2023 No. 045

Differential Item Functioning (DIF) Analysis for Kentucky's Career and Technical Education End-of-Program (CTE EOP) Assessments

Final Report

Prepared for: Kentucky Department of Education
Office of Assessment and Accountability
300 Sower Boulevard
Frankfort, KY 40601

Prepared under: Contract #1900004339

Authors: Gertrudes Velasquez
Meng Fan
Emily R. Dickinson

Date: April 14, 2023

Differential Item Functioning (DIF) Analysis for Kentucky's Career and Technical Education End-of-Program (CTE EOP) Assessments

Table of Contents

Table of Contents (Continued)

List of Tables

Table of Contents (Continued)

List of Figures

Differential Item Functioning (DIF) Analysis for Kentucky's Career and Technical Education End-of-Program (CTE EOP) Assessments

Chapter 1: Introduction

Background

As part of its Quality Control, Validation & Research Services contract with the Kentucky Department of Education (KDE), the Human Resources Research Organization (HumRRO) conducts research on Kentucky's Career and Technical Education (CTE) End-of-Program (EOP) Assessments. This study examined differential item functioning (DIF) among CTE EOP items. DIF occurs when students from different groups, but with the same level of ability, have different probabilities of answering a test item correctly (Camilli & Shepard, 1994). Items demonstrating DIF on an operational assessment are a threat to the validity of test score interpretations (Holland & Wainer, 1993) because item responses from some students are dependent upon factors other than the ability construct the test is intended to measure.

CTE in Kentucky is one of several options to support students to successfully transition into postsecondary life. CTE "prepares the workforce across a wide range of industries and occupations." [1] CTE is accessible to all students through their local high school or through their high school's partnering career and technical education center(s). KDE's CTE programs provide over 150 career pathways developed through collaboration between high schools, postsecondary institutions, and business and industry employers. The pathway courses emphasize foundational academic, employability, and occupational/pathway skills that lead to postsecondary degrees and/or industry recognized certifications and/or licensures. Completing a CTE pathway or earning an industry certification related to a CTE pathway allows students to earn early postsecondary credentials and/or credits.

The CTE EOP assessments are designed to measure the foundational academic, employability and technical skills of students in an approved CTE career pathway. KDE developed pathway-specific EOP assessments that align with standards identified by Kentucky employers. Students can take the assessment for a career pathway once they have completed two credits in that pathway. KDE provides opportunities for students to earn articulated credit toward postsecondary completion through articulation agreements between secondary and postsecondary schools. Passing the pathway-specific EOP assessment is one requirement of fulfilling an articulation agreement with a participating postsecondary institution. Student performance on CTE EOP assessments may also be used as an indicator of Postsecondary Readiness under Kentucky's District and School Accountability Model.

Description of Data

KDE provided HumRRO with student data from the 2018-2019 and 2021-2022 CTE EOP assessments to conduct this study. EOP assessments were given to high school students as one measure of the career readiness component of postsecondary readiness. 2018-2019 and 2021-2022 data consisted of 26 and 32 assessments, respectively. Each assessment consisted of operational (OP) and field test (FT) items. Performance on OP items contributed to students' test scores. FT items were administered to evaluate item quality before becoming operational for future administrations. In addition, OP and FT items on each assessment were classified

---

[1] https://education.ky.gov/CTE/Documents/What-is-CTE-Eng.pdf

according to whether they were intended to measure foundational academic (A), employability (E), or occupational/pathway (P) skills. Test items that measure students' academic and employability skills were common to all CTE EOP assessments. Items that measure students' occupational skills were test specific.

General Approach for Conducting Analyses

We applied the following steps to prepare data for analysis. For each assessment, we first used generic item identifiers (e.g., Q1sc – Q120sc), available on each assessment and from supplemental data on test items provided by KDE, to classify items by item type (OP, FT) and skill they intend to measure (A, E, P), and to match text in item stems to students' responses. We then merged across assessments using item text, item type, and skill type to create a comprehensive list of all items administered within a year. There were 1664 and 2558 unique items that were administered in the 2018-2019 and 2021-2022 assessments, respectively. We then assigned a unique identifier (item ID) to each item. IDs for 2018-2019 items ranged from i_5001 to i_6664. IDs for 2021-2022 items ranged from i_7001 to i_9558. We used these item IDs and item classification information to refer to items in subsequent analyses. Lists of all items administered within a year are provided as supplementary documents to this report (listed as *CTE_EOP_1819_Complete_Items_List.xlsx* and *CTE_EOP_2122_Complete_Items_List.xlsx* in Appendix A).

Having prepared data for analyses, we undertook two main approaches for analyzing assessments administered in each academic year. In the first approach, we analyzed each assessment individually, treating item sets in each assessment as units for analyses. We labeled this as the *individual assessment* approach. We also realized that test items that measure students' academic and employability skills were common to all assessments administered within each year. So, in the second approach, we aggregated students' responses on A and E items across assessments. Then, we analyzed these items together, treating aggregated students' responses on A and E items as a single assessment. We labeled this as the *aggregated assessment* approach. We did not include occupational/pathway (P) items in the latter approach because these items were assessment specific. For this approach, in scenarios where students provided item responses on multiple assessments (duplicate student IDs on different CTE EOP assessments), we selected students' responses based on their first administration date for inclusion in our analyses on the aggregated data matrix. We also examined the impact of maintaining all duplicate cases, removing all duplicate cases, and randomly selecting cases from duplicate cases on study results. We found that the study results were robust to our method for selecting duplicate cases. We present results based on sample selection by administration date only.

We conducted two primary analyses on assessments in each of the *individual assessments* and *aggregated assessments* approaches. We first conducted Classical Test Theory (CTT) item analysis by examining item difficulty, ability of item to differentiate between ability levels (item discrimination), and the impact of removal of each item on test reliability. We applied typically accepted criteria in the educational measurement literature to evaluate item quality based on their CTT item characteristics. In the second set of analyses, we examined whether students of similar ability on the construct of interest but from different demographic groups, such as gender (*Gender*), race (*Race*), number of credit hours taken (*Credit*), program completion status (*Completer*), English proficiency status (*LEP*), economically disadvantaged status (participation in free/reduced *Lunch* program), disability status (*SWD*), and work-based learning status (*WBL*) performed differently on test items. We refer to this as differential item functioning (DIF) analysis. We present summary statistics and graphical displays of key findings in this report,

and detailed results of our analyses for each assessment in separate documents supplemental to this report. Our supplementary documents contain flagged items that did not meet criteria established for our CTT item analysis and DIF analysis. We anticipate that this will be helpful to KDE for conducting item-level review based on our study findings. We present a brief overview of data available for conducting analyses in the following section.

Overview of Data available for Conducting Analyses

Table 1 presents summary statistics of the number of students who took 2018-2019 and 2021-2022 CTE EOP assessments. Tables 12 and 13 in Appendix A present further details on student sample sizes and on item types by skills/content strands for each of the assessment years. Of the 32 2021-2022 assessments available for analysis, one assessment (Agribiotechnology) was removed from all analyses because it consisted of only one observation. The mean number of students across both sets of assessments was 941 and 1,163, respectively. Student distributions for each assessment year display large variability in the number of students, as indicated by the respective standard deviation for each year. Sample sizes ranged from 30 (in Agribiotechnology) to 3,547 (in Admin Support) in 2018-2019 assessments and from 27 (in Production Crop) to 3,759 (Marketing) in the 2021-2022 assessments. Students took 100 OP and 20 FT items on each assessment. When these items were classified by skill, assessments consisted of 25% academic, 25% employability, and 50% occupational/pathway items. OP and FT items were randomly assigned to students in the 2021-2022 assessments, resulting in administration of numerous test forms per assessment. Due to the randomization process, the total number of OP and FT items administered in 2021-2022 to all students exceeded the total number of items each student received (Table 13 in Appendix A).

Table 1. Distribution of Students who took 2018-2019 and 2021-2022 CTE EOP Assessments

Assessment Year	Number of assessments	Minimum N of students	Median N of students	Mean N of students	SD	Maximum N of students
2018-2019	26	30.0	482.0	940.9	947.5	3,547
2021-2022*	31	27.0	641.0	1,162.3	1,128.6	3,759

Note. N= Number. SD= Standard deviation. Agribiotechnology is excluded from all analyses because it consisted of one observation.

Summary

The current study examined item quality in CTE EOP 2018-2019 and 2021-2022 assessments. We describe our study methodology, results, and present key findings from our CTT item analysis and DIF analysis for our *individual assessments* approach in Chapters 2 and 3 and from our *aggregated assessments* approach in Chapter 4. We present general recommendations in Chapter 5.

Chapter 2: Classical Test Theory (CTT) Analyses of Individual CTE EOP Assessments

Introduction

A well-constructed assessment typically consists of items that are not too easy or too difficult for examinees, that differentiate amongst examinees with different abilities on the construct being measured, and that consistently measure the ability construct of interest. Classical Test Theory (CTT) item analyses were first conducted to examine item quality (and, hence, evaluate test quality). Generated CTT item statistics included item difficulty (p-value) and item discrimination (pB). Item difficulty is the percentage of students who answered an item correctly. Item discrimination refers to the ability of an item to differentiate among students based on how well they know the material being tested. For a well-constructed item, students with higher abilities on the construct which the item measures should answer correctly more often than students with lower abilities. Point-biserial (pB) item-total correlation was used to evaluate item discrimination. Larger pB indexes indicate larger discrimination ability of an item. For each assessment, we sequentially removed each item from the test item pool and re-calculated Cronbach's alpha + to examine the impact of each item on the test's consistency in measuring the construct of interest. We refer to this reliability measure as *alphaIfDeleted* in our tables of CTT results.

Method

Items were flagged if their CTT item statistics did not meet recommended criteria used in educational measurement to evaluate item quality. Items that are too easy or too difficult are ineffective in distinguishing amongst levels of achievement over the ability range which the test is intended to measure (Ebel & Frisbie, 1991; Feldt, 1993). Mehrens and Lehmann (1991) noted that items with difficulty close to .05 or .95 make little or no contributions to reliability because of their low discrimination ability. Decreasing the range of item difficulty on a test increases discrimination ability of items and, hence, also increases reliability (Ebel & Frisbie, 1991; Feldt, 1993). Feldt (1991) showed that score reliability is enhanced when item difficulty lies between .57 and .67 in testing scenarios in which guessing is possible (e.g., on tests with multiple choice items). The effect on reliability is minimal when the range of item difficulty distributions increases from .27 to .79 (Feldt, 1991). Using averages of these lower and upper bounds for item difficulty described in the literature, we selected a range of .15 ($\cong$ (.05+.27)/2) to .85 ($\cong$ (.79+.95)/2) to flag items based on difficulty. We also used Ebel and Frisbie's (1991) recommendation for classifying items for revision or rejection if their item discrimination indexes are less than .20. All CTT item statistics were computed using the CTT package (v2.3.3; Willse 2018) in R statistical software (v4.2.1; R Core Team, 2022).

Results

In this section, we provide descriptions of observations and summary statistics from the CTT item analyses that we conducted on individual assessments (*individual assessments* approach). We provided KDE with additional files containing comprehensive results from our CTT item analyses (sheet labeled *CTT item analysis* in each file listed in Appendix B). These documents contain items that were flagged based on the CTT item analysis criteria described above.

We generated CTT item statistics for all CTE assessments. We were not able to generate reliability coefficients for 2021-2022 FT items because of the sparseness of the data matrix for these items in each assessment. CTT results for three assessments were based on small

sample sizes: 2018-2019 Agribiotechnology (n = 30), 2018-2019 Food Science Processing (n = 30), and 2021-2022 Production Crop (n = 27). We did not include these assessments in subsequent discussions on CTT item analyses results in this report and advise that results for these three assessments be interpreted with caution. Table 14 in Appendix A presents the number and percentages of 2018-2019 items flagged from the CTT item analyses. For operational items, the percentages of flagged items ranged from 21% (in Agribusiness and Marketing) to 46% (in Information Support Services). For field test items, the percentages of flagged items ranged from 35% (for Agribusiness) to 80% (Information Support Services). Table 15 in Appendix A presents percentages of 2021-2022 items flagged from the CTT item analyses. The percentages of flagged OP items ranged from 19.8% (in Ag Power Structured Tech) to 58.0% (in Aerospace Engineering). The percentages of flagged FT items ranged from 45.8% (in Retail Services) to 87.7% (in Aerospace Engineering).

