Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 26 Number 25, December 2021 ISSN 1531-7714

Embedded Accommodation and Accessibility Support Usage on a Computer-Based Statewide Achievement Test

Dukjae Lee, *University of Massachusetts Amherst*
Heather Buzick, *Educational Testing Service*
Stephen G. Sireci, *University of Massachusetts Amherst*
Mina Lee, *University of Massachusetts Amherst*
Cara Laitusis, *Educational Testing Service*

Although there has been substantial research on the effects of test accommodations on students' performance, there has been far less research on students' *use* of embedded accommodations and other accessibility supports at the item and whole test level in operational testing programs. Data on embedded accessibility supports from digital logs generated by computer-based assessment platforms are complex, and so decisions need to be made to make sense of the data with respect to appropriate and effective accommodation use. In this study, we explored different ways of defining students' use of accessibility supports and how to best summarize such use for accountability and other purposes. Examples of descriptive statistical indices and data visualizations are presented using mathematics and English language arts test data from a large statewide assessment. Such data are important for accommodations monitoring required by the United States Department of Education and for identifying schools and districts that may be over- or under-using these accommodations and supports.

Introduction

To promote fairness and access in educational assessment and valid interpretations of test scores, testing accommodations are typically provided to the students with disabilities (SWD) and English learners[1] (ELs) who need them. Testing accommodations are intended "to remove construct irrelevant barriers that otherwise would interfere with examinees' ability to demonstrate their standing on the target constructs" (American Education Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME], 2014, p. 67). For example, providing text-to-speech for reading directions aloud is expected to help some students with low decoding skills and those whose first language is not English to better understand what they are being asked to do on test items. Testing accommodations require approval prior to testing as a way to help ensure their appropriateness and minimize threats to validity.

The advent of computer-based assessments has opened up opportunities to provide some accommodations digitally and to integrate new digital supports that could expand both access to the test content and the population who uses the supports.

[1] Some researchers claim the term "English learners" (ELs) is deficit-minded and suggest using the term "emergent bilinguals" instead (García, 2011). We acknowledge this claim, but use the term ELs here because the specific context refers to students learning English, and we do not consider learning to be a deficit. That said, we also consider bilingualism and multilingualism to be important assets for all individuals.

Following principles of universal design for assessment (Thompson, Johnstone, & Thurlow, 2002), making appropriate supports available to all students who need them during the assessment is intended to make assessment more inclusive and fair, which in turn is expected to improve the validity of test-score based inferences (see also AERA et al., 2014). In K-12 education, many states have adopted a three-tiered approach to accessibility for their standardized accountability assessments to help educators and education teams make decisions about the appropriateness of specific accessibility supports for individual students (Shyyan et al., 2016). The Smarter Balanced Assessment Consortium (2014) defined these tiers as (a) universal tools, which are access features available to all students based on their preferences; (b) designated supports, which are features "available for use by any student for whom the need has been indicated by an educator or team of educators;" and (c) accommodations, which are "are changes in procedures or materials that increase equitable access during the...assessments by generating valid assessment results for students who need them and allowing these students to show what they know and can do" (p. 2).

Thus, on statewide assessments, some digital supports are available to any student who has a need identified prior to testing. Examples, from both Smarter Balanced and the Partnership for Accessing Readiness for College and Career (PARCC), include color contrast and text-to-speech on mathematics items (PARCC, 2017; Smarter Balanced Assessment Consortium, 2014). The rationale for expanding the use of accessibility supports to all students is the assumption that use of the tools, when familiar to students, does not interfere with the construct being measured and can help improve accessibility, as opposed to accommodations that may interfere with the construct measured for students who are not approved to use them. Where prior approval is required, the rationale is students should be familiar with the support through classroom use prior to testing. Other digital supports are available to all students without prior approval. Examples include highlighter, zoom, and English glossary. The rationale for providing these supports to all students without approval is the assumption that such supports are already familiar to students or are so straightforward to use that they improve accessibility without altering the construct intended to be measured by the assessment (Sireci & O'Riordan, 2020).

This tiered approach to accessibility focuses on who is eligible for particular accessibility features that make a test more accessible. This focus is described in technical documents with audiences concerned with assessment policy and the practice of assessment development and delivery (e.g., PARCC, 2017, Smarter Balanced, 2016). The expansion to designated supports and universal features is aligned with the perspective that flexibility in assessment can result in more fair assessment and more valid interpretations from test scores (e.g., Sireci, 2020).

Empirical support for accommodations policies and decisions had been limited when paper-based testing was the primary mode of administration. More widespread use of computer-based assessment has led to the accumulation of data that can be analyzed for the purpose of evaluating policies and decisions and collecting evidence to support the validity of test score interpretations when digital accessibility supports are offered. The data captured from computer-based assessments are complex, and there are a number of decisions that need to be made to make sense of the data to judge the appropriateness and effectiveness of accommodations and other supports (Abedi & Ewers, 2013). A particular concern is the degree to which the intended students are effectively using accommodations and designated supports. In fact, this concern led the U.S. Department of Education (USDE) to require states to monitor and report on testing accommodations used by SWD and ELs (USDE, 2018).

In the current study, we analyze data from statewide computerized-adaptive English Language Arts (ELA) and mathematics assessments administered in a large state to explore students' usage of embedded accessibility supports. Our focus is on the best means for summarizing usage data, obtained from digital logs of students' interactions with the assessment platform. We analyzed usage data at the item and test levels to provide guidance to researchers and policy-makers on how to define accommodation or digital support "use," and how to best visualize usage data for specific purposes such as evaluating appropriate use and providing guidance to schools and districts. We first provide some background on digital supports and accommodations and then describe our method and results. We end with a discussion of the implications of these results for future research and practice for improving educational assessments for all students.

Background

There is a large body of research on the effectiveness of the use of some testing accommodations, such as oral presentation and extended time accommodations (e.g., Buzick & Stone, 2014; Lovett; 2010), and how they may affect score comparability (e.g., Abedi & Ewers, 2013; Sireci, Banda, & Wells, 2018; Sireci & O'Riordan, 2020), but much less is known about the degree to which test accommodations and designated supports are appropriately provided to and used by students on large-scale, state standardized assessments (Crotts-Roohr & Sireci, 2017). That is because, in paper and pencil tests and early computer-based tests, data were not captured to describe how students used accessibility supports throughout the test. Until recently, most existing datasets included only information about whether individual students were permitted access to a particular testing accommodation on the test, or on a section (e.g., mathematics, ELA). With an increase in digital delivery of assessments, data from operational testing have become available on individual students' use of embedded accommodations and other digital accessibility supports at the item and test level. These new data, in the form of log files that capture students' inputs such as clicks and timing data, allow us to investigate whether embedded accessibility supports are actually used by students, and if so, to what extent.

Knowing whether testing accommodations, designated features, and universal features are being used by students who are eligible to use them can advance research on their appropriateness and effectiveness. For example, if it is found students are not using the accommodations we think are helpful, it is likely the accommodation would need to be redesigned. Finer grained information about the extent of use can also improve federal reporting requirements, outlined in the Every Student Succeeds Act (ESSA), which requires states to administer statewide annual assessments to their K-12 public school students and that,

> States must assess all students, including by offering appropriate accommodations for English learners and children with disabilities, and, to the extent practicable, must develop assessments using the principles of universal design for learning, which intentionally reduce barriers and improve flexibility in how students receive information or demonstrate knowledge. (USDE, 2017, p. 2)

The USDE also requires states "to monitor and report on testing accommodations used by special education students and Els and accessibility features available to students more generally" (Critical elements 5.1 and 5.4, USDE, 2018). The primary purpose of this monitoring is to ensure students who receive special education services are provided with appropriate accommodations as required under the Individuals with Disabilities Education Act (IDEA, 2004). A secondary goal is to ensure accommodations are not over- or under-used by some local school districts or subgroups of students, which would threaten the validity of inferences drawn from the test scores about students' knowledge, skills, and abilities. The monitoring has been done through a labor-intensive process of having trained observers travel to schools to observe randomly selected students (or classrooms) on test day and record if the student received any accommodations. Given this level of effort, most states have been cited as lacking evidence to support this monitoring requirement. In fact, two large states (Texas and California) were recently instructed to improve their monitoring.

Introduction to the Present Study

The raw data captured in digital logs of the assessment interface can help monitor and evaluate student use of embedded accessibility supports. However, these data are complex. In this study, we present different ways of analyzing summaries of log file data to display accommodation, designated support, and universal feature usage data at both the student and district levels. After we reviewed the literature and explored the data, it became evident there are multiple ways to define "use" of an embedded accessibility support on an assessment given digital log data on the number of times a button representing a particular accessibility support is clicked in the digital assessment interface. Based on these experiences, in this study we address the following research questions:

1. What is the best way to empirically define "use" of a digital accessibility support?
2. How much are accessibility supports used across students, items, and subgroups?
3. What types of supports are most often used by different subgroups of students?
4. What is the relationship between eligibility to use accommodations and designated supports and use of those supports?

5. How can we use eligibility and usage information to identify districts that may be overusing or underusing approved accessibility supports?

