ATEA

Accessibility for Technology-Enhanced Assessments

Report of Project Activities

Julia Shaftel

Sarah Benz

Elena Boeth

Jessica Gahm

Dan He

Jess Loughran

Michelle Mellen

Em Meyer

Erin Minor

Emily Overland

University of Kansas

November 16, 2015

The contents of this report were developed under a grant from the U.S. Department of Education. This report is solely the responsibility of the authors and does not necessarily represent the policy of the U.S. Department of Education. Do not assume endorsement by the federal government.

Table of Contents

Appendices

1.0 Executive Summary

The Accessibility for Technology-Enhanced Assessments (ATEA) project investigated the accessibility of innovative, computerized assessments for students with vision and motor disabilities. Innovative assessments, such as those developed by the major assessment consortia, include item types that may present particular challenges for students with these disabilities. This report summarizes the activities of the ATEA Enhanced Assessment Grant project and presents the findings regarding accessible item development and testing.

Project activities consisted of reviews of sample technology-enhanced (TE) test items by a panel of vision and motor experts, reviews of demonstration items in both original and accessible formats by panels of teachers from five of the ATEA partner states, cognitive labs with students of all ages in three states, and field tests and item tryouts in eleven states. Through these varied activities, project staff developed and tested a variety of accessible versions of TE items, from the first prototypes through items ready for student interaction and response scoring.

Initial prototype TE items were written based on items obtained in fall 2012 from major assessment consortia and other sources. A variety of innovative item types was prepared to cover the range of tasks and interactions that students would be expected to master. These item types included drag-and-drop items that require visual, motor, and hand-eye coordination skills. Other items featured click-to-select interfaces, constructed responses, or the use of radio buttons for single- or multiple-response selections. This report describes the barriers to accessibility identified for these tasks and explains how the project addressed these barriers through exploration of alternative item formats, accommodations, and special forms.

Initial Expert Review. During the first project activity in spring 2013, six experts in the instruction of students with visual and motor disabilities analyzed prototype TE items for accessibility barriers and provided suggestions to overcome the barriers. Reviewers received Universal Design for Computer-Based Testing (UD-CBT) Guidelines (Dolan, Burling, Rose et al., 2010) and Introduction to Webb's Depth of Knowledge Levels (Tennessee 3-5 Grade Band Training, ND). In terms of universal design and the accessibility of TE items, reviewers were encouraged to comment on six processing barriers that students with disabilities may encounter in online assessments using innovative items and tasks. Reviewers suggested accommodations to support access, and they provided their thoughts about how to deliver item content in alternative ways for these students. Reviewers were also asked to estimate the cognitive complexity that might be measured with various TE item formats.

Reviewers expressed concern that many original forms of TE item types were not accessible for blind or visually impaired students. Due to the visual nature of many TE items (e.g., interacting with a graphic, moving objects on a screen), students with low vision would have difficulty accessing many of the items. However, reviewers provided suggestions to allow students with visual impairments to navigate the computer-based testing environment independently.

Similarly, students with motor disabilities may have difficulty navigating some TE items due to limitations of assistive technology devices. Reviewers recommended several types of test items and tasks that would be more accessible for students with motor impairments. The use of these item types maintains the intended cognitive demands while allowing students with motor disabilities to interact with computer-based testing systems.

Teacher Panels. In the next round of project activities in fall 2013, teachers of students with vision and motor disabilities interacted with an item set that included original TE items and alternative versions online, in print, and in braille. These items encompassed the same general types that the experts had reviewed and included new item types developed after the expert review. These teacher panels occurred in five of the partner states and included 74 teachers, 5 occupational therapists, 1 orientation and mobility expert, and 25 special education administrators and state department of education staff.

Panelists agreed on the lack of accessibility of many of the TE item formats, primarily the drag-and-drop item types. Panelists recommended drop-down menus as accessible alternatives, particularly for students with low vision or motor impairments. Teachers commented on limitations of accommodations for TE items, as well. For example, audio delivery may be confusing for blind students, particularly when screen readers are misled by item layouts due to text boxes, columns, or lines on the screen. Braille-reading panelists also provided valuable criticism of the first pass at braille item presentation, convincing project staff that an entirely new approach was needed.

Cognitive Labs. Concurrently with the teacher panels, students of all ages in three states participated in cognitive labs in fall 2013. Twenty-eight students participated in cognitive labs with their teachers or ATEA project staff. In addition to student comments, teachers provided qualitative feedback on student interactions with online items adapted for accessibility and with items formatted in print and braille booklets.

Students themselves provided invaluable commentary on braille presentation, along with the most useful recommendations for improvement in item layout and numbering. Students also experimented with screen readers and identified a number of problems that had been previously unexplored. One exciting outcome of the cognitive labs was that students who had sufficient vision to interact with the onscreen display enjoyed responding to items online even when they accessed item content, such as lengthy texts, via braille or large print documents. Their responses confirmed the engagement potential of TE items for students who had not previously had access to online testing.

Field Tests. Two field tests in 2014 provided quantitative data on original and adapted item types. Field tests occurred as part of the Kansas Assessment Program summative testing in spring 2014. Embedding pairs of items with identical content but different screen layouts and response requirements in statewide field tests permitted the comparison of alternative item formats with large samples of students who did not use accommodations. Encouragingly, data on matched pairs of TE items in original and adapted formats, as administered to students without accommodations, established good concurrence for item difficulty and discrimination. Key outcomes included these results:

- Matrix items, which are fully accessible online and on special forms, showed exceptional equivalence in item difficulty to drag-and-drop TE items.
- Matching items, which are more difficult to present in braille and cannot be used for optical scan answer sheets, also displayed excellent equivalence to original TE items.
- Matrix and matching formats tended to provide better IRT-based item discrimination statistics than original TE formats.
- Item content tested in both matrix and matching formats was virtually identical in difficulty and discrimination.
- Drop-down items were slightly easier than traditional selected response items, particularly at lower levels of item difficulty. However, limitations to drop-down formats were revealed. Drop-down formats may not be fully accessible for online tests due to potential difficulties with audio presentation and switches. Drop-down items cannot be presented in paper-and-pencil or braille tests, though they can often be transformed into selected response items, which are fully accessible.
- Selected response items showed slightly better discrimination than drop-down items.
- Of 61 adapted TE items, only four showed moderate differential item functioning (DIF) and none showed large DIF.

Item Tryouts. Another series of smaller scale item tryouts took place in fall 2014 in ten states outside of Kansas, where students with vision and motor disabilities responded to adapted TE items with the goal of enabling statistical comparisons with matched samples of students without these disabilities. Because item tryouts took place early in the school year, these students took tests one grade level lower than their current grade placement. Their results were combined with those of Kansas students with vision and motor impairments who had participated in spring 2014 field tests.

Unfortunately, due to difficulty reaching and enrolling sufficient numbers of students in these lower incidence populations, the goal of statistical comparison was not attained.

Nevertheless, qualitative evaluation of items adapted for print and braille forms led to general conclusions about the accessibility of adapted TE items for these students:

- A total of 109 students with vision and motor disabilities in grades 3 through high school responded to altered TE items on print and braille booklets.
- Between 10 and 21 students took tests at each grade and subject.
- Of the 98 adapted items presented in the field tests and item tryouts, 81% had *p* values greater than .3.

Final Expert Review. All six original expert reviewers and an additional reviewer evaluated the culmination of the ATEA project on the secure test items administered in the field tests and item tryouts. As before, their feedback provided excellent guidance on item accessibility for students with vision and motor disabilities as well as on glitches or issues with the technology platform that required further investigation.

The technology platform used in this project, the Kansas Interactive Testing Engine (KITE) developed by the Center for Educational Testing and Evaluation (CETE) at the University of Kansas, continues to undergo growth and improvement. Enhanced accessibility, comprising better screen magnification quality, increased delivery of tests and directions with audio presentation, and improved response with keyboard and switch systems, has been at the forefront of development. Furthermore, CETE continues to develop new types of interactive tasks for students, which will come with their own access issues and raise entirely new and unforeseen challenges.

Beyond KITE, other test delivery platforms offer different types of TE items with their own accessibility trials and successes. Therefore, conclusions from this project may not apply to TE tasks or interfaces in other settings or for other uses. As always, accessibility and equity for students with disabilities is an ongoing challenge, and continued research will always be needed. This project has only succeeded in uncovering the very tip of the iceberg regarding technology enhancement and innovative computerized assessment for students with vision and motor needs.

2.0 Introduction

Six major federally funded assessment consortia prepared large-scale assessments for use beginning in 2014-2015. These are the Partnership for Assessment of Readiness for College and Careers (PARCC), SMARTER Balanced Assessment Consortium (SBAC), Dynamic Learning Maps Alternate Assessment System Consortium (DLM), National Center and State Collaborative Partnership (NCSC), Assessment Services Supporting ELs through Technology Systems (ASSETS), and the English Language Proficiency Assessment for the 21st Century Consortium (ELPA21). A major influence on the development of new assessments is the Common Core State Standards Initiative (CCSS; 2012). These standards dramatically changed the conceptualization of instructional goals. The CCSS are affecting how progress toward and proficiency on these goals is measured for America's youth.

A key feature of many of these new assessments is their use of computers and touch-screen technology for test delivery and student response collection. The six consortia were committed to exploring or using TE item types to enhance the validity of inferences from test scores. TE test items, also called innovative item types, involve computer-based functions and tasks. These items may require students to move objects or words on the screen, highlight text, plot points or lines on a coordinate grid, or label objects. Because of these features, TE items present greater accessibility barriers than traditional multiple-choice items, constructed response items, or paper-and-pencil test forms.

Two of the more difficult accessibility challenges are visual disabilities and motor impairments. Visual disabilities impede access to information presented in visual modalities, such as computer screens. Motor disabilities affect students' interaction with physical interfaces, such as keyboards and mice. For students with visual or motor disabilities, TE items present barriers such as difficulty seeing or manipulating keyboards, mice, or touch screens to engage with item content and enter responses.

The six consortia use several guiding principles to help ensure a test's reliability, maximize validity evidence, and improve accessibility for all students. These guiding principles include evidence-centered design (ECD) and universal design (UD). Under UD principles, accessibility is integral to the item/task development process. Using ECD principles, the consortia develop items/tasks so that the evidence gathered from a test (i.e., a student's test score or the observation of a student's behavior) relates directly to the knowledge and skills that the test aims to measure (Mislevy, Almond, & Lukas, 2003). ECD and UD principles will be described in more detail in the Conceptual Framework.

The purpose of the Accessibility for Technology-Enhanced Assessments (ATEA) project was to investigate the accessibility of next-generation computerized assessments such as those under development by the assessment consortia and test developers. The primary goal of the ATEA project was to determine whether or not computerized assessment items and tasks are actually accessible for students with blindness, low vision, or motor disabilities or if these items/tasks can be made accessible so that inferences from test scores for these students are valid and comparable to those of other students.

The activities described in this report are the series of tasks defined for the ATEA Enhanced Assessment Grant. The ATEA project obtained descriptions and prototypes or samples of the proposed TE assessment items and task types under development by the major consortia. The project reviewed the principles of ECD and UD used in item/task preparation by these consortia and methods intended to provide accessibility. Expert reviewers and teachers scrutinized item and task types to determine if they were accessible for students with vision and/or motor disabilities. For item and task types that were not accessible, this project examined whether viable and valid alternatives exist for the measurement of the target constructs for these students. Assessment items and tasks for both English language arts (ELA) and mathematics content were written and administered to students with and without vision or motor disabilities. Individual cognitive labs provided crucial information about the accessibility of technology-enabled features and accommodations for students with vision or motor disabilities. Large-scale data collection offered the opportunity for the ATEA project to assess score comparability of alternative item formats and layouts for students without disabilities. Outcomes include a catalog of specific accessibility recommendations organized by cognitive task and item type, exemplified by sample items for students with vision and motor disabilities.

Following a review of the current literature on assessments for students with vision and motor disabilities, this report describes the ATEA project's review activities and its teacher and student participants. The outcomes of the qualitative activities in the ATEA project are organized by TE item type. Discussion of the feedback from experts, teachers, and students defines the barriers presented by each TE item type and the steps taken to mitigate those barriers. Sample items in math and ELA illustrate the challenges of TE tasks at each grade level. Next, a description of the population of student participants provides detailed information about the characteristics of students targeted by this project. Finally, the report presents the results of student performance on field tests and item tryouts. These results elucidate the levels of accessibility and score comparability attained by changing item format in computerized tests as well as print and braille hardcopy tests.

2.1 Investigation of Accessibility for Technology-Enhanced Assessments

All students, including those with disabilities, must participate in challenging assessments of academic achievement, via either a general assessment or an alternate assessment. Students who are blind, who experience low vision, or who have motor disabilities make up a small proportion of participants in large-scale testing efforts. Less than 1% of children under 18 are blind or have low vision that is not corrected by eyeglasses (Leonard, 2002). About 1% of students experience physical disabilities, though the proportion of those who have arm and hand limitations that affect computer access has not been reported separately (National Science Foundation, 1996).

Traditionally, assessments have been altered extensively for students who are blind, who have low vision, or who have motor disabilities. An unspoken assumption may be that assessments are adequately accessible with assistive technology or the provision of alternate forms such as hard-copy braille tests. However, these students may be among the most difficult to accommodate with TE assessments. The visual nature of computerized assessments introduces barriers for blind students who read braille or students with low vision who require magnification. Students who read braille routinely use braille hardcopy assessments, and extra time is usually allotted for their completion because braille may take longer to read than printed text. Braille has also been delivered via refreshable braille displays, but these are unavailable to some schools and individuals because they are expensive and can be unreliable (American Foundation for the Blind, 2012b; Kamei-Hannan, 2008). Screen size, especially the smaller size of new touch-screen tablets, limits the amount of enlarged text that is visible without spilling over the edges of the display. This situation requires greater user memory and navigation capacity, especially if there are large areas of blank space (Kamei-Hannan, 2008). This challenge has been termed the "field navigation problem" (Zwern & Goodrich, 1996).

Assessment alterations are also common for students who have motor disabilities. Standard keyboards and mice raise challenges for students with motor disabilities unless they use assistive technology that plugs directly into the computerized assessment system. Older or less sophisticated assistive technology, such as that used in many schools (Brodin, 2010), may still require the assistance of a human facilitator to enable the student to respond to assessment items and tasks. Finally, cost can be a barrier to obtaining access to assistive technology (Uslan, 1992).

Assessment administrators may experience additional demands beyond the delivery of accommodations required by students. Responses to hardcopy braille versions of online assessments may require hand entry by teachers. When assistive technology is used, assessment of students with motor disabilities may necessitate human intervention for the purposes of delivering test items or obtaining and recording responses. Unlike research on common accommodations such as extra time or reading test items aloud, there is a paucity of research on the effects of these accommodations on test scores and score comparability. Computerized assessments with technology enhancements for both general and alternate assessments provide the means to investigate these issues. This project evaluated item accessibility for students across the ability continuum who would likely benefit from the standardized administration of online assessments and their associated accommodations, as well as students who require off-line assessment via hardcopy paper and pencil or braille test booklets.

Students with vision and/or motor disabilities, like other students with disabilities, have lower levels of post-secondary education and employment than do nondisabled learners (American Foundation for the Blind, 2012a; Capella-McDonnall, 2005). Educational failure for students with blindness, low vision, or motor disabilities, however, may not simply manifest in poor grades or school dropout, but may come in the form of limited opportunity that translates into the inability to transition from even a successful school experience to a career. Individuals with blindness are much less likely to interact with computer technology than are sighted people (Arlene R. Gordon Research Institute, 2012). At the same time, vocational opportunities for students with blindness would improve with higher levels of technological sophistication (Armstrong & Murray, 2010). Computer technology is equally vital for students with motor disabilities, but they may not find it sufficiently available in schools due to the lack of resources as well as the lack of staff's technical knowledge and competence (Brodin, 2010).

In addition, families with disabled members experience greater economic disadvantage than do families without disabled individuals. Families with disabled members require greater income to experience the same standard of living as comparable families without disabled members (Fujiura & Yamaki, 2000). Conversely, poverty increases the risk for disability (Rosano, Mancini, & Solipaca, 2009). Addressing the technological needs of students with disabilities will improve educational opportunity now, post-secondary vocational opportunity, and the ability to become employed, productive members of society later on.

Investigating methods by which students with vision or motor impairments can interact with innovative TE assessments improves the quality of inferences that can be made about performance and informs educational planning. Moreover, matching student needs to new assessment technologies enhances the identification and development of technology-based instructional supports and adaptations that advance educational and vocational opportunity. The ultimate result of this study is a set of guidelines and recommendations for valid accessible assessments that can provide the greatest score comparability and lead to sound inferences about achievement for students with vision and/or motor disabilities.

2.2 Significance

This investigation resulted in empirical knowledge about the requirements for access to TE assessments for students with vision and/or motor disabilities. The results of this project include a set of guidelines and recommendations for valid accessible assessments that provide the greatest score comparability and lead to sound inferences about achievement for these students as measured with TE items and tasks. These results were available to the six major assessment consortia through the participation of members of their technical-advisory committees on the National Advisory Board of the ATEA project and to member states through the project website. A major purpose of this undertaking is to identify the means to include students with vision and/or motor disabilities validly in assessments that are under development by those consortia and by other test developers. These new assessments represent a major shift in the application of technology and innovation to educational testing.

The project represents by far the largest and most significant effort to identify assessment access for students with vision and/or motor disabilities. Individualized cognitive labs, large-scale data collection, and focused assessment of students with vision and motor impairments, delivered evidence about the equivalence of item content presented in formats accessible to all students. These outcomes pertain to the students who have historically received the most individualized, and hence non-standardized, accommodations, probably the most difficult students for whom to ensure assessment access and adequacy.

The contribution of this project to knowledge about assessing students with vision and/or motor disabilities is novel in several respects. No previous studies have evaluated presentation or response requirements for innovative computerized test items. Accommodations for these new types of assessment tasks, whether online or offline, have not yet been developed or analyzed. While paper-and-pencil or braille tests have been accepted accommodations for blind and low vision students, adapting TE items for presentation on braille or print test forms is a new challenge. Finally, no studies have appraised score comparability for items in alternate formats, which will be required for online accommodations and supports and for offline delivery.

The outcomes of the ATEA project include guidelines and recommendations for technology-enabled accessibility features, tools, and accommodations. Alternative item formats were developed and compared for item and task types that were not accessible to students with vision and motor disabilities in their original online formats. Adapted items allow enhanced accessibility for online assessment supported by audio presentation or switch responses as well as the potential for delivery in offline tests for students whose instruction relies on paper-and-pencil or braille modalities. Student responses to sample TE items revealed the potential for students to engage with online TE items using paper-and-pencil or braille test forms as supports instead of as the primary testing format.

Finally, beyond access to computerized assessments, the investigation of technology-supported accessibility also added to knowledge about student requirements for ongoing educational access to computerized content and activities. Learning the means by which these students can interact with innovative technology improves the quality of inferences about performance and informs educational planning.

3.0 Conceptual Framework

Accessible assessments measure the same knowledge and skills as traditional assessments but without interference from students' disabilities. Accessibility must be evaluated empirically, but the goal of accessibility is valid measurement of the intended skills for all students (Russell et al., 2011; Thurlow et al., 2009). Accessible tests are constructed using ECD and UD principles, as well as available technologies and accommodations as needed to reach all students (Thurlow et al., 2009). ECD is a methodology for developing assessments in which the student response that demonstrates mastery of knowledge or skills is defined before assessment items/tasks are written (Mislevy & Haertel, 2006). Assessment items and tasks are prepared according to templates or patterns designed to define the relationship between instructional goals, student characteristics, and desired outcomes.