Brief Summary

In general, percentages of flagged operational and field test items varied by individual assessments because they did not meet accepted criteria for item quality. Between 35% to 80% of field test and 21% to 49% of operational items across 2018-2019 assessments were flagged for not meeting item difficulty or item discrimination criteria. Across 2021-2022 assessments, 45.8 to 87.7% of field test items and 19.8% to 69.4% of operational items were flagged using the same CTT criteria.

Examination of common academic (A) and employability (E) items across assessments indicate inconsistencies in flag assignment to items based on their CTT item statistics. For example, a flag may be assigned to the same A or E item in a particular assessment but may not be flagged in another. Based on similar observations on subsets of common A or E items, we realized that results from individual assessments (the *individual assessments* approach) depend on sample characteristics and the ability distribution of the test takers. Consequently, the same items may be flagged in one or several assessments but not in others. We recommend that the examination of results from these analyses be contained within individual assessments because results across assessments are not comparable. This limitation will be overcome in our *aggregated assessments* approach, in which we aggregate students' responses on A and E items across assessments.

Chapter 3: Differential Item Functioning (DIF) Analyses of Individual CTE EOP Assessments

Introduction

In addition to examining item characteristics to evaluate test quality, we also conducted analyses on whether students of comparable abilities perform differentially on test items. CTE EOP scores should reflect student performance on the construct of interest (i.e., academic, employability, and technical skills), and not on other irrelevant factors (e.g., gender, ethnicity/race, English language status, cultural relevance, socioeconomic status, experiences with some other topic). When different groups of students of comparable abilities on the construct of interest have different rates of success on an item because their performance depends on some other extraneous factor other than knowledge of the construct, we say that the item performs differentially for these groups (Camilli & Shepard, 1994). This is referred to as differential item functioning (DIF) in the educational measurement literature. Issues of test fairness and equity among subpopulations of tested students arise when scores are dependent on student characteristics other than ability. This study analyzes DIF for all assessed CTE EOP pathways using data from 2018-2019 and 2021-2022. In the next section, we describe the study methods in detail.

Methods

The Mantel-Haenszel (MH) method (Holland & Thayer, 1998; Mantel & Haenszel, 1959) is commonly used in large-scale state assessment programs to examine differences in performance on test items between or amongst groups of students who have the same ability levels. Examination of performance differences between groups with the same ability levels on the construct of interest is referred to as differential item functioning (DIF) analysis in the educational measurement literature. In MH DIF analysis, group comparisons are conducted between a reference group (typically the majority or normative group) and a focal group (typically the minority or historically disadvantaged group). The focal group is usually the group of interest while the reference group is the comparison group.

Our decision to proceed with conducting DIF analyses for groups formed on a particular demographic characteristic (e.g., gender) is informed by recommendations in the DIF literature on group sample sizes. MH DIF analyses are usually conducted for groups with minimum group sample sizes of 200 (Mazor et al., 1992). This recommended minimum group sample size is based on groups of comparable sample sizes, resulting in a balanced sample design for the study. However, in many testing scenarios, groups that are formed on demographic characteristics to conduct DIF analysis may result in asymmetric group sample sizes (e.g., groups formed based on race). In such unbalanced group samples, the focal group typically does not meet the minimum sample size requirement, so DIF analysis is not conducted. Paek and Guo (2011) showed through simulation, however, that similar or higher DIF detection rates and accuracy observed in balanced design can be achieved in unbalanced sample designs in which the sample size for the focal group does not meet the minimum sample size criterion. In such cases, the researchers recommend examining whether the sample size of the reference group meets the minimum sample size requirement stipulated by an inequality approximation derived for sample sizes in unbalanced designs necessary for achieving comparable DIF detection rates in balanced designs. We used Mazor and colleagues' (1992) and Paek and Guo's (2011) recommendations to inform our decision logic for conducting DIF analyses. We first applied Mazor and colleagues' minimum sample size criterion of 200 to both focal and reference groups under consideration for analysis. We proceeded with DIF analysis if groups

met this criterion. If the sample size for the focal group did not meet this sample size criterion, we then compared the sample size for the reference group to the minimum sample size suggested by Paek and Guo's inequality approximation. We proceeded with the analysis when the sample size for the reference group exceeded the minimum sample size recommended by the inequality approximation. We placed a flag ("Small group Ns: MH DIF result is unreliable") on each item that were analyzed for DIF under this condition.

In the MH DIF procedure, performance between focal and reference groups on dichotomously scored items are examined, after matching students on overall scores. Students in the focal and reference groups are matched on total test scores by dividing respondents in both groups into defined strata on those scores. Total scores are generated by summing item scores across all items. The MH DIF method generates a statistic known as a common odds ratio estimator $\hat{\alpha}_{MH}$. This statistic, when transformed onto a logarithmic scale, is referred to as MH delta, $\hat{\Delta}_{MH}$. The absolute value of MH delta, $|\Delta_{MH}|$, is also applied as a measure of effect size in DIF applications. Educational Testing Service (ETS) classified DIF effect sizes (Holland & Thayer, 1988) to aid in interpretation in applications: Class A denotes negligible magnitudes of DIF, when $|\Delta_{MH}| \leq 1.00$; Class B denotes moderate magnitudes of DIF, when $1.0 < |\Delta_{MH}| < 1.5$, and Class C denotes large magnitudes of DIF, when $|\Delta_{MH}| \geq 1.5$. We used the generalized Mantel-Haenszel (GMH; Penfield, 2001; Somes, 1986) DIF method to conduct DIF analysis for variables with more than two groups (e.g., race).

This study used the MH and GMH methods to examine DIF in items on the 2018-2019 and 2021-2022 CTE EOP assessments. We conducted DIF analysis by comparing group performance based on gender (*Gender*), race (*Race*), number of credit hours taken (*Credit*), program completion status (*Completer*), English proficiency status (*LEP*), economically disadvantaged status (participation in free/reduced *Lunch* program), disability status (*SWD*), and work-based learning status (*WBL*). Our analyses for the *Race* variable were restricted to groups of students classified as White, Black, and Hispanic because of small sample sizes for other available groups in the data. We categorized the *Credit* variable into three groups: students with 2 credit hours or less, between 2 and 3 credit hours, and greater than or equal to 4 credit hours. We applied ETS DIF criteria to classify items into A, B, and C DIF categories. Items were flagged for DIF if they were categorized into Class B or C. We used the generalized Mantel-Haenszel (GMH; Penfield, 2001; Somes, 1986) method to conduct uniform DIF analysis for variables with more than two groups (e.g., race, credit hours). All DIF analyses were conducted using the difR package (v5.1; Magis et al., 2010) in R statistical software (v4.2.1; R Core Team, 2022).

Results

In this section, we provide summary statistics from the DIF analyses that we conducted on individual assessments (*individual assessments* approach). We provided several Excel files to KDE with comprehensive results of our DIF analyses (sheets labeled *Gender DIF, Race DIF, Credit Hrs DIF, Completer DIF, LEP DIF, Econ. Disadvantaged DIF, SWD DIF, and WBL DIF* for each file listed in Appendix B). Each file contains multiple sheets, with each sheet containing item DIF classification, along with information necessary for identification (item ID, item type, skill, and item prompt). We provide summary tables along with key observations and findings of these results in this report.

Table 2 displays selected focal and reference groups for each variable. The Male group served as the reference group in DIF analyses based on gender. Students classified in the White group served as the reference group in DIF analyses based on race/ethnicity. Students who completed their career pathway (the *Yes* group) served as the reference group in DIF analyses based on pathway completion (*Completer*) status. Students who were English proficient (the *No* group) served as the reference group in DIF analyses based on English proficiency. Students who did not participate in the free/reduced *Lunch* program (the No group) served as the reference group in DIF analyses based on economically disadvantaged status. Students without disabilities (the *No* group) served as the reference group in DIF analyses based on disability status. Students who did not receive work-based learning (the *No* group) were the reference group in DIF analyses based on work-based learning status.

Table 2. Focal and Reference Groups in CTE EOP DIF Analyses

Variable	Focal Group	Reference Group
Gender (*Gender*)	Female	Male
Race (*Race*)	Others (Hispanic, Black, or African American)	White
Credit Hours (*Credit*)	Others (2 credit hours or less, between 2 and 3 credit hours)	4 or more credit hours
Completer (*Completer*)	No	Yes
Limited English Proficiency (*LEP*)	Yes	No
Economically disadvantaged (participation in free/reduced *Lunch* program)	Yes	No
Students with Disabilities (*SWD*)	Yes	No
Work-based Learning (*WBL*)	Yes	No

We had adequate sample sizes to conduct DIF analyses for 13 of 26 (50%) 2018-2019 assessments and 20 of 32 (62.5%) 2021-2022 assessments when we analyzed each assessment individually. We did not conduct DIF analyses based on race (*Race*) or English proficiency (*LEP*) status for 2018-2019 assessments because of small group sizes in individual assessments. We conducted DIF analyses for both OP and FT items for the other grouping variables (*Gender*, *Completer*, *Lunch*, *SWD*, *WBL*, and *Credit*). We indicated whether group sizes for individual items met the minimum sample size criterion for these DIF analyses in our comprehensive sets of results.

Table 3 displays percentages of 2018-2019 items, categorized by item type (OP and FT), which demonstrated moderate to large levels of DIF (B or C DIF) for these grouping variables. For assessments with items which displayed *Gender* DIF, Agribusiness has the most items (29%) with *Gender* DIF effects, followed by Horticulture (23%) and Financial Services (21%). The percentages of items displaying *Completer* DIF effects range from 1% to 7% (Accounting). Regarding DIF associated with economically disadvantaged status (participation in free/reduced *Lunch* program), Agribusiness has the most DIF items (30%), followed by Financial Services (10.0%) and Computer Programming (9%). 15% of OP items in Culinary Food Services showed *SWD* DIF effects. 9% of Business Management OP items showed *WBL* DIF effects. A single Business Management OP item displayed *Credit* DIF effects.

For FT items, four assessments contained items with *Gender* DIF effects, ranging from 10% (in Admin Support and Business Management) to 15% (in Accounting and Agribusiness). Additionally, 7% of items in Accounting displayed DIF effects associated with completer status (*Completer*). Computer Programming has 25% of FT items classified as DIF items associated with participation in the free/reduced *Lunch* program, followed by Agribusiness and Business Management (each with 15.0% DIF items). Moreover, 25% of Business Management and 10% of Admin Support FT items were classified as *WBL* DIF items.

For FT items, four assessments contained items with *Gender* DIF effects ranging from 10% (in Admin Support and Business Management) to 15% (in Accounting and Agribusiness). Additionally, 7% of items in Accounting displayed DIF effects associated with completer status (*Completer*). Computer Programming has 25% of FT items classified as DIF items associated with participation in the free/reduced *Lunch* program, followed by Agribusiness and Business Management (each with 15.0% DIF items). Moreover, 25% of Business Management and 10% of Admin Support FT items were classified as *WBL* DIF items.

Table 3. Percentages of 2018-2019 Operational and Field Test Items demonstrating Moderate to Large Levels of Differential Item Functioning DIF

CTE Pathway	Item Type	N of Items	Gender		Completer		Lunch		SWD		WBL		Credit	
			n	%	n	%	n	%	n	%	n	%	n	%'
Accounting	FT	20	3	15	1	5	**	**	**	**	**	**	**	**
	OP	100	14	14	7	7	8	8	**	**	**	**	**	**
Admin Support	FT	20	2	10	**	**	**	**	**	**	2	10	**	**
	OP	100	8	8	**	**	1	1	**	**	5	5	**	**
Ag Power Structured Tech	OP	100	**	**	1	1	1	1	**	**	4	4	**	**
Agribusiness	FT	20	3	15	**	**	2	10	**	**	**	**	**	**
	OP	100	29	29	**	**	15	15	**	**	**	**	**	**
Animal Science	OP	100	16	16	2	2	1	1	7	7	6	6	**	**
Business Management	FT	20	2	10	**	**	2	10	**	**	5	25	**	**
	OP	100	16	16	3	3	1	1	**	**	9	9	1	1
Computer Programming	FT	20	**	**	**	**	6	30	**	**	**	**	**	**
	OP	100	**	**	**	**	9	9	**	**	**	**	**	**
Consumer Family Mgmt.	OP	100	14	14	2	2	2	2	**	**	**	**	**	**
Culinary Food Services	OP	100	8	8	2	2	2	2	15	15	0	0	**	**
Early Childhood	OP	100	**	**	4	4	4	4	**	**	**	**	**	**
Financial Services	OP	100	21	21	**	**	10	10	**	**	**	**	**	**
Horticulture	OP	100	23	23	2	2	3	3	**	**	8	8	**	**
Marketing	OP	100	15	15	1	1	1	1	**	**	5	5	**	**

Note. OP = operational; FT = field test; SWD = disability status; WBL = work-based learning status; N= number

**=Not available due to small sample size of one or more group. Sample sizes were too small to evaluate DIF by racial/ethnic groups or limited English proficiency (LEP) status. Sample sizes were too small to evaluate DIF for any groups for Agribiotechnology, Cinematography, Digital Design, Environmental Science, Fashion Interior Design, Food Science Processing, Graphic Design, Hospitality Travel Tourism, Information Support Services, Network Administration, Production Crop, Retail Services, and Web Development.