Method

Data

The dataset for this study is from 6th-grade mathematics and ELA computer-based assessments from a large state, administered in spring 2018. The assessments were derived from the Smarter Balanced assessment program (Smarter Balanced Assessment Consortium, 2017). The tests in each subject area were computerized-adaptive tests (CATs), which means students took different sets of items, depending on how well they performed on earlier items. That is, in computerized-adaptative testing, if a student answers an item incorrectly, they are likely to be given an easier item and vice-versa (see Wainer, 1993 for an overview of CAT). The tests were administered without a time limit.

State assessment administrative data were merged with summary data on students' use of embedded accessibility supports (referred to as "supports" for brevity) derived from digital logs of test takers' interactions with the computer-based assessment interface. Variables available from the administrative dataset included unique student ID number, school district, student demographic information, IEP, 504 plan, English learner status, whether they were approved to use accommodations or designated supports, and types of accommodations and designated supports approved for each student to use on the assessment. Only usage data were available to the researchers—item and total test score data for students were not available. All tested students were included in the administrative dataset. Because SWD and ELs are often approved for accommodations and designated supports, we analyzed the data stratified by these student characteristics. General education students were also included because some were eligible to use designated supports, and the universal features were available to all students.

The summary data from the digital logs included the number of times a student used a support on an item, item ID, page (screen) number on which the support was used, test name, and whether the support was used on a stimulus (e.g., a reading passage) or an item (e.g., the question or answer options). Descriptions of the supports and who was eligible to use them are in Table 1. There were five supports in the database: highlighter, line reader, masking, text-to-speech (TTS), and print. TTS was of two types: students could choose to have the entire entity read aloud (TTS-entire, e.g., an entire question, including every answer choice) or only a selection (TTS-selection, e.g., a word or sentence). These supports represent a subset of the embedded accessibility supports that were available to students and include all those the digital platform owners deemed as having sufficient data quality for research. The supports cover the range of accessibility support categories, namely, universal features, designated supports, and accommodations. The percentage of students in each of four groups, SWD, EL, students with disabilities who are English learners (SWD-EL), and general education students (GenEd, i.e., students without disabilities who are not ELs) who were eligible to use each accessibility support is shown in Table 2.

Data Cleaning and Analyses

In evaluating the data, there was a large positive skew in support use that appeared to be due to data capture glitches or some students haphazardly and repeatedly clicking on an accessibility support tool button. To focus on what appeared to be only legitimate usage, we deleted any student action that involved using the same support more than 10 times on an item[2]. This process resulted in the deletion of about 4% of the student-by-item interactions.

Data were analyzed at both student and district levels. We computed descriptive statistics, stratified by eligibility status and student subgroup. We aggregated results up to the district level. We used data tables and visualizations to describe support use and identify aberrant districts.

Results

Describing and Defining Accessibility Support Usage

To address our first research question, "What is the best way to define 'use' of an embedded accessibility

[2] This decision was made based on analysis of the skewed distributions of use actions, with 10 or more uses representing the highest 5% of the number of uses distribution. Note that we eliminated the student record for a single item when this occurred, not the entire student record.

Table 1. List of Embedded Digital Supports Studied, Definitions, And Eligibility for Use

Support	Definition	Eligibility for Use on Assessments
Highlighter	Mark text with color	*Universal feature.* Any student can use, no prior approval needed
Line reader	Shade a full line of text.	*Universal feature.* Any student can use, no prior approval needed
Masking	Black out text.	*Designated support.* Available to any student with a need for it identified prior to testing by an educator or team familiar with the student.
Text-to-speech	Chosen text read aloud via embedded digital text-to-speech technology. Students could choose to have the entire entity read aloud (*TTS-entire*), for example, an entire question, including every answer choice, or only a selection (*TTS-selection*), such as a word or sentence.	*Designated support* for mathematics and ELA items and for mathematics stimuli. Available to any student with a need for it identified prior to testing by an educator or team familiar with the student. *Accommodation* for ELA passages. Available only to students on an IEP or 504 accommodation plan whose need to use it is identified in their plan prior to testing.
Print	Passages, stimuli, or questions are printed for students on demand during testing. Students can request paper copies of any or all of these in combination on any page.	*Accommodation.* Available only to students on an IEP or 504 accommodation plan whose need to use it is identified in their plan prior to testing. Test coordinator must obtain prior approval for security reasons and to set up printing equipment.

Table 2. Percent of Students Eligible to Use Each Accessibility Support by Subgroup

	ELA				Math			
Accessibility Support	EL n=59,709	SWD-EL n=18,159	SWD n=46,452	GenEd n=350,740	EL n=60,421	SWD-EL n=18,267	SWD n=46,765	GenEd n=355,829
Line reader	100%	100%	100%	100%	100%	100%	100%	100%
Highlighter	100%	100%	100%	100%	100%	100%	100%	100%
Text-to-speech (designated support)	16%	65%	43%	5%	16%	65%	43%	5%
Text-to-speech (accommodation)	-	46%	30%	-	--	--	--	--
Masking	1%	8%	7%	-	1%	8%	7%	-
Print	-	1%	< 1%	-	-	1%	< 1%	-

Note. A dash (-) represents student groups who were not eligible for a particular support.

support?" we computed frequencies of use based on (a) the number of times students accessed a support on an item, (b) whether students accessed a support at least once on any item, and (c) the number of items on which students accessed the support.

In Table 3 we present a frequency distribution based on the first definition (i.e., the number of times an accessibility support was used on the ELA or mathematics test). For these frequencies, if the support was used more than once on an item, we counted it more than once. These frequencies represent the total number of times students accessed the support across items and stimuli on the test. This broad measure, computed as use per student, times number of students, times number of items, can be used to compare the extent of use across supports within an assessment. The line reader and TTS-entire were the most frequently used supports in both subject areas for both item and stimuli (e.g., reading passages).

Table 4 follows the second usage definition by presenting the numbers and percentages of students who used each specific support at least once on any item in each subject area. The number of times a student used the accessibility support is not factored in. For example, if a student used the line reader five times on 20 items, that student is just counted once for using this support.

There are two percentages reported in Table 4. One is based on all students who took each test; the other is based only on students who were approved to use each accessibility support in advance. Text-to-speech, masking, and print are accessibility supports that required approval before the exams; therefore, the total number of approved students for these supports is smaller than that of total students. In contrast, line-reader and highlighter are universal supports that are allowable for all students without pre-approval. Therefore, the total number of approved students for universal tools is equal to that of the total students.

When we computed the percentages based on all students who took the tests, line reader was the most frequently used accessibility support for both ELA and mathematics, followed by highlighter, text-to-speech, masking, and print. In contrast, when the percentages were computed based on students who were approved to use each support, TTS-entire was the most frequently used support for both subjects; this was followed by line reader, highlighter, masking, and print. Percentages of using both TTS-entire and TTS-selection increased when computed based on approved students. More than half of the students who were approved to use TTS actually used it at least once during the test.

A third way to describe accessibility support use is to consider the number of items on which students used them. In Table 5, we present the means and standard deviations (SD) for the number of items on which a student accessed each specific support. For both subject areas, TTS-entire was used on the largest number of items (more than 11 items for ELA and more than 12 for mathematics, on average). Print, highlighter, and TTS-selection had the next highest averages for both subject areas; but it should be noted, the print tool had very small sample sizes (Table 3), because it was available to only a small number of students.

Table 3. Accessibility Support Use Frequencies

Accessibility Support	Type	Item		Stimulus	
		ELA	Mathematics	ELA	Mathematics
Line reader	Universal Tool	731,928	443,599	348,218	35,858
TTS-entire	Designated Support/Accommodation[1]	648,081	593,338	102,129	40,609
Highlighter	Universal Tool	431,847	312,693	598,808	61,246
TTS-selection	Designated Support/Accommodation	20,854	13,898	7,240	1,694
Masking	Designated Support	3,249	3,321	1,851	334
Print	Accommodation	55	108	129	42

Note: The numbers in each cell represent the total number of times an accessibility support was accessed on any item or stimulus by any student. ELA = English language Arts. Entire entity = the entire stimulus or item was read aloud. Entity selection = a portion of the text was read aloud. [1]TTS is an accommodation for ELA stimuli and a designated support for everything else

Table 4. Number and Percentage of Students Accessing Each Accessibility Support

	ELA (n=475,060)			Math (n=481,282)		
Accessibility Support	n	% of Total Students	% of Approved Students	n	% of Total Students	% of Approved Students
Line reader	247,027	52%	--	176,862	37%	--
Highlighter	161,644	34%	--	59,834	12%	--
TTS-entire (designated support)	39,487	8%	67%	36,570	8%	61%
TTS-selection (designated support)	6,416	1%	11%	4,072	1%	7%
TTS-entire (accommodation)	18,577	4%	84%	--	--	--
TTS-selection (accommodation)	3,533	<1%	16%	--	--	--
Masking	959	<1%	14%	728	<1%	11%
Print	92	<1%	28%	42	<1%	12%

Note. "n" refers to the number of students who accessed the accessibility support on at least one item. "% of Total Students" represents the number of students who used the accessibility support at least once on any item divided by the number of students who took each test. "% of Approved Students" represents the number of students who used the accessibility support at least once on an item divided by the number of students who were identified prior to the assessment as needing the support.