Figure 3.0.1 displays a simplified version of the ECD framework. ECD follows a "backwards design" in which test developers first specify the broad-level content domains (content standards). Based on the content standards, test developers outline the claims (i.e., inferences) they wish to make about students' knowledge and skills (i.e., the assessment target). Test developers then outline the specific types of evidence they wish to elicit from students, or what students can do to show evidence of their knowledge and skills. Items/tasks that will elicit the targeted responses from students are then designed.

Figure 3.0.1. Evidence-Centered Design Framework (Smarter Balanced Assessment Consortium, 2012).

ECD methods are crucial for students with disabilities because their performance may be constrained by a disability condition that limits interaction with test items or tasks and may alter outcomes. The six assessment consortia base their item and task development on ECD principles. For example, to keep item/task development in line with the ECD framework, ELPA21 item writers use item specifications that relate to the types of evidence they wish to elicit from students. In the development of the DLM Alternate Assessment System, the DLM consortium uses a framework that is “a variant of ECD” (Dynamic Learning Maps Consortium, 2013). In the development of the NCSC Alternate Assessment based on Alternate Achievement Standards (AA-AAS), ECD principles were the foundation of task development for an item family, which was designed to cover a range of abilities and levels of cognitive and communication skills (NCSC, 2015).

Universal design is a second essential underpinning for accessible tests. UD encompasses the general idea that features that improve access for groups of individuals, such as those with certain disabilities, may improve access for all in ways that are unforeseen. Therefore, minimizing barriers of all kinds should be a guiding theme in the development of new products. Initially, the Center for Universal Design (1997) conceptualized UD as a set of guiding principles for architecture that encompassed ease and flexibility of use, perceptibility of information, adequate size and space, tolerance for error, and low physical effort. As an example of how UD may benefit many, even when designed for a few, curb cuts now allow ease of access for strollers and rolling luggage in addition to wheelchairs. These principles have propagated into other fields, including learning and assessment.

According to the Center for Applied Special Technology, UD for learning involves three networks: recognition networks for the “what” of learning, strategic networks for the “how” of learning, and affective networks for the “why” of learning (CAST, 2011). CAST’s recommendations for UD for all learners include providing multiple means of representation, multiple means of action and expression, and multiple means of engagement to access the three learning networks respectively. Representation consists of perception, language and symbol use, and comprehension of instructional materials. Representation encompasses alternate presentations to match the learner's perceptual and receptive communication skills and abilities. Action and expression include expressive communication, physical action, and executive functioning, which comprise alternate methods of expressing what the learner knows and can do. Engagement refers to maintaining interest, effort, and persistence with learning tasks along with self-regulation. This learning network is closely involved with motivating and encouraging optimal responses from the learner consistent with the learner's ability to perform. This conceptualization of UD is equally applicable to assessment.

Building on UD for Learning, the National Center on Educational Outcomes (NCEO) developed UD guidelines for assessment (Thompson, Johnstone, & Thurlow, 2002, p. 6):

1. Inclusive assessment population
2. Precisely defined constructs
3. Accessible, non-biased items
4. Amenable to accommodations
5. Simple, clear, and intuitive instructions and procedures
6. Maximum readability and comprehensibility
7. Maximum legibility

As Thompson and Johnstone, et al. (2002, p. 5) explained, "universally designed assessments are not intended to eliminate individualization, but they may reduce the need for accommodations and various alternative assessments by eliminating access barriers associated with the tests themselves". Therefore, UD as a foundational premise does not eliminate the need to make individualized accommodations or adaptations for students who need them. NCEO (2011) has recently addressed the relationship between technology-based assessments, accommodations, and UD:

> Technology-based assessment platforms offer new opportunities and ways for accommodations to be provided to students who need them, but they will not

> eliminate the need for accommodations. Technology-based assessments can be developed with all students in mind from the beginning (universal design) so that the assessments are accessible to the greatest number of students right from the start. Yet even with the best-designed test, some students still will require accommodations.

Furthermore, while technology-based assessments offer enhanced opportunities to meet individual needs through built-in options for accessibility, the need for additional accommodations may be created.

> Decisions [sic] makers also should be made aware of which accommodations may need to be provided in addition to those embedded in the assessment. . . Technology-based tests may create a need for new accommodations. For example, students with some physical disabilities that affect coordination may be able to take a paper and pencil test without accommodations, but may need accommodations to navigate a technology-based assessment. In addition, some technology-based assessments may require the use of more working memory than paper-based tests. For example, less information may be visible on a screen than on a page in a test booklet (NCEO, 2011).

TE assessments developed with UD principles and features cannot be viewed as a panacea for the assessment of students with disabilities. Further effort is needed to identify the supports that will be required for valid inferences about the achievement of students with vision and/or motor disabilities, even with next-generation technology platforms for assessment.

Technology-enabled accessibility and accommodations. One of the exciting new advances in technology is the development of the Question and Test Interoperability (QTI2) and Accessible Portable Item Protocol (APIP) standards by the IMS Global Learning Consortium (2012; Russell et al., 2011). The procedures and functionality described by these standards aim toward achieving interoperability of assessments on different technological systems for the seamless performance of items and tasks and their interface with student access profiles across platforms.

Concepts integral to QTI2 and APIP are default, alternate, and supplemental content (Russell et al., 2011). Default content is the test item or task as developed for presentation to students without specific access needs. Alternate content refers to alternate representations of an item or task to meet individual needs, such as presentation of a translated item or alternate forms of graphics or images. Supplemental content is additional content, such as braille text or audio files, that is available in addition to the default content. Under the model of the APIP standards, an item file would contain pointers to alternate content and embedded supplemental content to be accessed upon demand or as triggered by information contained in an individual test taker’s Personal Needs and Preferences profile.

Assessments developed using QTI2 and APIP standards are also expected to use UD principles and procedures in order to minimize the alterations necessary for individualized access. This means that alternate and supplemental content to meet individualized access needs should be defined before items and tasks are created, not as a post hoc activity. Furthermore, the order and type of the delivery of item content is specified during item writing. Finally, at the time of item presentation, access tools such as screen magnification, contrast, masking, and highlighting are made available for student use on demand. However, APIP and QTI2 standards do not currently address TE item types, although they are anticipated to be flexible as new item types are developed. Therefore, even with QTI2 and APIP, there is much room for technological enhancement and additional accommodation.

Accommodations, including technology-enabled access features, can be categorized as presentation, response, setting, scheduling, and special tools options (American Foundation for the Blind, 2005). Setting and scheduling accommodations refer to the environment in which and timing when a student takes a test, such as a quiet or private location, shorter or multiple test sessions, and frequent breaks from testing. Onscreen text, a printed page, and a page of braille are presentation options. Writing by hand, typing on a keyboard, or using a braillewriter are response options (Christensen, Braam, Scullin, & Thurlow, 2011). Students with blindness or low vision may use both presentation and response accommodations when they access braille test booklets or refreshable braille displays, respond orally to a scribe, or use a braillewriter. Students may also require the use of tools, such as screen magnification, an abacus, a braille ruler, and tactile graphing materials. Students with motor disabilities who do not experience vision disabilities frequently rely on response options involving individualized assistive technology that are matched to their motor skills and age- or grade-level needs, either for producing a response for a scribe or as a direct interaction with a computerized system.

Accommodations can be categorized into methods by which they are made available to students. A teacher may deliver accommodations such as signing into the hands of a deaf-blind student or provide that student with mathematics manipulatives or special tools, such as raised-grid graph paper, a braille ruler, or a compass. Computer or technology-delivered accessibility tools include magnified onscreen font size, increased contrast, and auditory calming, which may include calming music or white noise delivered through headphones. These tools are delivered to students along with alternate or embedded content as described by the QTI2 and APIP standards. An individual's assistive technology, such as an eye-gaze system, switches, or enhanced keyboard, can function as a technology-delivered response-option accommodation when linked directly to the computer. If an individual student's assistive technology system is of lower technological sophistication, it may serve as an interface with a human facilitator who then records the student's response on the technology platform.

Assistive technology. According to the Individuals with Disabilities Education Improvement Act of 2004, section 300.5, "assistive technology device means any item, piece of equipment, or product system, whether acquired commercially off the shelf, modified, or customized, that is used to increase, maintain, or improve the functional capabilities of a child with a disability.” Given this broad definition, and the constant modification, improvement, and enhancement of assistive technologies, the topic of assistive technology in assessment settings is a moving target.

Watts, O’Brian, and Wojcik (2004) reviewed the state of assistive technology in student assessment and concluded that, “due to the limited research in the field of assistive technology in school settings, the cross-referencing of assistive technology practices with educational assessment reveals an imbalance between a historically rich research base and one that is relatively new. Only recently has the field of assistive technology begun to contemplate the issues of applying educational assessment practices to the assistive technology consideration process.” In 2004, Watts et al. approached the selection of assistive technologies for students with disabilities from the perspective of the demands of educational assessment. Now, with ten years of advancement in the form and purpose of educational assessment, the reverse consideration applies: how can contemporary, particularly computerized, educational assessments be formulated and delivered so that students who use various assistive technologies and methods can experience equitable access? The myriad tools that students with disabilities use must be part of the consideration of access and validity going forward. However, the remainder of this literature review references assistive technology as an overarching concept without further definition.

3.1 Students with Vision Disabilities

Common standardized testing accommodations for low vision or blind students include braille, large print versions of the test, assistive magnifying devices, and teacher scripts for reading aloud (Landau, Russell, & Erin, 2006). For students with low vision who require magnification, standard large print size is 18 point font (Allman, 2006), though students may need to magnify text well beyond that size. While students with vision disabilities are expected to meet the same standards as other students, some of the test may need to be modified in order to be translatable to braille format. Modifications may include word substitutions, reformatting the layout of the item, and replacing untranslatable items with others of equal weight, content, and difficulty (Allman, 2006).

Tactile graphics are raised images that can be deciphered by the braille reader in order to gather the same information a sighted reader would get from an image (Hasty, n.d.). While some images, such as photographs, are not generally effective as tactile graphics, many diagrams and figures can be successfully included in an assessment (Hasty, n.d.). As discussed by Beck-Winchatz and Riccobono (2008), certain fields, such as science, rely more on visual representations of information and this reliance can create a disadvantage for students with visual disabilities. Some organizations are attempting to improve the situation by developing tactile books and curriculum materials for these students; however, tactile graphics can present difficulties for individuals if not developed properly. A literature review by Lebaz and Picard (2012) found that identifying tactile graphics could be a difficult task with much variation in accuracy. The review found that a few characteristics that tended to improve accuracy were less complex images, prior semantic information about the images, guided exploration, and heat sensitive paper instead of plastic film. In addition, images must have a high degree of contrast to be interpretable by students with low vision (Allman, 2006). Individuals may also be better able to discriminate, identify, and comprehend images when emphasis is on the boundaries of images, as in raised-line graphics (Krufka & Barner, 2006). Issues include the complexity of the image, access to supplemental information about the graphic available in text, and the methods used by the individual to explore the image.

Other challenges and limitations of tactile graphics include the development of the graphics and the subsequent training that professionals receive. Issues in the development of tactile graphics include an understanding of how the tactile graphic will be used by children with visual impairments. The purpose of tactile graphics is to communicate an idea or information, not to reproduce a visual image in tactile form; a tactile graphic cannot be as complete as a visual picture and thus cannot be understood as instantly and completely as a visual image. Tactile illustrations present unique issues for children with visual impairments because tactile graphics are interpreted piecemeal, not at a glance as visual pictures can be. The viewing area is limited to what is beneath the fingertip; a child using a tactile graphic must think about all of the separate parts and put them together cohesively in order to understand the task. Additionally, a visually impaired child's experience with a tactile image is much different from a child's experience with the object itself. This presents a challenge to visually impaired children who must formulate an image of the object in their minds. Thus, it is crucial that all objects be placed in context to minimize confusion (Wright, 2008).

Rosenblum and Herzberg (2011) found that people trained to prepare mathematics materials for tactile graphics did not feel as though they had enough time and were not able to proofread them for accuracy. Preparers in the sample also reported feeling as though they did not have the appropriate amount of developmental training, particularly in braille, for their students. Many individuals in the sample were also the only people in their school districts qualified to create materials for low-vision students and thus felt pressured to produce those materials. Another issue with preparation concerns matching braille textbooks with print graphics. In a study by Smith & Smothers (2012), the researchers found that in the math and science textbooks used in the sample, there were numerous discrepancies with the braille books, such as omission of tactile graphics. This may be more of an issue for textbook developers, but it is a valid concern that should be recognized and confronted if tactile graphics are going to be used appropriately.

With the increasing use of computerized testing platforms comes a new wave of TE tests with computer-provided accommodations and supports. A 2008 study by Kamei-Hannan investigated the accessibility of the widely used, computer-adaptive Measurements of Academic Progress (MAP) test developed and distributed by the Northwest Evaluation Association. In this study, vision participants were offered screen magnification software as a test accommodation. The results revealed that increased magnification levels increased item completion time. This finding is supported by the most recent National Center on Educational Outcomes report on the effects of test accommodations (Cormier, Altman, Shyyan, & Thurlow, 2010). In addition, because such a small amount of text is shown on the screen at a time when the magnification level is high, the study participants needed to have strong visual efficiency and hand-eye coordination skills to navigate the text.

Blind participants in the Kamei-Hannan study accessed the assessment with refreshable braille displays. Refreshable braille displays translate on-screen text to the tactile braille format one line at a time using moving pins (Abbott, 2005). The study revealed limitations of the refreshable braille software due to the original HTML coding of the test items. Many of the items were unanswerable because of untranslatable HTML elements such as long scroll bars, images, and underlined words (Kamei-Hannan, 2008). Another limitation of the refreshable braille displays is their cost. Each unit costs between $3,500 and $15,000 depending on its complexity and number of characters it can display at once (American Foundation for the Blind, 2012b). According to the Oregon Department of Education website (2012), the state of Oregon has crossed these barriers of refreshable braille displays and currently offers real-time braille accommodations for their adaptive statewide achievement tests. The items are printed as students work through the test using refreshable braille displays in conjunction with braille embossers. If the item requires Nemeth code or tactile graphics, it is automatically sent to the braille embosser. If not, it is displayed on the refreshable braille delivery system (Oregon Department of Education, 2012).

Oral reading of test directions and other allowable portions of the test by a person, audiocassette, CD, or a computer text reader are common ways of increasing test accessibility for students with visual disabilities. This method can be employed in an effort to reduce test-taking time for blind or visually disabled students (Allman, 2006). A small study by Kim (2012) examined the effect of providing a read-aloud accommodation for a reading comprehension test to students with visual impairments and students without visual impairments. The results indicated students with visual impairments given the read-aloud accommodation had higher scores than those given other accommodations (e.g., more time, braille, or large print). There was not a significant difference between the test scores of those with and without the read-aloud accommodation for students without visual impairments (Kim, 2012). However, issues with the mode of delivery are encountered when the construct being tested can no longer be credibly measured with a reader, for example, reading comprehension or silent reading skills.

Other than the reading accommodations mentioned above, technology advancements have brought new accessibility solutions for students with vision disabilities through listening. In 2002, researchers Hansen, Lee, and Forer at Educational Testing Service (ETS) examined the effectiveness of a self-voicing version of the Test of English as a Foreign Language (TOEFL) for test takers with visual disabilities. A self-voicing test has the audio embedded in the test delivery as opposed to using a peripheral text-to-speech program. Along with the built-in audio for item text, the test platform had audio descriptions of images and navigation cues. To navigate and respond to the items, the subjects used simple keystrokes on a keyboard[1]. The benefits of this self-voicing TE test included increased independence, standardization of delivery, and privacy for the test taker. Most of the study's participants indicated that they would "highly recommend" this type of accommodation over a human reader. One concern that came out of this study was the quality of the electronic voice. Some words were difficult to understand, which is especially problematic for the population taking the TOEFL (Hansen et al., 2002).

Another interactive listening solution is the Talking Tactile Tablet. In 2006, results of a study conducted by Landau, Russell, and Erin indicated the usefulness of the Talking

---

[1] There are many available keyboard modifications for low vision or blind students, including large print and tactile braille stickers for the keys.

Tactile Tablet as a test accommodation for students with vision disabilities. The tablet is a separate device that allows the user to interact with the computer display of standardized test items and is especially useful for items with complex graphics. Students use their hands to navigate the test on the tablet and they can cue audio voice recordings to hear about text, features, and graphics in the items. Students can replay the voice recordings as many times as needed. While the study did not find that this tool significantly affected test scores for students who were blind or had low vision, the researchers suggested that its benefits to the intended population included increased speed of test completion, increased independence during testing, and increased standardization of test delivery (Landau et al., 2006). Hansen, Shute, & Landau (2010), discussed students' positive reports after using the Talking Tactile Tablet. Students stated that the voice was clear, they liked having the audio with the tactile graphics, and that the overlay sheets were easy to switch. However, development of technology alone is not enough; schools must also give students with visual disabilities opportunities to utilize this technology.

In the development and pursuit of assistive technology for students with visual disabilities, accommodations will need a level of usability that is appropriate for school-aged children. A secondary analysis of the National Longitudinal Transition Study-2 found that over three data collection years (2000-2001, 2002-2003, and 2004-2005), only 42%, on average, of high school students who read braille or large print were using high-tech assistive technology (Kelly, 2011). High school students with visual impairments might be at a disadvantage because of the lack of exposure and familiarity to assistive technology.

Three common test-taking tools for students with visual disabilities are physical manipulatives, talking calculators, and abaci. Physical manipulatives, such as blocks, money, and geometric shapes, may convert some information that is visually represented on the test into accessible physical representations (Allman, 2006). A talking calculator is a tool designed specifically for low vision and blind students. It articulates everything that the user punches into it to ensure accuracy and reads all output (Learning, Sight & Sound Made

Easier, 2011). Simple talking calculators are affordable and are available for purchase at around $12 to $25 per unit (LS&S, 2011). On items that do not permit calculators, blind students can use abaci as substitutes for paper-and-pencil calculations (Allman, 2006).

Although changes in testing platforms and accommodations are being explored, state testing policy and procedure manuals do not appear to currently address or include these developments. While all states acknowledge the necessity of braille and large print forms for their students with visual disabilities, less than half of states allow other accommodations such as a talking calculator or a personal magnification device. However, in 2012, over 30% of states had policies that did not require them to state specific accommodations for students with visual disabilities. This policy allows flexibility in choosing which accommodations to use (Amato & Smith, 2012).

3.2 Students with Motor Disabilities

Accommodations for students with motor disabilities generally involve manipulating the means of student response, rather than the presentation of the test. The Minnesota Department of Education (2009) suggested the following response accommodations for students with motor disabilities: "Express response to a scribe through speech, pointing or by using an assistive communication device [such as a mouth stick or head wand (Thompson, Thurlow, & Moore, 2003)], voice-activated computers, type on or speak to word processor, speak into tape recorder, or use thick pencil or pencil grip" (Minnesota Department of Education, 2009, p. 37). Thompson et al. (2003) suggested individualizing the setting of computerized test taking if the response input method could be distracting to other students. Beyond the information presented in this paragraph, research that details appropriate assessment accommodations and their effects for students with motor disabilities could not be identified. While the field of assistive technology is highly developed, evaluation of its use for assessment beyond the individual level is nonexistent.

3.3 Score Comparability of Accommodated Assessments

According to the New York State Education Department (2006), the purpose of test accommodations is to make assessments accessible to students with disabilities. Accommodations are not intended to modify the tested content or give an advantage to any group of students (New York State Education Department, 2006). To ensure that the accommodations are achieving their purpose, it is important to understand their effects on students' scores.