Table 4 displays percentages of items in 2018-2019 assessments, categorized by skill type (academic, employability, and pathway), which demonstrated moderate to large levels of DIF (B or C DIF). For items which displayed *Gender* DIF, 6.7% to 36.7% (of 30 items) measured academic skills, 3.3% to 20% (of 30 items) measured employability skills, and 5.0% to 26.7% (of 60 items) measured occupational/pathway skills. Accounting, Agribusiness, Business Management, Consumer Family Management, Financial Services, and Horticulture had larger percentages of *Gender* DIF items measuring these three skills. For items which displayed *Completer* DIF, 3.3% to 16.7% (of 30 items) measured academic skills and 3.3% to 10% (30 items) measured employability skills. Accounting and Business Management had larger percentages of *Completer* DIF items measuring these three skills. For items displaying DIF associated with economically disadvantaged status (participation in free/reduced *Lunch* program), 3.3% to 23.3% (of 30 items) measured academic skills, 3.3% to 26.7% (of 30 items) measured employability skills, and 3.3% to 11.7% (of 60 items) measured occupational/pathway skills. Agribusiness and Computer Programming had larger percentages of items classified as DIF items associated with participation in the free/reduced *Lunch* program. 13.3% of (30) items in Animal Science and 30% of (30) items in Culinary Food Services displayed DIF associated with disability status (*SWD*) measured academic skills. For items that displayed DIF associated with work-based learning (*WBL*), 3.3% to 23.3% (of 30 items) measured academic skills, 3.3% to 20% (of 30 items) measured employability skills, and 1.7% to 5.0% (of 60 items) measured occupational/pathway skills. Horticulture and Business Management had larger percentages of items with *WBL* DIF effects.

Table 4. Percentages of 2018-2019 Items categorized by Skill demonstrating Moderate to Large levels of Differential Item Functioning DIF

CTE Pathway	Skill	Number of Items	Gender n	%	Completer n	%	Lunch n	%	SWD n	%	WBL n	%	Credit n	%
Accounting	A	30	9	30.0	5	16.7	3	10.0	**	**	**	**	**	**
	E	30	5	16.7	1	3.3	3	10.0	**	**	**	**	**	**
	P	60	3	5.0	2	3.3	2	3.3	**	**	**	**	**	**
Admin Support	A	30	5	16.7	**	**	1	3.3	**	**	4	13.3	**	**
	E	30	1	3.3	**	**	**	**	**	**	**	**	**	**
	P	60	4	6.7	**	**	**	**	**	**	3	5.0	**	**
Ag Power Structured Tech	A	30	**	**	1	3.3	**	**	**	**	1	3.3	**	**
	P	60	**	**	**	**	1	1.7	**	**	3	5.0	**	**
Agribusiness	A	30	11	36.7	**	**	6	20.0	**	**	**	**	**	**
	E	30	5	16.7	**	**	8	26.7	**	**	**	**	**	**
	P	60	16	26.7	**	**	3	5.0	**	**	**	**	**	**
Animal Science	A	30	7	23.3	**	**	1	3.3	4	13.3	1	3.3	**	**
	E	30	2	6.7	1	3.3	**	**	1	3.3	3	10.0	**	**
	P	60	7	11.7	1	1.7	**	**	2	3.3	2	3.3	**	**
Business Mgmt.	A	30	7	23.3	**	**	2	6.7	**	**	7	23.3	**	**
	E	30	5	16.7	3	10.0	1	3.3	**	**	6	20.0	1	3.3
	P	60	6	10.0	**	**	**	**	**	**	1	1.7	**	**
Computer Programming	A	30	**	**	**	**	7	23.3	**	**	**	**	**	**
	E	30	**	**	**	**	3	10.0	**	**	**	**	**	**
	P	60	**	**	**	**	5	8.3	**	**	**	**	**	**

Table 4. (Continued)

CTE Pathway	Skill	Number of Items	Gender		Completer		Lunch		SWD		WBL		Credit	
			n	%	n	%	n	%	n	%	n	%	n	%
Consumer Family Mgmt.	A	30	2	6.7	2	6.7	1	3.3	**	**	**	**	**	**
	E	30	1	3.3	**	**	**	**	**	**	**	**	**	**
	P	60	11	18.3	**	**	1	1.7	**	**	**	**	**	**
Culinary Food Services	A	30	3	10.0	1	3.3	1	3.3	9	30.0	**	**	**	**
	E	30	**	**	**	**	**	**	1	3.3	**	**	**	**
	P	60	5	8.3	1	1.7	1	1.7	5	8.3	**	**	**	**
Early Childhood	A	30	**	**	2	6.7	1	3.3	**	**	**	**	**	**
	E	30	**	**	1	3.3	1	3.3	**	**	**	**	**	**
	P	60	**	**	1	1.7	2	3.3	**	**	**	**	**	**
Financial Services	A	30	6	20.0	**	**	2	6.7	**	**	**	**	**	**
	E	30	6	20.0	**	**	1	3.3	**	**	**	**	**	**
	P	60	9	15.0	**	**	7	11.7	**	**	**	**	**	**
Horticulture	A	30	9	30.0	**	**	2	6.7	**	**	4	13.3	**	**
	E	30	5	16.7	2	6.7	**	**	**	**	4	13.3	**	**
	P	60	9	15.0	**	**	1	1.7	**	**	**	**	**	**
Marketing	A	30	7	23.3	**	**	1	3.3	**	**	2	6.7	**	**
	E	30	4	13.3	1	3.3	**	**	**	**	1	3.3	**	**
	P	60	4	6.7	**	**	**	**	**	**	2	3.3	**	**

Note. A = academic; E = employability; P = occupational/pathway; SWD = disability status; WBL = work-based learning status.

**=Not available due to a small sample size of one or more groups. Sample sizes were too small to evaluate DIF by racial/ethnic groups or limited English proficiency (LEP) status. Sample sizes were too small to evaluate DIF for any groups for Agribiotechnology, Cinematography, Digital Design, Environmental Science, Fashion Interior Design, Food Science Processing, Graphic Design, Hospitality Travel Tourism, Information Support Services, Network Administration, Production Crop, Retail Services, and Web Development.

Table 5 displays percentages of OP items in 2021-2022 assessments which demonstrate moderate to large levels of DIF (B or C DIF). We did not conduct DIF analyses based on English proficiency (*LEP*) status or disability (*SWD*) status because of small group sizes. For the other grouping variables (*Gender*, *Race*, *Completer*, *Lunch*, *WBL*, and *Credit*), we conducted DIF analyses for OP items only because group sample sizes for FT items did not meet the minimum sample size criterion. We indicated whether group sizes for individual items met the minimum sample size criterion for these DIF analyses in our comprehensive sets of results. 14 of 31 assessments included OP items with *Gender* DIF effects. Environmental Science has the most items (16.7%) with *Gender* DIF effects, followed by Computer Programming and Graphic Design (each with 13.1%). Additionally, 1.7% of items in Business Management and 1.5% of items in Consumer Family Services were identified with *Race* DIF effects. Moreover, 5.1% of items in Computer Programming displayed *Completer* DIF effects. Regarding DIF items associated with economically disadvantaged status (participation in the free/reduced *Lunch* program), Automation Engineering and Mechanical Engineering had the most DIF items (each with 9.5%), followed by Financial Services (9.0%) and Cinematography (8.5%). Lastly, 2.3% of items in Consumer Family Services and Early Childhood displayed *WBL* DIF effects. Agribusiness and Marketing each had an item that displayed DIF associated with the number of credit hours taken (*Credit*).

Table 5. Percentage of 2021-2022 Operational Items demonstrating Moderate to Large Levels of Differential Item Functioning DIF

CTE Pathway	* Item Type	Number of items	Gender n	Gender %	Race n	Race %	Completer n	Completer %	Lunch n	Lunch %	WBL n	WBL %	Credit n	Credit %
Accounting	OP	131	11	8.4	**	**	**	**	4	3.1	**	**	**	**
Admin Support	OP	164	7	4.3	**	**	**	**	**	**	1	0.6	**	**
Ag Power Structured Tech	OP	131	**	**	**	**	1	0.8	2	1.5	**	**	**	**
Agribusiness	OP	127	11	8.7	**	**	**	**	2	1.6	**	**	1	0.8
Animal Science	OP	131	10	7.6	**	**	**	**	**	**	2	1.5	**	**
Automation Engineering	OP	126	**	**	**	**	**	**	12	9.5	**	**	**	**
Business Management	OP	172	9	5.2	3	1.7	**	**	2	1.2	1	0.6	**	**
Cinematography	OP	129	11	8.5	**	**	**	**	11	8.5	**	**	**	**
Civil Engineering	OP	126	**	**	**	**	**	**	7	5.6	**	**	**	**
Computer Programming	OP	137	18	13.1	**	**	7	5.1	3	2.2	**	**	**	**
Consumer Family Services	OP	132	14	10.6	2	1.5	**	**	**	**	3	2.3	**	**
Culinary Food Services	OP	132	8	6.1	**	**	1	0.8	**	**	1	0.8	**	**
Digital Design	OP	138	**	**	**	**	**	**	8	5.8	**	**	**	**
Early Childhood	OP	132	**	**	**	**	3	2.3	2	1.5	3	2.3	**	**
Environmental Science	OP	126	21	16.7	**	**	**	**	5	4	**	**	**	**
Financial Services	OP	134	16	11.9	**	**	**	**	12	9	**	**	**	**
Graphic Design	OP	130	17	13.1	**	**	**	**	8	6.2	**	**	**	**
Horticulture	OP	130	11	8.5	**	**	**	**	**	**	**	**	**	**
Marketing	OP	137	6	4.4	**	**	**	**	**	**	**	**	1	0.7
Mechanical Engineering	OP	126	**	**	**	**	**	**	12	9.5	**	**	**	**

Note. OP = operational; WBL = work-based learning.

*= Sample sizes were too small to evaluate DIF for FT items for the indicated assessments in the table.

**= Not available due to a small sample size of one or more group. Sample sizes were too small to evaluate DIF by limited English proficiency (LEP) and disability (SWD) status. Sample sizes were too small to evaluate DIF for all groups for Aerospace Engineering, Electrical Electronics Engineering, Engineering Design, Fashion Interior Design, Food Science Processing, Hospitality Travel Tourism, Information Support Services, Network Administration, Production Crop, Retail Services, and Web Development.

Table 6 displays percentages of items in 2021-2022 assessments, categorized by skill type (academic, employability, and occupational/pathway), which demonstrated moderate to large levels of DIF (B or C DIF). For items that displayed *Gender* DIF, 6.1% to 18.2% measured academic skills, 3.0% to 18.2% measured employability skills, and 2.0% to 18.5% measured occupational/pathway skills. Cinematography, Computer Programming, Consumer Family Services, Environmental Science, Financial Services, and Graphic Design Horticulture had larger percentages of *Gender* DIF items measuring these skills. Business Management and Consumer Family Management each contained a single item that displayed *Race* DIF effects. For items with *Completer* DIF effects, Ag Power Structured Tech contained a single item that measures academic skills, Early Culinary Food Services contained a single item that measures occupational/pathway skills, and Early Childhood contained three items that measure occupational/pathway skills, and Computer Programming contains three items that measure all three skills. For items displaying DIF associated with economically disadvantaged status (participation in the free/reduced *Lunch* program), 3.0% to 12.1% measured academic skills, 3.0% to 21.2% measured employability skills, and 1.4% to 7.9% measured occupational/pathway skills. Accounting, Automation Engineering, Cinematography, Civil Engineering, Digital Design, Financial Services, Graphic Design, and Mechanical Engineering Agribusiness and Computer Programming had larger percentages of items classified as DIF items associated with participation in the free/reduced *Lunch* program.