Table 5. Mean (SD) Number of Items on which Designated Support was Used

Accessibility Support	ELA Mean (SD)	Mathematics Mean (SD)
TTS-entire	11.43 (12.10)	12.19 (11.96)
Print	3.21 (7.18)	5.06 (9.45)
Highlighter	2.62 (4.49)	2.69 (4.04)
TTS-selection	2.45 (3.24)	2.37 (2.93)
Line reader	2.17 (2.21)	1.58 (1.19)
Masking	1.66 (1.18)	1.60 (1.18)

Note. ELA = English language arts. SD = Standard deviation

In Table 6, we illustrate the rank order of the use of accessibility supports using the three different definitions of use. TTS-entire ranks in the top 3 most frequently used supports across all three definitions. Line reader was the first or second most commonly used, but not used frequently over test items. The differences in definition are important when considering the frequency of accessibility support use. The second definition illustrates that larger percentages of students are using the highlight text tool, relative to TTS-entire. The first definition masks this percentage because the fewer students who used TTS-entire used it more often. The average number of items on which a support was used is also interesting for gauging frequency of accessibility support use. Thus, the latter two definitions (based on using a support at least once or the average

Table 6. Rank-Order of Most Frequently Used Supports Based on Different Definitions of Use

Accessibility Support	ELA: Number of times a support was accessed on any item	ELA: Whether support was accessed on any item	ELA: Average number of items on which support was used	Mathematics: Number of times a support was accessed on any item	Mathematics: Whether support was accessed on any item	Mathematics: Average number of items on which support was used
Line reader	1	1	5	2	1	6
TTS-entire	2	3	1	1	3	1
Highlight	3	2	3	3	2	3
TTS-selection	4	4	4	4	4	4
Masking	5	5	6	5	5	5
Print	6	6	2	6	6	2

Note. Numbers in each cell represent the rank order of the use of each support based on each definition.

number of items on which a support was used) may be best for understanding and defining usage in the aggregate, for example, across school districts.

Although we did not have the data to identify the order in which students responded to the items, the data did include the "page" (screen) on which an item was presented to a student. Because more than one item could appear on a page, we do not know the order in which the items appeared on the screen and so do not know the order in which a student responded to items on the same page. However, by looking at support use across pages and knowing that items on pages with lower numbers were presented before items on pages with higher numbers, we can get some idea of student use of the supports as the test progressed. These data are summarized in Figures 1 (ELA) and 2 (mathematics), respectively. Similar to Crotts-Roohr and Sireci (2017), students' use of the supports was noticeably higher earlier in the test, particularly for the first few pages. The TTS supports had the most sustained use across pages relative to the other supports, for both subject areas.

Support Use by Subgroup

To answer our second research question ("How much are accessibility supports used across students defined by disability and English learner status?"), and third research question ("What types of supports are most often used by these different student groups?), we computed the numbers and percentages of students who used each accessibility support, broken down by the four subgroups of students. These data, summarized in Table 7, represent students who used the accessibility support at least once on at least one item. All percentages were computed using the total number of approved students who are allowed to use each accessibility support as denominators. For ELA, line reader was used the most for ELs and GenEd, while TTS-entire was the most common support for SWD-EL and SWD. The percentages in Table 7 include students who did not use any accessibility supports throughout the tests.

With respect to the number of items on which students used supports, Table 8 shows the means and standard deviations by subgroup. TTS-entire was used on the most items for all subgroups on both ELA and mathematics. SWD-EL used this support on the most items (just over 13 items on average), followed by SWD (just under 13 items), ELs (about 10 items), and GenEd (about 8 items).

These data can also be presented visually at the district level. As an example, in Figure 3, we present box plots and distributions of the average number of times on which TTS-entire was used on the ELA test within districts, broken down by student subgroup. For the boxplot, the sizes of bubbles represent the size of districts, with larger bubbles representing larger districts.

Figure 1. Percentage of Eligible Students Who Used Accessibility Supports on Each "Page" on the ELA Assessment

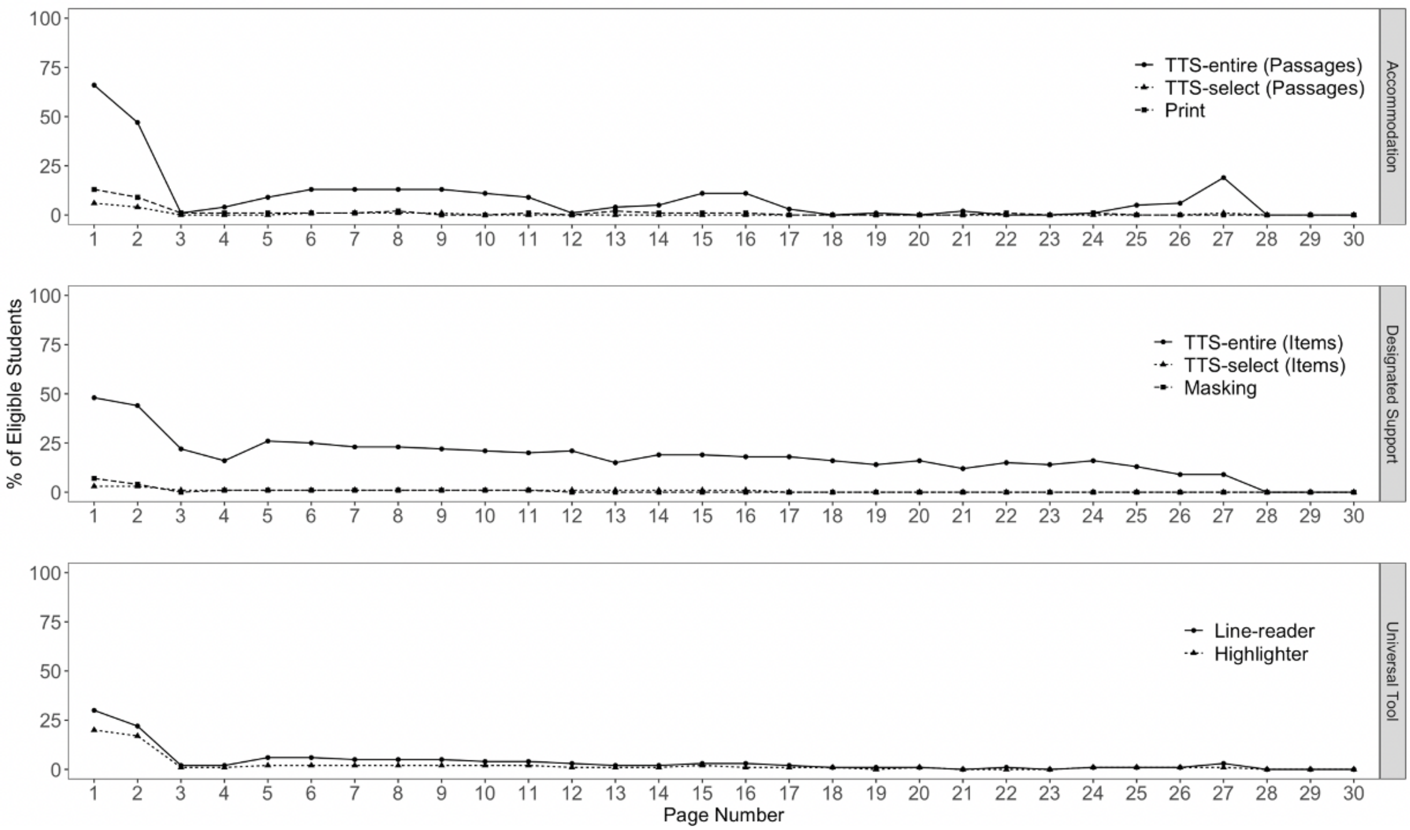


Figure 2. Percentage of Eligible Students Who Used Accessibility Supports on Each "Page" on the Mathematics Assessment