Ideally, an appropriate test accommodation raises the score of a qualifying student with disabilities while having no effect on the scores of students without disabilities. This phenomenon is what Sireci, Scarpati, and Li (2005) called the *interaction hypothesis* in their analysis of 150 research studies on the effects of test accommodations. They discovered that the majority of the studies reported score gains for all students due to test accommodations but with significantly greater gains for students with disabilities. This finding is consistent with Fuchs and Fuchs's (2001) concept of differential boost. Sireci et al. argued that the finding that scores for all students tended to improve does not imply that test accommodations are unfair but perhaps that current testing conditions are too strict for all students. More specifically, Sireci and colleagues (2005) identified differential boosts for the extended time and oral presentation accommodations. The authors found that receiving oral presentation for the math section of the tests (whether from a person, a computer, or an audio device) significantly improved the scores for students with disabilities, but this accommodation had no effect on scores in other subject areas (Sireci et al., 2005). Due to the great diversity in students with disabilities and the types of accommodations they receive, there is limited scholarly research dedicated to generalizing the effects of test accommodations.

The use of alternative formats and accommodations for individuals with visual disabilities raises concerns about fairness and validity. Specifically, two groups of students (e.g., students with and without visual disabilities) with matched ability should possess the same probability of answering a test item correctly. Presence of significant difference in this probability is indicative of differential item functioning (DIF). Zebehazy, Zigmond, and Zimmerman (2012) investigated DIF of test items on Pennsylvania's Alternate System of Assessment (PASA) for students with visual impairments and severe cognitive disabilities and attempted to identify reasons for any noted differences. Results indicated DIF among the functional vision groups when compared to a matched group of sighted students. Researchers noted 17 instances in reading and 22 instances in math, with 14 skills in reading and 13 skills in math emerging as harder for students with visual impairments. Potential reasons for the differences included the following: the need for better orientation to the test materials; the influence of lucky guesses based on distractor characteristics; and the influence of accommodations, such as the substitution of objects.

In contrast, Stone, Cook, Laitusis, and Cline (2010) conducted a study of students at grades 4 and 8. They used DIF procedures to compare item-level results of students without vision disabilities to results of students with visual impairments. The students without vision disabilities took a test under standard conditions, and the students with visual impairments took the same test with large print or braille formats. Only one item at each grade showed large DIF favoring students without visual impairments, supporting the accessibility and validity of alternate formats for students with visual disabilities.

As statewide assessment continues to dictate classroom instruction, there is a need to enhance the use of assistive technology and subsequent testing procedures. However, there are several issues, such as the paucity of research and the bias and misconceptions of teachers regarding the fairness of using accessible accommodations in testing, which have impeded the usage of these accommodations. Several strategies that will aid in this endeavor are to increase knowledge regarding accessible accommodations and their uses with students with disabilities, to educate professionals to increase the connection between curriculum and technology, and to improve the dissemination of key information to the appropriate audiences (Parette et al., 2006).

Need for more research and development of accessible assessments. While instructional technology has boomed in the classroom, TE assessments have progressed more slowly (Bechard et al., 2010). While many states offer computerized, large-scale, standardized assessments, the forms are parallel to their paper-and-pencil counterparts in presentation and item type. In 2010, Bechard and colleagues published a research agenda for technology-enabled assessments as a result of their Invitational Research Symposium on Technology-Enabled and Universally Designed Assessments. The agenda highlights the need for development of technologically advanced, interactive assessments, and draws attention to the importance of validity research, especially for students with disabilities.

The ATEA project addressed accessibility requirements for vision and/or motor disabilities through both technology-delivered methods and offline assessment using paper-and-pencil or braille tests. As recognized by the NCEO (2011), technology-enabled access features may not be sufficient for full access for students with vision and/or motor disabilities. This investigation considered technology-enabled accessibility enhancements, such as those designed for APIP-compliant systems and accommodations options currently in use. As the variety of assistive technology methods for students with motor disabilities is vast and constantly changing, a catalog of those individualized methods is beyond the scope of this report. However, methods in use by participating students were included as part of this project.

3.4 Scope of the Accessibility for Technology-Enhanced Assessments Project

This project resulted in a clearer understanding of the types of TE tasks that are and are not accessible for students with blindness, low vision, or motor disabilities. Research questions for this project included:

a. What types of TE test items or tasks in ELA and mathematics are proposed or currently under development by the major assessment consortia?
b. Which items or tasks are accessible without alteration or accommodation to students who are blind or have low vision and to students with motor disabilities?

c. Are there constructs, items, or tasks that are not accessible to these students? If so, what is required to make these constructs, items, and tasks accessible? Are alternative methods available to assess the same or equivalent constructs?

d. How do altered or accommodated items and tasks perform in terms of construct measurement and score comparability?

e. What are the characteristics of students with vision and/or motor disabilities that affect their engagement with TE assessments?

To address these questions, the ATEA project first engaged experts in the fields of vision and motor disabilities to review prototype TE tasks. Next, cognitive labs with individual students and examination by panels of teachers assessed the accessibility of item types with existing technologies and planned accommodations. When accessible item alternatives were not feasible, as when an online item type mimicked a constructed response task, alternate methods for construct measurement were proposed and tested. Based on qualitative feedback from experts, teachers, and students, accessible online TE items were created by altering layout or presentation and response modalities without changing item content.

This project then investigated score comparability with original TE items through large-scale data collection in Kansas using students without vision or motor disabilities. Next, students with vision and motor impairments responded to assessment items and tasks in adapted formats presented on paper-and-pencil and braille test forms. Because of the nature of offline assessments, these items could not include technology enhancements but did match the layouts of the adapted online items. The project used the Kansas Interactive Testing Engine developed by the Center for Educational Testing and Evaluation (CETE) for data collection across the consortium member states. Finally, detailed information about the characteristics of this population of students was obtained via an online survey.

3.5 Use of Evidence-Centered Design by Assessment Consortia

A review of the test development processes provided by the six major federally funded assessment consortia revealed affirmation for the use of ECD in test development. The Partnership for Assessment of Readiness for College and Careers (PARCC) uses the ECD framework during assessment design and development and expects benefits that include strong validity arguments, ease of sharing expertise regarding item development, and long-term cost reduction through the use of reusable design tools (2013). The Smarter Balanced Assessment Consortium (SBAC) designed item specifications that align with the ECD framework with the goal of matching the content, rigor, and performance of the Common Core State Standards (2012). The Dynamic Learning Maps project incorporates the principles of ECD into item writing (2015). The NCSC assessment system depends on ECD as the foundation for developing alternate assessments based on alternate achievement standards (AA-AAS) in the conceptual, design, and existence proof phases (NCSC, 2015). The Assessment Services Supporting ELs through Technology Systems (ASSETS) project stresses its reliance on principles of ECD to operationalize English Language Proficiency (ELP) standards (2013). The English Language Proficiency Assessment for the 21st Century (ELPA21) relies on an ECD framework to ensure that items align with ELP Standards (2014).

3.6 Use of Universal Design Principles by Assessment Consortia

Publicly available literature for the six major assessment consortia also described their use of UD in assessment design. PARCC expressed its commitment to provide equitable access to their assessment (2015). They stated that their goal is to implement the UD principles throughout every stage of the assessment development process (initial design, item development, field testing, and implementation) and to minimize the need for individual accommodations. PARCC UD requirements include but are not limited to consideration of the diverse assessment population, clear instructions and procedures, and accessible features or format changes that maintain the construct and item difficulty (2014). SBAC reported that their development of accessible assessments involves an Access by

Design approach to maximize student access. Access by Design includes the application of UD principles in the assessment creation process (2012). By considering possible adaptations to meet access needs in the initial stages of task and item development, SBAC hopes to minimize and support specialized extensions. DLM uses UD item-writing guidelines in the development of assessments for students with significant cognitive disabilities (2015). NCSC described how UD was incorporated into the development of task templates and design patterns, which are elements of the ECD process for test development, to span levels of content difficulty in item families (NCSC, 2015). ASSETS intends to create accessible items by using the principles of UD with the support of computer-embedded resources that meet student needs (2013). ELPA21 commented that valid results are contingent upon students having appropriate access to the assessment (2014).

3.7 Definition of Technology-Enhanced Assessment Items

The ATEA project's definition of technology enhancement for assessment tasks will be useful before continuing to the presentation of prototype items and discussion of the qualitative feedback obtained about them. TE items introduce innovative features in several aspects of item presentation and construct measurement. A TE item should be more than simply a multiple-choice item presented on a computer screen with enhanced graphics. If the enhancements available with technology do not enrich item presentation and response or improve construct measurement, then the item might be better presented in a traditional selected or constructed response format.

The tasks comprising well-designed TE items cannot be duplicated with traditional item types. Technology enhancement can create a more interesting and engaging experience for the examinee by including greater access to content and resources and a variety of response options. Construct measurement can be improved through increased cognitive demands on the examinee, improved efficiency of measurement by incorporating the content of several traditional items in a single stimulus, or assessing broader or deeper aspects of the construct. Furthermore, universally designed TE items enable broader access via built-in accommodations and available tools and supports.

The ATEA project utilized the following definition of TE items synthesized from literature on the potential benefits of computerized testing (Jodoin, 2003; Parshall, Davey, & Pashley, 2000; Parshall & Harmes, 2009; Wendt, Kenny, & Marks, 2007). This definition will certainly require modification as TE items, tasks, and assessment continue to mature. A TE item need not contain all or even most of the attributes or potential benefits listed here. However, an initial project goal was to define crucial features of TE items that would distinguish them from traditional assessment tasks. This breakdown presents TE item characteristics in two major categories: item features available to the user during testing and potential psychometric benefits of TE items.

1. Item Features
    a. Presentation
        i. Computer-delivered features such as online tools, color, animation, alternative fonts, color overlays, reverse contrast, screen magnification, and interactive graphics
        ii. Supplemental content comprising linked or embedded audio and video files, data files, dictionaries, and other resources
        iii. Alternate content including braille-ready text or tactile graphics files, verbal descriptions of graphics, text-to-speech audio, language or keyword translations, and human or avatar sign language interpretation
    b. Engagement
        i. Opportunity to manipulate content interactively
        ii. Authentic, real-world experience and application
        iii. Motivation for the examinee
    c. Response

i. Demanding response requirements, such as manipulation of content onscreen, graphing, text entry, and multiple responses

ii. Alternate platforms such as touch screen tablets and assistive technologies in addition to keyboards and mice

iii. Capture of response latency, time spent on each item, and history of changed responses

2. Construct Measurement
    a. Sophisticated scoring procedures such as partial credit and other complex scoring algorithms
    b. Increased cognitive complexity
    c. Increased measurement efficiency

4.0 Initial Expert Review

The purpose of the expert review was to obtain preliminary feedback on the accessibility issues and barriers of prototype TE items availably publicly. The project used this feedback to guide the development of sample TE items for the teacher panels and cognitive labs. A benefit of the expert review was the chance to expose vision and motor experts to new types of online assessment tasks that their students would be facing. This exposure may have motivated the experts to enhance their involvement with assessment development in a way that they had not previously been able to do.

4.1 Expert Review Procedures

Six experts agreed to serve as reviewers for this first pass of analyzing accessibility for TE items. These reviewers included three experienced teachers of students with vision impairments (TVIs) working at public schools and the state school for the blind, one blind reviewer also certified as a TVI, one parent of a child with vision and motor impairments also certified as a TVI, and one occupational therapist working as a university trainer.

Expert reviewers evaluated prototype TE items drawn from several sources, including sample items prepared by CETE content area item developers. While the review focused on students with visual and motor disabilities, these experts had a diverse range of experience in working with students with a variety of sensory, linguistic, and cognitive challenges. The review aimed to capture the breadth of these reviewers’ knowledge and experience. Therefore, the project asked experts to consider six different categories of processing when considering each test item: perceptual, linguistic, cognitive, motoric, executive, and affective. Each of these processes is defined in Universal Design for Computer-Based Testing (UD-CBT) Guidelines (Dolan, Burling, Rose et al., 2010).

Perceptual processing includes vision, hearing, and other sensory modalities. Vision impairment was of greatest concern to this group. Motoric processing comprises patterns of action used in expression; these abilities are a diverse set of skills that were the second major focus of the ATEA project. Linguistic processing involves language development and use in any modality—oral, written, sign, and braille. Cognitive processing refers to intelligence and to the cognitive effects of perception and language on constructing meaning from interaction with the environment. Executive processing incorporates the abilities used to set and reach goals, such as engagement, attention, and motivation. Affective processing encompasses the interactions between the task and an individual's states and moods. Reviewers were given descriptions of TE items, a list of questions to consider, links to sample items and login information for the Kansas Interactive Testing Engine (KITE), and a blank review template for each item (Appendix A).

Depth of Knowledge. Reviewers also evaluated the level of cognitive demand (i.e., depth of knowledge) of the prototype test questions. Reviewers were provided with Webb's Depth of Knowledge (DOK) descriptors and were asked to identify the DOK level they thought most appropriately described the task. Table 4.1 shows cognitive demand levels with brief descriptions for each.

Table 4.1

Webb's Depth of Knowledge (DOK) Levels and Descriptors

Level	Description
Level 1 (Recall)	Recall a fact, definition, or term (rote response)
Level 2 (Skill/Concept)	Make a decision about how to approach a problem
Level 3 (Strategic Thinking)	Engage in reasoning, planning, etc.; support reasoning
Level 4 (Extended Thinking)	Engage in complex reasoning, planning, and developing; engage in thinking over an extended period of time

When considering accessibility, reviewers evaluated the item types to determine whether DOK was equitable for students across disabilities. Consistency of DOK is a major concern when items appear in different formats or with accommodations. If the DOK level is different for different groups of students, the measured construct may also differ. Experts recorded their evaluations of the items on templates labeled with the item name and key information about that item (Appendix A).

4.2 Expert Review Results

Experts reviewed example items of the following types using drag-and-drop, click-to-select, and constructed response interfaces. Because this was the first research activity of the ATEA project, no accommodations or altered formats had yet been implemented for these items. Sighted reviewers accessed items online. One reviewer, blind since birth, worked in partnership with a sighted reviewer to access and discuss the items and to provide recommendations for accessibility.

4.2.1 Drag-and-Drop Items

Categorization. Categorization items ask the user to choose the correct category for each element on the left side of the screen. In the drag-and-drop version, the user clicks or touches each element while dragging it into position and then releases it. Figure 4.2.1.1 illustrates this item type. Categorization items using a click-to-select interface instead of the drag-and-drop interface look precisely the same. In the click-to-select version, the user clicks or taps an element on the left and then clicks or taps the desired category label on the right. The expert reviewers believed that using the click-to-select option with categorization items would make these items accessible to students with motor disabilities who use switch systems or keyboard commands such as Tab and Enter.

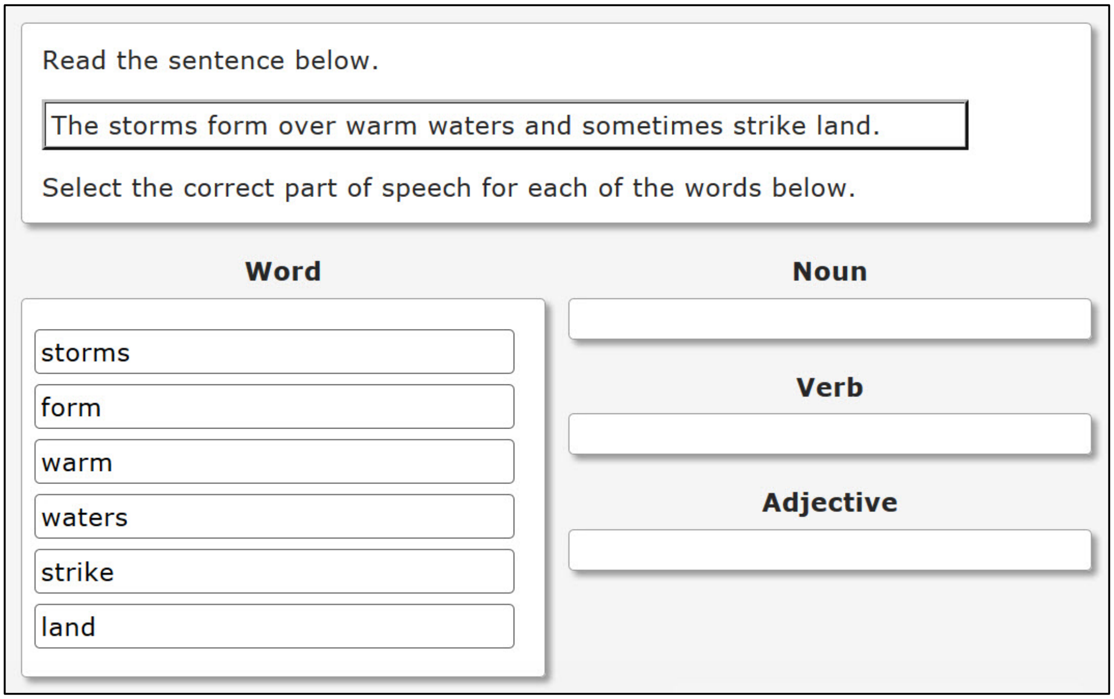


Figure 4.2.1.1. Categorization item with three categories.

Venn diagram. Venn diagram items are another type of categorization task that uses the familiar Venn diagram as a visual aid. Figure 4.2.1.2 shows an example of the type of Venn diagram item that expert reviewers evaluated.

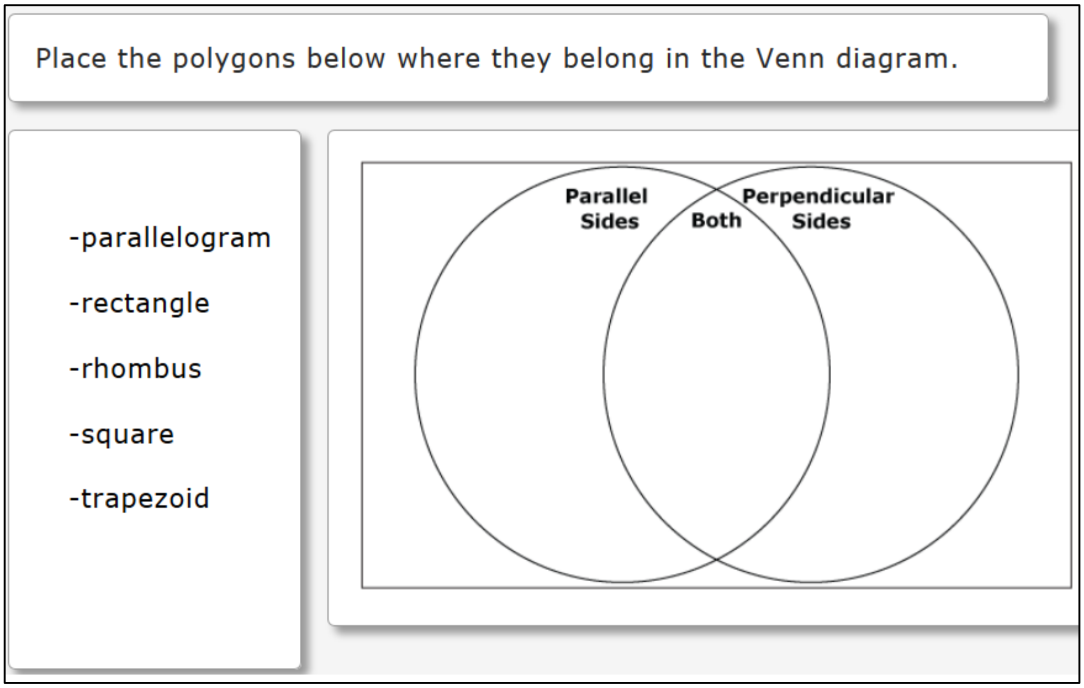


Figure 4.2.1.2. Venn diagram item.

A clear barrier to accessing Venn diagrams exists for students with motor and visual disabilities. In general, blind students and students with motor disabilities that interfere with using a mouse or touchscreen will not be able to click, drag, and drop the elements of these items. The experts' recommendations for Venn diagram items was similar to their recommendations for other drag-and-drop items. The experts recommended using formats with radio buttons, drop-down menus, or multiple-choice selections.