Admin Support, Animal Science, Business Management, and Consumer Family Services each contained a single item with *WBL* DIF effects that measure employability skills. Culinary Food Services and Early Childhood had one and three items with *WBL* DIF effects, respectively, that measure occupational/pathway skills. Agribusiness had a single item with *Credit* DIF effects that measure employability skills.

Table 6. Percentages of 2021-2022 Items categorized by Skill demonstrating Moderate to Large Levels of Differential Item functioning DIF

CTE Pathway	Skill	Number of OP Items*	Gender		Race		Completer		Lunch		WBL		Credit	
			n	%	n	%	n	%	n	%	n	%	n	%
Accounting	A	33	2	6.1	**	**	**	**	1	3.0	**	**	**	**
	E	33	3	9.1	**	**	**	**	3	9.1	**	**	**	**
	P	65	6	9.2	**	**	**	**	**	**	**	**	**	**
Admin Support	A	33	3	9.1	**	**	**	**	**	**	**	**	**	**
	E	33	2	6.1	**	**	**	**	**	**	1	3.0	**	**
	P	98	2	2.0	**	**	**	**	**	**	**	**	**	**
Ag Power Structured Tech	A	33	**	**	**	**	1	3.0	1	3.0	**	**	**	**
	E	33	**	**	**	**	**	**	1	3.0	**	**	**	**
Agribusiness	A	33	3	9.1	**	**	**	**	**	**	**	**	**	**
	E	33	4	12.1	**	**	**	**	**	**	**	**	1	3.0
	P	61	4	6.6	**	**	**	**	2	3.3	**	**	**	**
Animal Science	A	33	3	9.1	**	**	**	**	**	**	**	**	**	**
	E	33	2	6.1	**	**	**	**	**	**	2	6.1	**	**
	P	65	5	7.7	**	**	**	**	**	**	**	**	**	**
Automation Engineering	A	33	**	**	**	**	**	**	4	12.1	**	**	**	**
	E	33	**	**	**	**	**	**	4	12.1	**	**	**	**
	P	60	**	**	**	**	**	**	4	6.7	**	**	**	**
Business Management	A	33	3	9.1	1	3.0	**	**	2	6.1	**	**	**	**
	E	33	2	6.1	1	3.0	**	**	**	**	1	3.0	**	**

Table 6. (Continued)

CTE Pathway	Skill	Number of OP Items*	Gender		Race		Completer		Lunch		WBL		Credit	
			n	%	n	%	n	%	n	%	n	%	n	%
Cinematography	P	106	4	3.8	1	0.9	**	**	**	**	**	**	**	**
	A	33	1	3.0	**	**	**	**	2	6.1	**	**	**	**
	E	33	5	15.2	**	**	**	**	4	12.1	**	**	**	**
Civil Engineering	P	63	5	7.9	**	**	**	**	5	7.9	**	**	**	**
	A	33	**	**	**	**	**	**	3	9.1	**	**	**	**
	E	33	**	**	**	**	**	**	3	9.1	**	**	**	**
Computer Programming	P	60	**	**	**	**	**	**	1	1.7	**	**	**	**
	A	33	5	15.2	**	**	3	9.1	1	3.0	**	**	**	**
	E	33	7	21.2	**	**	3	9.1	1	3.0	**	**	**	**
Consumer Family Services	P	71	6	8.5	**	**	1	1.4	1	1.4	**	**	**	**
	A	33	6	18.2	1	3.0	**	**	**	**	**	**	**	**
	E	33	2	6.1	**	**	**	**	**	**	1	3.0	**	**
Culinary Food Services	P	66	6	9.1	1	1.5	**	**	**	**	2	3.0	**	**
	A	33	3	9.1	**	**	**	**	**	**	**	**	**	**
	E	33	2	6.1	**	**	**	**	**	**	**	**	**	**
Digital Design	P	66	3	4.5	**	**	1	1.5	**	**	1	1.5	**	**
	A	33	**	**	**	**	**	**	3	9.1	**	**	**	**
	E	33	**	**	**	**	**	**	2	6.1	**	**	**	**
	P	72	**	**	**	**	**	**	3	4.2	**	**	**	**

Table 6. (Continued)

CTE Pathway	Skill	Number of OP Items*	Gender		Race		Completer		Lunch		WBL		Credit	
			n	%	n	%	n	%	n	%	n	%	n	%
Early Childhood	A	33	**	**	**	**	**	**	1	3.0	**	**	**	**
	P	66	**	**	**	**	3	4.5	1	1.5	3	4.5	**	**
Environmental Science	A	33	5	15.2	**	**	**	**	1	3.0	**	**	**	**
	E	33	4	12.1	**	**	**	**	2	6.1	**	**	**	**
	P	65	12	18.5	**	**	**	**	2	3.1	**	**	**	**
Financial Services	A	33	4	12.1	**	**	**	**	4	12.1	**	**	**	**
	E	33	6	18.2	**	**	**	**	5	15.2	**	**	**	**
	P	68	6	8.8	**	**	**	**	3	4.4	**	**	**	**
Graphic Design	A	33	4	12.1	**	**	**	**	3	9.1	**	**	**	**
	E	33	5	15.2	**	**	**	**	3	9.1	**	**	**	**
	P	64	8	12.5	**	**	**	**	2	3.1	**	**	**	**
Horticulture	A	33	3	9.1	**	**	**	**	**	**	**	**	**	**
	E	33	1	3.0	**	**	**	**	**	**	**	**	**	**
	P	64	7	10.9	**	**	**	**	**	**	**	**	**	**
Marketing	A	33	3	9.1	**	**	**	**	**	**	**	**	**	**
	E	33	2	6.1	**	**	**	**	**	**	**	**	**	**
	P	71	1	1.4	**	**	**	**	**	**	**	**	**	**
Mechanical Engineering	A	33	**	**	**	**	**	**	2	6.1	**	**	**	**
	E	33	**	**	**	**	**	**	7	21.2	**	**	**	**
	P	60	**	**	**	**	**	**	3	5.0	**	**	**	**

Note. A = academic; E = employability; P = occupational/pathway; OP = operational; WBL = work-based learning.

*= Percent calculations based on number of operational items.

**= Not available due to a small sample size of one or more group. Sample sizes were too small to evaluate DIF by limited English proficiency (LEP) and disability (SWD) status. Sample sizes were too small to evaluate DIF for all groups for Aerospace Engineering, Electrical Electronics Engineering, Engineering Design, Fashion Interior Design, Food Science Processing, Hospitality Travel Tourism, Information Support Services, Network Administration, Production Crop, Retail Services, and Web Development. Sample sizes were too small to evaluate DIF for FT items for the indicated assessments in the table.

We also examined the number of times items were flagged for moderate or large DIF (B or C DIF) across assessments. Figures 1 and 2 display counts of items that were flagged at least twice in the 2018-2019 and 2021-2022 assessments, respectively. Each bar contains information on item type (OP or FT) and measured skill (academic and employability). Complete item descriptions including item prompts, along with assessments in which items were flagged, are presented in separate files (listed as *CTE_EOP_Summary_Tables_20182019_Total DIF.xlsx* and *CTE_EOP_Summary_Tables_20212022_Total DIF.xlsx* in Appendix B).

Figure 1. Total number of DIF flags for 2018-2019 items by item type and skill
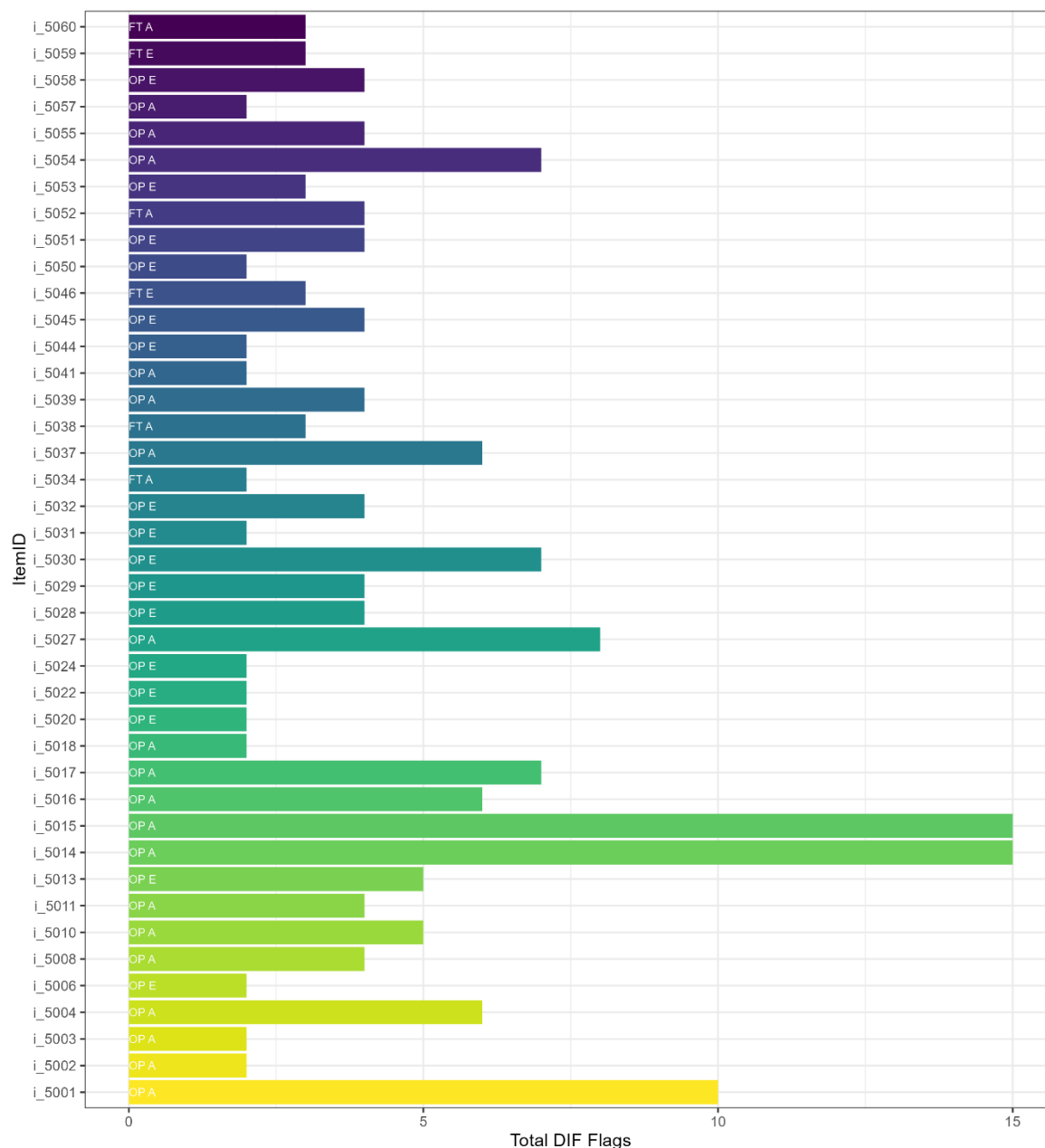
Figure 2. Total number of DIF flags for 2021-2022 items by item type and skill
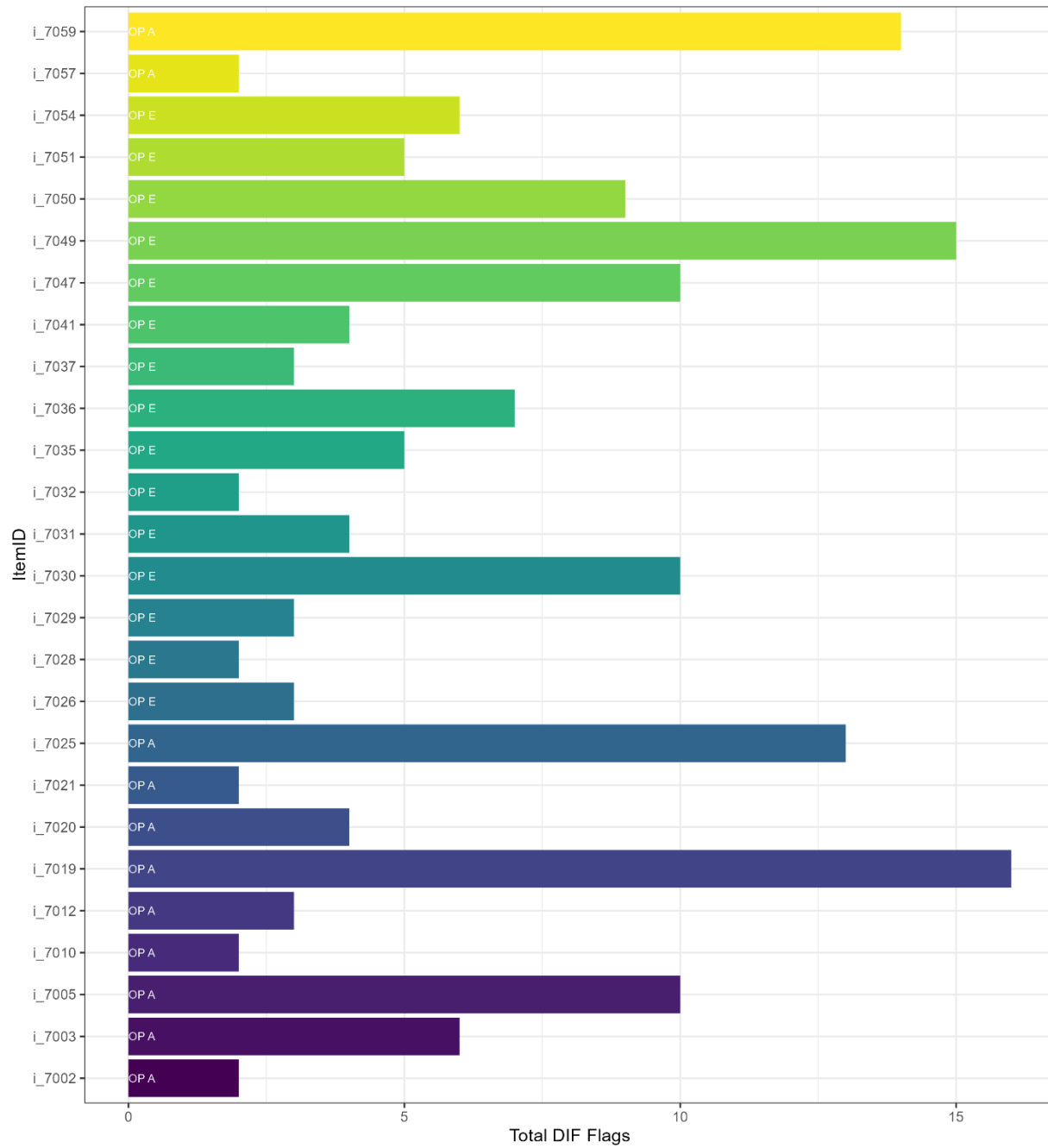
Brief Summary