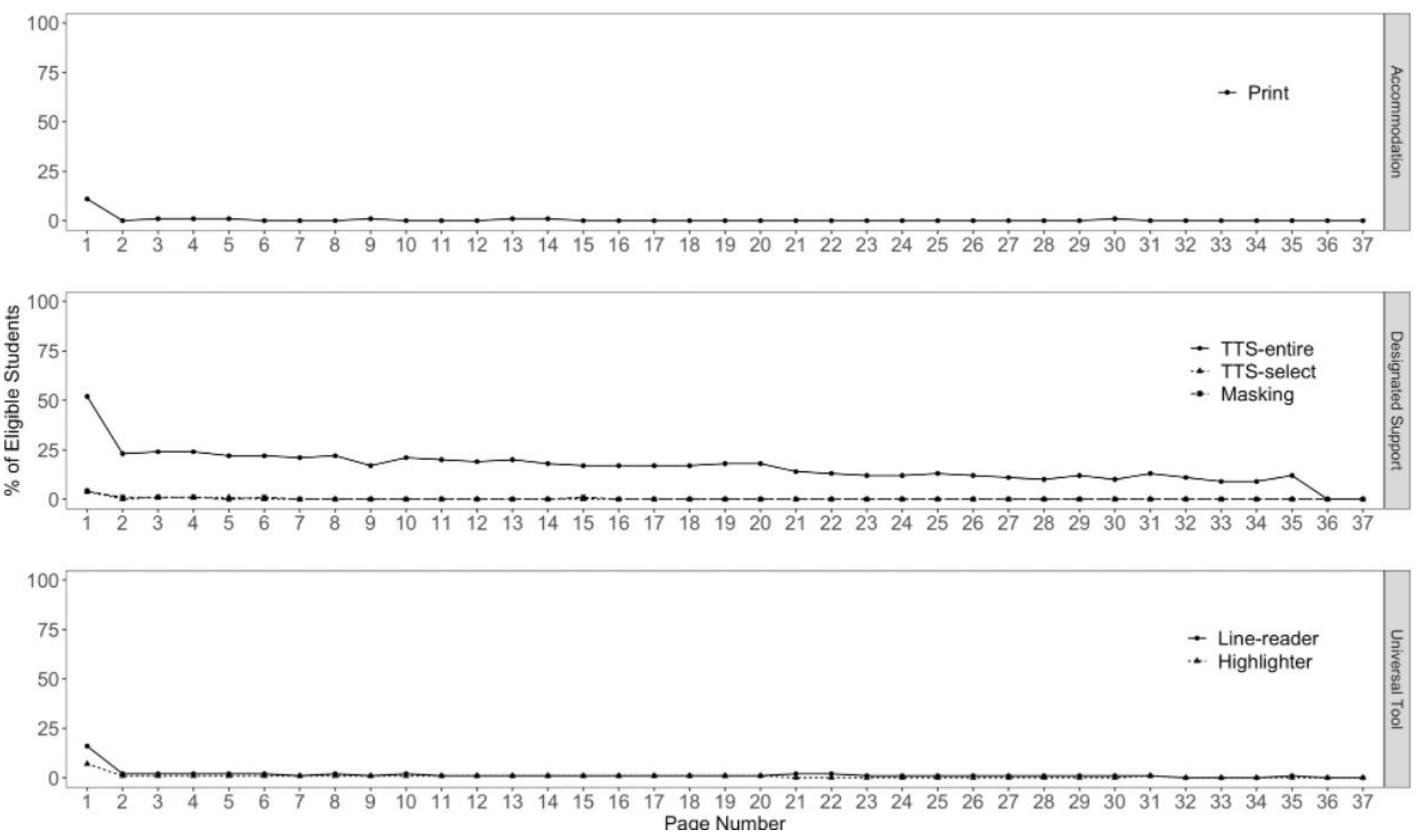


Table 7. Percent of Students Using Accessibility Supports by Subgroup

Accessibility Support	ELA				Math			
	EL n=59,709	SWD-EL n=18,159	SWD n=46,452	GenEd n=350,740	EL n=60,421	SWD-EL n=18,267	SWD n=46,765	GenEd n=355,829
Line reader	56.29%	48.60%	47.25%	52.07%	40.69%	35.69%	34.23%	36.46%
TTS-entire (designated support)	56.11%	80.76%	76.86%	51.57%	53.92%	74.02%	68.52%	48.57%
TTS-selection (designated support)	8.70%	14.07%	12.80%	7.62%	6.63%	8.01%	7.27%	5.63%
Highlighter	37.13%	26.57%	25.31%	35.04%	13.90%	10.03%	9.37%	12.71%
TTS-entire (accommodation)	75.00%	85.75%	82.81%	--	--	--	--	--
TTS-selection (accommodation)	20.00%	16.79%	15.45%	--	--	--	--	--
Masking	17.40%	11.21%	12.23%	20.28%	14.67%	7.42%	9.21%	15.23%
Print	--	25.23%	29.47%	--	--	14.95%	10.48%	--

Note. The numbers in each cell represent the percentage of students who used each accessibility support across four subgroups. The denominators were the total students who were approved to use each accessibility support. Print on demand was approved only for students with disabilities, so the percentages of Print for EL and GenEd students are left blank.

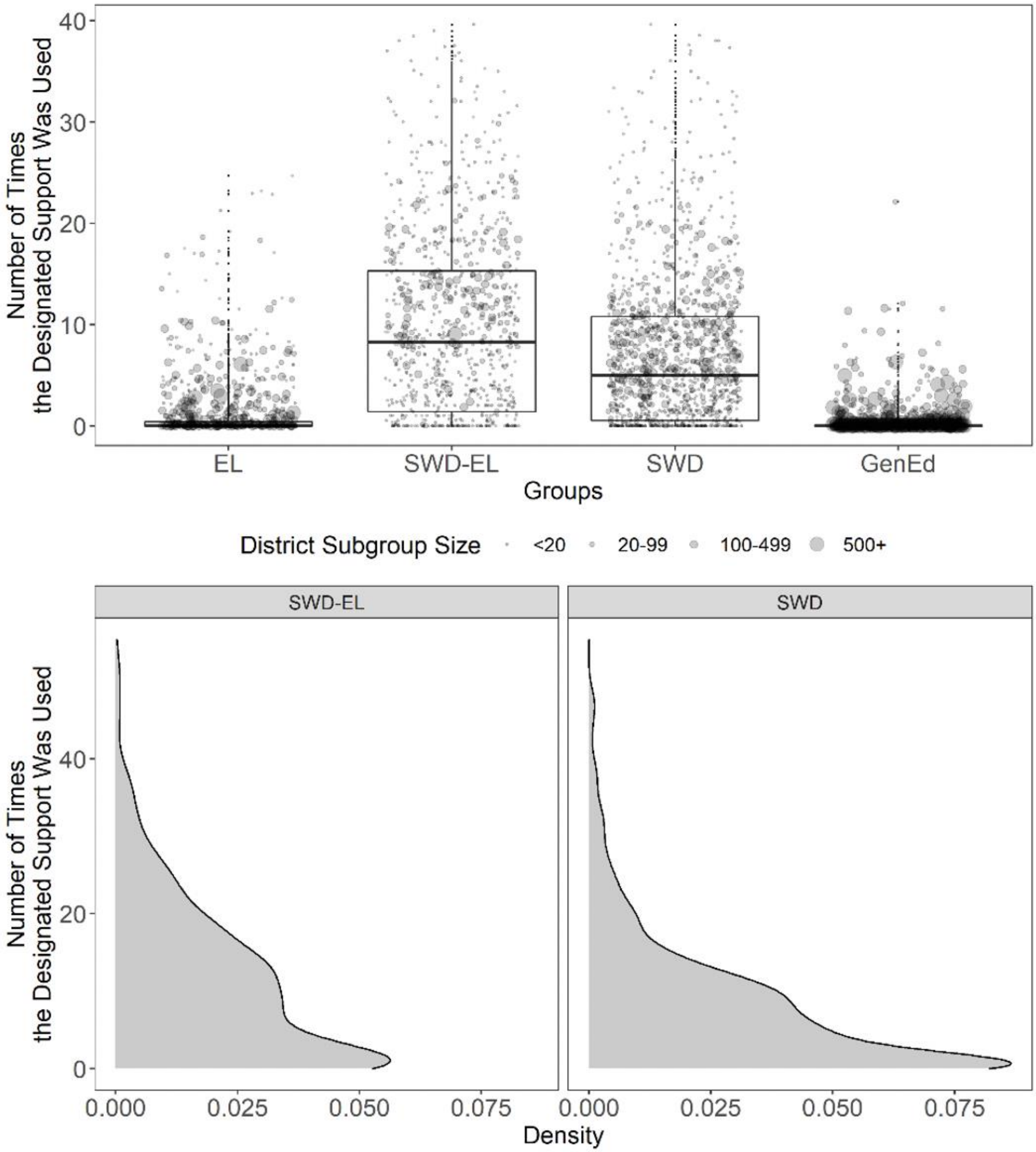
Table 8. Mean (SD) Number of Items on which Accessibility Support was Used by Subgroup

Accommodation	ELA				Math			
	EL	SWD-EL	SWD	GenEd	EL	SWD-EL	SWD	GenEd
TTS-entire	10.03 (10.92)	13.33 (13.12)	12.82 (12.90)	8.41 (9.62)	9.73 (10.51)	14.95 (12.60)	14.03 (12.59)	7.85 (9.32)
Highlighter	2.67 (4.49)	2.17 (3.34)	2.20 (3.51)	2.66 (4.61)	2.78 (4.19)	2.41 (3.60)	2.40 (3.66)	2.72 (4.07)
Line reader	2.31 (2.24)	2.09 (2.03)	2.12 (2.17)	2.15 (2.21)	1.70 (1.29)	1.67 (1.38)	1.65 (1.47)	1.55 (1.13)
TTS-selection	2.24 (2.60)	2.68 (3.98)	2.56 (3.38)	2.23 (2.60)	2.16 (2.37)	2.73 (3.77)	2.41 (2.90)	2.08 (2.23)
Masking	1.72 (1.07)	1.53 (0.94)	1.68 (1.29)	1.67 (1.19)	1.63 (1.52)	1.66 (1.29)	1.59 (0.98)	1.56 (1.16)
Print	-	1.75 (1.50)	3.80 (8.51)	-	-	2.00 (3.00)	8.11 (12.63)	-

Note. Print on demand is not available for ELs or GenEd students.

The vertical lines indicate the average number of times TTS-entire was used for the respective subgroup within the district. The distribution of TTS-entire usage is more spread out for SWD-ELs and SWD. EL and GenEd students were extremely positively skewed: students in these groups rarely used TTS-entire for the ELA test, because most were not approved to use it as a designated support.

Figure 3. Distribution for the Use of TTS-Entire (Designated Support) Across all ELA Items



Note. EL and GenEd were extremely positively skewed, and so only SWD-EL and SWD are presented.