Labeling. Labeling items require the user to move each label into the correct position on a grid. Figure 4.2.1.3 shows an example of one-to-one labeling in which each label matches a single element on the right.

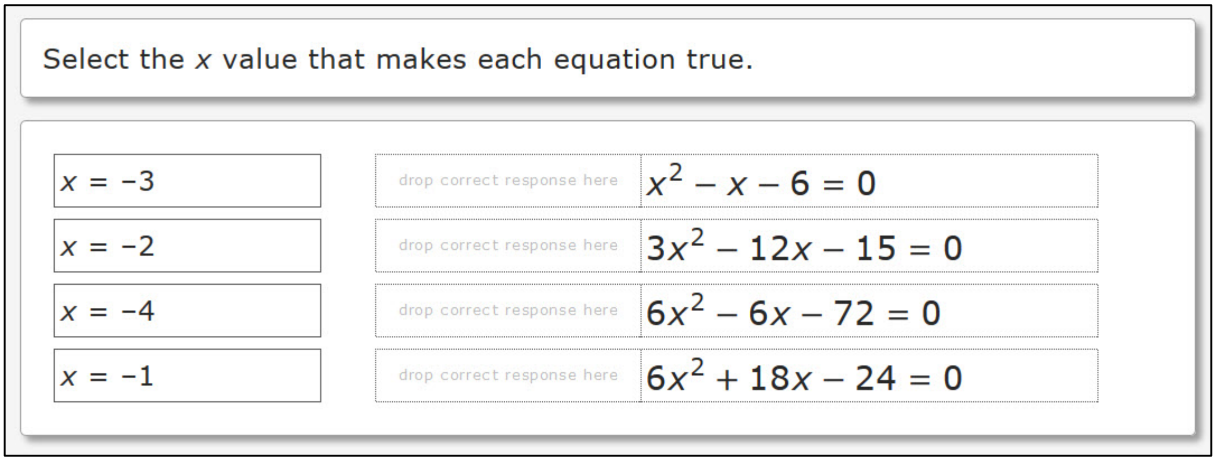


Figure 4.2.1.3. One-to-one labeling item.

Figure 4.2.1.4 shows a second type of labeling item in which the labels can be used as many times as necessary to complete the item.

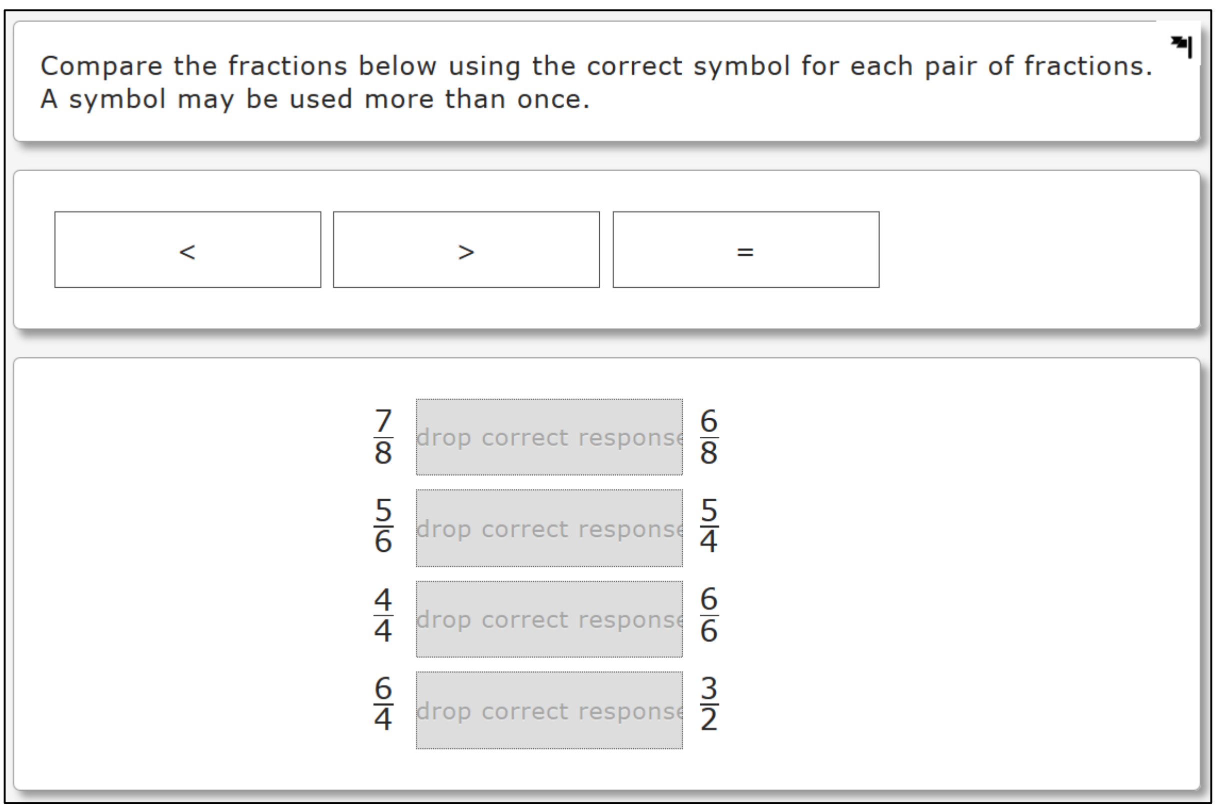


Figure 4.2.1.4. Unlimited labeling item.

Both of these labeling tasks have the same shortcomings as the drag-and-drop categorization tasks and Venn diagram. Experts suggested altering these items using drop-down menus or radio buttons.

Graphic labeling. Graphic labeling, as shown in Figure 4.2.1.5, requires the user to label elements of an onscreen image.

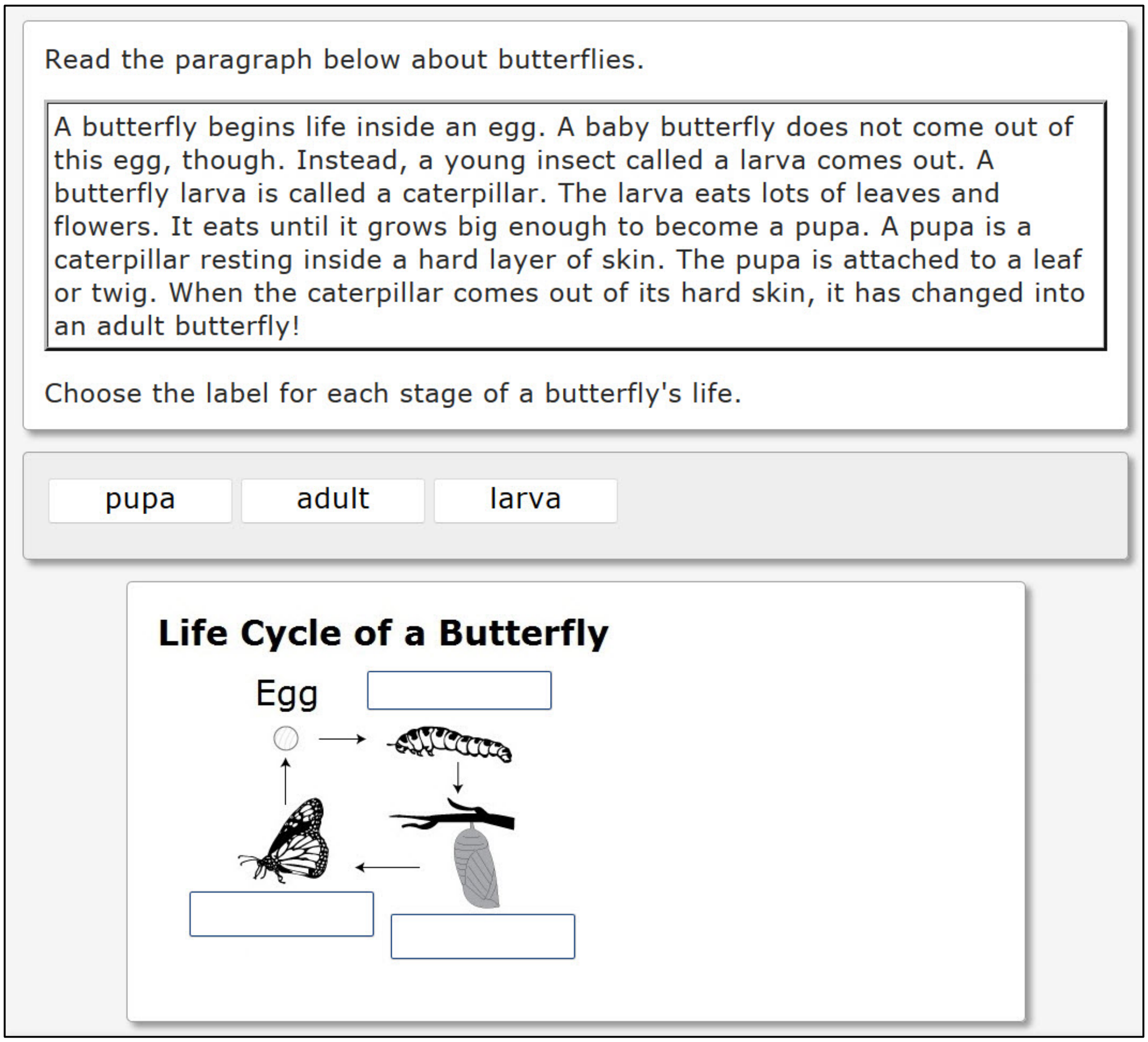


Figure 4.2.1.5. Graphic labeling with images.

A second type of graphic labeling task uses a number line as the background graphic, as shown in Figure 4.2.1.6.

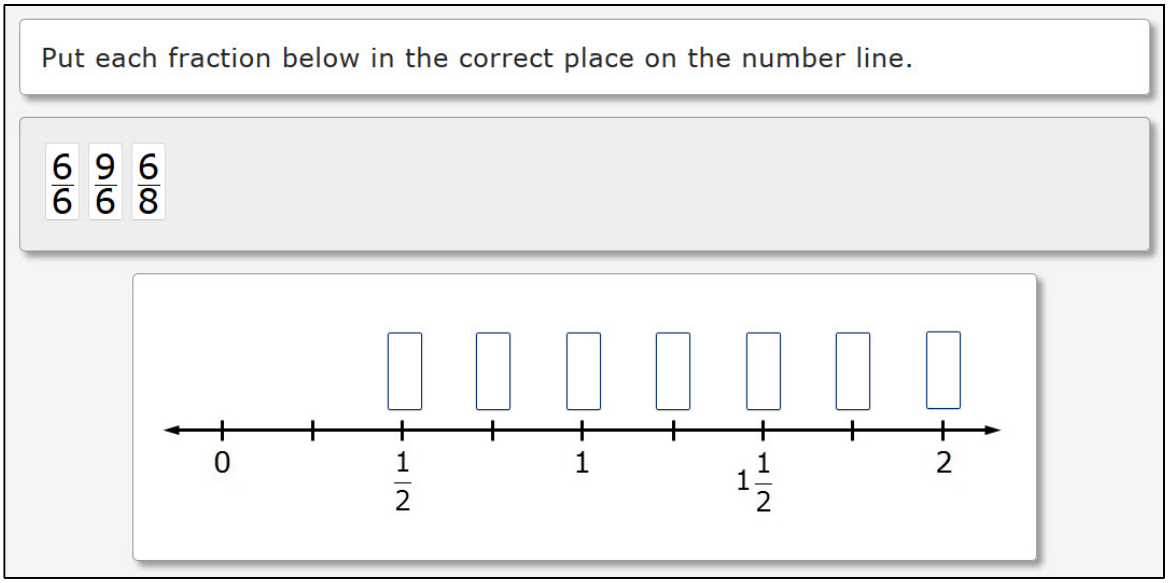


Figure 4.2.1.6. Graphic labeling with a number line.

The barriers of these tasks are similar to those of the drag-and-drop labeling items that do not include graphics. Students with motor or visual disabilities will have difficulty using the drag-and-drop feature of these items. For students with low vision, locating and linking the graphic with the labels when using screen magnification may make this task significantly more difficult. Students with low vision may not be able to see the whole graphic at once and that could potentially inhibit understanding. A unique challenge when considering recommendations to make this item type more accessible is the principle of asking the user to interact with the graphic. As with the recommendations for the labeling tasks, the question can be reformatted into a drop-down menu or radio button format. However, this takes the user away from the context of the graphic. Placing a drop-down menu on top of a graphic would alleviate the demand for students with motor difficulties; however, the complexity would increase for students with visual disabilities. This is an example of the potential for accommodations designed for one group of students to interfere with performance for another group of students. In many cases, items with specific interaction demands may have to be adapted in multiple ways to meet the needs of students with different characteristics and abilities.

Ordering. Ordering items are another TE task that often uses a drag-and-drop interface. Figures 4.2.1.7 and 4.2.1.8 show two items that require the user to move elements into the correct order, either from top to bottom or from left to right.

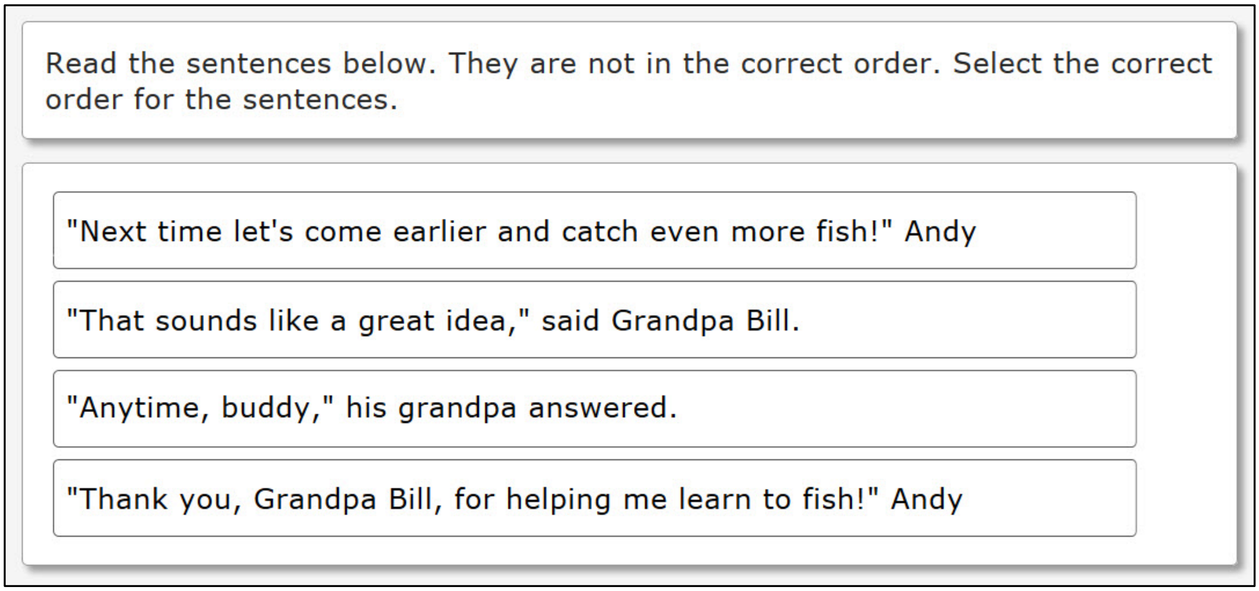


Figure 4.2.1.7. Ordering item for English language arts.

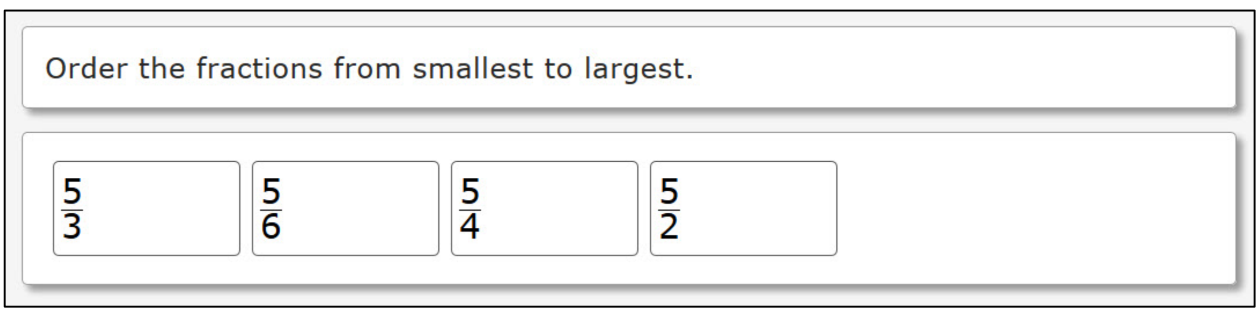


Figure 4.2.1.8. Ordering item for math.

The expert reviewers recommended a variety of ways that students could interact with this type of question without the demand of clicking and dragging. For example, the user could order items using numbers or letters.

4.2.2 Click-to-Select Items

Graphing. Graphing items, such as the example shown in Figure 4.2.2.1, appear to the student to be constructed response items even though they can be scored electronically. One of the advantages of technology enhancement is the variety of item types that require the student to perform tasks of greater complexity while retaining the benefits of computerized testing.

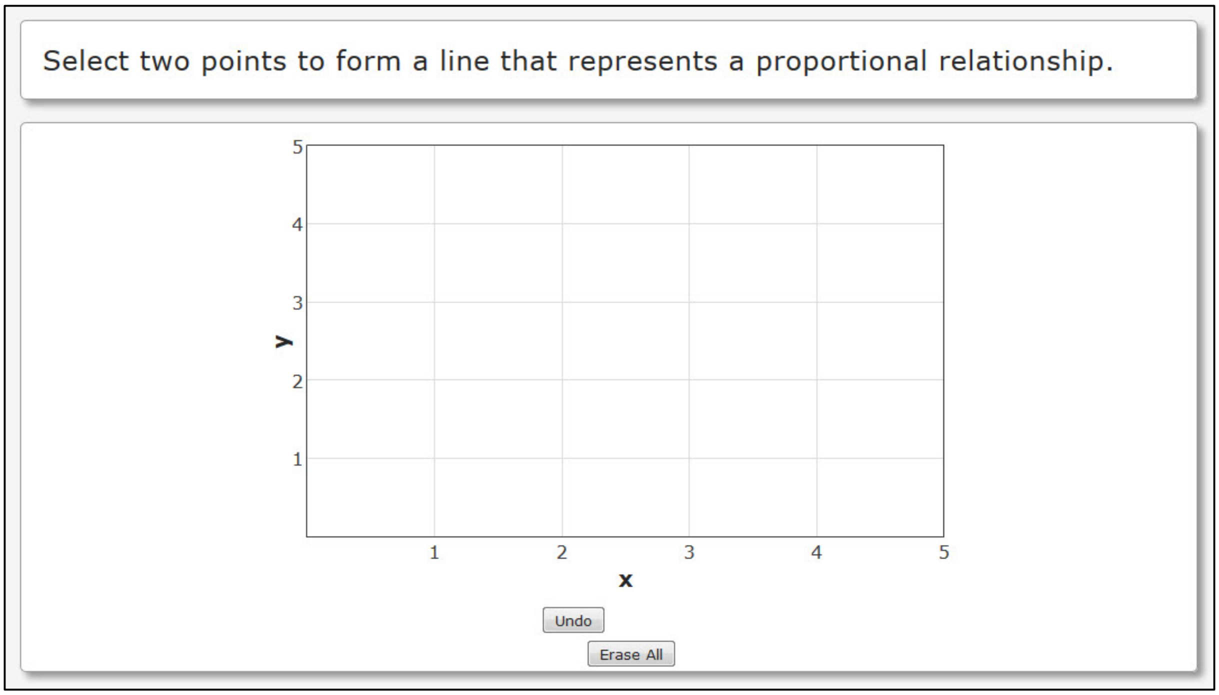


Figure 4.2.2.1. Line graph item.