When we analyzed assessments separately, we could not conduct several types of DIF analyses because of insufficient group sample sizes in certain grouping variables to make the results trustworthy. We were unable to conduct DIF analyses associated with English proficiency status for 2018-2019 and 2021-2022 assessments. We also were not able to conduct DIF analyses associated with race for 2018-2019 assessments and with disability status for 2021-2022 assessments. Our analyses were limited to only examining operational items in 2021-2022 assessments.

DIF analyses associated with gender, economically disadvantaged status, and work-based learning status were the most common analyses conducted across 2018-2019 assessments. 10% to 15% of field test items and 8% to 29% of operational items displayed DIF effects associated with gender. Additionally, 10% to 30% of field test items and 1% to 15% of operational items displayed DIF effects associated with economically disadvantaged status. Moreover, 10% to 25% of field test items and 4% to 9% of operational items displayed DIF effects associated with work-based learning status.

DIF analyses associated with gender, completer status, economically disadvantaged status, and work-based learning status were the most common analyses conducted across 2021-2022 assessments. 4.3% to 16.7% of operational items displayed DIF effects associated with gender. 0.8% to 5.1% of operational items displayed DIF effects associated with completer status. Additionally, 1.2% to 9.5% of operational items displayed DIF effects associated with economically disadvantaged status. Moreover, 0.6% to 2.3% of operational items displayed DIF effects associated with work-based learning status.

Like the findings from our CTT item analysis, DIF flags were inconsistently assigned across assessments to common academic (A) and employability (E) items. For example, a DIF flag based on gender may be assigned to the same A or E item in a particular assessment but may not be flagged in another. As with results from our CTT item analysis on individual assessments which are sample dependent, we recommend that the examination of these DIF results be contained within each assessment because results across assessments are not comparable. This limitation will be overcome in our *aggregated assessments* approach in which we combine students' responses on A and E items across assessments.

Chapter 4: CTT and DIF Analyses of Common Academic and Employability Items in CTE EOP Aggregated Assessments

Introduction

In addition to conducting CTT and DIF analyses on individual assessments (described in Chapters 2 and 3 of this report), we also aggregated students' responses on items that measure academic (A) and employability (E) skills across assessments because these items were administered across all assessments within each year. We analyzed these items together, treating aggregated students' responses on these items as if they were from a single assessment (the *aggregated assessment* approach). We did not include occupational/pathway (P) items in these analyses because these items were assessment specific.

Methods

We implemented the same processes for conducting CTT item analyses and DIF analyses that we employed on individual assessments to aggregated students' responses on A and E items. We present results for both analyses below.

Results

In this section, we describe observations and summary statistics from the CTT item analyses and DIF analyses conducted on responses to items that measure academic and employability skills aggregated across assessments. We also provided KDE with detailed results of these analyses in supplementary documents to this report (*CTE_EOP_1819_Aggregated_Results.xlsx* and *CTE_EOP_2122_Aggregated_Results.xlsx* listed in Appendix C). These documents contain flagged items based on the same CTT item analysis and DIF criteria we established for individual assessments.

Table 7 presents summary statistics based on CTT item analysis of 2018-2019 and 2021-2022 academic and employability items aggregated across assessments. For 2018-2019 A and E items, 10 of 50 (20%) OP items and 1 of 10 (10%) FT items were flagged because they did not meet item difficulty criteria. 6 of 50 (12%) OP items and 8 of 10 (80%) of items did not meet item discrimination criteria. For 2021-2022 A and E items, 18 of 66 (27.3%) OP items and 15 of 56 (26.8%) FT items were flagged because they did not meet item difficulty criteria. Additionally, 6 of 66 (9.1%) OP items and 25 of 56 (44.6%) items did not meet item discrimination criteria. We could not generate reliability coefficients for 2021-2022 FT items because of the sparseness of the data matrix for these items.

Table 7. Summary Statistics based on CTT Item Analysis of Academic and Employability Items in 2018-2019 and 2021-2022 Aggregated Assessments

Year	Item Type	Minimum Sample Size	Maximum Sample Size	CTT Item Difficulty	CTT Item Discrimination	CTT Reliability
2018-2019	OP	19788	19816	.491 - .927	-.082 - .496	.865 - .875
	FT	11829	11976	.399 - .945	.092 - .213	.935 - .978
2021-2022	OP	13891	28130	.282 - .920	.009 - .522	.853 - .861
	FT	4877	5794	.050 - .941	-.012 - .375	

Tables 8 and 9 detail flagged 2018-2019 and 2021-2022 items based on their CTT item statistics. We highlight (in the green shade in Tables 8 and 9) three items that have negative item discrimination indexes: for 2018 – 2019 aggregated assessments, item i_5042's (type = OP, skill = E) discrimination index was - .082; for 2021 – 2022 aggregated assessments, item i_7066 (type = FT, skill = E) discrimination index was -.092 and item i_7188's (type = FT, skill = E) discrimination index was -.016. These items cannot differentiate the performance of students with high ability from students with low ability. Additionally, the negative discrimination index indicates that the probability of correctly answering an item decreases as ability increases.

Table 8. Flagged 2018-2019 Items based on CTT Item Statistics

Item ID	Item Type	Skill	Sample Size	CTT Item Difficulty	CTT Item Discrimination	CTT Reliability	CTT Flag
i_5001	OP	A	19807	.896	.398	.866	**
i_5003	OP	A	19805	.857	.318	.867	**
i_5008	OP	A	19802	.927	.173	.869	*
i_5011	OP	A	19811	.868	.295	.868	**
i_5016	OP	A	19806	.878	.412	.866	**
i_5017	OP	A	19812	.917	.269	.868	**
i_5027	OP	A	19809	.891	.496	.865	**
i_5033	OP	A	19814	.546	.167	.870	*
i_5035	OP	A	19809	.600	.162	.870	*
i_5041	OP	A	19798	.631	.187	.870	*
i_5042	OP	E	19805	.500	-.082	.875	*
i_5044	OP	E	19808	.510	.194	.870	*
i_5051	OP	E	19812	.908	.454	.866	**
i_5053	OP	E	19790	.902	.382	.867	**
i_5054	OP	A	19805	.924	.466	.866	**
i_5057	OP	A	19796	.884	.371	.867	**
i_5034	FT	A	11921	.834	.154	.950	*
i_5038	FT	A	11976	.843	.125	.950	*
i_5040	FT	A	11840	.771	.191	.956	*
i_5046	FT	E	11968	.886	.203	.943	**
i_5048	FT	E	11875	.724	.155	.961	*
i_5049	FT	E	11928	.399	.092	.978	*
i_5052	FT	A	11954	.922	.183	.939	*
i_5059	FT	E	11898	.945	.197	.935	*
i_5060	FT	A	11829	.884	.150	.944	*

Note. A = academic; E = employability; P = occupational/pathway; OP = operational; FT = field test.

* = item discrimination, *pB* < .20

**= item difficulty, *p*-value < .15 or *p*-value > .85

Table 9. Flagged 2021-2022 Items based on CTT Item Statistics

Item ID	Item Type	Skill	Sample Size	CTT Item Difficulty	CTT Item Discrimination	CTT Reliability	CTT Flag
i_7003	OP	A	28089	.858	.482	.853	**
i_7005	OP	A	28092	.911	.494	.854	**
i_7006	OP	A	13921	.907	.429	.857	**
i_7011	OP	A	28087	.460	.165	.858	*
i_7016	OP	A	28095	.362	.009	.861	*
i_7023	OP	A	14022	.881	.373	.858	**
i_7025	OP	A	28081	.869	.439	.854	**
i_7030	OP	E	28124	.866	.451	.854	**
i_7038	OP	E	13949	.888	.459	.857	**
i_7039	OP	E	14014	.866	.409	.857	**
i_7040	OP	E	14037	.863	.404	.858	**
i_7046	OP	E	14017	.856	.522	.857	**
i_7049	OP	E	28086	.905	.479	.854	**
i_7050	OP	E	28093	.927	.487	.854	**
i_7054	OP	E	28087	.908	.470	.854	**
i_7055	OP	A	28093	.609	.165	.858	*
i_7062	OP	A	14074	.898	.362	.857	**
i_7063	OP	A	13961	.880	.439	.857	**
i_7067	OP	A	14011	.282	.081	.860	*
i_7082	OP	A	14178	.853	.377	.858	**
i_7091	OP	A	13903	.297	.185	.859	*
i_7179	OP	E	14025	.880	.340	.858	**
i_7200	OP	E	14098	.584	.181	.860	*
i_7202	OP	E	13891	.878	.516	.857	**
i_7015	FT	A	4939	.616	.103		*
i_7022	FT	A	5082	.494	.053		*
i_7034	FT	E	4960	.404	.120		*
i_7043	FT	E	5145	.870	.356		**
i_7044	FT	E	4942	.601	.094		*
i_7045	FT	E	5015	.876	.292		**
i_7064	FT	A	5050	.616	.195		*
i_7065	FT	A	4949	.319	.093		*
i_7066	FT	A	5184	.050	-.092		*
i_7070	FT	A	5127	.624	.173		*

Table 9. (Continued)

Item ID	Item Type	Skill	Sample Size	CTT Item Difficulty	CTT Item Discrimination	CTT Reliability	CTT Flag
i_7071	FT	A	5025	.618	.128		*
i_7072	FT	A	5082	.853	.319		**
i_7073	FT	A	4997	.151	.059		*
i_7074	FT	A	4908	.261	.058		*
i_7075	FT	A	5002	.505	.171		*
i_7076	FT	A	5153	.595	.193		*
i_7081	FT	A	5102	.922	.359		**
i_7084	FT	A	4998	.503	.074		*
i_7085	FT	A	5006	.892	.302		**
i_7086	FT	A	4918	.387	.174		*
i_7088	FT	A	5045	.506	.150		*
i_7089	FT	A	5086	.922	.335		**
i_7090	FT	A	4890	.302	.026		*
i_7180	FT	E	4914	.863	.269		**
i_7181	FT	E	4984	.752	.100		*
i_7182	FT	E	4927	.921	.322		**
i_7184	FT	E	5035	.902	.282		**
i_7185	FT	E	5048	.550	.150		*
i_7186	FT	E	4976	.879	.337		**
i_7187	FT	E	5029	.852	.273		**
i_7188	FT	E	5059	.250	-.012		*
i_7189	FT	E	5017	.547	.168		*
i_7191	FT	E	5040	.397	.150		*
i_7193	FT	E	5194	.594	.183		*
i_7195	FT	E	5089	.792	.152		*
i_7198	FT	E	4907	.921	.375		**
i_7203	FT	E	5120	.395	.088		*
i_7204	FT	E	5039	.863	.309		**
i_7205	FT	E	5050	.941	.318		**
i_7209	FT	E	5033	.871	.266		**

Note. A = academic; E = employability; P = occupational/pathway; OP = operational; FT = field test.

