indicators on Blueprint	Percent of Domain Coverage
Reading Literary Text	13	8	61.5%
Reading Informational Text	12	8	66.7%
Writing Communication Text Dependent Analysis	7 15 *	4 2 1	57.1% 13.3%
Inquiry	12	3	25.0%

Note: * = TDA item is reported as a separate area category in the EOCEP English 2 blueprint, but the stated item indicator falls under the Writing domain.

The Test Blueprint provides guidance of the number of possible items included on the assessment and how these relate to the English 2 Standards. Table 3 evaluates the percentage of the test allotted to each English 2 domain and the two testing sessions. The Reading session involves the most standards (16 total) and includes the most items to cover these standards. The Writing session tests four areas (Writing, Communication, Inquiry, and

TDA). It covers fewer standards and a smaller range of items. However, these items are more involved (e.g., essay) and at a higher cognitive level, requiring fewer items.

In summary, the test appears to balance the number of items that are devoted to Reading and Writing, with more of the test content and percentage devoted to reading content. This is partly due to the complexity of tasks required to assess writing as compared to assessment of reading and processing information. The blueprint information is acceptable to inform stakeholders of what is expected on the EOCEP English 2, in terms of domain coverage and possible range of items.

Table 3. Review of Test Blueprint Information, EOCEP English 2 Examination

Area	Number of Standards/Indicators on Test Blueprint	Range of Items to be Included	Percentage of Test
Reading Literary Text	8	16 – 26	29 - 47%
Reading Informational Text	8	18 – 25	33 – 45%
Writing Communication Text Dependent Analysis	4 2 1	6 – 12 2 – 6 1	11 – 22% 4 – 11% 2%
Inquiry	3	4 – 8	7 – 15%
EOCEP English 2 Total		**55 items**	

Evaluation: DOK. The test blueprint also includes a breakdown of the DOK levels included on the EOCEP English 2 test. Three of the four DOK levels (Levels 1-3) are included. As stated on the test blueprint, at DOK Level 1 it is estimated that the percentage of items is between a minimum of 0% of the test to a maximum of 15%, Level 2 between 55% and 75% of the assessment, and between 25% and 45% at Level 3.

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 EOCEP English 2 items at DOK 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 EOCEP English 2 assessment between (roughly) 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 an academic year of participation with English 2 content.

A.3. Test Scoring and Test Performance

Scoring. The EOCEP English 2 score contributes 20 percent in calculation of a students' final course grade. Information from the EOCEP is used statewide as part of federal accountability requirements. At the school and district levels, EOCEP scores from Algebra 1 and English 1 (note: to be replaced by English 2 starting with the academic year 2021-22) are currently used in calculation of school accountability ratings which are reported to the state and

the federal government; grades of C or better on the gateway courses included in EOCEP are reported to stakeholders on school/district report cards.

Evaluation: Scoring. The EOCEP English 2 test score is provided by the responses to close-ended items and the essay (TDA) item. The EOCEP English 2 items are largely closed-response, objective items which can be machine scored. These items are generally worth 1 point for a correct answer.

The open-ended essay question is hand-scored scored by trained raters using a pre-established rubric. A copy of the rubric and definitions at each scoring level is provided on the SCDE website: https://ed.sc.gov/tests/tests-files/eocep-files/eocep-tda-scoring-rubric/. The TDA rubric is a 4-point holistic rubric. The rubric is detailed, providing raters and test stakeholders a description of the skills that should be demonstrated by examinees at a particular level and context of how examinees at one level differ in ability from those at other levels. A strength of using a holistic rubric is that persons reviewing rubric scores can clearly see strengths of students rated at a given level and what skills may be reinforced to advance to the next level (Mertler, 2016). As there is one TDA question, this item is weighted by a factor of 4 to contribute to the overall score.

For this report, the SCDE provided a Standard Setting report written by DRC staff (DRC, 2019) detailing the development of cut scores into four achievement categories. These categories describe the Performance Level Descriptors (PLDs) across the continuum of scores, using categories of: Does not meet, Marginally Meets, Meets, Exceeds. The percentage of Meets and Exceeds is also reported, this aligns with the reporting of EOCEP scores of C or better for Federal Accountability and School Report Card purposes.

Table 4. Percentage of EOCEP English 2 Students Scoring in Each Performance Level Descriptor

	Percent of Students by PLD				
	Does not Meet	Minimally Meets	Meets	Exceeds	Meets + Exceeds
Expected Percent	38.2%	12.5%	26.9%	25.4%	52.3%

There is limited information on the SCDE website to show stakeholders how scores are reported. The Testing Administration Manual does note that scores will be posted for Reading and Writing domains (along with a Total Score); however, this source may not be the first to come to minds of some groups of stakeholders (e.g., parents, students) when looking for scoring information. Relatedly, there is not yet documentation (e.g., technical manual) to report how scores are transformed to a total score. There are technical manuals for SCDE tests; however, the latest technical manual for the EOCEP English 2 is for the 2014-15 English 1 EOC. Once operational, updated technical information for the EOCEP English 2 assessment would be a useful addition to the test documentation on the SCDE website.

Test Performance. All students enrolled in credit-bearing courses are expected to participate in the EOCEP English 2 assessment. The test performance resources are defined as specific test materials (other than the Blueprint and English 2 Standards) which are provided to

teachers and the population of English 2 test takers to prepare for the test. These include information such as the Teacher's Guide (which includes practice objective response format items), sample TDA items, sample responses with scoring protocols, the TDA rubric and Test Review reports. As the EOCEP English 2 test is delivered online, the Online Tools Training site simulates the online testing situation and allow students to practice using the testing interface's online tools.

The SCDE's Office of Assessment conducts annual committee meetings, where district-level curriculum experts review the item results data for state testing programs, including the EOCEP tests; findings from these meetings are detailed on the SCDE website (https://ed.sc.gov/tests/tests-files/eocep-files/english1-test-results-data-review-2019/). The reviews provide teachers guidance regarding specific standards/indicators exhibiting performance deficiencies and suggestions for how these areas may be developed for subsequent test administrations.

Evaluation: Test Performance. To assist teachers and students with test performance, practice information is easily accessible on the EOC website. Twenty sample objective items are provided for practice (https://ed.sc.gov/tests/tests-files/eocep-files/2019-eocep-sample-release-items-for-english-2/). These items include information about the alignment of items to standards, DOK level, and estimated item difficulty. There are also two sample essay questions for response practice to the TDA along with the link to the rubric used to score the responses. South Carolina student responses to the Writing-TDA questions are provided along with annotations, which describe the reasoning behind the rubric scores given to the responses. A TDA Checklist (English 1) is provided to help craft responses to the essay (https://ed.sc.gov/tests/tests-files/eocep-files/eocep-english-1-writer-s-checklist/).

While the information reviewed was for the (currently) operational English 1 end-of-course test, it is assumed that the information is similar for responding to the TDA on the EOCEP English 2. The Online Tools Training gives students an opportunity to become familiar with use of online tools (e.g., drag and drop) which may be required during testing. In summary, there are many materials available for examinees to become more familiar with the test questions and testing format to help test takers understand the types of questions and responses expected.

The SCDE website states that the EOCEP English 2 test is similar in structure and content to the EOCEP English 1 test. Therefore, the data review information may be useful to school personnel until the data are updated to include reviews focused on the English 2 test. The information provides a mechanism for learning from previous results and enhancing test performance. These materials help provide transparency for teachers and students regarding EOCEP English 2 test content and procedures. Detailed information about what information is included on the test, access to practice questions, and use of previous test result data can enhance training and ultimately, student performance.