Expert reviewers found this item type to be nearly impossible to accommodate online for students with vision or motor difficulties. One suggestion was to allow students who use switch systems to tab to each grid intersection and then select the locations of the points. However, this solution would be cumbersome and time consuming. In general, experts reported that tasks that function like constructed response items should be administered in paper-and-pencil or braille formats and then hand scored. Constructed response items are fully accessible if students are allowed to respond in the manner they would typically use in the classroom to accomplish these tasks.

Fraction selection. Similar to graphing items, the fraction selection item uses a click-to-select interface to create a shaded fraction in an image (Figure 4.2.2.2). In a variation this item type, the student first creates the number of partitions in the rectangle by adding horizontal and vertical lines. Then the student shades in the correct fraction by clicking on individual boxes.

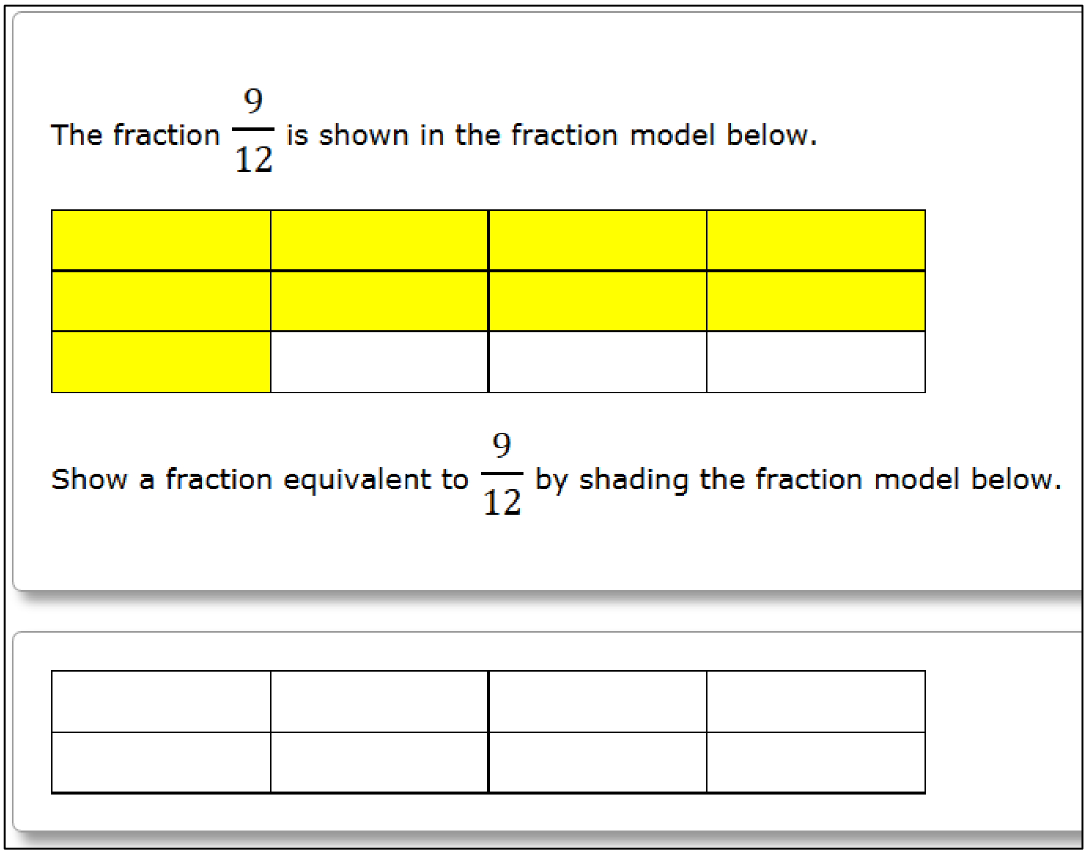


Figure 4.2.2.2. Click-to-select fraction item.

In addition to standard accommodations such as screen magnification or a scribe, experts recommended tactile graphics for students who are blind and the ability to tab to each box for students who use switch systems.

Matching. Matching items have a click-to-select interface (Figure 4.2.2.3). When the student selects an element on the left and a corresponding element on the right, a line appears between the selections.

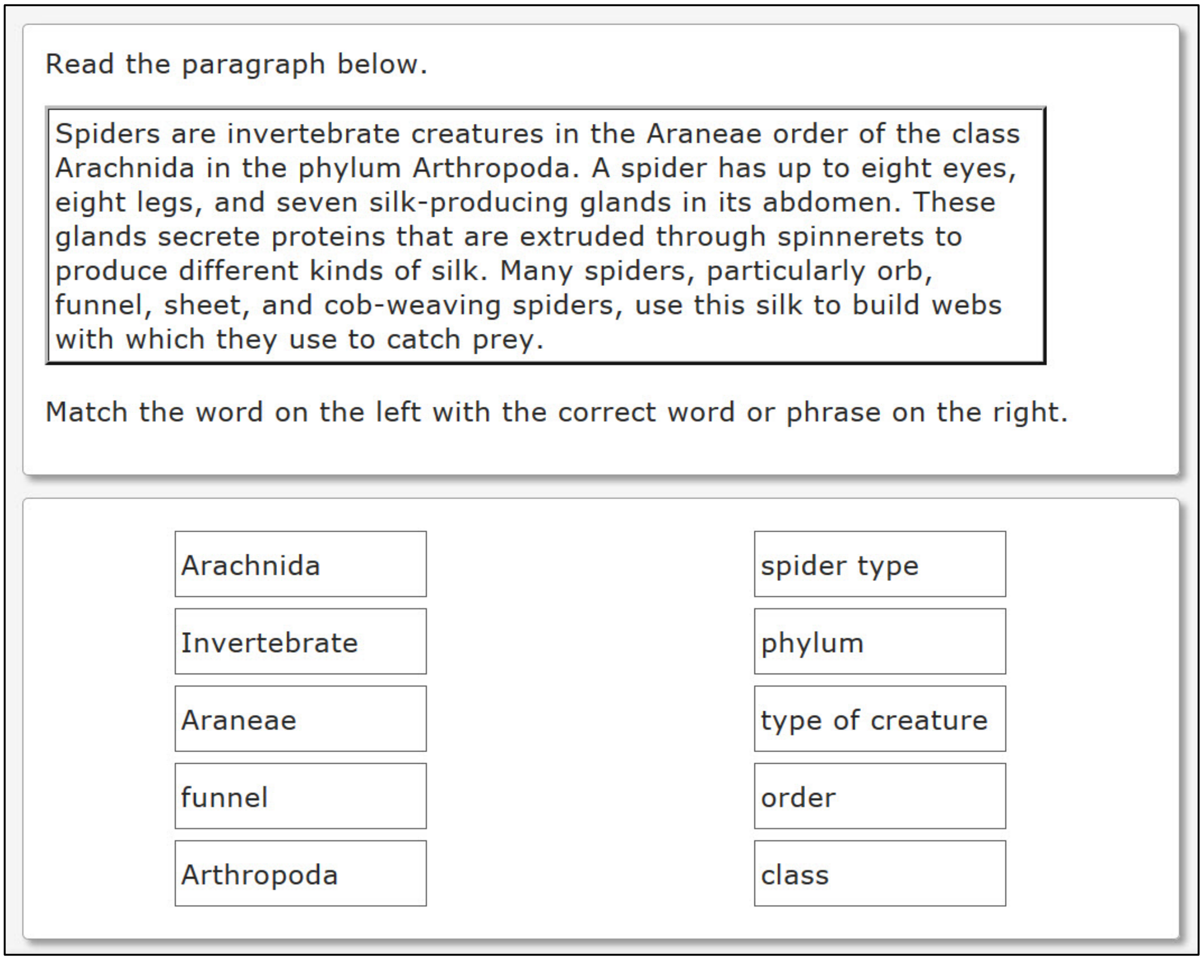


Figure 4.2.2.3. Matching item.

Experts stated that matching items should generally be accessible to students with vision and motor impairments and could be used frequently in a variety of subjects. However, students with low vision, even with screen magnification, will have difficulty seeing both columns simultaneously. Students with motor disabilities will have difficulty navigating the mouse, even to click. Recommendations included using switches to navigate this item type.

Select text. In a select text item, the user can select one or more words, phrases, or complete sentences from within a text excerpt. When answer options are specific words within the text, they are shown highlighted so the student knows which words are available for selection (see Figure 4.2.2.4). After the student selects his or her responses, a red box outlines the words.

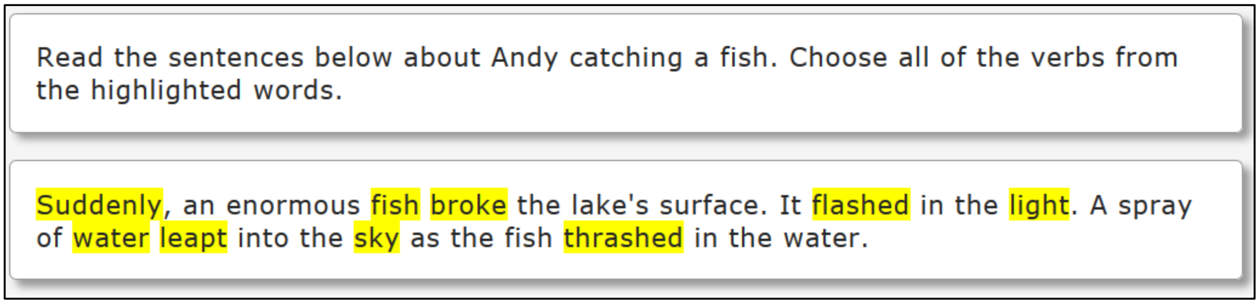


Figure 4.2.2.4. Select text item.

After reviewing sample items, members from the expert panel noted barriers to accessing select text items. The experts noted that when the highlighting feature is activated, students with visual and perceptual processing issues might have difficulty reading the selections or seeing which sentence or word is highlighted. Additionally, in order to activate the highlighting, the student would have to hover the cursor over the intended selection and then click the mouse. Without the use of an assistive technology device, hovering a cursor over an onscreen selection is difficult, if not impossible, to perform for students with motor disabilities.

The experts made recommendations to increase the accessibility of these items. In order to accommodate blind and visually impaired students, the panel recommended a screen reader for improved access to the onscreen text and a text box to allow students to enter their responses using a keyboard or braillewriter. Another suggestion was to enable the tab key to cycle through words or sentences so the student can access the materials utilizing the keyboard and make their selection with the enter key. For switch selection, the experts also recommended that students should have the ability to scan the passage multiple times. Without such a device, it will be difficult for students to hover a cursor over their selection.

Matrix. Matrix items use radio buttons in an array of rows and columns, similar to common survey questions that may be familiar to most adults. A matrix item is essentially several yes/no, true/false, or multiple choice items bundled into a package with a single stimulus. Figure 4.2.2.5 shows a matrix format item.

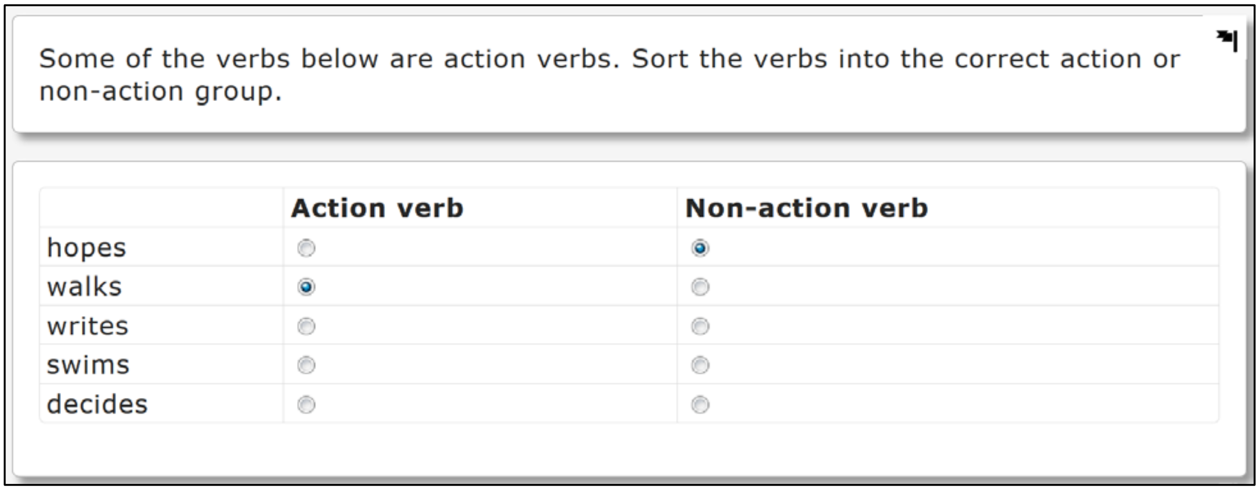


Figure 4.2.2.5. Matrix item.

4.2.3 Constructed Response Items

Short answer. Text-entry items that require short answers require the student to type or dictate a response to an open-ended question. Short-answer text-entry items that require a numeral or a single word can easily be computer scored. Figure 4.2.3.1 illustrates a typical text-entry item in math.

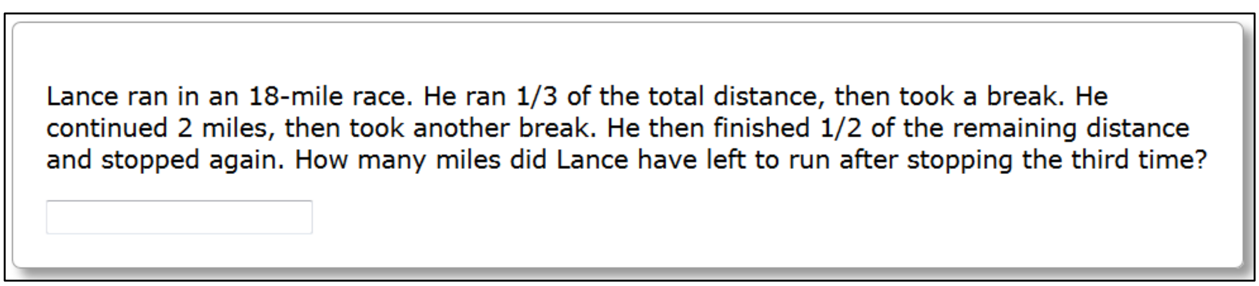


Figure 4.2.3.1. Short-answer text-entry item.

A concern for students who are blind is their ability to see the box or guide the mouse to the location to type in the text. Students with low-vision issues may have difficulty seeing a graphic or seeing the directions and the text box at the same time, particularly with screen magnification activated. Students with motor difficulties find typing on a keyboard to be a significant challenge.

For students who are blind, the experts suggested a screen reader or refreshable braille display. For students with low-vision issues, on-screen magnification would enable the student to see the text box and access this item. Students with motor disabilities could access this item utilizing a joystick, touch pad, or on-screen keyboard.

Extended response. Extended response items include essays or worked arithmetic problems as well as items with shorter but variable responses, such as sentences, that must be hand scored. Though sample extended response items were not presented, reviewers noted that blind or low vision users with keyboarding skills can enter sentences and essays online, and braille users can use a braillewriter to record their responses. Students with motor disabilities can use dictation software or a scribe. Because these tasks require human scoring, they can also be administered offline, making them accessible to students who use paper and pencil, large print, or braille formats.

4.3 Conclusions

In general, expert reviewers responded to the overall presentation and response demands of items rather than to the various item types individually. Reviewers simplified the conceptualization of item barriers by grouping inaccessible presentations and response requirements and suggesting alternative formats they deemed to be accessible. Reviewers recommended that standard accommodations, such as onscreen magnification for low-vision students, large print and braille tests, auditory presentation, and the availability of switch systems, would continue to be useful and necessary for students with vision and motor disabilities.

Reviewers recognized that standard accommodations may also introduce additional task demands. Students with visual disabilities would be required to memorize facts in the ordering and Venn diagram tasks when presented auditorially, unless the text of the item was available in print or braille. This would increase the level of difficulty for these students if the question and cognitive depth of knowledge were to remain constant. Students with low vision who are presented with on-screen magnification may have difficulty seeing the entire graphic for the labeling and Venn diagram tasks. Again, this would increase the task demand for that population of students.

In addition to recommending the continued use of well-known accommodations, reviewers responded to specific features of TE items with suggestions for alternative formats and presentations. Experts deemed the drag-and-drop interface used in many TE items inaccessible to students with vision impairments regardless of the item type in which it appeared. Expert reviewers pointed out that radio buttons are fully accessible in a variety of formats and onscreen with switches. Radio buttons are familiar to students from multiple-choice items and are amenable to audio presentation as well as use in print or tactile test forms. To alleviate the motor demands of dragging and dropping, experts suggested using alternative formats with radio buttons, keyboard commands or switches, or a scribe to enter responses. These would significantly increase the accessibility of these items for students with motor disabilities. Reviewers also thought that drop-down menus, matching, and multiple-choice formats would offer greater accessibility by virtue of enabling keyboard responses or switches rather than requiring a mouse or touchscreen. These forms maintain the fidelity of the question but allow the user to interact with the test with greater independence.

For items that require text entry, the reviewers suggested a providing a text box for users who are blind, as long as the student has sufficient keyboarding knowledge. Other constructed response items, such as the graphing example in Figure 4.2.2.1, were inaccessible online. Items such as these, even though they are machine-scored in an online format, would require human scoring if presented as constructed response items in paper-and-pencil or braille tests. Nonetheless, expert reviewers supported the use of constructed response items as they are accessible to students in offline formats such as print and braille.

Reviewers also provided feedback on the DOK they believed that these item types might attain. While the content of each item is the most important factor in the DOK, some item types may tend to limit or enhance cognitive complexity. In many cases, reviewers selected three or even four DOK indicators for individual items. However, there were some general outcomes regarding DOK levels.

Reviewers consistently identified the graphic labeling item shown in Figure 4.2.1.5 and the categorization item shown in Figure 4.2.1.1 at DOK level 1, Identify/Recall. Most categorization items were rated DOK level 1, Arrange/Sequence and DOK level 2, Skill/Concept. Because categorization items require analyzing and sorting items, experts suggested that they can involve higher-level thinking and can be used for level 4 questions as well.

The DOK for ordering and labeling ranged from level 1 to level 3. Depending on which task the student must perform, select text questions can be as complex as level 3 DOK. Matrix items were similarly assessed to have DOK levels of 2 and 3. The click-to-select fraction shown in Figure 4.2.2.2 was identified as DOK 2 while graphing items were rated at DOK 3. Constructed response items requiring either short or longer text entry were evaluated at DOK 3, Strategic Thinking, and DOK 4, Extended Thinking.

5.0 Teacher Panels and Cognitive Labs

Teacher panels and cognitive labs provided the opportunity for in-depth examination of TE items online, in print, and in braille, in their original formats and with the alterations and accessibility features suggested by the expert reviewers. Teacher panels consisted of teachers of students with vision and motor disabilities who met for half a day in five of the ATEA partner states. Teachers evaluated and compared items in multiple formats. In addition, students with vision or motor disabilities nominated by their teachers responded to accommodated TE items online, in print, and in braille in cognitive labs. Teacher panels and cognitive labs occurred during the fall of 2013.

5.1 Teacher Panel Procedures

Teacher panels included 74 teachers, 5 occupational therapists, 1 orientation and mobility expert, and 25 district and state department of education administrators, many of whom were also special education teachers. The process of eliciting teacher feedback consisted of training in TE items, demonstrations of inaccessible and accessible TE items, and then a review session during which pairs of participants logged onto sample tests and evaluated TE items in several formats. TE items were grouped into six tests, one math test and one ELA test at the elementary, middle school, and high school levels. Each test consisted of 12 to 14 items. In many cases, the same item content was presented to teachers in multiple online formats so that teachers could evaluate the accessibility of each format. Online items appeared in as many as three different layouts, such as a drag-and-drop version, a radio button version, and a matching or drop-down version. Support materials included braille and print forms of the tests with accessible item layouts. Teachers provided feedback by answering the following questions on templates containing the items on each test:

> For each group of test items that you review, please provide feedback on the different item types and their accessibility for students with vision or motor disabilities.

a. Based on your expertise in vision or motor disabilities, what are the barriers to accessibility for this item? How can these barriers be overcome?

b. If there is more than one item in the group, which is most accessible and best measures the construct? How can the best item be improved?