* = item discrimination, $pB < .20$

** = item difficulty, p-value $< .15$ or p-value $> .85$

In Table 10, we present results on academic and employability items in the 2018-2019 assessments that displayed DIF based on gender, race, number of credit hours taken, English language proficiency status, participation in free/reduced *Lunch* program, and disability status. We also present the total number of times each item was flagged for DIF. We also present results for DIF based on gender, race, English language proficiency status, and disability status, along with counts for items in the 2021-2022 assessments in Table 11.

Table 10. Type and Frequency of DIF displayed by Academic and Employability Items in 2018-2019 Aggregated Assessments.

Item ID	Item Type	Skill	Gender DIF	Race DIF	Credit DIF	LEP DIF	Lunch DIF	SWD DIF	Total DIF Flags
i_5001	OP	A	x						1
i_5003	OP	A				x			1
i_5004	OP	A	x						1
i_5005	OP	E		x		x			2
i_5007	OP	E		x					1
i_5011	OP	A		x					1
i_5012	OP	E		x					1
i_5013	OP	E		x		x			2
i_5014	OP	A	x	x		x		x	4
i_5015	OP	A	x	x		x	x	x	5
i_5016	OP	A						x	1
i_5017	OP	A				x			1
i_5018	OP	A		x					1
i_5019	OP	A		x					1
i_5022	OP	E		x					1
i_5023	OP	E		x					1
i_5029	OP	E		x					1
i_5030	OP	E		x		x			2
i_5033	OP	A		x		x			2
i_5035	OP	A		x					1
i_5036	OP	A		x	x				2
i_5037	OP	A		x					1
i_5039	OP	A		x					1
i_5042	OP	E		x					1
i_5044	OP	E		x					1
i_5053	OP	E		x					1
i_5055	OP	A		x					1
i_5056	OP	E		x					1

Table 11. Type and Frequency of DIF displayed by Academic and Employability Items in 2021-2022 Aggregated Assessments.

Item ID	Item Type	Skill	Gender DIF	Race DIF	LEP DIF	SWD DIF	Total DIF Flags
i_7003	OP	A	x				1
i_7005	OP	A		x	x		2
i_7010	OP	A		x			1
i_7016	OP	A		x			1
i_7017	OP	A		x			1
i_7019	OP	A	x	x	x	x	4
i_7021	OP	A		x			1
i_7025	OP	A	x				1
i_7026	OP	E		x	x		2
i_7027	OP	E		x			1
i_7030	OP	E		x	x		2
i_7032	OP	E		x			1
i_7035	OP	E		x			1
i_7036	OP	E		x			1
i_7037	OP	E		x			1
i_7041	OP	E		x			1
i_7047	OP	E		x			1
i_7050	OP	E			x		1
i_7055	OP	A		x			1
i_7057	OP	A		x	x		2
i_7058	OP	E		x			1
i_7059	OP	A	x				1

Brief Summary

Aggregating students' scores on common academic and employability items assisted us with overcoming sample size limitations that we encountered when we conducted DIF analyses for individual assessments. Because of overcoming this limitation, we were able to analyze items for DIF based on all demographic groups of interest, in particular for groups formed by students' race. Our current analyses revealed that most flagged items were flagged for DIF based on this demographic characteristic.

Chapter 5: Conclusions and Recommendations

This study separately analyzed sets of 2018-2019 and 2021-2022 CTE EOP assessments from each academic year based on the following reasoning. Assignment of test items to students differed across years. 2018-2019 assessments consisted of a single test form in which all could respond to all test items. Each 2021-2022 assessment, however, consisted of multiple test forms, so students responded to subsets of all possible items on an assessment. We treated missing responses on assessments with multiple test forms as missing completely at random (MCAR) based on assessment administration design. Because we did not apply any statistical methods for equating or linking scores across test forms for an assessment or linking the same assessment across years, we did not attempt to compare scores across test forms within and across assessments in this study because scores from these forms are not comparable. Rather, we treated each assessment administered within a given year holistically.

We undertook two main approaches for analyzing assessments administered in each academic year. In the first approach, labeled as the *individual assessments* approach throughout the study, we analyzed each assessment individually, treating item sets in each assessment as units for analysis. In the second approach, labeled as the *individual assessments* approach, we aggregated students' responses on A and E items across assessments and then analyzed them together, treating them as if they were administered in a single assessment.

We conducted two primary analyses for each of the *individual assessments* and *aggregated assessment* approaches. We first generated CTT item statistics to examine characteristics related to item difficulty and item discrimination. We also examined the impact of each item on the test's consistency in measuring the construct of interest by sequentially removing each item from the test item pool and re-calculating Cronbach's alpha as a measure of test reliability after each removal. We then examined whether items functioned differently for groups based on gender (*Gender*), race/ethnicity (*Race*), completer status (*Completer*), economically disadvantaged status (participation in the free/reduced *Lunch* program), English proficiency status (*LEP*), work-based learning status (*WBL*), disability status (SWD), and the number of credit hours taken (*Credit*). We used the Mantel-Haenszel DIF method to classify items into A, B, and C categories using ETS DIF criteria. We described salient observations from the CTT and DIF analyses for each approach in separate chapters in this report. We also provided comprehensive results of these analyses in separate documents (listed in Appendices at the end of this report).

Recommendations on How to Use Study Results

We make some general recommendations below on using these results for decision-making regarding item quality on the CTE EOP assessments.

An important consideration when examining items based on their CTT item statistics and DIF statistics is that these statistics are dependent on student sample characteristics. Item difficulty, for example, is essentially the percentage of students that got an item correct. However, if the sample of students who took a set of items in an assessment are homogeneous on ability at either the lower or upper end of the ability distribution, then these items are going to be difficult for one sample of students but easy for another group. Similarly, calculating item discrimination and calculating MH DIF require that students be placed into strata based on ability. Again, depending on sample characteristics, items may be flagged based on their item discrimination indexes or may be flagged for DIF in one assessment but not in another because ability distributions of the samples to which the assessments were administered are not comparable.

This will lead to conflicting results when we evaluate item quality by comparing item statistics generated by analyzing assessments separately across assessments.

In contrast, aggregating students' responses across all assessments on common (academic and employability) items resulted in a larger pool of students of different ability levels from which meaningful evaluation of item quality based on CTT and DIF analyses can be conducted. Student subgroups based on ability levels were no longer isolated to particular assessments as a result of aggregating responses across assessments. As a result, observed conflicting results from analyzing assessments individually were no longer problematic. This approach also assisted us with overcoming challenges with obtaining minimum group sample sizes for DIF analyses when each assessment was analyzed individually. For example, we could proceed with conducting DIF analyses based on groups formed by students' race because group sample sizes met Mazor and colleagues' (1992) and Paek and Guo's (2011) recommendations for balanced and unbalanced sample designs. Our DIF results based on aggregating students' responses demonstrated that 23 of all 28 A and E items that were flagged for DIF in the 2018-2019 assessments (Table 10) and 18 of all 22 A and E items that were flagged for DIF in the 2021-2022 assessments (Table 11) were also flagged for DIF based on race. Apart from resolving issues related to sample dependency when we analyzed assessments separately, we would not have been able to make these observations regarding DIF based on race if we had only analyzed assessments using the former approach.

Based on the above discussion, we recommend the following decision logic when examining the results presented in this study to evaluate item quality and test quality. We do not intend our recommendations to be prescriptive, but rather to serve as a guide when navigating the volume of, sometimes seemingly conflicting, results generated from the *individual assessments* approach in this study.

We recommend examining results from aggregating students' responses across assessments (presented in Chapter 4 and in documents listed in Appendix C) as the first step. Information provided in Tables 8 and 9 on 2018-2019 and 2021-2022 A and E items that were flagged based on CTT item statistics can be helpful with conducting basic item review, including review of item prompts and distractors, and possible item mis-keying. Information provided in Tables 10 and 11 on 2018-2019 and 2021-2022 A and E items that were flagged based on DIF statistics can be helpful to content experts to examine reasons why different groups perform differently on these items, to suggest changes to item content, and/or to recommend possible removal of items from item banks. Of particular concern may be items flagged for DIF across multiple groups (e.g., items i_5014 and i_5015 in the 2018-2019 assessments, and item i_7019 in the 2021-2022 assessments). After completing this first step, we recommend using results from individual assessments (presented in Chapters 2 and 3 and in documents listed in Appendix B) to assist with the triangulation of observations made from results from the aggregated results for a specific assessment. We demonstrate this process using results from DIF analyses on individual and aggregated 2018-2019 assessments, presented in Figure 3 below and Table 16 in Appendix D.

The vertical axis in Figure 3 lists all academic and employability items flagged for DIF in individual or aggregated assessments. The rightmost panel (labeled *Flagged in Both Assessment Approaches*) consists of items that were flagged in both approaches so that when these items are combined with items in the leftmost panel (*Flagged in Individual Assessments only*) or in the middle panel (*Flagged in Aggregated Assessments only*), we obtain a subset of items that were flagged for DIF in each of the assessment approaches. We recommend beginning the review process with items flagged in the *aggregated assessments* approach (the subset of items formed by combining items in the middle and rightmost panels of the figure).

Table 10 provides information on the types of DIF each item displayed. Some of these items were also flagged in individual assessments, confirmed by the frequency of DIF flags for each item in Figure 1. After conducting this first set of reviews, we recommend reviewing items that were flagged in individual assessments only (the subset of items in the leftmost panel of Figure 3). We recommend a similar approach for reviewing 2021-2022 academic and employability items flagged for DIF in individual or aggregated assessments. DIF results for these items are presented in Figure 4 below and Table 17 in Appendix D.

Figure 3. Academic and Employability Items that were flagged for DIF in 2018-2019 Individual and Aggregated Assessments

Figure 4. Academic and Employability Items that were flagged for DIF in 2021-2022 Individual and Aggregated Assessments

Recommendations for Fairness in Testing

Test scores of students with the same ability on the measured construct should be comparable, regardless of background (e.g., gender, race, English proficiency). Issues of test fairness and equity among subpopulations of tested students arise when scores depend on student characteristics other than ability. Consequently, issues of score interpretation, which is associated with test validity, also arise. According to the *Standards for Educational and*

Psychological Testing (American Educational Research Association [AERA] et al., 2014), validity of test scores is regarded as one of the most important attributes of an assessment. Assessment scores must reflect the underlying construct of interest to be considered valid. For all subpopulations of students taking an assessment, examining whether test scores equitably reflect the proficiency on the construct that an assessment intends to measure is critical.

While the *Standards* emphasizes the role of DIF analyses in addressing fairness issues in testing, it notes that items are not necessarily unfair to certain student subpopulations when these groups perform differently on these items. The identification of DIF indicates the existence of a latent trait besides the one of primary interest (multidimensionality). Fairness can still be established subsequently if the detected secondary latent trait is intentionally related to the primary latent trait. In line with recommendations from the *Standards*, we recommend that items in the present study which display DIF undergo additional review by qualified stakeholders (e.g., subject matters experts, members of the comparison subpopulation) to determine if additional actions are necessary for these items. We also recommend that any items that are edited because of this additional scrutiny should be field tested before subsequent operational use.