A.4. Summary: Test Regulations, Construction, and Performance

In summary, materials detailing construction of the EOCEP English 2 are available and easy to access from the SCDE website. The test appears to balance the number of items that are devoted to Reading and Writing, considering complexity of tasks. The blueprint information is acceptable to inform stakeholders of what is expected on the EOCEP English 2, in terms of domain coverage and possible range of items. Information from the DOK levels reported on the blueprint help stakeholders understand the complexity of the test. There are many materials available for examinees to become more familiar with the test questions and testing format to help test takers understand the types of questions and responses expected.

Updated technical information regarding score calculations (session scores and total score) and a test review may be helpful to include on the SCDE website once the EOCEP English 2 becomes operational.

Section B

Evaluation of EOCEP English 2 Test Items

The EOCEP English 2 assessment was structured similarly to the English 1 end of course examination, consisting of 55 total items across the Reading and Writing sections. Items on the EOCEP English 2 include a variety of formats. The Reading section includes items which are largely objective response (i.e., closed response) test questions which require selection of the answer(s) to achieve full credit. This item 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). The majority of the EOCEP English 2 assessment items are Multiple Choice (or selected response) format, where respondents select the correct response from four possible alternatives.

There are a few objective response items per session that are of a different format than multiple choice. These formats include Multiple Selection items, where students are prompted 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 steps in correct sequence, place multiple punctuation marks correctly, 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 paper/pencil tests).

The Writing section includes a Text Dependent Analysis (TDA). This is a constructed response item, where examinees are provided a prompt and then construct their answer. For the EOCEP English 2, students read a piece of text and draw upon the passage to provide an extended written response, supporting the essay with evidence from the text. The response is scored by raters using the TDA rubric.

This section provides a review of test items to ensure that the items are constructed following best practices in the psychometric field. Specifically, this includes reviewing items to ensure match to content standards/indicators, are unbiased, and are error free in terms of grammar, etc. This is a preliminary review of EOCEP English 2 test content. A more intensive item review is planned for fall of 2020 utilizing school personnel familiar with the targeted student population and the English 2 standards.

As there are 16 EOCEP English 2 field test forms available, one test form was created and reviewed; however, it is assumed that the items included on this form are representative of

the content included on other test forms. The test reviewed for this analysis was provided by the SCDE and is hard copy of Form 110, field tested in the Winter of 2019. To adhere to regulations of test security and confidentiality, only item identification number information was reported in the evaluation; if particular item characteristics were discussed (e.g., percentage of 'A' responses), all information was reported in the aggregate. While it is recognized that the suggestions here are relevant to the form reviewed, these may be transferred to other forms as items typically appear on more than one of the 16 forms used in practice

B.1. Content Review: Item Alignment to English 2 Standards

In accordance with test construction principles, the EOCEP English 2 test should be aligned to the applicable content standards for English 2. These content standards are what teachers use to plan instruction and guide student learning in the course. As an initial content review, alignment of the end of course test content was compared with the English 2 Standards to review the accuracy of the test content to the test blueprint materials.

Item alignment to English 2 Standards was conducted for all items by Session administration and across the total assessment. Item alignment was conducted using the Standard and Indicator numbers. However, to adhere to test security and confidentiality practices, responses were aggregated across entire test and only the broader Standards category is reported and compared to the test blueprint information.

Evaluation: Item Alignment to English 2 Standards. English 2 Standards detailing assessment content are easy for stakeholders to find on the SCDE website under the Tests tab of the website, under the High School section of the website (https://ed.sc.gov/tests/tests-files/eocep-files/english-2-assessment-boundaries/). All test items were evaluated and compared to their stated Standard and indicator. On face value, the items appear to be aligned with the respective content. No mis-match between indicators and test content were apparent with the review materials. Items were reviewed to determine that the number of items, percentage of items, and standards tested were in line with the information reported by the test blueprint. Review information is presented in Table 5.

All test content areas were in line with information reported in the blueprint. As expected, the Writing subtest includes the items measuring Writing Standards, Communication, and the TDA. The Reading subtest includes items aligned with the two reading areas, Reading Literary Text and Reading Informational Text. Inquiry items (which are included in the total EOCEP English 2 test score) are split evenly across both tests.

Table 5. Item Alignment to Standards, EOCEP English 2 Items

English 2 Content Domain	Number of Indicators on Test Blueprint	Alignment of Items to Percentage Range	Alignment of items to Stated Indicators
Reading Literary Text	8	Yes	Yes
Reading Informational Text	8	Yes	Yes
Writing Communication Text Dependent Analysis	4 2 *	Yes Yes Yes	No Yes
Inquiry	3	Yes	Yes

Note: * = Test Blueprint includes the TDA item as a separate category; in the item alignment review, this was included with the Writing Domain.

The information on the test was examined to determine if items within a domain were matched to the indicators listed on the Test Blueprint. The majority of domains did test all of the stated indicators; however, one area did not. Writing listed four indicators on the Test Blueprint (1.1., 2.1, 4.1, 5.2). From review of the item characteristics, indicator 1.1, may be also included in the TDA, but only one indicator is associated with the item description. The last Writing indicator, 5.2, did not appear to be included on the version of the field test reviewed; however, this indicator may be included on a different field test version. In sum, the EOCEP English 2 items align with the standards and what is reported in the Test Blueprint (posted on the SCDE website). Percentages of the actual items on the field test was in concordance with the percentage of items to expect by content domain, as stated in the Test Blueprint.

B.2. Complexity Review: Item DOK Levels

Depth of Knowledge (DOK) alignment was reviewed to ensure that the EOCEP English 2 possesses the targeted complexity levels as noted by the purpose of the test. Items were examined and matched to the stated DOK levels reported by the test contractor, DRC. Item DOK evaluation was conducted across items, Session administration, and for the Total assessment. Again, to adhere to test security and confidentiality practices, responses were aggregated across the entire test. It is noted that this evaluation provides an initial review of the DOK level information; a more extensive evaluation of item DOK levels is planned for fall 2020.

Evaluation: Item DOK Levels. Individual items appeared aligned with their stated complexity levels. Of the 54 selected response items, there were no items that seemed to mis-state the complexity level of the item. The test blueprint reports that the examination includes few items at the lowest DOK level (Level 1). As expected, the test was largely comprised of items at DOK Levels 2 and 3. Considering the test overall, the DOK levels reported in the test blueprint were in line with the percentages reported on the test. Table 4 reports the test blueprint information alongside the percentages by complexity level. In sum, the EOCEP English 2 item DOK levels align with the standards and what is reported in the Test Blueprint on the SCDE website.

Table 6. Item Alignment to DOK, EOCEP English 2

DOK	Min/Max	EOCEP English 2 Percent
Level 1	0 - 15%	2%
Level 2	55 - 75%	65%
Level 3	25 - 45%	33%

B.3. Adherence to Item Writing Guidelines

All EOCEP English 2 reading passages and items were reviewed to determine if test content aligned with best practices for test construction (e.g., Greene, 2009; Mertler, 2016). This included review of item stems for clarity, grammatical and spelling errors, providing clues to the correct answer. Item options were reviewed to ensure that the options made sense to examinees with partial knowledge of the content area and were plausible. Correct answers to items were reviewed to ensure that the answer key did not form a pattern or have the correct option (e.g., "D") repeated excessively. In addition, items were examined to ensure that the language was appropriate for student test takers, used standards-based vocabulary, and were written to support research-based instructional.

Evaluation: Adherence to Item Writing Guidelines. Of the total 55 items, the Session 1 assessment included fewer multiple-choice questions and TDA constructed response item. Session 2 included more items, but all were selected response. Test items were primarily multiple choice; however, other item formats included (roughly 7%) evidence based selected response, multiple answer, and technology enhanced. Both sessions include reading passages, where items relate to passage content (i.e., testlet). The number of items per testlet vary between four and nine items. Reading passages were clear and interesting, varying content from fiction and non-fiction. Items were clear and easy to understand. In terms of content, items did not exhibit any problems related to fairness in terms of content presented for examinees, items (e.g., "trick" questions) or response options (e.g., deliberately non-plausible or humorous response alternatives).