5.2 Cognitive Lab Procedures

Cognitive labs were planned to include as many as 60 students, but as it turned out only 28 students participated in three states. Teachers initially submitted information and parent permission for 42 students, including several students who usually participated in an alternate assessment. Due to lack of subsequent response from teachers or teacher decision not to continue with the cognitive labs, 14 of these students did not participate. Test format and grade breakdown for the participating students are shown in Table 5.2.

Table 5.2

Cognitive Lab Participants by Grade Level and Test Format

Format	Elementary	Middle	High	Total
Online	0	0	4	4
Large print	3	6	3	16
Braille	2	2	8	12
Total	5	8	15	28

All students had access to the online tests in addition to large print and braille booklets. Many students, including some who used braille and print forms for support, accessed the online system to experiment with the TE items and to enter their responses. Students were assessed in their schools and used their typical instructional equipment.

Teachers or project staff assessed students individually on their ability to access and interpret TE items in a variety of formats. All students received logins to the KITE online testing engine. Students also received print, large print, or braille test booklets as requested by their teachers. The procedures included an introduction to the cognitive lab activities (Appendix B, Cognitive Lab Protocol) and an explanation of each test item, with any necessary supports, such as oral presentation, provided by the teacher. As students read and responded to each item, notes were made of any questions or difficulties they encountered. Students were queried about any confusion they had with each test item and how they would prefer to have that item presented. Teachers recorded student feedback for each item using the following questions for guidance:

a. What were the barriers to accessing the content of the task?
b. What were the barriers to understanding how to complete the task?
c. What were the barriers to responding to the task?
d. What would improve this task for this format (online, print, or braille)?

High school students who used braille forms were particularly helpful in designing alternative spatial organization and labeling for some of the new item types.

5.3 Results of Teacher Panels and Cognitive Labs

Teacher panelists and students were exposed to a variety of item and task types that varied by subject and grade level. Each item and task type fit into one of three major interfaces: drag and drop, click to select, and constructed response. Drag-and-drop test items require the student to click/tap on an element in the item, sustain pressure on the mouse/screen, drag the element, and then drop the element in a new position by releasing pressure on the mouse/screen. Click-to-select test items require students to use one click/touch or a series of clicks/touches to indicate their response. Depending on the specific item type, click-to-select test items can work with switches, Tab and Enter keys, step scanning systems, and/or single touches on a touch-screen tablet. While drag-and-drop and click-to-select test items both contain answer options from which the student must select a response, constructed response items require the student to enter a response in a text box or on a graphic background such as a coordinate plane.

Because different tasks within the same interface category require similar types of interaction, comments about one task type are applicable to other types. For example, drag-and-drop ordering tasks require similar motor outputs to drag-and-drop labeling tasks. Both tasks require the student to interact with the test item by selecting, holding, and dragging the computer mouse or a finger on a touch-screen tablet. Therefore, similar issues exist for students with vision and motor disabilities for these item types. The following section explains each item type according to interface category. Summaries of feedback from teacher panels and cognitive labs follow, with general comments and conclusions after the discussion of individual TE tasks.

5.3.1 Drag-and-Drop Items

Participants in the teacher panels were exposed to several types of drag-and-drop items. Cognitive labs did not include drag-and-drop item types because they are not accessible to students with vision and motor impairments, so they were not included on online assessments or special forms. Teacher panelists identified several barriers to accessibility with drag-and-drop items. Barriers to accessibility were similar for ordering, labeling, graphic labeling, and categorization item types within the drag-and-drop interface category. Therefore, a summary of feedback from the teacher panelists will follow the descriptions of these item types.

Categorization. Categorization items require students to organize given elements into categories. In the drag-and-drop version of categorization items, the user must select an element from the left side or top of the screen and sustain pressure to drag the element into the category. The user must click or touch and hold each element, drag it into position, and then release it. Examples of these item types are displayed in Figures 5.3.1.1 and 5.3.1.2.

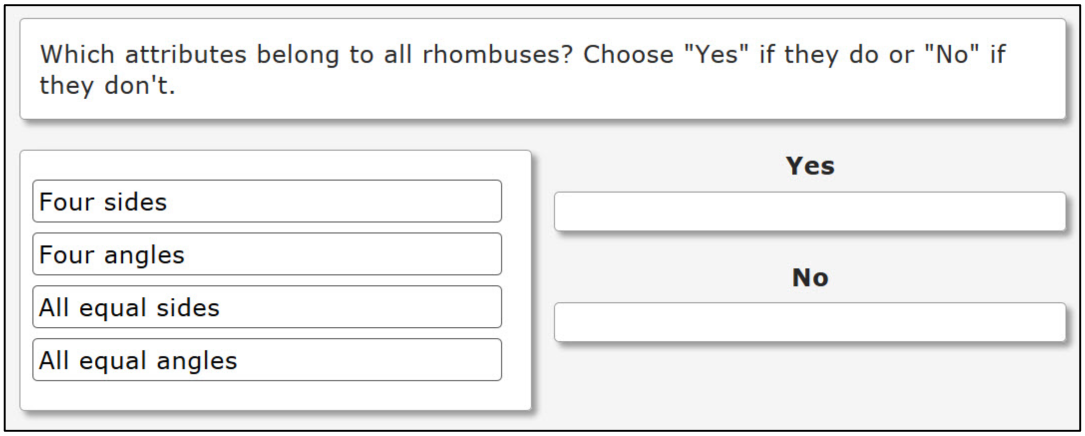


Figure 5.3.1.1. Math categorization item with two categories.

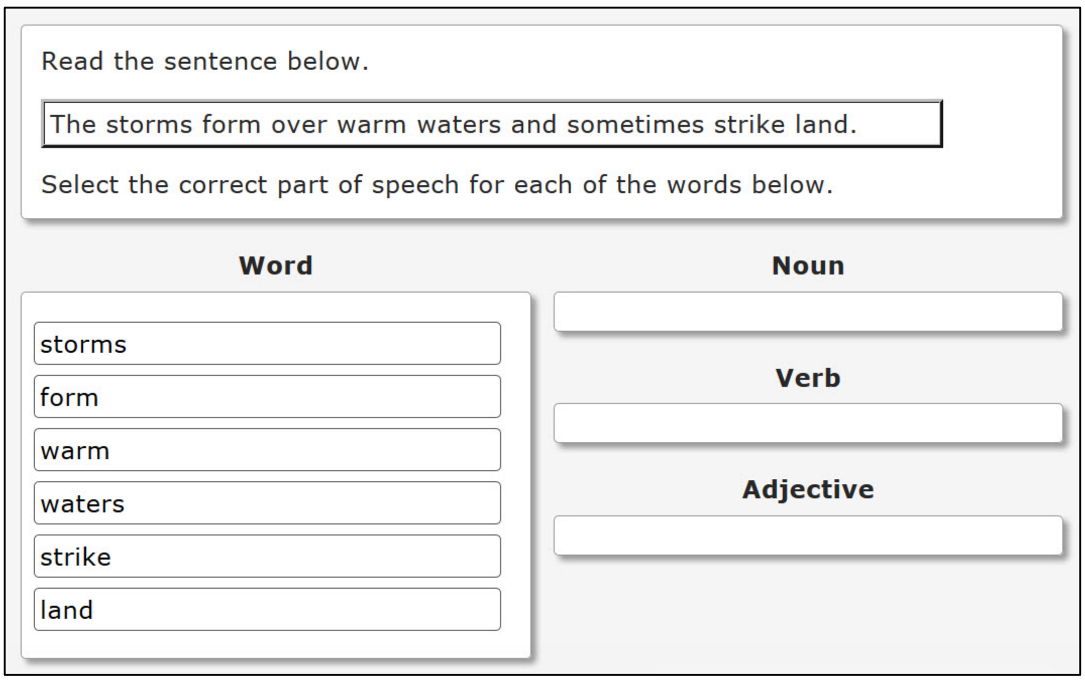


Figure 5.3.1.2. ELA categorization item with three categories.

Venn diagrams. As a type of categorization task, Venn diagram items are similar to other drag-and-drop items (i.e., ordering, labeling, and categorization). With Venn diagram items, students use the mouse to click and drag attributes or descriptions into the appropriate section of the diagram. However, on the Venn diagrams that were displayed for teacher panels, when the student clicks on an answer choice and drags it to the diagram, the part of the diagram over which the student hovers becomes shaded in a rectangular yellow box (see Figure 5.3.1.3).

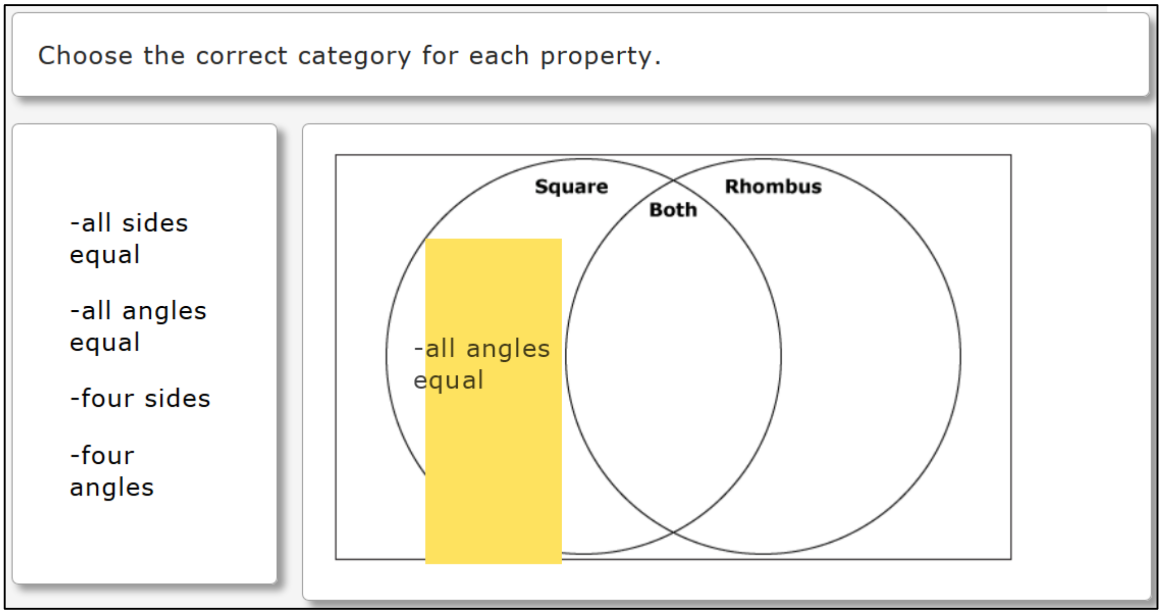


Figure 5.3.1.3. Venn diagram item viewed while student hovers an answer choice over the diagram.

Labeling. Labeling items require students to select a label from a list at the top or left side of the screen and drag the label to the correct location. Labeling items in math, for example, might require students to move an x value next to an equation to indicate that the x value would make the equation true (see Figure 5.3.1.4). There are two types of labeling items: one-to-one labeling and unlimited labeling. One-to-one labeling items allow students to use each label only once (unless they change their answer, which requires that they first drag the label back to its original position). Unlimited labeling items allow students to use each label more than once.

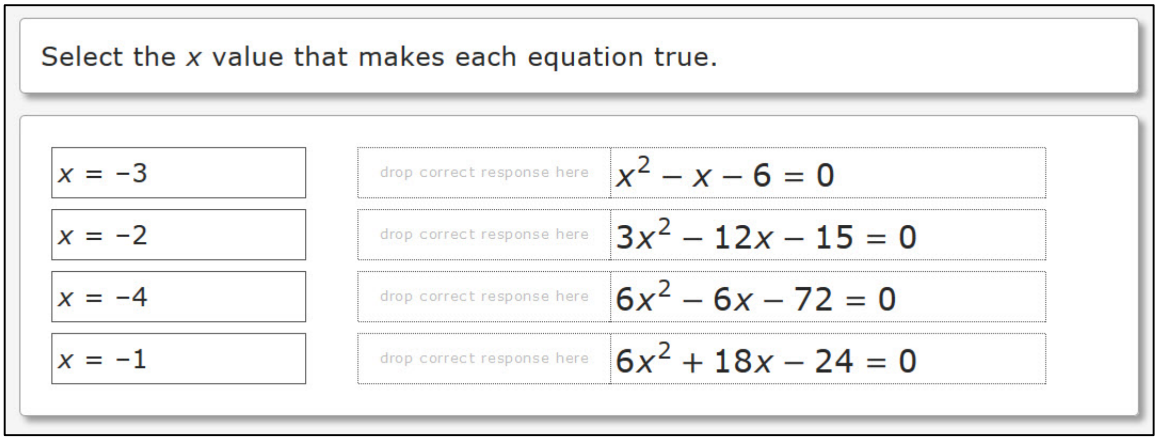


Figure 5.3.1.4. Math labeling item (one-to-one).

Graphic Labeling. Graphic labeling items include a background graphic onto which students label parts, place elements into a sequence, or plot points on a graph. Graphic labeling items are functionally similar to other labeling items that do not include a background graphic. However, graphic labeling items require students to examine some type of visual or schematic figure (e.g., drawing, chart, or diagram) in order to answer the item. Figure 5.3.1.5 is an example of a graphic labeling item.

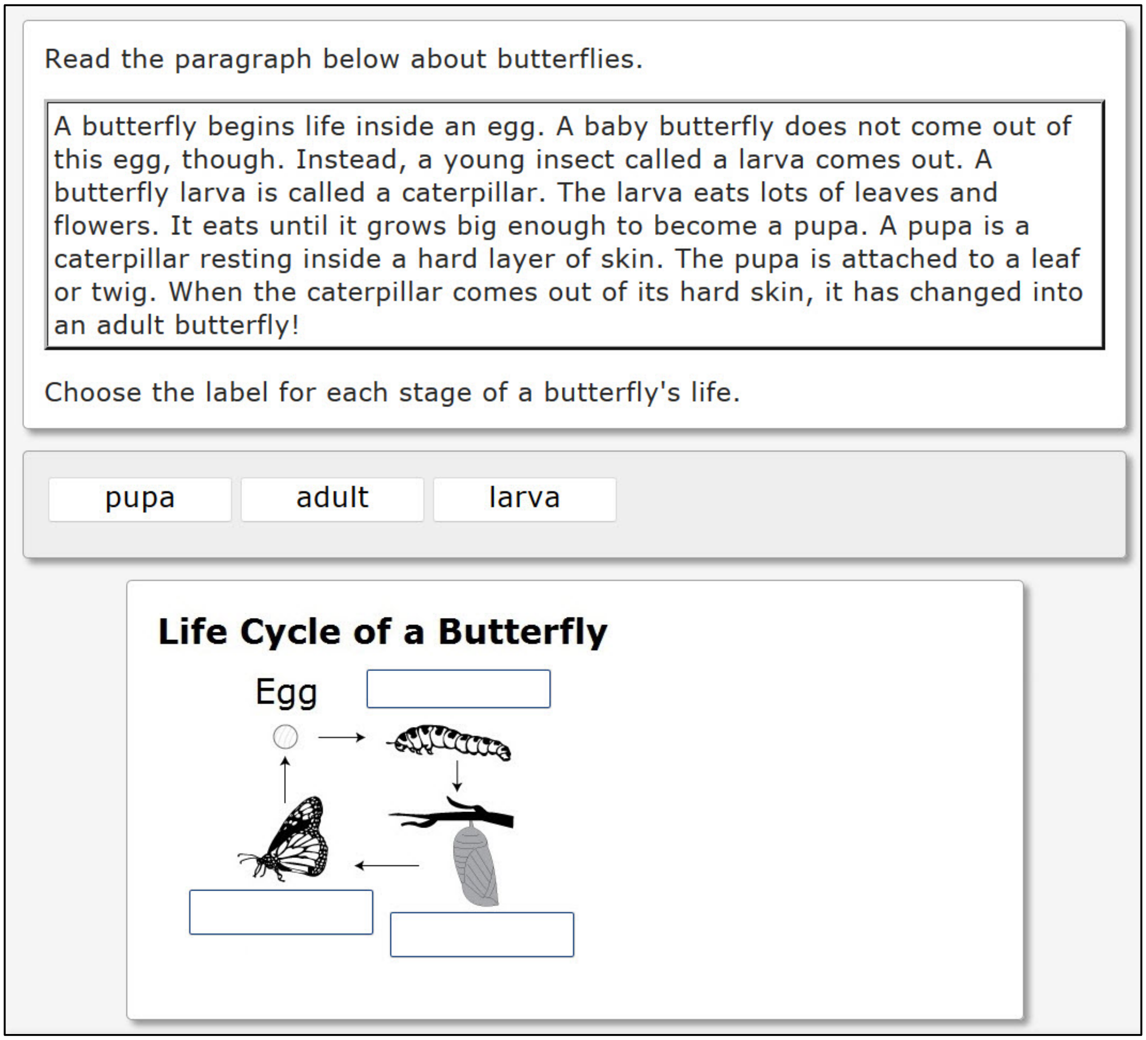


Figure 5.3.1.5. Elementary ELA graphic labeling item.

Ordering. Ordering items require students to place elements into a sequence. Ordering items in ELA, for example, might require students to move a series of sentences into the correct order to form a paragraph. Ordering items in math might require students to place elements in numerical order or to indicate the correct order of operations (see Figures 5.3.1.6 and 5.3.1.7).

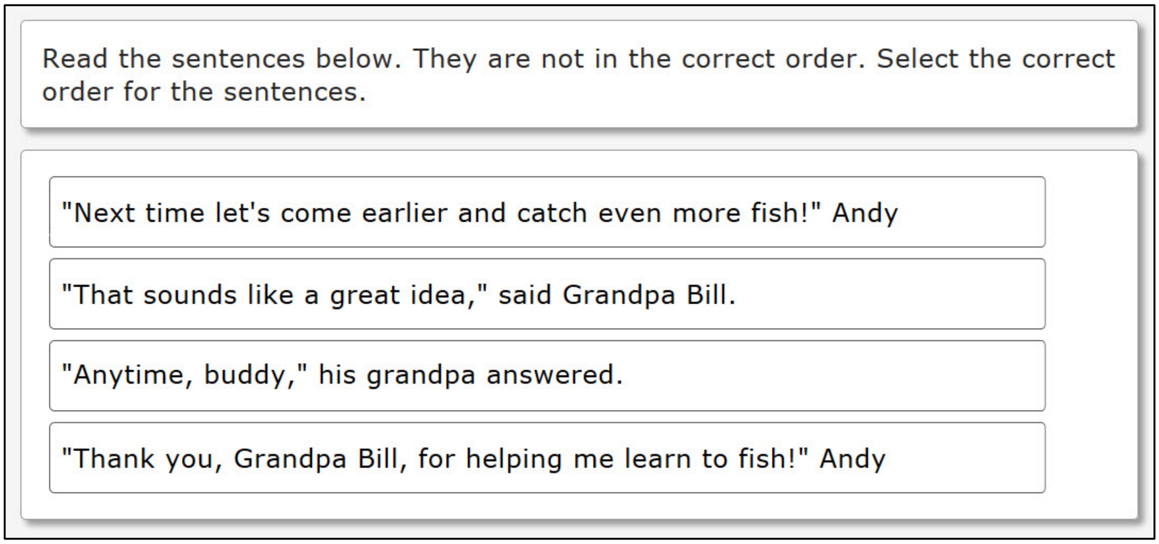


Figure 5.3.1.6. ELA ordering item.