Additional Recommendations

The Kentucky Department of Education (KDE) included additional operational and field tests, in the 2021-2022 assessments to build its item banks. These items were randomly assigned to students so that each student took a subset of 100 operational and 20 field items, of all available items. This randomization process resulted in the administration of numerous test forms per assessment. As a result, each assessment consisted of a sparse data matrix with students' responses, leading, in some instances, to group sample sizes that were too small to conduct DIF analyses. In assessments where DIF analyses were conducted, we indicated whether the group sample sizes for each item met the recommended minimum sample size criterion. We provide results for 2021-2022 operational and field test items for these assessments in separate documents. In general, group sample sizes for field test items in the 2021-2022 assessments did not meet the sample size criterion. We recommend caution in interpreting DIF results for items in which the sample size criterion was not met. The results presented in this report are based on only operational items in those assessments which met the minimum sample size criterion and for which DIF analyses were conducted. We also recommend limiting the number of generated test forms for each assessment to decrease data matrix sparseness and to increase number of responses per item necessary to meet sample size criterion for DIF analyses, while simultaneously meeting test specifications requirements and maximizing the number of fielded operational and field test items. Sets of common field test items administered across forms randomly assigned to students may be helpful with achieving these goals.

References

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

Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items.* Sage.

Ebel, R. L. & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Prentice Hall.

Feldt, L. S. (1993). The relationship between the distribution of item difficulties and test reliability. *Applied measurement in education*, *6*(1), 37-48.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel–Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129-145). Erlbaum.

Holland, P. W., & Wainer, H. (1993). *Differential Item Functioning.* Lawrence Erlbaum Associates.

Magis, D., Beland, S., Tuerlinckx, F., & De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods*, 42, 847-862.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the national cancer institute*, *22*(4), 719-748.

Mazor, K. M., Clauser, B. E., & Hambleton, R. K. (1992). The effect of sample size on the functioning of the Mantel-Haenszel statistic. *Educational and Psychological Measurement, 52*, 443-451.

Mehrens, W. A., & Lehmann, I. J. (1991). *Measurement and evaluation in education and psychology* (4th ed.). Holt, Rinehart & Winston.

Paek, I., & Guo, H. (2011). Accuracy of DIF estimates and power in unbalanced designs Using the Mantel–Haenszel DIF detection procedure. *Applied Psychological Measurement, 35*(7), 518–535. https://doi.org/10.1177/0146621611420559

Penfield, R. D. (2001). Assessing differential item functioning among multiple groups: A comparison of three Mantel-Haenszel procedures. *Applied Measurement in Education*, 14, 235-259. 10.1207/S15324818AME1403_3

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/

Somes, G. W. (1986). The generalized Mantel-Haenszel statistic. *The American Statistician, 40*, 106-108. doi: 10.2307/2684866

Willse, J. T. (2018). CTT: Classical test theory functions. R package version 2.3.3, https://CRAN.R-project.org/package=CTT

Appendix A

Supplementary files with comprehensive list of all items administered in each academic year:

- CTE_EOP_1819_Complete_Items_List.xlsx

- CTE_EOP_2122_Complete_Items_List.xlsx

Table 12. Sample Size and Number of Administered Items by Type and Skill for 2018-2019 CTE EOP Assessments

CTE Pathway	Sample Size	Item Type	Academic Skills	Employability Skills	Pathway Skills	Total
Accounting	1032	FT	5	5	10	20
	1032	OP	25	25	50	100
Admin Support	3547	FT	5	5	10	20
	3547	OP	25	25	50	100
Ag Power Structured Tech	1500	FT	15	15	20	50
	1500	OP	25	25	50	100
Agribiotechnology	30	FT	5	5	10	20
	30	OP	25	25	50	100
Agribusiness	506	FT	5	5	10	20
	506	OP	25	25	50	100
Animal Science	2541	FT	15	15	20	50
	2541	OP	25	25	50	100
Business Management	1844	FT	5	5	10	20
	1844	OP	25	25	50	100
Cinematography	396	FT	5	5	10	20
	396	OP	25	25	50	100
Computer Programming	719	FT	5	5	10	20
	719	OP	25	25	52	102
Consumer Family Management	1580	FT	15	15	20	50
	1580	OP	25	25	50	100
Culinary Food Services	1881	FT	15	15	20	50
	1881	OP	25	25	50	100
Digital Design	431	FT	5	5	10	20
	431	OP	26	25	51	102
Early Childhood	1472	FT	15	15	30	60
	1472	OP	25	25	50	100
Environmental Science	458	FT	15	15	20	50
	458	OP	25	25	50	100

Table 12. (Continued)

CTE Pathway	Sample Size	Item Type	Academic Skills	Employability Skills	Pathway Skills	Total
Fashion Interior Design	184	FT	5	5	10	20
	184	OP	25	25	50	100
Financial Services	592	FT	15	15	20	50
	592	OP	25	25	50	100
Food Science Processing	81	FT	5	5	10	20
	81	OP	25	25	50	100
Graphic Design	302	FT	5	5	10	20
	302	OP	25	25	50	100
Horticulture	1882	FT	15	15	20	50
	1882	OP	25	25	50	100
Hospitality Travel Tourism	69	FT	15	15	20	50
	69	OP	25	25	50	100
Information Support Services	303	FT	5	5	10	20
	303	OP	26	25	49	100
Marketing	2428	FT	15	15	20	50
	2428	OP	25	25	50	100
Network Administration	281	FT	5	5	10	20
	281	OP	26	25	49	100
Production Crop	30	FT	5	5	10	20
	30	OP	25	25	50	100
Retail Services	101	FT	5	5	10	20
	101	OP	25	25	50	100
Web Development	273	FT	5	5	10	20
	273	OP	25	25	50	100

Note. OP = operational; FT = field test.

Table 13. Sample Size and Number of Administered Items by Type and Skill for 2021-2022 CTE EOP Assessments

CTE Pathway	Sample Size	Item Type	Academic Skills	Employability Skills	Pathway Skills	Total
Accounting	801	FT	28	28	23	79
	801	OP	33	33	65	131
Admin Support	3277	FT	28	28	42	98
	3277	OP	33	33	98	164
Aerospace Engineering	391	FT	28	28	17	73
	391	OP	33	33	65	131
Ag Power Structured Tech	1678	FT	28	28	35	91
	1678	OP	33	33	65	131
Agribusiness	844	FT	28	28	30	86
	844	OP	33	33	61	127
Animal Science	3420	FT	28	28	40	96
	3420	OP	33	33	65	131
Automation Engineering	572	FT	28	28	32	88
	572	OP	33	33	60	126
Business Management	3387	FT	28	28	41	97
	3387	OP	33	33	106	172
Cinematography	812	FT	28	28	23	79
	812	OP	33	33	63	129
Civil Engineering	643	FT	28	28	16	72
	643	OP	33	33	60	126
Computer Programming	1555	FT	28	28	28	84
	1555	OP	33	33	71	137
Consumer Family Services	2789	FT	28	28	31	87
	2789	OP	33	33	66	132
Culinary Food Services	2327	FT	28	28	29	85
	2327	OP	33	33	66	132
Digital Design	510	FT	28	28	33	89
	510	OP	33	33	72	138
Early Childhood	1781	FT	28	28	36	92
	1781	OP	33	33	66	132
Electrical Electronics Engineer	371	FT	28	28	31	87
	371	OP	33	33	64	130
Engineering Design	544	FT	28	28	15	71
	544	OP	33	33	60	126

CTE Pathway	Sample Size	Item Type	Academic Skills	Employability Skills	Pathway Skills	Total
Environmental Science	753	FT	28	28	40	96
	753	OP	33	33	65	131
Fashion Interior Design	282	FT	28	28	20	76
	282	OP	33	33	64	130
Financial Services	450	FT	28	28	33	89
	450	OP	33	33	68	134
Food Science Processing	134	FT	28	28	47	103
	134	OP	33	33	60	126
Graphic Design	633	FT	28	28	49	105
	633	OP	33	33	64	130
Horticulture	2167	FT	28	28	41	97
	2167	OP	33	33	64	130
Hospitality Travel Tourism	364	FT	28	28	21	77
	364	OP	33	35	64	132
Information Support Services	283	FT	28	28	30	86
	283	OP	33	33	66	132
Marketing	3759	FT	28	28	21	77
	3759	OP	33	33	71	137
Mechanical Engineering	641	FT	28	28	33	89
	641	OP	33	33	60	126
Network Administration	403	FT	28	28	31	87
	403	OP	33	33	68	134
Production Crop	27	FT	28	28	17	73
	27	OP	33	33	58	124
Retail Services	134	FT	28	28	16	72
	134	OP	33	34	65	132
Web Development	300	FT	28	28	41	97
	300	OP	33	33	63	129

Note. OP = operational; FT = field test.

Table 14. Percentage of 2018-2019 Items Flagged for Classical Test Theory (CTT) Statistics

CTE Pathway	Item Type	Number of Items	Item Difficulty		Item Discrimination		Total	
			n	%	n	%	n	%
Accounting	FT	20	7	35	4	20	11	55
	OP	100	19	19	22	22	41	41
Admin Support	FT	20	4	20	6	30	10	50
	OP	100	17	17	16	16	33	33
Ag Power Structured Tech	FT	50	5	10	20	40	25	50
	OP	100	12	12	19	19	31	31
Agribiotechnology*	FT	20	5	25	5	25	10	50
	OP	100	15	15	32	32	47	47
Agribusiness	FT	20	3	15	4	20	7	35
	OP	100	10	10	11	11	21	21
Animal Science	FT	50	8	16	21	42	29	58
	OP	100	12	12	18	18	30	30
Business Mgmt.	FT	20	4	20	6	30	10	50
	OP	100	17	17	10	10	27	27
Cinematography	FT	20	6	30	7	35	13	65
	OP	100	19	19	19	19	38	38
Computer Programming	FT	20	11	55	6	30	17	85
	OP	102	23	22.6	20	19.6	43	42.2
Consumer Family Mgmt.	FT	50	8	16	16	32	24	48
	OP	100	17	17	7	7	24	24
Culinary Food Services	FT	50	11	22	22	44	33	66
	OP	100	14	14	12	12	26	26
Digital Design	FT	20	5	25	8	40	13	65
	OP	102	16	15.7	21	20.6	37	36.3
Early Childhood	FT	60	11	18.3	31	51.7	42	70
	OP	100	23	23	14	14	37	37
Environmental Science	FT	50	7	14	27	54	34	68
	OP	100	20	20	8	8	28	28
Fashion Interior Design	FT	20	5	25	5	25	10	50
	OP	100	19	19	21	21	40	40
Financial Services	FT	50	13	26	18	36	31	62
	OP	100	23	23	8	8	31	31

			Item Difficulty		Item Discrimination		Total	
Food Science Processing	FT	20	6	30	7	35	13	65
	OP	100	12	12	20	20	32	32
Graphic Design	FT	20	6	30	9	45	15	75
	OP	100	26	26	16	16	42	42
Horticulture	FT	50	4	8	25	50	29	58
	OP	100	10	10	17	17	27	27
Hospitality Travel Tourism	FT	50	6	12	18	36	24	48
	OP	100	13	13	26	26	39	39
Information Support Services	FT	20	4	20	12	60	16	80
	OP	100	17	17	29	29	46	46
Marketing	FT	50	9	18	21	42	30	60
	OP	100	11	11	10	10	21	21
Network Administration	FT	20	6	30	7	35	13	65
	OP	100	17	17	22	22	39	39
Production Crop*	FT	20	7	35	8	40	15	75
	OP	100	20	20	26	26	46	46
Retail Services	FT	20	6	30	4	20	10	50
	OP	100	22	22	18	18	40	40
Web Development	FT	20	7	35	4	20	11	55
	OP	100	19	19	30	30	49	49

Note. OP = operational; FT = field test.

* = Based on sample size of 30.