All Item stimuli and options were reviewed to determine adherence to item writing guidelines. The EOCEP English 2 test displays best practices of item writing principles including:

- use of spacing, where item stimuli is separated from the item alternatives,
- formatting to focus reader's attention (e.g., bold, underlying),
- complete thoughts or sentences for the item alternatives,
- plausible options for multiple choice item alternatives,
- correct response is not always the longest option.

With selected response items, letters associated with a correct answer should not form a pattern or include one option an excessive number of times over the course of a test (or testlet). The EOCEP English 2 selected response answer key was examined by testlet, test session, and across the entire test for alignment with item writing guidelines (note: to ensure test security item responses are aggregated across the test and only for the four main options, A-D).

No patterns or continued correct option letters were observed. Table 3 provides the distribution of correct responses across the 54 selected response questions included on the EOCEP English 2. As shown in Figure 1, the percentages are roughly balanced by across the four options.

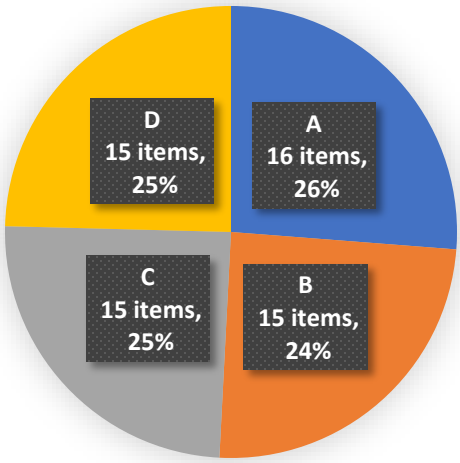

Figure 1. Distribution of Correct Response Options, EOCEP English 2 Field Test

Finally, all reading passage items were reviewed to ensure student-appropriate language. The information was acceptable, with appropriate language and readability level which was acceptable given the purpose of the examination and the target population. Use of standards-based vocabulary was apparent in the items. The content standard vocabulary was largely included in the item stems (e.g., evaluating points of view, use of context clues to decipher meaning, cite textual evidence). The items were written to support research-based instructional technology, as teaching the content skills could be approached from a wide variety of methods based in empirical support.

B.4. Summary of EOCEP English 2 Test Items

The EOCEP English 2 items aligned with the English 2 content standards. Also, the Test Blueprint accurately represented the percentage of items to be expected by content domain and DOK level. Items illustrated best practices of industry standards, were error free and appeared unbiased. Items used appropriate content-based language and written to the target population. No recommendations are needed; the EOCEP English 2 test items, blueprint alignment, and adherence to best practices of item construction appear sound.

Section C

EOCEP English 2 Test Administration Procedures

As a state-wide standardized test, the EOCEP English 2 follows state and district regulations related to test procedures including test security, 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 EOCEP English 2 examination.

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. Information for the analysis was obtained from archival documents on the SCDE website and discussions with SCDE personnel. The test engine delivery was not able to be evaluated due to Spring 2020 postponement of testing from COVID-19. However, the documents provided on the website provide sufficient evidence for review of the test administration protocol.

C.1. EOCEP English 2 Test Administration Procedures

The EOCEP English 2 test 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. 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.

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 (e.g., https://ed.sc.gov/tests/tests-files/eocep-files/eocep-online-test-administration-manual-for-spring-2020/) and all testing personnel at a school also receive a paper copy of the manual. The TAM clearly describes testing instructions, including a listing of steps to be taken before testing, during testing, and after testing.

Evaluation: EOCEP English 2 Test Administration Procedures. 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) .

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.

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.

C.2. Summary of Test Administration Procedures

The test administration procedures provide clear directives to deliver the EOCEP English 2 properly and with fidelity. Clear, objective information that is followed by all district/school testing personnel helps to ensure uniform testing procedures delivered to all English 2 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 EOCEP English 2 scores for use with accountability and decision making.

Section D

Test Calibration, Equating, and CTT Item Analysis

The EOCEP English 2 field test includes 16 forms, each of which is scored to produce item- and person-statistics (i.e., calibrated parameters). The calibration process is performed by DRC using the Rasch measurement model and the item- and person-statistics can be evaluated and compared within and across forms. Scores across the forms are equated by using common items, a small subset of items included on different forms. Common items provide a mechanism to allow scores across tests to be put on a common metric.

This section provides a review of the procedures used to link and equate the EOCEP English 2 examination. The data for the review came from were largely archival documents which were obtained from the SCDE website and DRC. The evaluation provides information about the timeliness of the scoring process for providing test assessment results to teachers and students.

For the item analysis, item statistics were calculated using Classical Test Theory (CTT) techniques and modern test theory techniques. All statistics were calculated by DRC and contained statistical information for the EOCEP English 2 Spring 2019 Field Test Data across the 16 forms. Summaries of item statistics (e.g., item difficulty, average point biserial correlation) were summarized for the set of items across all forms (i.e., all items in the field test pool) and by form. Each test will have 55-items; however, when analyzing item statistics there were fewer than 55-items per form as the information here (in general) represents a pool of available field test items. In the evaluation, these are still referred to as "Forms" with the understanding that they are not an active EOCEP English 2 test. Ten non-performing items were removed by DRC for operational form consideration; these items were also removed from review analyses. Besides calculation of summary statistics (e.g., mean values, standard deviations), no additional estimation procedures were performed. Item analysis information is presented for review of classical test theory (CTT) indices.

D.1. Test Calibration, Equating, and Scoring

DRC uses the Rasch measurement model to provide EOCEP English 2 parameter estimates. The Rasch model is a general name for a family of measurement 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 EOCEP English 2 test is computer scored for all dichotomous items using the Rasch model. The TDA item is on a four-point scale and is scored using Master's Partial Credit model. The Rasch model estimates the probability of a correct response given the examinee's ability level and the difficulty of the item. The partial credit model is similar but estimates the probability that a person will be observed in a specific category of the rubric (1 to 4), given the person and item characteristics.

After Rasch calibration, scores on the different EOCEP English 2 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.

For reporting of scores, the EOCEP English 2 TAM provides a timeline for receipt of Reading or Writing (Preliminary) Score Reports and the Assessment Schedule provides the date of delivery of data and paper reports to schools. Both documents are available on the SCDE website.

Evaluation: Test Calibration, Equating, and Scoring. The Rasch model is a popular measurement model for use with statewide testing programs. 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 the 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 English 2 assessment.

The EOCEP English 2 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. As a check of the stability of the process, the sum of the link constants should be zero. The implementation of test networks leads to banks of commonly calibrated items far larger in number and far more dispersed in difficulty than appropriate for any one test. Sums across link constants were provided by DRC and the SCDE in a summary email. The information reported stability of link constants, with all five link constants approximating 0.0 (within ±2 standard error of estimate).

The EOCEP 2020 TAM details when preliminary scores can be expected. Objective test questions for the EOCEP English 2 are quickly scored by the DRC INSIGHT online system; the TDA is scored by trained raters and 10-days are allocated for scoring the essay. With the English 1 test, preliminary score reports were available for review on the online system within 36 hours after the Reading domain assessment, 10 days after the Writing domain assessment, and the total score can be quickly computed after both components are provided. This information provides timely feedback for teachers and schools to guide instruction.

D.2. CTT Based Item Analysis

Two Classical Test Theory (CTT) indices were included in the dataset: item difficulty and adjusted point-biserial. 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 that 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 between students (Bandalos, 2018).