Eligibility Versus Use

As described by our fourth research question ("What is the relationship between eligibility to use accessibility supports and use of those supports?"), we were interested in describing whether students who were eligible for accommodations or designated supports used them and to check whether students who were not eligible used them. We were also interested in describing the extent of universal tool use. Figure 4 presents the percentages of students who actually used the accessibility supports they were eligible to use during testing. There were no students who used a support they were not eligible to use. The usage patterns varied by the type of supports. For example, of those who were approved to use TTS-entire on passages, 84% used it on at least one passage. About 67% of those eligible to use TTS-entire on ELA items used the support on at least one item. For the line reader, about 52% used the support on at least one item. For mathematics (Figure 5), again there were no unapproved students who used TTS, and about 61% of the students approved for TTS-entire used it. Similarly, about 37% of the students used a line reader during the mathematics test.

The breakdowns of approved versus usage results by subgroup for both ELA and mathematics are presented in Figures 6 and 7. For TTS-entire and line reader, a considerable number of students within each subgroup used these supports for both ELA and mathematics. TTS-entire had the highest usage rates for SWD-EL (ELA: 80%; Math: 74%) and SWD (ELA: 76%; Math: 69%). Line reader was also frequently used across all studied groups. Almost half the students in all subgroups used line reader for ELA, and about 40% across all groups used it for mathematics.

Figure 4. Student Use Among Those Eligible to Use: ELA

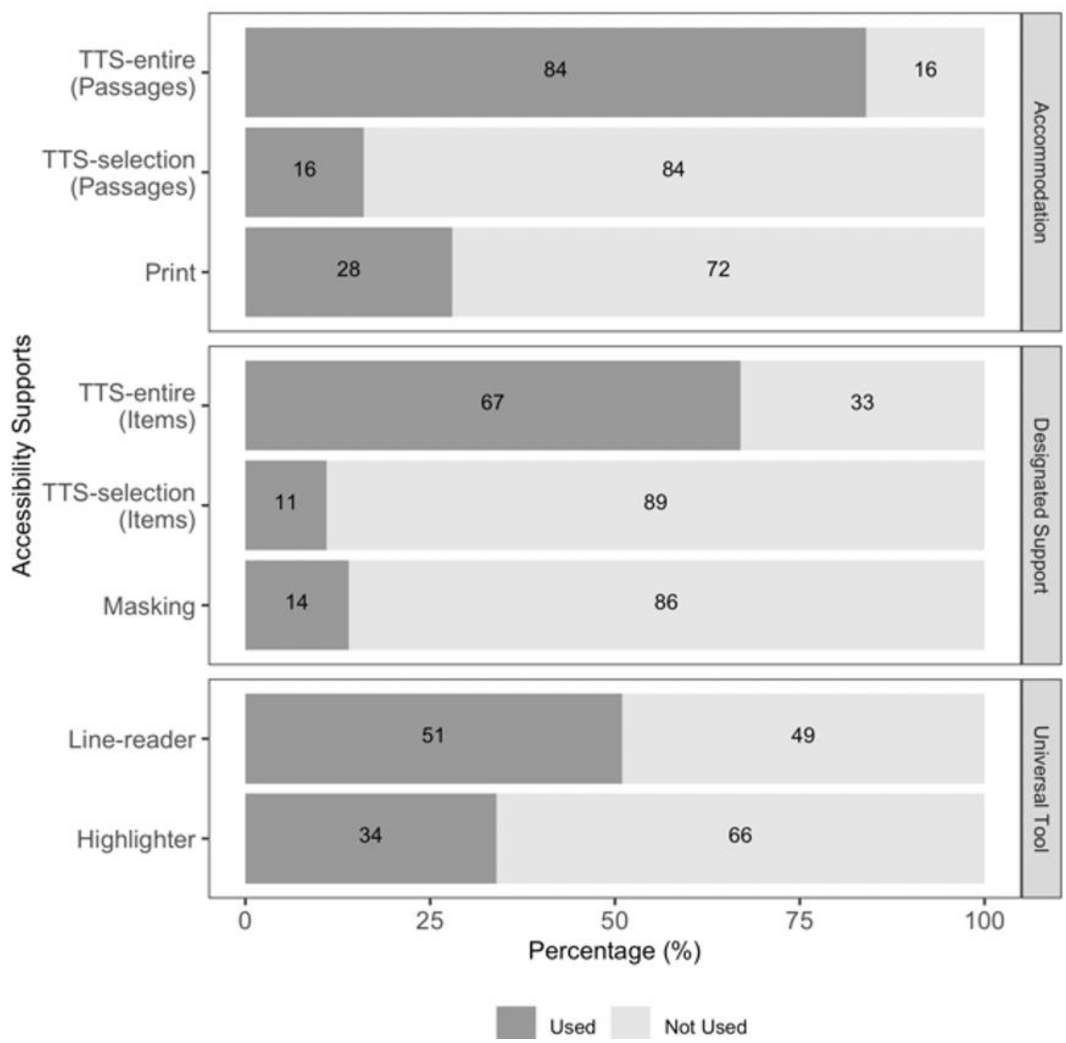
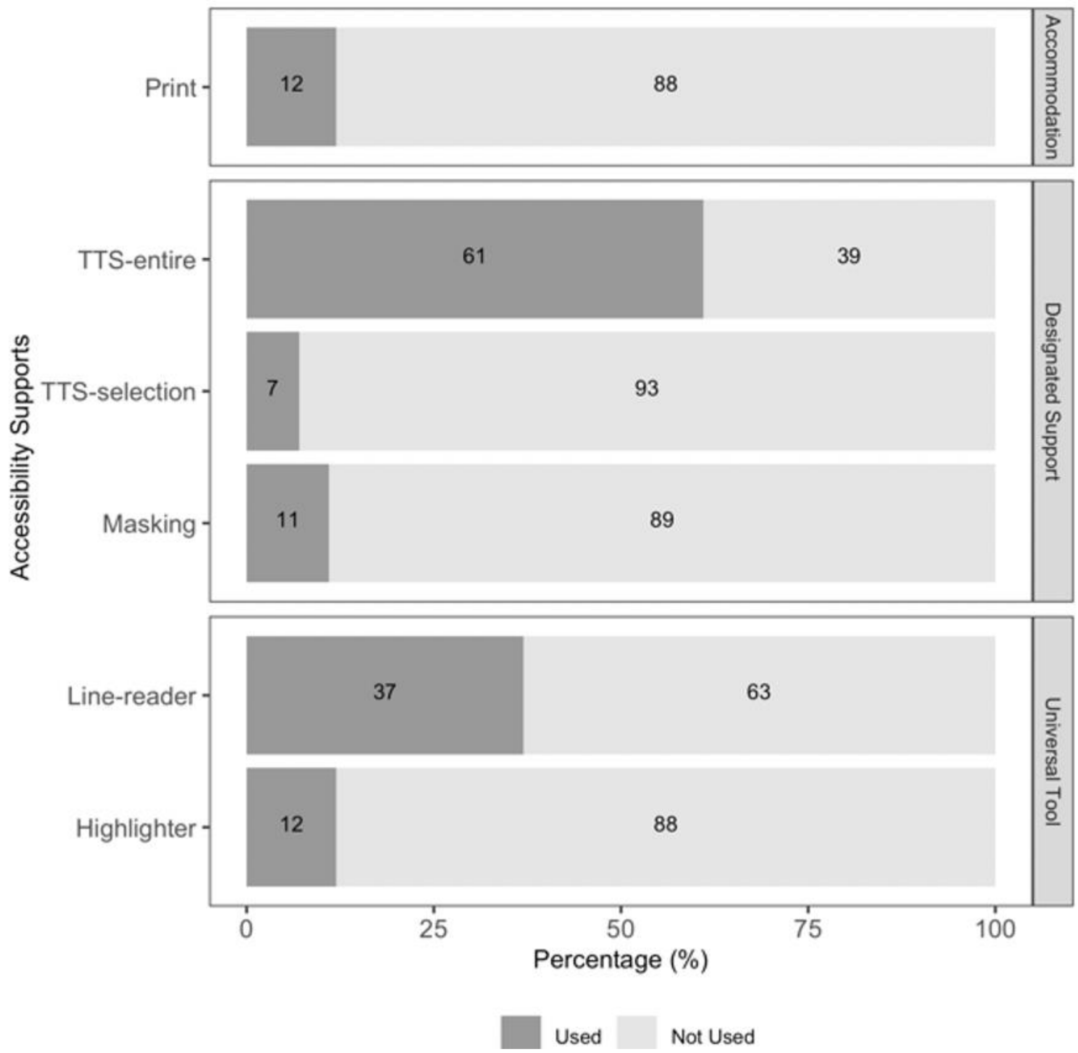


Figure 5. Student Use Among Those Eligible to Use: Mathematics



Identifying Aberrant Districts

Our fifth and final research question ("How can we use assignment and usage information to identify districts with students who may be overusing or underusing the supports?") was addressed by computing the proportion of designated support and accommodation use data for subgroups within each district. Here we defined underuse as students who are eligible to use the supports not using them at all, and overuse as students who are not eligible using the support. The latter could occur d to computer glitches or test security issues such as supports being turned on "illegally" for students in a particular district or school. For this question, we focused on TTS-entire since it showed the highest usage rate for eligible students. In Figure 8, we present a "bubble plot" of district proportions for TTS-entire eligibility and usage for the ELA items, stratified by student subgroup. The location of the bubble in each plot represents the proportion of students approved for the support (horizontal axis) and the proportion of students who used the support on at least one item (vertical axis). The size of the "bubbles" reflects district size.