Table 15. Percentage of 2021-2022 Items Flagged for Classical Test Theory (CTT) Statistics

CTE Pathway	Item Type	N	Item Difficulty		Item Discrimination		Total	
			n	%	n	%	n	%
Accounting	FT	79	13	16.5	34	43.0	47	59.5
	OP	131	28	21.4	12	9.2	40	30.5
Admin Support	FT	98	21	21.4	36	36.7	57	58.2
	OP	164	35	21.3	22	13.4	57	34.8
Aerospace Engineering	FT	73	16	21.9	48	65.8	64	87.7
	OP	131	60	45.8	16	12.2	76	58.0
Ag Power Structured Tech	FT	91	7	7.7	40	44.0	47	51.6
	OP	131	4	3.1	22	16.8	26	19.8
Agribusiness	FT	86	12	14.0	34	39.5	46	53.5
	OP	127	16	12.6	22	17.3	38	29.9
Animal Science	FT	96	16	16.7	36	37.5	52	54.2
	OP	131	18	13.7	22	16.8	40	30.5
Automation Engineering	FT	88	18	20.5	38	43.2	56	63.6
	OP	126	40	31.7	10	7.9	50	39.7
Business Management	FT	97	17	17.5	35	36.1	52	53.6
	OP	172	28	16.3	25	14.5	53	30.8
Cinematography	FT	79	18	22.8	28	35.4	46	58.2
	OP	129	32	24.8	13	10.1	45	34.9
Civil Engineering	FT	72	19	26.4	32	44.4	51	70.8
	OP	126	49	38.9	15	11.9	64	50.8
Computer Programming	FT	84	24	28.6	28	33.3	52	61.9
	OP	137	51	37.2	18	13.1	69	50.4
Consumer Family Services	FT	87	17	19.5	28	32.2	45	51.7
	OP	132	31	23.5	11	8.3	42	31.8
Culinary Food Services	FT	85	14	16.5	36	42.4	50	58.8
	OP	132	29	22.0	16	12.1	45	34.1
Digital Design	FT	89	13	14.6	40	44.9	53	59.6
	OP	138	32	23.2	18	13.0	50	36.2
Early Childhood	FT	92	27	29.3	35	38.0	62	67.4
	OP	132	37	28.0	17	12.9	54	40.9
Electrical Electronics Engineer	FT	87	14	16.1	48	55.2	62	71.3
	OP	130	45	34.6	19	14.6	64	49.2

			Item Difficulty		Item Discrimination		Total	
Engineering Design	FT	71	14	19.7	30	42.3	44	62.0
	OP	126	19	15.1	7	5.6	26	20.6
Environmental Science	FT	96	15	15.6	50	52.1	65	67.7
	OP	131	17	13.0	16	12.2	33	25.2
Fashion Interior Design	FT	76	14	18.4	34	44.7	48	63.2
	OP	130	37	28.5	22	16.9	59	45.4
Financial Services	FT	89	15	16.9	37	41.6	52	58.4
	OP	134	23	17.2	11	8.2	34	25.4
Food Science Processing	FT	103	13	12.6	40	38.8	53	51.5
	OP	126	16	12.7	17	13.5	33	26.2
Graphic Design	FT	105	17	16.2	46	43.8	63	60.0
	OP	130	30	23.1	13	10.0	43	33.1
Horticulture	FT	97	7	7.2	52	53.6	59	60.8
	OP	130	9	6.9	26	20.0	35	26.9
Hospitality Travel Tourism	FT	77	8	10.4	31	40.3	39	50.6
	OP	132	11	8.3	20	15.2	31	23.5
Information Support Services	FT	86	10	11.6	45	52.3	55	64.0
	OP	132	26	19.7	26	19.7	52	39.4
Marketing	FT	77	14	18.2	28	36.4	42	54.5
	OP	137	25	18.2	10	7.3	35	25.5
Mechanical Engineering	FT	89	14	15.7	44	49.4	58	65.2
	OP	126	51	40.5	14	11.1	65	51.6
Network Administration	FT	87	16	18.4	41	47.1	57	65.5
	OP	134	27	20.1	19	14.2	46	34.3
Production Crop*	FT	73	7	9.6	57	78.1	64	87.7
	OP	124	22	17.7	64	51.6	86	69.4
Retail Services	FT	72	17	23.6	16	22.2	33	45.8
	OP	132	30	22.7	15	11.4	45	34.1
Web Development	FT	97	12	12.4	61	62.9	73	75.3
	OP	129	26	20.2	26	20.2	52	40.3

Note. OP = operational; FT = field test.

*=Based on sample size of 27.

The following are provided as separate files supplemental to this report.

Results for 2018-2019 individual assessments:

- CTE_EOP_20182019_Accounting_DIF.xlsx

- CTE_EOP_20182019_Admin Support_DIF.xlsx

- CTE_EOP_20182019_Ag Power Structured Tech_DIF.xlsx

- CTE_EOP_20182019_Agribiotechnology_DIF.xlsx

- CTE_EOP_20182019_Agribusiness_DIF.xlsx

- CTE_EOP_20182019_Animal Science_DIF.xlsx

- CTE_EOP_20182019_Business Management_DIF.xlsx

- CTE_EOP_20182019_Cinematography_DIF.xlsx

- CTE_EOP_20182019_Computer Programming_DIF.xlsx

- CTE_EOP_20182019_Consumer Family Management_DIF.xlsx

- CTE_EOP_20182019_Culinary Food Services_DIF.xlsx

- CTE_EOP_20182019_Digital Design_DIF.xlsx

- CTE_EOP_20182019_Early Childhood_DIF.xlsx

- CTE_EOP_20182019_Environmental Science_DIF.xlsx

- CTE_EOP_20182019_Fashion Interior Design_DIF.xlsx

- CTE_EOP_20182019_Financial Services_DIF.xlsx

- CTE_EOP_20182019_Food Science Processing_DIF.xlsx

- CTE_EOP_20182019_Graphic Design_DIF.xlsx

- CTE_EOP_20182019_Horticulture_DIF.xlsx

- CTE_EOP_20182019_Hospitality Travel Tourism_DIF.xlsx

- CTE_EOP_20182019_Information Support Services_DIF.xlsx

- CTE_EOP_20182019_Marketing_DIF.xlsx

- CTE_EOP_20182019_Network Administration_DIF.xlsx

- CTE_EOP_20182019_Production Crop_DIF.xlsx

- CTE_EOP_20182019_Retail Services_DIF.xlsx

- CTE_EOP_20182019_Web Development_DIF.xlsx

Results for 2021-2022 individual assessments:

- CTE_EOP_20212022_Accounting_DIF.xlsx

- CTE_EOP_20212022_Admin Support_DIF.xlsx

- CTE_EOP_20212022_Aerospace Engineering_DIF.xlsx

- CTE_EOP_20212022_Ag Power Structured Tech_DIF.xlsx

- CTE_EOP_20212022_Agribusiness_DIF.xlsx

- CTE_EOP_20212022_Animal Science_DIF.xlsx

- CTE_EOP_20212022_Automation Engineering_DIF.xlsx

- CTE_EOP_20212022_Business Management_DIF.xlsx

- CTE_EOP_20212022_Cinematography_DIF.xlsx

- CTE_EOP_20212022_Civil Engineering_DIF.xlsx

- CTE_EOP_20212022_Computer Programming_DIF.xlsx

- CTE_EOP_20212022_Consumer Family Services_DIF.xlsx

- CTE_EOP_20212022_Culinary Food Services_DIF.xlsx

- CTE_EOP_20212022_Digital Design_DIF.xlsx

- CTE_EOP_20212022_Early Childhood_DIF.xlsx

- CTE_EOP_20212022_Electrical Electronics Engineer_DIF.xlsx

- CTE_EOP_20212022_Engineering Design_DIF.xlsx

- CTE_EOP_20212022_Environmental Science_DIF.xlsx

- CTE_EOP_20212022_Fashion Interior Design_DIF.xlsx

- CTE_EOP_20212022_Financial Services_DIF.xlsx

- CTE_EOP_20212022_Food Science Processing_DIF.xlsx

- CTE_EOP_20212022_Graphic Design_DIF.xlsx

- CTE_EOP_20212022_Horticulture_DIF.xlsx

- CTE_EOP_20212022_Hospitality Travel Tourism_DIF.xlsx

- CTE_EOP_20212022_Information Support Services_DIF.xlsx

- CTE_EOP_20212022_Marketing_DIF.xlsx

- CTE_EOP_20212022_Mechanical Engineering_DIF.xlsx

- CTE_EOP_20212022_Network Administration_DIF.xlsx

- CTE_EOP_20212022_Production Crop_DIF.xlsx

- CTE_EOP_20212022_Retail Services_DIF.xlsx

- CTE_EOP_20212022_Web Development_DIF.xlsx

Frequencies of Flagged DIF Items

- *CTE_EOP_Summary_Tables_20182019_Total DIF.xlsx*

- *CTE_EOP_Summary_Tables_20212022_Total DIF.xlsx*

Appendix C

The following are provided as separate files supplemental to this report.

- CTE_EOP_1819_Aggregated_Results.xlsx

- CTE_EOP_2122_Aggregated_Results.xlsx

Appendix D

Table 16. List of Academic and Employability Items that were flagged for DIF in 2018-2019 Individual Assessments only, Aggregated Assessments only, and on both Types of Assessments

Item ID	Item Type	Skill	DIF Flag for Individual Assessments only	DIF Flag for Aggregated Assessments only	DIF Flag for Both Assessment Approaches
i_5001	OP	A			x
i_5002	OP	A	x		
i_5003	OP	A			x
i_5004	OP	A			x
i_5005	OP	E			x
i_5006	OP	E	x		
i_5007	OP	E		x	
i_5008	OP	A	x		
i_5009	OP	E	x		
i_5010	OP	A	x		
i_5011	OP	A			x
i_5012	OP	E			x
i_5013	OP	E			x
i_5014	OP	A			x
i_5015	OP	A			x
i_5016	OP	A			x
i_5017	OP	A			x
i_5018	OP	A			x
i_5019	OP	A		x	
i_5020	OP	E	x		
i_5021	OP	E	x		
i_5022	OP	E			x
i_5023	OP	E		x	
i_5024	OP	E	x		
i_5025	OP	A	x		
i_5026	OP	A	x		
i_5027	OP	A	x		
i_5028	OP	E	x		
i_5029	OP	E			x
i_5030	OP	E			x
i_5031	OP	E	x		

Table 16. (Continued)

Item ID	Item Type	Skill	DIF Flag for Individual Assessments only	DIF Flag for Aggregated Assessments only	DIF Flag for Both Assessment Approaches
i_5032	OP	E	x		
i_5033	OP	A		x	
i_5034	FT	A	x		
i_5035	OP	A		x	
i_5036	OP	A			x
i_5037	OP	A			x
i_5038	FT	A	x		
i_5039	OP	A			x
i_5041	OP	A	x		
i_5042	OP	E			x
i_5043	OP	E	x		
i_5044	OP	E			x
i_5045	OP	E	x		
i_5046	FT	E	x		
i_5047	FT	E	x		
i_5050	OP	E	x		
i_5051	OP	E	x		
i_5052	FT	A	x		
i_5053	OP	E			x
i_5054	OP	A	x		
i_5055	OP	A			x
i_5056	OP	E			x
i_5057	OP	A	x		
i_5058	OP	E	x		
i_5059	FT	E	x		
i_5060	FT	A	x		

Table 17. List of Academic and Employability Items that were flagged for DIF in 2021-2022 Individual Assessments only, Aggregated Assessments only, and on both Types of Assessments

Item ID	Item Type	Skill	DIF Flag for Individual Assessments only	DIF Flag for Aggregated Assessments only	DIF Flag for Both Assessment Approaches
i_7002	OP	A	x		
i_7003	OP	A			x
i_7005	OP	A			x
i_7008	OP	A	x		
i_7010	OP	A			x
i_7012	OP	A	x		
i_7016	OP	A		x	
i_7017	OP	A		x	
i_7019	OP	A			x
i_7020	OP	A	x		
i_7021	OP	A			x
i_7025	OP	A			x
i_7026	OP	E			x
i_7027	OP	E			x
i_7028	OP	E	x		
i_7029	OP	E	x		
i_7030	OP	E			x
i_7031	OP	E	x		
i_7032	OP	E			x
i_7035	OP	E			x
i_7036	OP	E			x
i_7037	OP	E			x
i_7041	OP	E			x
i_7047	OP	E			x
i_7049	OP	E	x		
i_7050	OP	E			x
i_7051	OP	E	x		
i_7052	OP	E	x		
i_7054	OP	E	x		
i_7055	OP	A		x	
i_7057	OP	A			x
i_7058	OP	E		x	
i_7059	OP	A			x