The adjusted point biserial correlation (r_{pb}) is a measure of association, illustrating how well an item discriminates between high performing and low performing examinees. The value is calculated as the correlation between item scores (correct/incorrect) and the total score, with the

item in question removed from the total 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, or negative, indicating that the item is not discriminating as intended. 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; negative values indicate the tendency for high ability students to answer incorrectly and low ability students to answer correctly. A high point-biserial coefficient means that students selecting the correct response are students with higher total scores; students selecting incorrect responses to an item have lower total scores, meaning the item can discriminate between low-performing examinees and high-performing examinees. In general, 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.

Removed Items and Items to be Re-fielded. The data file included information on 415 items which are used across the 16 EOCEP English 2 test forms. However, not all items from the field-test pool were considered items to advance to the operational forms. Ten items were removed from the item pool due to poor performance, and these 10 were not included in the descriptive summaries. In addition, SCDE documentation stated that items with marginal performance have been designated to be re-fielded. The items to be re-fielded were included in the analysis.

Evaluation: CTT Difficulty. The average CTT-difficulty value across the 405-item pool was p= .55, meaning, on average, students answered 55% of the EOCEP English 2 items correctly. This is a moderately difficult level, and also the value approximates the value to maximize differentiation among individuals. Figure 2 provides a histogram of difficulty values. Across the forms, the set of values had a minimum difficulty of p= .14 (14% of examinees answering the item correctly) to a maximum of p = .89 (89% of examinees answering the item correctly). As shown in Figure 2, the EOCEP English 2 tests include a mixture of items noted as “harder” and “easier”, in terms of CTT-difficulty values.

Item difficulty values were reviewed to determine the number of items per form that were challenging for students, where p < .50. Tests are at a slightly easier level of difficulty, with roughly 61% of the items (246 of 405 items) at or above a difficulty level of p ≥.50 and 39% (159 of 405) seen as more difficult (p-values < .50). The majority of the EOCEP English 2 items were less difficult for the population of test-takers.

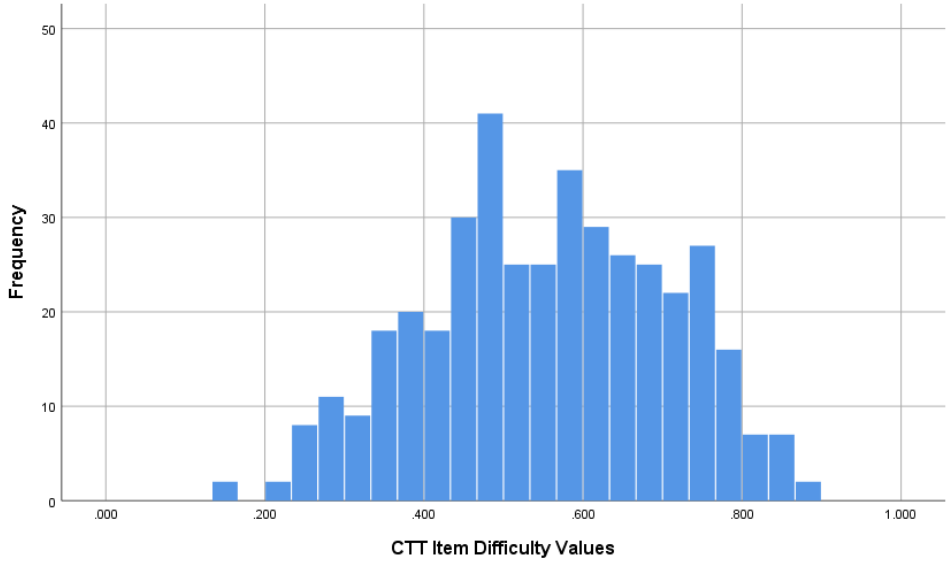

Figure 2. Distribution of EOCEP English 2 CTT-Based Difficulty Values

Table 8 provides summary statistics for the difficulty values by EOCEP English 2 Test forms. Average difficulty values ranged from p= .46 to p = .63. Standard deviation values showed variability in the values. CTT-based difficulty values are generally within a similar range, with the exception of one form, which displayed a greater range of values (SD = .21). Average values for individual forms are close to the overall average CTT difficulty value of .55, (roughly within ± .10); however, there are individual forms with greater difference in average difficulty (e.g., greater than ± .15, the standard deviation for the entire set of items). Inclusion of other items from the pool (i.e., anchor items) may help alleviate average differences among forms.

Table 8. Descriptive Statistics for CTT Difficulty Values, EOCEP English 2 Forms

English 2 Form	Number of Unique Items	Mean	Std. Deviation	Minimum	Maximum
1	47	.60	.10	.44	.83
2	25	.62	.16	.34	.90
3	26	.58	.16	.21	.84
4	24	.54	.14	.31	.79
5	25	.59	.14	.29	.85
6	22	.53	.21	.16	.83
7	27	.63	.12	.33	.80
8	21	.55	.15	.24	.76
9	25	.46	.15	.24	.78
10	24	.48	.14	.14	.77
11	26	.53	.14	.26	.81
12	22	.46	.12	.26	.69
13	28	.56	.16	.24	.86
14	21	.52	.18	.21	.85
15	21	.60	.13	.33	.77
16	21	.55	.17	.25	.88
All items	**405**	**.55**	**.15**	**.14**	**.89**

CTT difficulty values were examined by item types; descriptive statistics are provided in Table 9. As expected, the TDA was the most difficult item type on average for EOCEP English 2 test-takers, with the lowest average difficulty (p = .36); multiple choice questions were the "easiest" item type, with the highest p-value reported among item formats (p = .57). Evidence-based and Technology-enhanced items reported some very difficult items, with low CTT-based difficulty values.

Table 9. Descriptive Statistics for CTT Difficulty Values, EOCEP English Item Types

Item Type	N	Mean	Std. Deviation	Minimum	Maximum
Evidence Based	24	.40	.20	.14	.59
Multiple Choice	367	.57	.15	.21	.89
Multiple Selection	3	.52	.08	.45	.60
TDA	6	.36	.02	.34	.39
Technology Enhanced	5	.44	.25	.16	.74

Over the 405 EOCEP English 2 field-test items, the item difficulty values appear to be acceptable given the purpose of the test. Average values generally report a test of moderate difficulty. Unique items on each form show that the item difficulty values are reasonable; with differences among forms warranting a closer review. Difficulty values by item types performed largely as expected.

Evaluation: CTT Discrimination. Across all items, the average discrimination value (r_{pb} = 0.39), illustrates that the set of test items are discriminating acceptably between examinees of different English 2 ability levels. Generally, EOCEP English 2 examinees with lower total test scores chose incorrect responses and higher ability students chose correct responses; however, as the r_{pb} is at a moderate correlation value, there are some inconsistencies. The range of r_{pb} values, from .001 to .66 shows a spread of values; however, the majority of item discrimination values are between .30 and .50.

Of the set of 405 field test items, 34 (8%) were at or below a point biserial correlation value of .20. These items may be candidates for examination, revision, and field re-testing with future EOCEP English 2 administrations.