For SWD and SWD-EL, there is an expected linear trend between eligibility and use for some, but not all districts. In all districts, the percentage of students who used the designated support is never higher than the percentage eligible. The nonlinearities are due to some districts with lower usage proportions relative to eligible proportions. The data for mathematics are

Figure 6. Student Use Among Those Eligible, by Subgroup: ELA

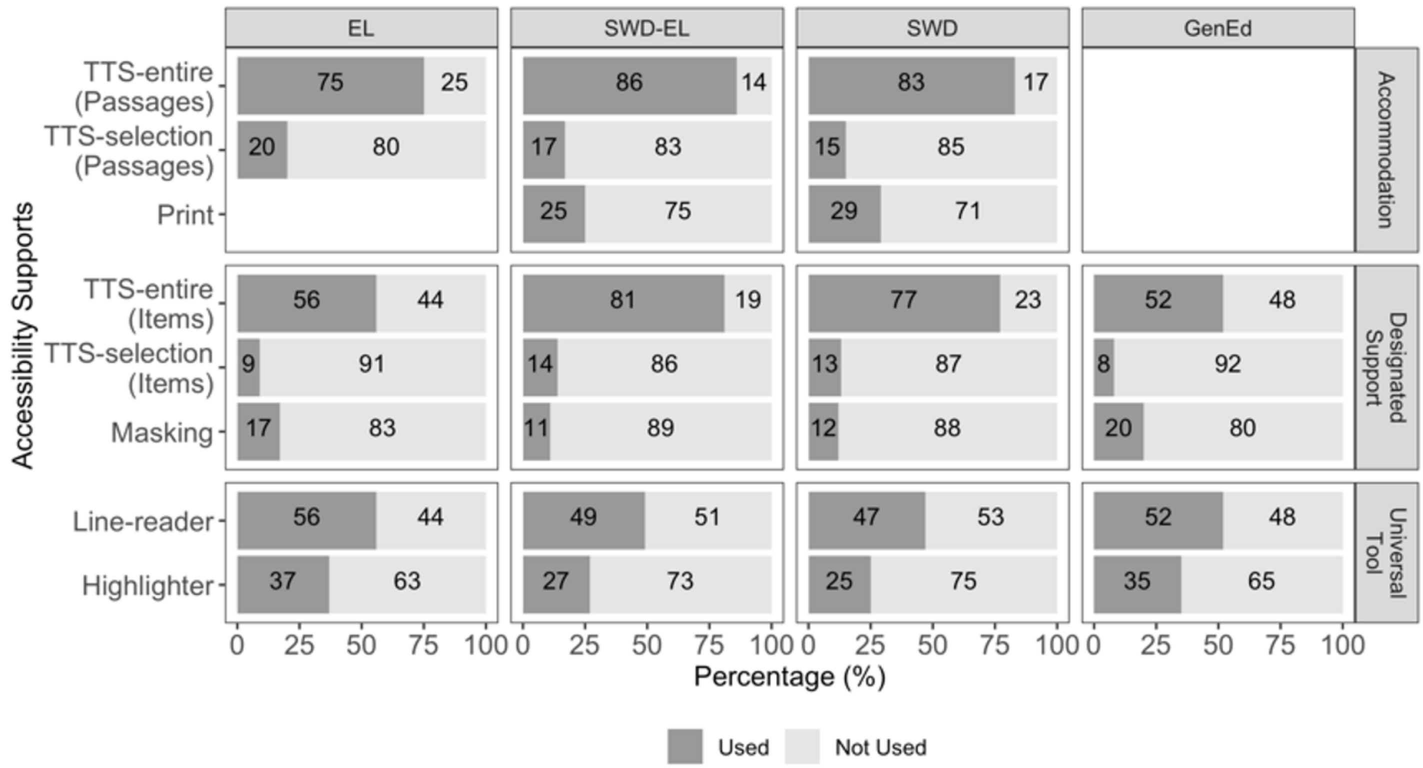


Figure 7. Student Use Among Those Eligible, by Subgroup: Mathematics

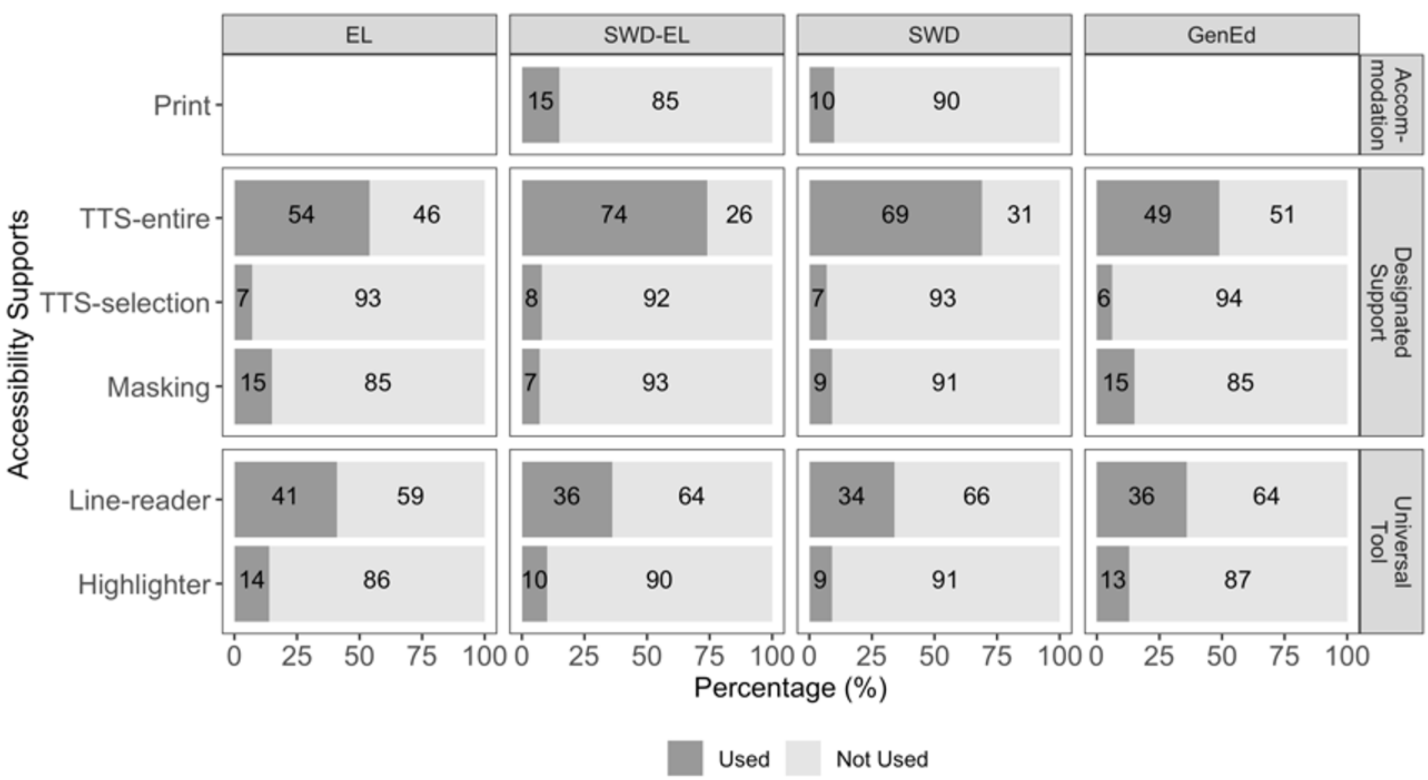
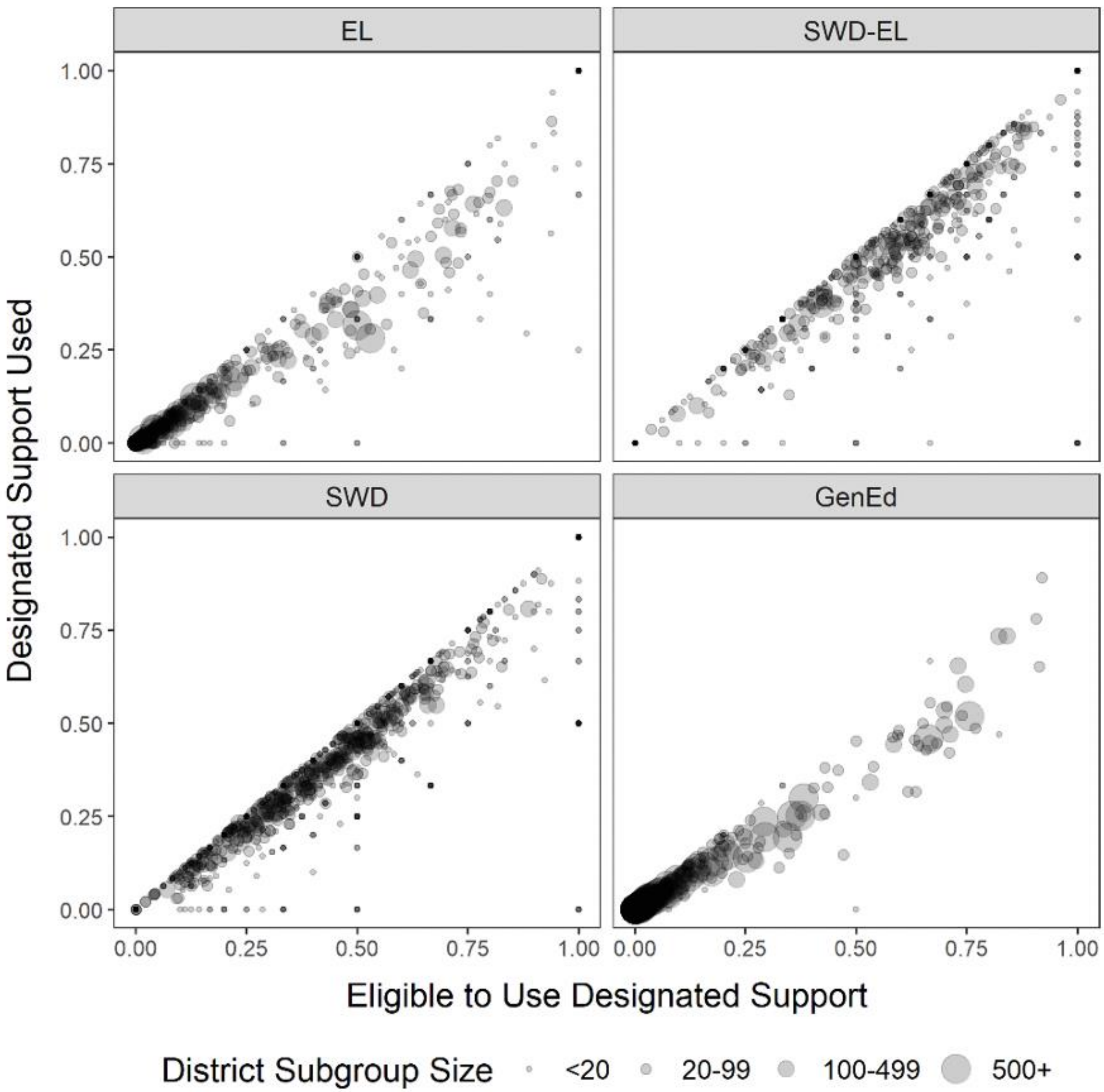