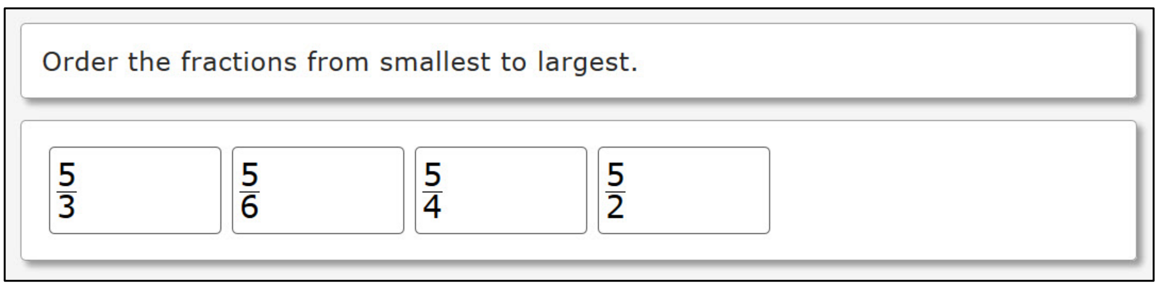


Figure 5.3.1.7. Math ordering item.

Teacher panels with drag-and-drop items. In general, teachers contended that some categorization items were unnecessarily complicated for students and could present barriers to accessibility for both visually impaired and motor impaired students. Since many of these items require students to classify words, phrases, sentences, or numerical expressions into two categories, the panelists suggested check boxes or radio buttons as better formats for these tasks.

Drag-and-drop item types can be difficult to navigate using a screen reader. Additionally, visually impaired students should have some way to check their answers on drag-and-drop items. For students with motor impairments, it may be difficult to sustain pressure on the mouse to drag and drop the answer choices. Dragging and dropping is especially difficult on items with small answer-choice locations.

Teachers were concerned about drag-and-drop items that required students to scroll to view all parts of the item. As an example, one middle school math item presented a series of two-dimensional shapes at the beginning of the item and a series of three-dimensional shapes at the end of the item. The student had to match each two-dimensional shape to its corresponding three-dimensional shape by dragging and dropping each shape. Since the shapes were too large to view on one screen, even without magnification, the student had to scroll between the item stem and the answer choices. As one panelist stated, "The item requires the child to select a shape, scroll to the correct place in the item, and then drop the shape. This is very difficult for *any* student."

The teachers recommended a variety of ways that students could still interact with this type of item without the demand of clicking and dragging. For ordering items, students could indicate order by typing numbers (e.g., 1, 2, 3) or letters (e.g., A, B, C) next to each answer choice in a text box. For labeling and categorization items, teachers recommended using radio buttons, drop-down menus, or multiple-choice items, which would maintain the fidelity of the question while allowing the user to interact with the test with greater independence.

Some labeling items include an answer box with the instructions "drop correct response here" (see Figure 5.3.1.8). The panelists believed this text is helpful for students. However, they also expressed concern that the grey text used in these boxes is too faint for students with visual impairments. They suggested making this text darker.

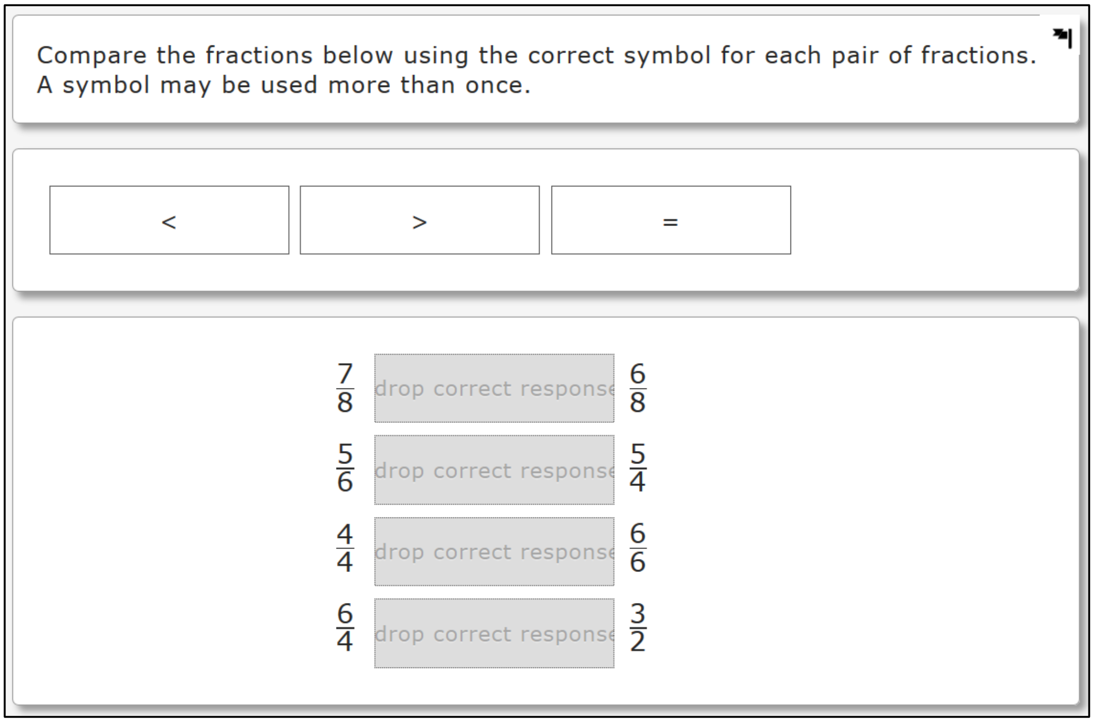


Figure 5.3.1.8. Labeling item with instructions to "drop correct response here".

Number line items present unique challenges for students with visual impairments and students with motor impairments. With this item type, the student drags and drops answer choices onto a number line (see Figure 5.3.1.9).

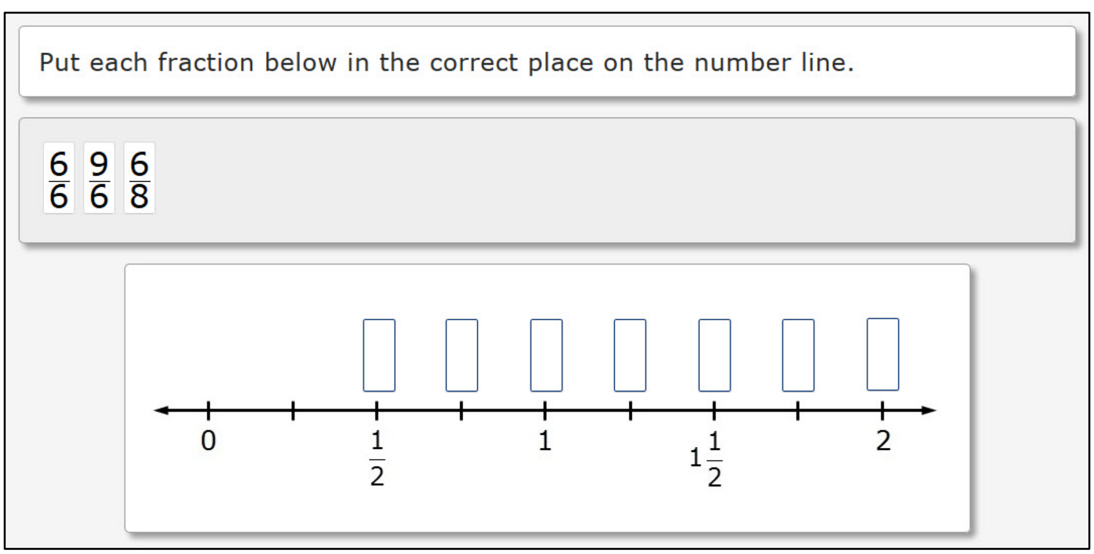


Figure 5.3.1.9. Number line item viewed before student has placed answer choices on the number line.

Panelists expressed concerns about the accessibility of this item type. Panelists stated that the item does not seem to be compatible with a screen reader, and blind students would require a tactile version of the number line. To make the item more accessible to students with low vision, the numbers, hash marks, and answer boxes on the number line could be made larger and darker. The answer boxes could also be better distinguished from one another and could be moved closer to the number line or connected to the number line in some way.

In the item in Figure 5.3.1.9, the answer boxes in which the fractions are listed are quite small. Manipulating these boxes might require fine motor skills, making the item particularly difficult for students with motor impairments. The panelists suggested making both the boxes and hot spots much larger to make the item more accessible.

Panelists also had accessibility concerns that were specific to the Venn diagram item type (Figure 5.3.1.10), especially for visually impaired students. Blind students may not have previous experience using Venn diagrams, and these items might not have meaning for blind students. Visually impaired students may also have difficulty locating the correct part of the diagram.

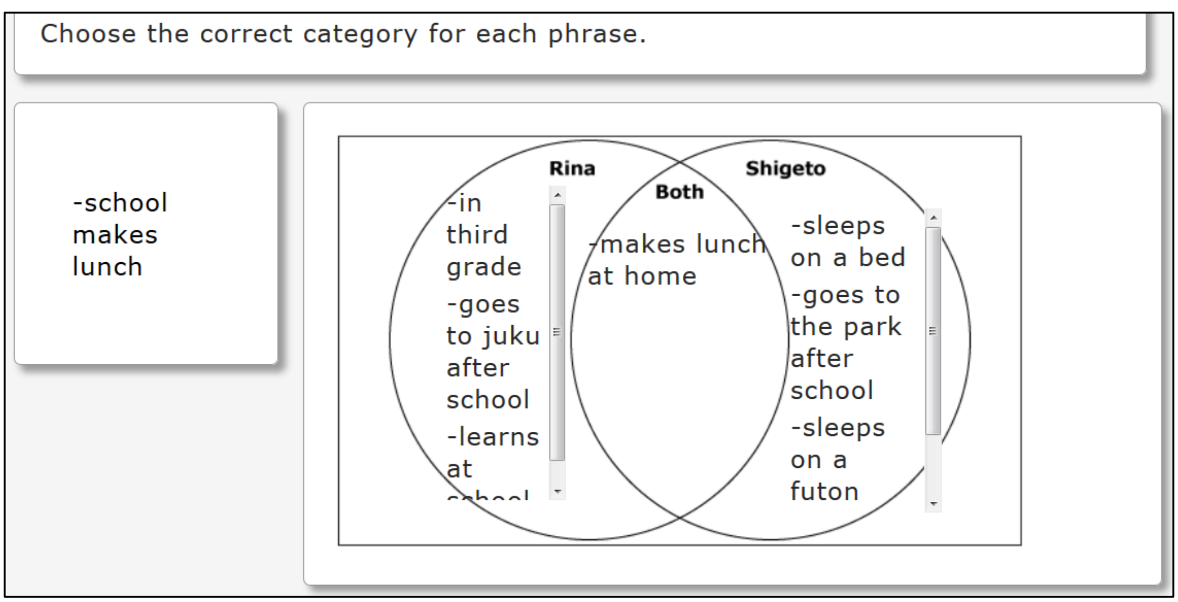


Figure 5.3.1.10. Venn diagram item viewed with answer choices inserted into the diagram.

The teacher panelists expressed concerns that the Venn diagram item type would not work well with assistive technologies. When screen readers are used, students with visual disabilities would be required to memorize information from the Venn diagram tasks when that information is read to them. This would increase the level of difficulty for these students. Students with low vision who are presented with on-screen magnification may have difficulty viewing the entire diagram for the labeling and Venn diagram tasks. Again, this would increase the task demand for that population of students.

Teacher panelists overwhelmingly described the Venn diagram items as "visually cluttered" and "confusing," especially when answer choices are inserted into the diagram. Panelists generally described the diagram as being too small, and they suggested making the circles larger and more separated. The small size of the diagram is especially problematic after the student drags and drops answer choices. As Figure 5.3.1.10 shows, when a student drops his or her answer choices into the Venn diagram, the text adjusts to fit and a scroll bar appears inside the diagram. It can be difficult to read answer choices once they are dropped into the diagram.

Panelists expressed other visual concerns with Venn diagram items. The answer choices listed on the left side of the screen could be better distinguished from one another. For example, each answer choice could be presented in its own box. For students with small visual fields, the panelists suggested that each part of the diagram be color-coded and/or placed inside a thick border. The yellow highlighting of the rectangular box (see Figure 5.3.1.3) could also be distracting for students with vision impairments.

The panelists also expressed accessibility concerns about Venn diagram items for students with motor impairments. The panelists thought that the drag-and-drop function was too difficult to control in this item type.

The panelists provided general comments about the Venn diagrams, as well. They stated that the item instructions were unclear, and the instructions did not indicate that the student should drag and drop the elements. Given the issues with this item for visually impaired students, several teams of panelists suggested a radio button item as an alternative to Venn diagram item types.

Barriers to accessibility for number line and other background graphics items are similar to barriers for all drag-and-drop items. Students with motor or visual disabilities might have difficulty using the click-and-drag feature. This feature is especially problematic when 1) the student has to click and hold the answer choice as he/she scrolls to the graphic at the bottom of the screen, 2) when the hot spot is not sensitive enough or is too sensitive, and 3) when parts of the graphic are too close to each other.

Furthermore, blind students require tactile versions of the graphics. For students with low vision, locating the graphic and linking it with item information when using screen magnification may make this task significantly more difficult. In addition, students with low vision may not be able to see the whole graphic at once, which could potentially inhibit understanding.

Improving the accessibility of background graphic items presents unique challenges. The presence of a graphic in an item can increase the authenticity of the task. Some of the panelists' recommendations involved removing interaction with the graphic in favor of using drop-down menus or radio buttons. However, such changes would remove the student from the context of the graphic.

There are also tradeoffs between accommodating students with visual impairments and accommodating students with motor impairments. Placing a drop-down menu on top of a graphic would alleviate motor demands for students; however, such a change might increase item difficulty for students with visual disabilities.

5.3.2 Click-to-Select Items

Click-to-select categorization. Categorization items require the user to choose the correct category for each of the given elements. Click-to-select categorization items appear identical to drag-and-drop categorization items onscreen, but the motor requirements to responds to the item are simpler. In the click-to-select interface, the student clicks on or touches an element from a list on the left side of the screen, which selects and highlights that element. Then, the student clicks on a category on the right side of the screen and the selection snaps into place.

Teacher panels with click-to-select categorization items. The panelists overwhelmingly favored click-to-select categorization items. They commented that click-to-select categorization items were less visually confusing and easier to navigate than drag-and-drop interface items (e.g., ordering and labeling). The panelists suggested a few ways to make click-to-select categorization items even more accessible to students. First, the directions should clearly explain that the student needs to click in the left column and then click a category on the right to provide a response. Second, the process of changing an answer should be simplified. Students should be able to move answer choices from one box to another within the right-hand column.

For items that involve categorization, the panelists commented that click-to-select categorization would be more accessible to students with motor disabilities than drag-and-drop item types. This would be especially true for students who use switch systems or keyboard commands, such as the tab and enter keys.

Cognitive labs with click-to-select categorization items. For elementary students, the cognitive labs included two click-to-select categorization items. With the first item (Figure 5.3.2.1), one barrier to accessibility was that students had to scroll across the screen to view all item content. However, this did not seem to hinder students' ability to complete the task.

One student completing this item in braille complained about the blanks and dots in the matrix format of the item and was unsure about how to respond whereas another student did not find this layout confusing. These students had several suggestions: place instructions at the top of the item instead of halfway down the page, number the items, and clarify the instructions.

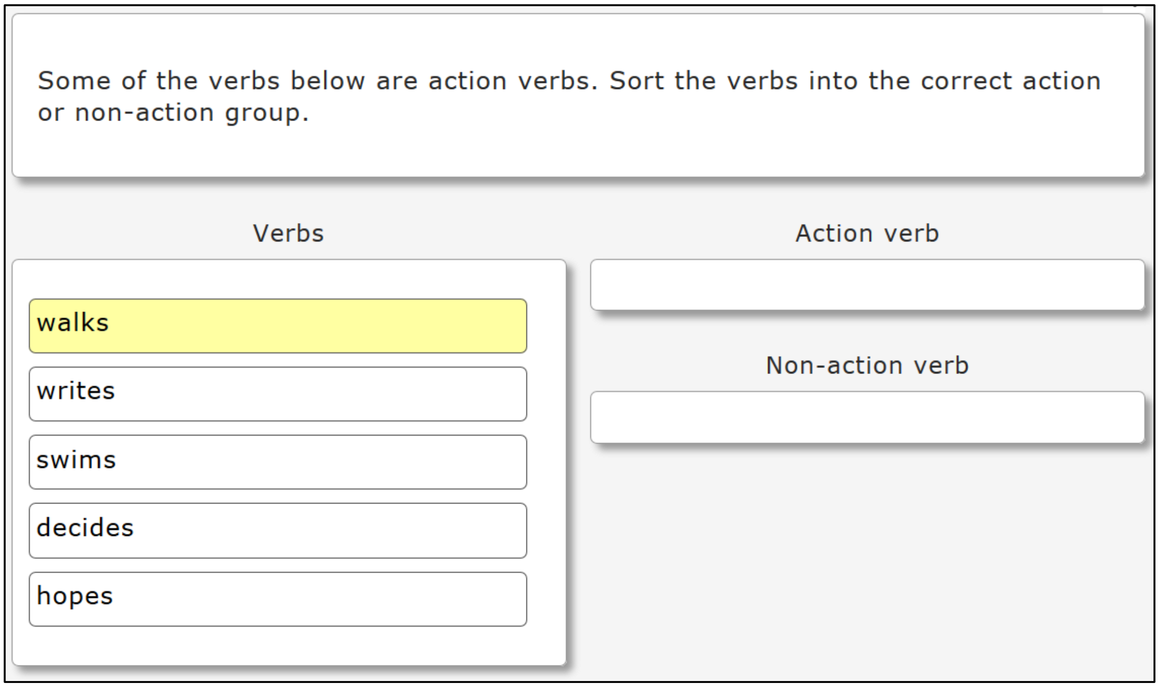


Figure 5.3.2.1. Elementary ELA click-to-select categorization item.

The second click-to-select categorization item for elementary students (Figure 5.3.2.2) presented several barriers to access for online test takers. Students offered several suggestions for improving this item: provide a picture of a square and a rhombus, add the ability to change answers, place the answer choices closer together, and provide more detailed instructions about how to respond to the task. Students who completed this item in braille did not report any issues with the item.

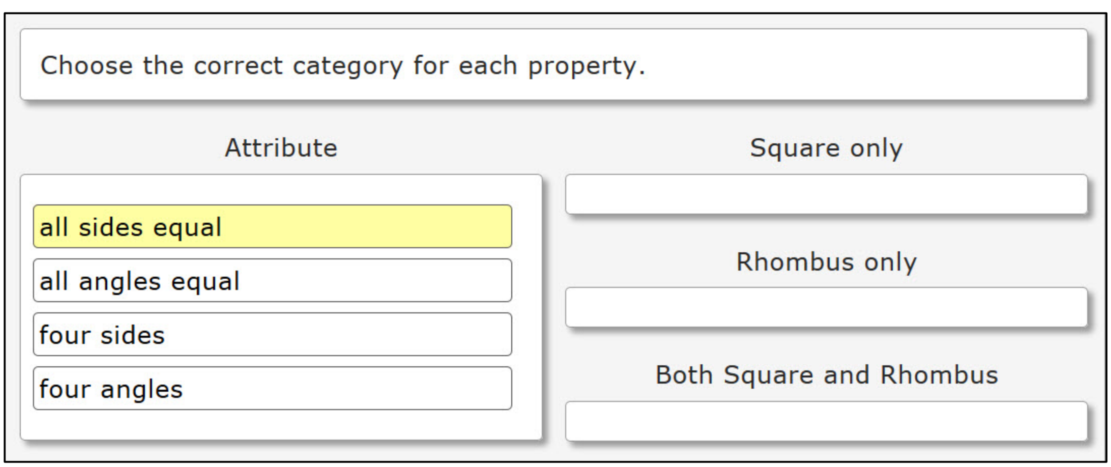


Figure 5.3.2.2. Elementary math click-to-select categorization item.