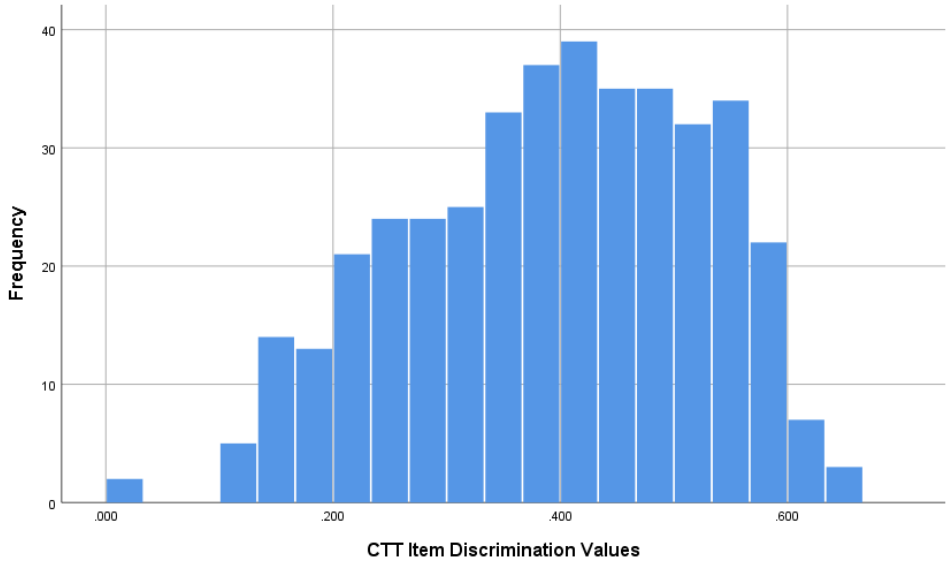

Figure 3. Distribution of EOCEP English 2 CTT-Based Discrimination Values

Table 10 provides summary statistics for the discrimination index across EOCEP English 2 forms; mean values by form were close to the overall average. Across the set, the highest form discrimination value was r_{pb} = 0.46 and the lowest form r_{pb} = discrimination value was .31. While there may not be differences in discrimination between forms and the pool average, there were a few forms with a difference in point biserial values greater or equal to r_{pb} of .10 between forms. These larger differences are noted for forms with at least one non-discriminating item, with a r_{pb} value less that .15 (as shown by the minimum r_{pb} value column). These differences between forms are very likely to become smaller as the items scheduled for field re-testing are retested.

Table 10. Descriptive Statistics for CTT Discrimination Values, EOCEP English 2 Forms

Form	N	Mean	Std. Deviation	Minimum	Maximum
1	47	.46	.11	.19	.63
2	25	.44	.11	.28	.63
3	26	.40	.12	.20	.66
4	24	.39	.12	.21	.60
5	25	.39	.13	.18	.62
6	22	.36	.14	.11	.60
7	27	.44	.09	.23	.60
8	21	.43	.12	.20	.60
9	25	.31	.12	.10	.52
10	24	.36	.13	.15	.66
11	26	.32	.14	.001	.57
12	22	.33	.15	.03	.62
13	28	.40	.11	.12	.59
14	21	.35	.13	.14	.60
15	21	.44	.11	.21	.64
16	21	.39	.12	.16	.58
All items	**405**	**.39**	**.13**	**.001**	**.66**

Considering EOCEP English 2 item formats, mean values for the TDA items and the Multiple Selection items suggested that these formats were the most difficult for examinees. As these two item formats require students to conduct more analysis and/or creating skills, the higher discrimination values are appropriate. The multiple-choice item format contains the items with lower discriminations (i.e., under the typically used guideline of ≤ .2). Review of item content, item stems and response options may help these items prior to field re-testing.

In summary, the EOCEP English 2 items are, on average, moderately discriminating between students with higher and lower skill levels. This level is appropriate for the purpose of the assessment and values are in line with other state-wide examinations. As plans for revising future field-test versions progresses, the 34 items with lower point biserial values (r_{pb} ≤ .20) can be reviewed and revised as needed to produce items which accurately distinguish between students.

Table 11. Descriptive Statistics for CTT-Based Discrimination Values, EOCEP English Item Types

Item Type	N	Mean	Std. Deviation	Minimum	Maximum
Evidence Based	24	.48	.11	.23	.66
Multiple Choice	367	.38	.13	.001	.62
Multiple Selection	3	.61	.04	.59	.66
TDA	6	.60	.03	.58	.64
Technology Enhanced	5	.41	.15	.21	.62

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 by drawing the test takers away from the correct answer. This is an important component, as distractors that are not effective are virtually useless. As a result, there is a greater possibility that students will be able to select the correct answer by guessing, as the plausible options have been reduced. Our intention in distractor analysis is to identify distractors that would seem to be the correct answer to weaker students. In addition, item omissions were examined to see if there were items which were "skipped" by many examinees. The number of omissions per item and the r_{pb} per distractor were examined relative to the correct answer to assess if the distractors were functioning appropriately.

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 r to analyze its functioning and, as needed, to alert users that an option may need refined to increase effectiveness. However, instead of expecting a positive, high r_{pb} value, a negative correlation is of interest, illustrating students with lower ability select the option instead of the correct answer. Distractors which may be partially correct or appealing to higher ability students can be identified.

Evaluation: Distractor Analysis. The number of omissions for selected response items was not a concern as omitted counts were low across all 405 items in the EOCEP English 2 field-test pool. The highest number of omissions was 170 (roughly .005% of field test examinees). The difficulty values for the items with over 100 omissions reported values of p = .43 or greater, showing that characteristics of the item (e.g., excessive hardness) was not an issue for omission.

The distractor analysis reviewed distractor information for the 367 multiple choice items. For each correct answer, the remaining three options were examined to determine if incorrect options yielded negative discrimination values with a positive discrimination for the correct value. Figure 4 presents results for the 367 multiple choice items and distractors. For the set of items in the field test pool, 39 (roughly 11%) yielded at least one option with an r_{pb} value greater than the point biserial value for the correct option (illustrating that more of the higher ability students were selecting a distractor than the correct option). The majority of these items (36 of 39) reported a point-biserial value less than .20; only three of the items reported a r_{pb} value greater than .20. All of these items may be candidates for re-examination of options and field re-testing in future administrations of the EOCEP English 2 examination.

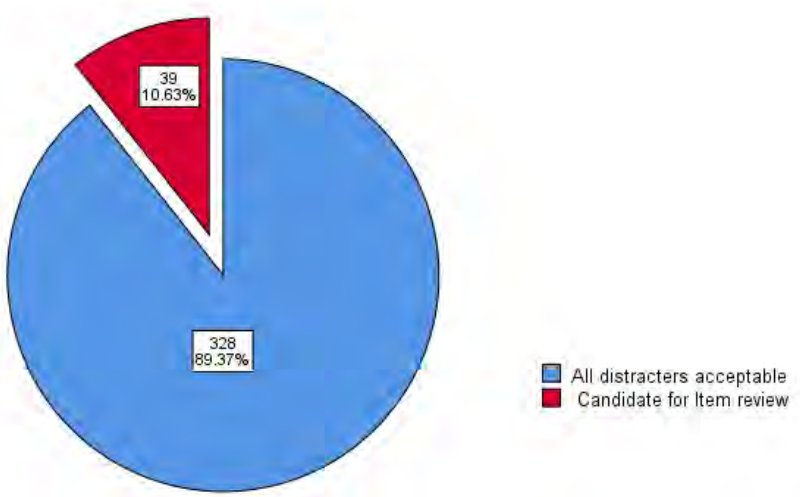

Figure 4. Distractor Analysis EOCEP English 2 Multiple Choice Items

D.3. Summary: Test Calibration, Equating, and CTT Based Item Analysis

Test calibration using the Rasch measurement model is an appropriate paradigm to use to analyze and score the EOCEP English 2 test. In addition, the linking and equating procedures appear appropriate. Additional information about the procedures may be helpful for stakeholders and can be provided in a future technical manual.

As the test is delivered online, selected response items are quickly scored and preliminary results returned within 36 hours. The TDA does take longer for return of scores (10-days); however, this is reasonable given the intensity needed with scoring an open-ended response. The information via preliminary reports provides timely feedback for teachers and schools to guide instruction.

CTT-based difficulty and discrimination values were examined for the 405 items in the EOCEP English 2 field-test pool, item difficulty values showed an average level of difficulty at the moderate level with the range of item difficulty values acceptable given the purpose of the test. Difficulty values by item types performed largely as expected. Concerning discrimination, 34 items (8% of the field-test pool) were at or below a point biserial correlation value of .20. Finally, the distractor analysis illustrated that multiple choice options were largely functioning as intended. Roughly 11% of items yielded one distractor with an r_{pb} value greater than the point biserial value for the correct option (illustrating that more of the higher ability students were selecting a distractor than the correct option). Most of these items were noted as problematic by CTT-based discrimination. These items may be candidates for further examination, revision, and re-field testing with future EOCEP English 2 administrations.