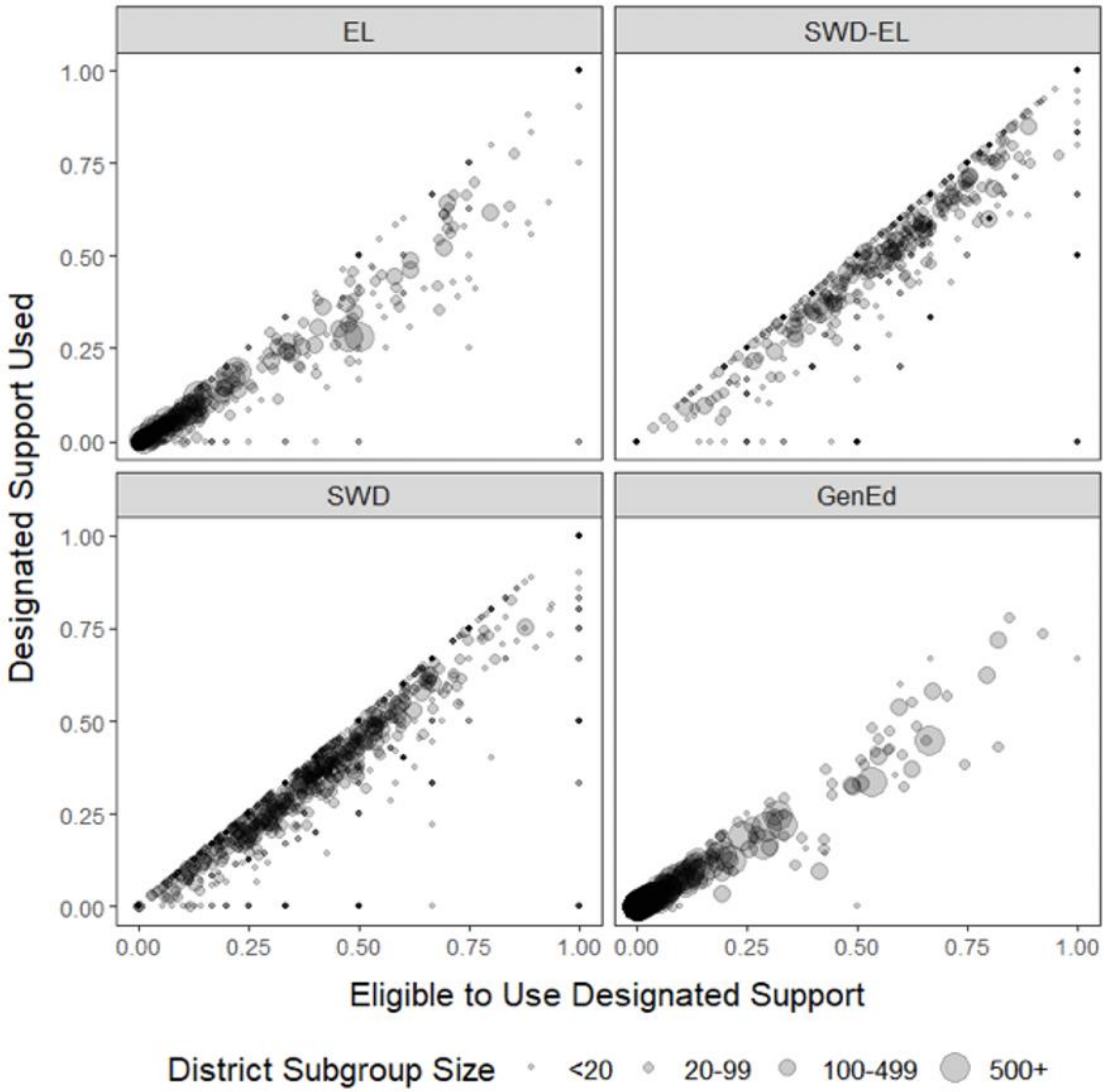
Figure 8. TTS-Entire Use vs. Eligibility, ELA Items, Average District Proportions



Note. Use is defined as used TTS-entire at least once on at least one item. Opaque dots represent multiple districts with the same percentages.

presented in Figure 9. For both subjects, there are many districts where the percentage of students using a designated support is much lower than the percentage of students approved to use it. These districts could be flagged for further review and, if necessary, training on how to ensure students are aware of and use the supports designed for them.

Figure 9. TTS-Entire Use vs. Eligibility, Mathematics Items, Average District Proportions



Note. Use is defined as used TTS-entire at least once on at least one item. Opaque dots represent multiple districts with the same percentages.

Discussion

In this study, we analyzed the relatively unexplored data derived from digital logs of test taker actions on sixth-grade computerized-adaptive statewide assessments of ELA and mathematics to better understand how students use embedded accessibility supports. One issue we addressed is how to define the "use" of designated supports for purposes of aggregate reporting over students. We explored three ways of describing use: (a) computing proportions of use based on the number of times a support was accessed on any and all items (for approved and unapproved students), (b) computing percentages based on whether a support was used on any item, and (c) computing the average number of items on which each specific support was used.

The different approaches provided somewhat different information. The first definition provides information about the overall use of each support, the second approach provides information on the numbers of students who used a support at least once, and the third approach provides information regarding students' use of a support throughout the test. All three variables are likely to be of interest to researchers and practitioners interested in evaluating the degree to which accessibility supports are being effectively used. The first definition can lead to the conclusion a support is being widely used, when in fact it is being used often by relatively few students. Thus, the second definition may be best for summarizing information on a lack of support use. The third definition may be best for evaluating support use across all items on a test. Thus, there is no "best" way to define use overall, but rather different ways to define it for specific questions and purposes. In Table 9, we suggest applications for different definitions of use and suggest that other researchers try out, revise, and expand these definitions and applications.

Table 9. Definitions of "Use" of Accessibility Supports and Suggested Applications

Definition	Applications
Number of items a support was accessed on any and all items	Comparing supports, snapshot of use.
Student used support on at least 1 item/ did not use support at all.	Subsetting the data, aggregating state, district, and subgroup data for further analyses.
Percentage of students using support on an item.	Item analysis, evaluation of the support.
Number of items on which each student uses a support.	Evaluation of the support, aggregating state, district, and subgroup data for descriptive purposes or further analyses.
Used on a particular item.	Item level statistical models for estimating effectiveness and/or predicting use.
Typical patterns of use over the sequence of items on the test (latent classes).[1]	Predicting use patterns, describing use for research purposes.
How used on a particular item.[1]	Inference about problem solving strategies, effectiveness of support.

[1]The data available for this study did not support the use of these definitions.