Section E

Rasch-Based Indices and Assessment of Impact

The Rasch measurement model relates person and item characteristics to the probability of choosing a correct response (or placement in a given category). 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). The evaluation purposefully focuses on application and is a non-technical presentation. Formulas for the Rasch model, computation of difficulty value estimates, and and calculation of fit indices may be found in many excellent texts on measurement and/or Rasch modeling (e.g., Bandalos, 2018; Bond & Fox, 2007; Smith & Smith; 2004).

For scoring, DRC uses the Rasch model with dichotomous items (i.e., selected response) and the Rasch Partial Credit Model with the TDA (4-category) item. EOCEP English 2 field test data were calibrated to obtain item parameters, item fit information, and estimation of score impact. This section examines Rasch-based psychometric indices to evaluate characteristic of item difficulty (i.e., location), item fit to the Rasch model, differential item functioning, and estimation of impact. Data for analyses was provided by the SCDE and all estimates and fit indices were computed by DRC. No individual item statistics were computed, only summary information, such as means and standard deviations of indices, were computed. Impact data and cut scores reviewed were acquired from the Standard Setting draft (DRC, 2019).

E.1. Rasch-based Item Difficulty and Item Fit

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 an item is .5 (or 50%). Difficulty values are typically within the range –3 < = *difficulty* < = +3. (Baker, 2001). Item difficulty parameters can be interpreted relative to ability level. As stated in Baker (2001, p. 34-35) "an item whose difficulty is –1 functions 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."

Both Infit and Outfit are Rasch-based fit statistics, indicate how accurately the data fit to the Rasch model. 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. 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.

Evaluation: Rasch Based Difficulty Indices. Rasch item parameters provide a model-based item difficulty. For dichotomously scored (e.g., 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. Difficulty values for all 399 objective response items (multiple choice, evidence based, technology enhanced, and multiple selection) are discussed first.

For the set of objective response items, the mean Rasch difficulty value was .55, meaning the set of items was targeted just above the average position on the latent variable of ability. As shown in Figure 5, the difficulty values cover a wide range of ability levels, ranging from a minimum value of -1.73 to a maximum value of 3.01. The distribution of values shows more values under a latent ability value of 1.0, meaning (for a 50% chance of getting the item correct) the items in the EOCEP English 2 field test pool are generally targeted toward lower than average to slightly higher than average ability examinees.

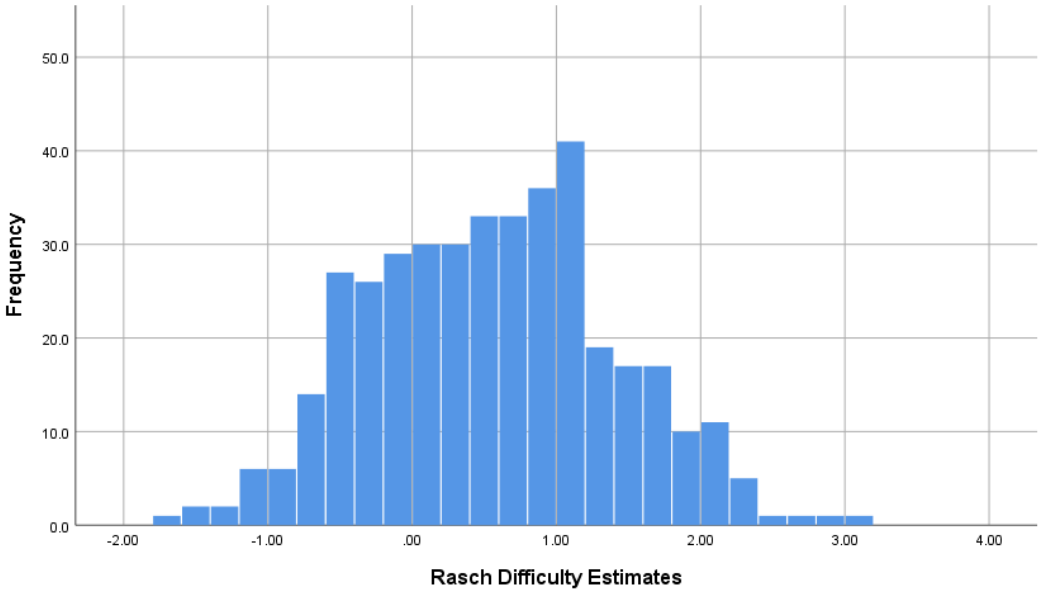

Figure 5. Rasch Difficulty Estimates, EOCEP English 2 Objective Response Items

Examining the distribution of Rasch-based difficulty values in Figure 5, the majority of items are located at an ability level of 0.0 up to a value of 1.0. These items are (generally) targeted to examinees with average to slightly above average knowledge of English 2. Approximately 69% of the test items are targeted under a Rasch ability estimate of 1.0. There are roughly 31% of the items in the EOCEP English 2 field test pool 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 English 2. Table 12 provides a frequency chart, by category, of item location (difficulty) values for the set of 399 objective response items.

Table 12. Frequency Table of Rasch-Based Difficulty Estimates, EOCEP English 2 Field Test Items

Item Location	Frequency	Percent	Cumulative Percent
-2.00 up to -1.00	11	2.8	2.8
-1.00 up to 0.0	102	25.6	28.3
0.0 up to 1.0	162	40.6	68.9
1.0 up to 2.0	104	26.1	95.0
2.0 up to 3.0	19	4.8	99.7
3.0 and Higher	1	.3	100.0
Total	399	100.0	

Rasch-based item difficulty values were examined across forms. As before, it is noted that these are not complete EOCEP English 2 test forms with 55-items but are a selection of unique items that appear on a form. Form 1 has the most items and yielded an average difficulty value (location parameter) close to an ability level of 0.0, or targeted toward 0 (average) assuming a normal distribution of examinee knowledge. Instead, Forms 2 through 16 are compared as they have similar numbers of items. Form 2 reports the lowest average difficulty value of .20 and Form 12 reports the highest average difficulty, 1.03. This is a wide discrepancy between forms, with Form 2 targeted at (approximately) the average ability (Form 2) and the others at higher ability levels, leading up to 1 standard deviation above average (Form 12). It is reiterated that the forms below are not the final test forms, as additional items will be added to any one form to create the 55-item test while also including common items across test forms. However, it is noted that care should be taken to create EOCEP English 2 operational test forms that are balanced in terms of form difficulty.

Table 13. Descriptive Statistics for Rasch-Based Difficulty Estimates, by Form

Form	N	Mean	Std. Deviation	Minimum	Maximum
1	47	.05	.52	-1.06	.96
2	24	.20	.93	-1.73	1.75
3	25	.43	.90	-1.12	2.60
4	23	.67	.75	-.67	2.02
5	25	.44	.79	-1.27	2.04
6	22	.72	1.15	-1.07	2.89
7	27	.27	.70	-.76	1.91
8	21	.64	.81	-.51	2.38
9	25	1.02	.80	-.68	2.21
10	24	.99	.74	-.54	3.01
11	26	.69	.70	-.81	2.06
12	22	1.03	.59	-.09	2.10
13	28	.57	.90	-1.36	2.34
14	20	.77	1.03	-1.13	2.51
15	20	.35	.73	-.62	1.93
16	20	.58	.96	-1.45	2.32
All items	**399**	**.55**	**.85**	**-1.73**	**3.01**

Rasch-based item difficulty values were investigated across item formats. As expected, the TDA items are targeted to the highest ability level, with an average item difficulty value located at 1.89. Multiple-choice items yielded the lowest average difficulty, with a mean value slightly above the average difficulty level of 0. Evidence-based and Technology Enhanced items also appeared difficult for examinees, with these item formats reporting average difficulty levels above 1.0. However, there are fewer of the alternate item types and more multiple-choice items on a given EOCEP English 2 test form. This will help to ensure that there is a mixture of "less" difficult items along with item formats noted as "harder" for examinees. In general, the Rasch-based item difficulty values were as expected across the different item types included on the EOCEP English 2 examination.

Table 14. Descriptive Statistics for Rasch-Based Difficulty Estimates, by Item Format

Item Format	N	Mean	Std. Deviation	Minimum	Maximum
Evidence Based	24	1.44	.69	.43	3.01
Multiple Choice	367	.48	.82	-1.73	2.51
Multiple Selection	3	.84	.36	.44	1.13
TDA	6	1.89	.19	1.62	2.19
Technology Enhanced	5	1.26	1.44	-.43	2.89

Evaluation: Rasch Based Fit Indices. Tables 15 and 16 provides the mean square values for Rasch Infit and Outfit measures by form and for the entire field test item pool. For both infit and outfit mean square values, mean values suggest adequate fit. All items used on the field tests yielded average Infit and Outfit vaues close to the expected value of 1. No values were outside of the recommended bounds. The information indicates that the Rasch model provides an acceptable fit to the field test items used to create the EOCEP English 2 forms.

Table 15. Average Standardized Infit Values, by EOCEP English 2 Form

Form	N	Mean	Std. Deviation	Minimum	Maximum
1	47	1.00	.17	.75	1.43
2	25	1.02	.14	.79	1.28
3	26	1.03	.15	.73	1.29
4	24	1.03	.14	.79	1.23
5	25	1.02	.16	.77	1.32
6	22	1.01	.15	.79	1.36
7	27	1.00	.13	.81	1.29
8	21	1.00	.15	.77	1.32
9	25	1.06	.12	.86	1.26
10	24	1.03	.14	.72	1.25
11	26	1.05	.14	.79	1.28
12	22	1.05	.15	.77	1.35
13	28	1.02	.13	.81	1.30
14	21	1.05	.12	.85	1.23
15	21	1.03	.13	.81	1.30
16	21	1.05	.13	.88	1.30
Total	405	1.03	.14	.72	1.43

Table 16. Average Standardized Outfit values, by EOCEP English 2 Form

Form	N	Mean	Std. Deviation	Minimum	Maximum
1	47	1.00	.25	.60	1.72
2	25	1.07	.24	.72	1.47
3	26	1.04	.24	.66	1.48
4	24	1.06	.23	.67	1.38
5	25	1.02	.25	.63	1.44
6	22	1.04	.30	.66	1.58
7	27	.98	.21	.69	1.46
8	21	1.00	.23	.65	1.46
9	25	1.11	.20	.82	1.44
10	24	1.05	.21	.65	1.51
11	26	1.08	.22	.71	1.70
12	22	1.09	.22	.70	1.55
13	28	1.03	.25	.68	1.66
14	21	1.13	.23	.72	1.62
15	21	1.07	.25	.60	1.75
16	21	1.12	.26	.79	1.74
Total	405	1.05	.24	.60	1.75

E.2. Differential Item Functioning

Test items are typically reviewed for differential item functioning (DIF). Examinations of DIF examine the actual test performance of examinees in different demographic groups, where examinees are matched in terms of their ability level (i.e., Theta level). If examinees different groups perform differently as related to an item, a characteristic about the question could be unfairly causing a difference to appear. Here, DIF 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).

For the EOCEP English 2 field test results, DRC performed a DIF analysis based on demographic groups of gender (male vs. female participants) and race/ethnicity (Caucasian vs. African American participants). The groups are termed focal and reference groups, where disadvantaged individuals are categorized as the focal group (e.g., female, African Americans), and the advantaged ones are categorized as the reference group (males, Caucasian students).

The standard in the psychometric industry (i.e., Mantel-Hanzel test statistic) was used to examine 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 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, where a negative sign (e.g., C-, B-) shows the presence of DIF against the focal group; a positive sign (e.g., C+, B+) illustrates the presence of DIF against reference group. In other words, positive DIF values mean that the question is more difficult for members of the reference group (along the ability continuum) than for matched members of the focal group and vice versa for negative DIF items.

Any assessment will ideally be comprised of category A questions if the test pool is sufficient. Category B questions may be used, with preference for questions with smaller DIF values (all other aspects, including content coverage, etc. equal). Questions exhibiting category C level DIF should not be used, if possible.

Evaluation: Differential Item Functioning. For the EOCEP English 2 ELA tests, DIF measures were investigated for the 405 items available in the item pool by running frequency tables of DIF classification indices computed by DRC.

Considering DIF across gender groups, the majority of EOCEP English 2 items (roughly 98%) exhibited negligible DIF. Only 10 items demonstrated slight to moderate DIF. In general, it can be assumed that the tests are free of gender DIF.

Table 17. DIF Investigation by Gender, EOCEP English 2 Field Test Items

DIF Classification	Frequency	Percent	Cumulative Percent
A-	174	43.0	44.4
A+	215	53.1	97.5
B-	6	1.5	99.0
B+	4	1.0	100.0
Total	**405**	**100.0**	

Note: Male = focal group; Female = reference group

Race/ethnicity comparisons showed slightly more items exhibiting DIF. The majority of the items, roughly 95%, yielded negligible DIF (i.e., A level). However, there were more items with DIF at the B level (slight to moderate DIF), with roughly 4.5% of the items in the field test pool showing a low level of DIF by race/ethnic groups. In addition, one C- level DIF item was observed, where this item was more difficult for the focal group (African American) than reference group (Caucasian) examinees. This item needs reviewed, and possibly re-fielded due to the presence of DIF as well as other reason psychometric indices showing substandard values (e.g., low CTT-based difficulty, higher r_{pb} value for distractor than correct option).

Table 18. DIF Investigation by Race/Ethnicity, EOCEP English 2 Field Test Items

DIF Classification	Frequency	Percent	Cumulative Percent
A-	220	54.3	55.8
A+	160	39.5	95.3
B-	16	4.0	99.3
B+	2	.5	99.8
C-	1	.2	100.0
Total	405	100.0	

Note: Caucasian = focal group; African American = reference group

DIF was examined for gender and race/ethnicity by item type. There was little DIF observed by gender over all field test items. Across genders, all items with Category B classification (i.e., slight to moderate DIF) were constrained to questions of multiple choice format. Considering race/ethnicity, items with Category B classification were of multiple choice and evidence-based formats. The one item with Category C DIF was a multiple choice question.

E.3. Estimates of Impact

EOCEP information is used for accountability evidence at the federal and local levels. Scores from the English 2 examination are categorized into performance levels for accountability purposes. The Standard Setting Draft Technical Report (DRC, 2019) reports the four achievement levels (Does Not Meet Expectations, Minimally Meets Expectations, Meets Expectations, and Exceeds Expectations) used to categorize students' test performance. The performance level is related to a student's ability, as provided by the Rasch person measure. Considering a normal distribution of ability (i.e., Theta), the 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. As the ability score (Theta) is used to create a student's EOCEP English 2 score, different cut scores produce different letter grades.

To judge impact of the EOCEP English 2 cut scores, the assessments should be able to categorize students into different ability levels, according to the amount of knowledge students possess. Using the Rasch-calibrated estimates, these raw scores (on the Theta metric) may be transformed and categorized for accountability reporting. Data evaluated in this section was taken directly from DRC Standard Setting documentation for the final cut-scores. 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 report (DRC, 2019).