The USDE requires states to monitor and report accommodation and accessibility support use by SWD and ELs. Although USDE does not specify how use should be defined or how the monitoring should be done, the results of our study illustrate how log data can be used for monitoring embedded accessibility supports. Our analyses can be useful for monitoring at both the state and local level, and perhaps even during a testing window for a district or state. Thus, in addition to these analyses being useful for a state's formal monitoring process, they can also be useful for intervening when districts may be over- or under-using accommodations.

Although we analyzed these data to evaluate the use of each embedded accessibility support, such usage data could also be analyzed to provide information about the students. We looked at student differences to some extent by breaking down support usage by student subgroup. These analyses indicated the accommodations most appropriate for each student group tended to be used most by that student group. For example, SWD and SWD-ELs used TTS-entire much more frequently than ELs without disabilities and GenEd students. The degree to which actual accommodation and designated support use matched accommodation and designated support eligibility is encouraging. We also found there was virtually no "illegal" use of designated supports and accommodations. That is, only students who were approved for such use actually used them.

On the other hand, we observed a lack of use of supports for many students who were approved to use them, both at the individual student level where students did not use the accessibility support on any items or only used on a few items and at the district level where some districts had much lower proportions of use than others. To highlight those instances, we used illustrations to present the data in ways in which we hope district personnel and policymakers can easily understand. The bubble charts, for example, show which districts tended to have relatively lower usage rates for a subgroup. We believe clear visualizations are important for communicating accessibility support use, and for acting upon the results. These results should help move the field forward with respect to understanding and using test accessibility support usage data and for meeting the demands of USDE's Peer Review process.

Although there has not been a lot of published research on the degree to which students use accommodations and other supports, our results are consistent with the one previous study we identified by Crotts-Roohr and Sireci (2017). They found the use of accommodations designed for ELs was low, and that use decreased as the test progressed. That study, and ours, suggests more work needs to be done to engage students in the use of supports, or design better supports that students will use.

We believe research in this area will promote the validity of educational assessments by providing students as optimal a test experience as we can, with respect to supports and access. Log data traditionally can be used to acquire validity evidence based on response processes (AERA et al., 2014; Padilla & Benitez, 2014). A reviewer of an earlier version of this paper pointed out our analysis of log data may fit into that category of validity evidence in that we can confirm what supports students used (or did not use) when responding to items. Further exploration of the degree to which such use confirms the assessment measures the intended cognitive skills would enhance validity evidence based on response processes for the assessment.

Limitations and Suggestions for Future Research

Although we believe the results of this study will be useful to practitioners, our study has several limitations. First, our data come from a single state and focus on only two subject areas for one grade level. Sixth grade was chosen because students at this age have greater familiarity with both assessments and technology than students at lower grade levels, and the proportion of ELs is larger than subsequent grades. That said, future research should consider other grade levels, and it would be good to replicate the findings across other states and subject areas.

Another limitation is that our interpretations were based on analysis of digital log data, rather than direct observations or direct probing of students while they interacted with the assessment. Furthermore, these log data were limited in what was collected while students interacted with the assessment. We did not have information about particular test items, such as text complexity or test-taker characteristics, such as types of disabilities or English language proficiency level. We also did not have complete information regarding the order in which items were presented to students, and so we could not fully evaluate how well students used supports from the beginning to the end of the assessment. We also did not have access to complete data on all available supports. As such, and because we noticed that many students did not use the supports throughout the whole test, we did not explore students' use of bundled accommodations and supports. Thus, the use of bundled accommodations remains an area for future research.

Although knowing the degree to which students used supports is important, future research should explore reasons why students did or did not use the supports, as well as how well such use impacted performance on items and the test as a whole (c.f., Lazarus et al, 2021, pp.46-48). As a reviewer of an earlier version of this manuscript pointed out, a student's familiarity with a support provided will be a major factor in the degree to which the student uses it. Thus, districts that have higher rates of embedded accessibility support use may have done a thorough job in familiarizing students with the support in the classroom. The 4% of student actions we deleted seem to be the result of testing the system, but also could be due to students' frustration or unfamiliarity with a support. Thus, future research may dive deeper into aberrant use behaviors that may be valid student responses. In addition, we recommend future research explore differences in support usage across additional student characteristics such as sex, race/ethnicity and poverty, and by achievement level. Such breakdowns of the data could be illuminating if differences are found across these important student characteristics. In addition, future research should consider how to best provide data from digitally delivered assessments to support decisions regarding how to best provide supports to promote student achievement.

Our analyses were conducted at a "high" level and illustrate strategies states can use to understand overall accessibility support use, and potentially flag districts who may deviate from common practices. However, the data reporting techniques we used could also be used at the district level to evaluate the degree to which students within their schools are effectively using supports and accommodations, and the degree to which test administrators and IEP teams are properly trained in orienting students to these supports. Future research should investigate the types of visualizations most helpful to state and local administrators for best understanding effective and appropriate accessibility support use.

Conclusions

Analysis of the log data from computer-based testing programs can provide illuminating information about the degree to which different types of students use different types of assessment supports. Calculating the numbers and proportions of students who use specific accessibility supports allows us to confirm the intended students are receiving and using the intended supports, and aggregation of these data can be used for monitoring, training, and program accountability purposes. Future research should explore how these data can be used to improve the effectiveness and use of such supports to serve the education of all students.

References

Abedi, J. & Ewers, N. (2013). Accommodations for English learners and students with disabilities: A research based decision algorithm. Smarter Balanced Assessment Consortium. Downloaded December 31, 2019 from https://portal.smarterbalanced.org/library/en/accommodations-for-english-language-learners-and-students-with-disabilities-a-research-based-decision-algorithm.pdf

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, DC: Author.

Buzick, H.M., & Stone, E. A. (2014). A meta-analysis of research on the read aloud accommodation. *Educational Measurement: Issues and Practice, 33*(3), 17-30.

Crotts-Roohr, K., & Sireci, S. G., (2017). Evaluating computer-based test accommodations for English learners. *Educational Assessment, 22*, 35-53. DOI: 10.1080/10627197.2016.1271704.

García, O. (2009). Emergent bilinguals and TESOL: What's in a name? *TESOL Quarterly, 43*, 322-326. DOI: 10.1002/j.1545-7249.2009.tb00172.x.

Lazarus, S., Goldstone, L., Wheeler, T., Paul, J., Prestridge, S., Sharp, T., Hochstetter, A., and Warren, S. (2021). *CCSSO Accessibility Manual: How to Select, Administer, and Evaluate Use of Accessibility Supports for Instruction and Assessment of All Students*. The Council of Chief State School Officers (CCSSO).

Lovett, B. J. (2010). Extended time testing accommodations for students with disabilities: Answers to five fundamental questions. *Review of Educational Research*, *80*(4), 611-638.

National Center on Education Outcomes. (2016). *Participation in AA-AAS*. Minneapolis: University of Minnesota.

Partnership for Assessment of Readiness for College and Careers (2017). *PARCC Accessibility Features and Accommodations Manual 2017–2018*. PARCC Inc. Washington, DC: PARCC Assessment Consortia.

Shyyan, V., Thurlow, M., Christensen, L., Lazarus, S., Paul, J., and Touchette, B. (2016). *CCSSO accessibility manual: How to select, administer, and evaluate use of accessibility supports for instruction and assessment of all students.* Washington, DC: CCSSO.

Sireci, S. G., & Banda, E., & Wells, C. S. (2018). Promoting valid assessment of students with disabilities and English learners. In Elliott, S. N., Kettler, R. J, Beddow, P. A., & Kurz, A., (Eds.), *Handbook of Accessible Instruction and Testing Practices: Issues, Innovations, and Application* (pp.231-246). Sage.

Sireci, S. G., & O'Riordan, M. (2020). Comparability issues in assessing individuals with disabilities. In A.I. Berman, E.H. Haertel, & J.W. Pellegrino (Eds.), *Comparability Issues in Large-Scale Assessment: Issues and recommendations* (p. 117-204). Washington, DC: National Academy of Education Press.

Smarter Balanced Assessment Consortium (2016, January). Smarter Balanced Assessment Consortium accommodations and accessibility framework. Author. Available at https://portal.smarterbalanced.org/library/en/accessibility-and-accommodations-framework.pdf.

Smarter Balanced Assessment Consortium (2017). Smarter Balanced Assessment Consortium: Summative Technical Report. Author. Available at https://portal.smarterbalanced.org/library/en/2017-18-summative-assessment-technical-report.pdf

U.S. Department of Education (2018, September). *A state's guide to the U.S. Department of Education's assessment peer review process*. Washington, DC: Author.

U.S. Department of Education (2017, December). Every Student Succeeds Act Assessments under Title I, Part A & Title I, Part B: Summary of final regulations. Available at https://www2.ed.gov/policy/elsec/leg/essa/essaassessmentfactsheet1207.pdf.

Wainer, H. (1993). Some practical considerations when converting a linearly administered test to an adaptive format. *Educational Measurement: Issues and Practice*, 12, 15-20.

Citation:

Lee, D., Buzick, H., Sireci, S. G., Lee, M., & Laitusis, C. (2021). Embedded Accommodation and Accessibility Support Usage on a Computer-Based Statewide Achievement Test. *Practical Assessment, Research & Evaluation*, *26*(25). Available online: http://pareonline.net/getvn.asp?v=26&n=25

Corresponding Author

Dukjae Lee
Senior Research Assistant
University of Massachusetts Amherst
email: dukjae.lee.125 [at] gmail.com