DRC and SCDE personnel held a workshop in the summer 2019 to recommend performance standards for the EOCEP English 2 assessments (DRC, 2019). The July workshop involved 21 educators and stakeholders from across the state. The purpose of the meeting was to develop cut scores for the EOCEP English 2 assessment to divide students into four achievement levels: Does Not Meet Expectations, Minimally Meets Expectations, Meets

Expectations, and Exceeds Expectations. The performance level descriptors and grade associated grade level(s) are reported in Table 19.

Table 19. Generic Description of EOCEP English 2 Performance Level Descriptors

PLD	Description of EOCEP English 2 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: Estimates of Impact. EOCEP English 2 scores are provided in line with the state's uniform grading policy, including numerical scores relating to letter grades bounded by A, B, C, D, F. Using information from discussions over three rounds of the Bookmarking procedure, educators constructed cut-scores for the ability (i.e., Theta) distribution of EOCEP English 2 examinees. As five "grades" are needed, four cut-points (i.e., cut-scores) in the ability distribution were required. The SCDE website allows examination of the percentage of students scoring letter grades A through F by district and high school (https://ed.sc.gov/data/test-scores/state-assessments/end-of-course-examination-program-eocep/).

For accountability purposes, the distribution of Rasch ability scores (i.e., the Theta distribution), cut scores were identified which would break the Theta distribution into ordered performance levels. The Standard Setting Draft Technical Report (DRC, 2019) reports the four cut scores which divide the latent domain of English 2 ability into letter grades. Table 20 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 Overall, the EOCEP English 2 ability estimates appears to be within adequate ranges; the categorization of students into different performance levels allows for differentiation of students at different ability levels.

Table 20. Cut Scores on the Theta Metric and Associated Grade, EOCEP English 2

English 2 Ability Distribution Cut-Scores			
D	**C**	**B**	**A**
-0.1173	0.6975	1.4614	2.2507

Note: cut-scores based on the unstandardized Theta metric

Impact data illustrates the effect of using the "cuts" on the percentage of EOCEP English 2 students that would receive a given letter grade. The discussions outlined in the Standard Setting draft report detail the procedures used to arrive at the final cut scores, including review of other percentages per category for other tests (e.g., EOCEP English 1, SC READY) and review of standard errors surrounding scores. The final cut scores providing the percentage of students per category is in line with previous test data and are acceptable and appropriate for use.

Table 21. Impact Data for South Carolina EOCEP English 2, Percent of Examinees by Letter Grade

Letter Grade					
F	**D**	**C**	**B**	**A**	**Percentage C or Higher**
28.3%	19.4%	18.4%	19.0%	15.0%	52.4%

E.4. Summary: Rasch-based Item Difficulty and Item Fit

The distribution of Rasch-based difficulty estimates for the EOCEP English 2 field test pool are generally targeted toward lower than average to slightly higher than average ability examinees. In general, the Rasch-based item difficulty values were as expected across the different item types included on the EOCEP English 2 examination. Forms may be reviewed prior to distribution to show that difficulty values are comparable across forms.

Items statistics showed that items had acceptable fit to the Rasch model. No alterations are needed. Also, EOCEP English 2 tests appear free of gender DIF. One item exhibiting racial/ethnicity DIF item was observed, where this item was more difficult for the focal group (African American) than reference group (Caucasian) examinees. Finally, impact data showed the EOCEP English 2 ability estimates were within acceptable limits; the categorization of students into different performance levels allows for differentiation of students at different ability levels.

Section F

Summary and Recommendations

This report summarized the results from the spring 2019 field test of the South Carolina End of Course Educational Program, English 2 examination (EOCEP English 2). The EOCEP English 2 is a requirement for all students enrolled in public school programs (unless noted by IEPO and counts as 20% of a student's course grade as well as for local and federal accountability purposes. This study reviewed item and form data from the spring 2019 field test, which were computed by the test contractor, Data Recognition Corporation. Based on the results, the following summary information and recommendations are provided.

1. Test Regulations, Construction, and Performance

In summary, materials detailing construction of the EOCEP English 2 are detailed and easy to access from the SCDE website. The test appears to balance the number of items that are devoted to Reading and Writing, considering complexity of tasks. The blueprint information is acceptable to inform stakeholders of what is expected on the EOCEP English 2, in terms of domain coverage and possible range of items. Information from the DOK levels reported on the blueprint help stakeholders understand the complexity of the test. There are many materials available for examinees to become more familiar with the test questions and testing format to help test takers understand the types of questions and responses expected.

Recommendation: Updated technical information regarding scoring and a test review will be helpful to include on the SCDE website once the EOCEP English 2 becomes operational.

2. Alignment to Content and Standards

The EOCEP English 2 items aligned with the English 2 content standards. Also, the Test Blueprint accurately represented the percentage of items to be expected by content domain and DOK level. Items illustrated best practices of industry standards, were error free and appeared unbiased. Items used appropriate content-based language and written to the target population.

3. Test Administration

The test administration procedures provide clear directives to deliver the EOCEP English 2 properly and with fidelity. Information provided for district/school personnel are clear and detailed. Objective information that is followed by all district/school testing personnel helps to ensure uniform testing procedures delivered to all English 2 examinees across the state. Information is easy to find on the SCDE website, helping to ensure that all testing coordinators are well-informed, and have appropriate training, and follow relevant security procedures. Access to uniform testing procedures can help ensure validity associated with EOCEP English 2 scores for use with accountability and decision making.

4. Test Calibration, Equating, and CTT Item Analysis

Test calibration using the Rasch measurement model is an appropriate paradigm to use to analyze and score the EOCEP English 2 test. In addition, the linking and equating procedures

appear appropriate. As the test is delivered online, selected response items are quickly scored and preliminary results returned within 36 hours. The TDA does take longer for return of scores (10-days); however, this is reasonable given the intensity needed with scoring an open-ended response. The information via preliminary reports provides timely feedback for teachers and schools to guide instruction.

CTT-based difficulty and discrimination values were examined for the 405 items in the EOCEP English 2 field-test pool, Item difficulty values showed an average level of difficulty at the moderate level with the range of item difficulty values acceptable given the purpose of the test. Difficulty values by item types performed largely as expected. Concerning discrimination, 34 items (8% of the field-test pool) were at or below a point biserial correlation value of .20. Finally, the distractor analysis illustrated that multiple choice options were largely functioning as intended. Roughly 11% of items yielded one distractor with an r_{pb} value greater than the point biserial value for the correct option (illustrating that more of the higher ability students were selecting a distractor than the correct option).

Recommendations: Additional information about test calibration technical procedures may be helpful for stakeholders. This information may be provided in a future technical manual.

Items noted as problematic by CTT-based indices can be reviewed, revised, and re-field testing with future EOCEP English 2 administrations.

E. Rasch-Based Indices and Assessment of Impact

In general, the Rasch-based item difficulty values were as expected across the different item types included on the EOCEP English 2 examination. Forms may be reviewed prior to distribution to show that difficulty values are comparable across forms. Items statistics showed acceptable fit to the Rasch model. No alterations are needed. Also, EOCEP English 2 tests appear free of gender DIF. Finally, impact data showed the EOCEP English 2 ability estimates were within acceptable limits; the categorization of students into different performance levels allows for differentiation of students at different ability levels.

Recommendation: Review the one item with C level racial/ethnicity DIF to see if revisions and/or re-testing can help alleviate problems with differential functioning across groups.

Overall, the EOCEP English 2 field test data is appropriate and provides a test with good psychometric 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.

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

Data Recognition Corporation (2019). *South Carolina End-of-Course Examination Program (EOCEP) English 2, Standard Setting Draft Technical Report.* Information retrieved from the South Carolina State Department of Education.

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

Jones, E. (2018). Memorandum retrieved from the South Carolina State Department of Education.

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, August 28, 2007.

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

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

South Carolina Department of Education (2007). Information retrieved from http://www.ed.sc.gov, August 28, 2007.

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