Cognitive labs for middle school students included the click-to-select categorization item in Figure 5.3.2.3. For this item, some online test takers did not understand the click-to-select function and were therefore unable to provide responses to the item. Students suggested that, rather than click to select, this item should be formatted with check boxes under each prompt. Students who completed this item in braille requested a print form to accompany the braille form but had no problems completing the task.

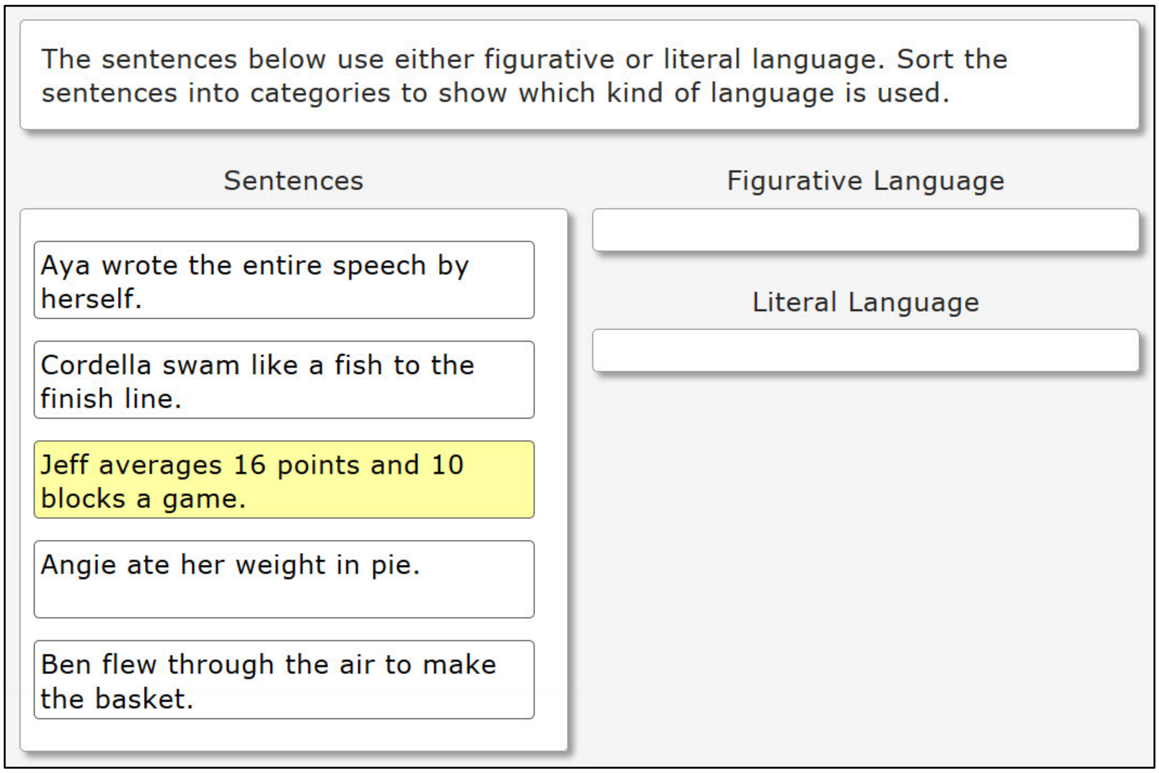


Figure 5.3.2.3. Middle school ELA click-to-select categorization item.

The second click-to-select categorization item for middle school is shown in Figure 5.3.2.4. Students who completed the online version had difficulty viewing the “Noun,” “Verb,” and “Adjective” boxes on the right side of the screen with enlargement. Students were also confused about how to provide a response to the item. Students requested that the instructions be clarified.

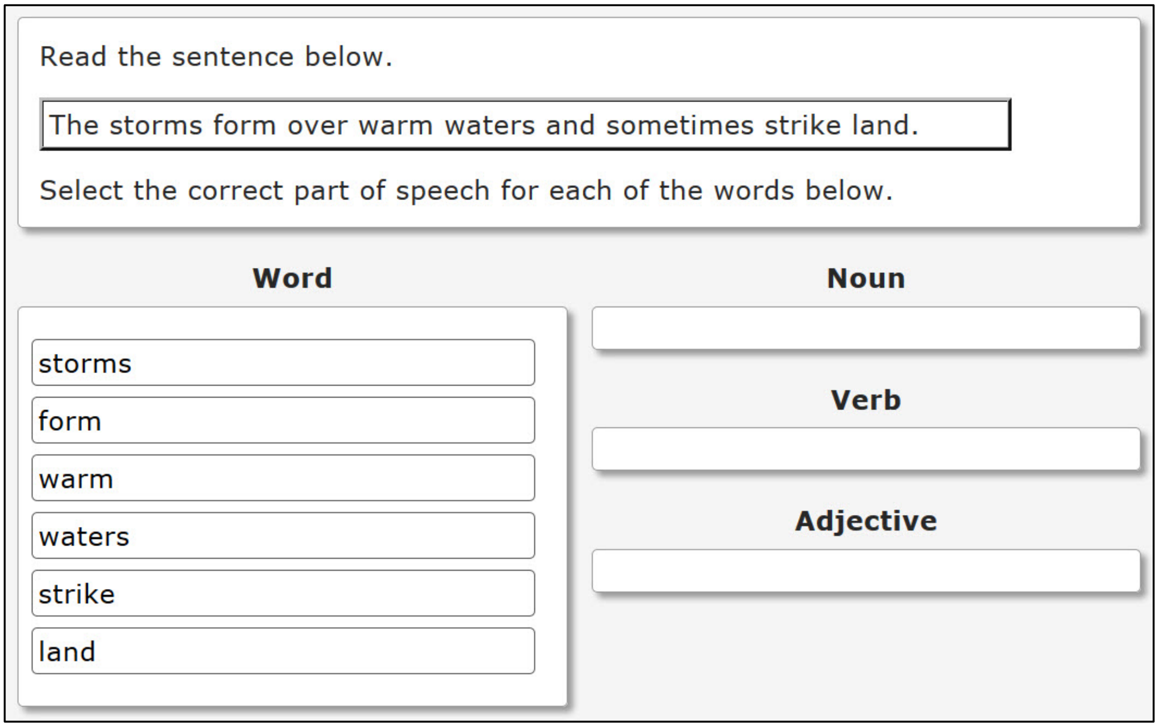


Figure 5.3.2.4. Middle school ELA click-to-select categorization item.

Middle school students who completed the item in braille had several suggestions for improving item accessibility: number the items, do not list the words *noun*, *verb*, and *adjective* after each word, and provide the words *noun*, *verb*, and *adjective* before the list of words.

One high school level click-to-select categorization item is shown in Figure 5.3.2.5. For online users, scrolling to access all item content was problematic. Test takers also said that it was difficult to see the highlighting in the item that appeared when a selection was made.

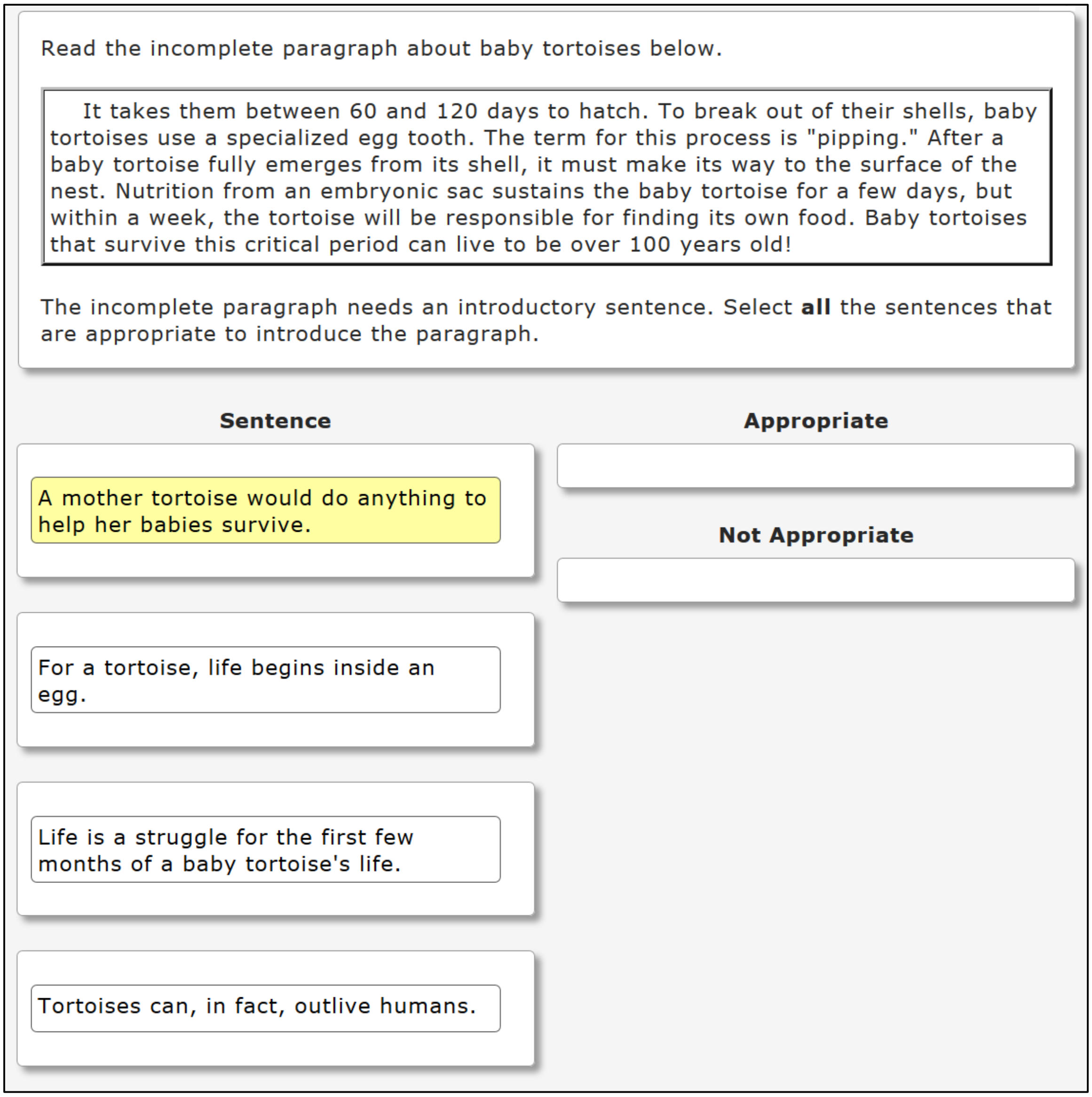


Figure 5.3.2.5. High school ELA click-to-select categorization item.

On the paper-and-pencil and braille versions of this item, the sentences were listed below the text and the student was asked to mark all of the appropriate introductory sentences. One braille reader who also used the online form preferred a drop-down list for this task. Another said the item would be easier to access if they could underline sentences within the text. On the print version, one student was confused about what to do with the sentences that were “not appropriate.”

Another high school level click-to-select categorization item is shown in Figure 5.3.2.6. Most students who completed the item online had no barriers. One student had difficulty accessing the equations and understanding how to respond to the task.

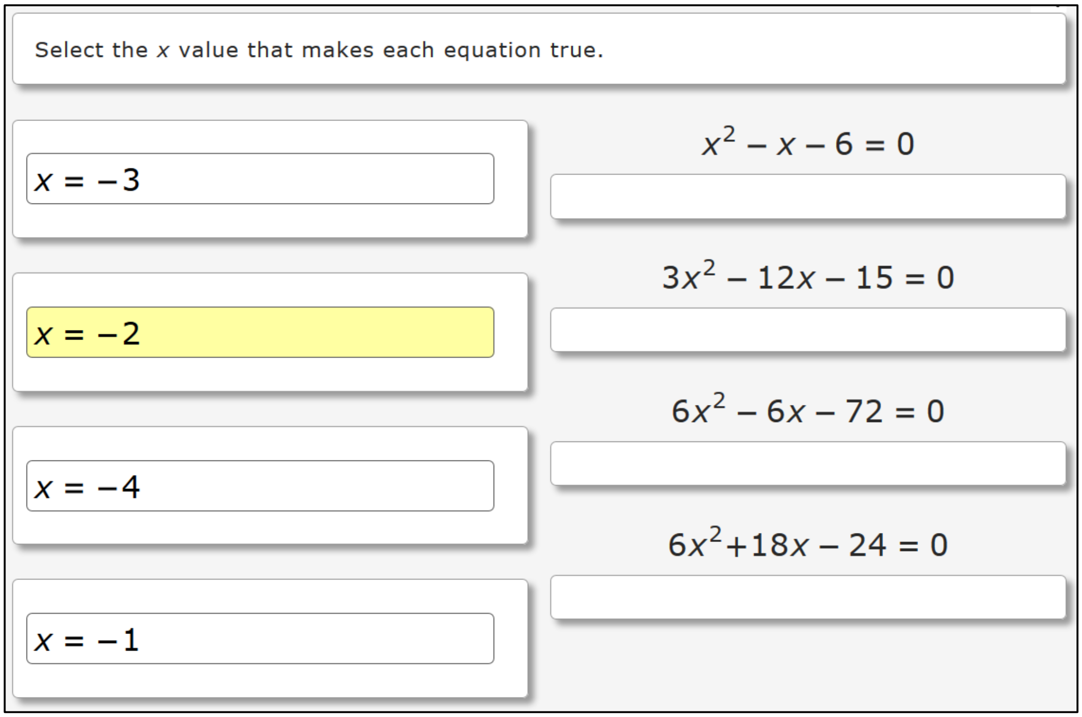


Figure 5.3.2.6. High school math click-to-select categorization item.

To improve the task, students provided the following suggestions: provide clearer, more detailed instructions about how to use the click-to-select function or change the item format to a drop-down item in which the values of *x* are listed under each equation. Another student suggested reformatting the item into a table (see Figure 5.3.2.7).

Answer Options	Value of X	Equation
$x^2 - x - 6 = 0$	-1	
$3x^2 - 12x - 15 = 0$	-2	
$6x^2 - 6x - 72 = 0$	-3	
$6x^2 - 18x - 24 = 0$	-4	

Figure 5.3.2.7. A student’s suggestion for improving the layout of the item in Figure 5.3.2.6.

Students who completed the item in braille as a matching item did not provide much feedback on the item in Figure 5.3.2.6. As with other matching items in braille, one student stated that the item was confusing and the task instructions should be clearer. A student who completed this item in print reported that he liked the item's format.

Matching. Matching items require the student to create lines that connect two corresponding elements, one from each column. Once a student has matched items in the left and right columns, a line appears to connect the two choices. An example of a matching item is shown in Figure 5.3.2.8.

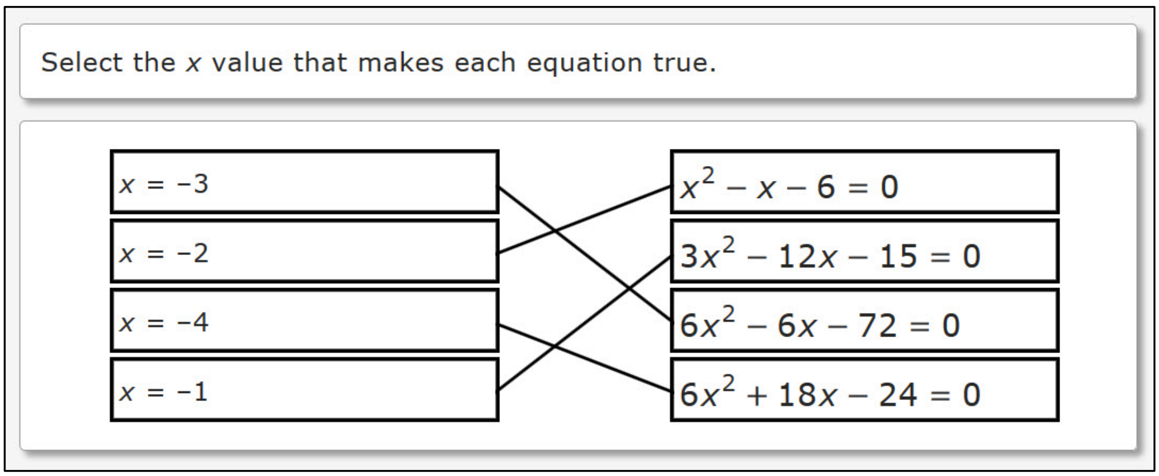


Figure 5.3.2.8. Matching item after student has selected answers.

Teacher panels with matching items. The matching items used in the teacher panels and cognitive labs used color to emphasize the matches selected by the student. Panelists were concerned about the colors in the items and the amount of space between the left and right columns (see Figure 5.3.2.9). They thought the colored lines between the two columns might be confusing to students. This would be especially problematic when a student changed an answer. In this case, two lines could appear in the same color. These comments, along with problems that could result when a student uses reverse contrast or online color overlays, resulted in a revision to remove the color from matching items.

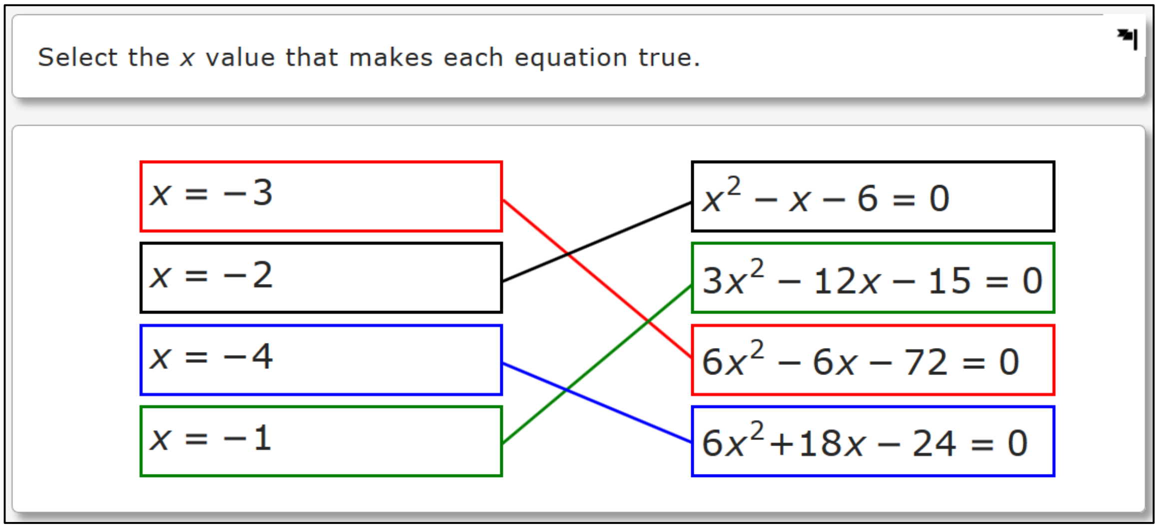


Figure 5.3.2.9. Matching item after student has selected answers.

Additionally, looking back and forth between the left and right columns requires visual tracking. Students would also have to scan up and down to view all of the information in the item and then scan from left to right to provide answers. All of these item features may present difficulty for visually impaired students. Likewise, maneuvering between the two columns could be difficult for students with motor disabilities.

Cognitive labs with matching items. Elementary students completed three matching items during the cognitive labs. On the first item (Figure 5.3.2.10), students expressed concern that the left and right columns were too far apart. Other students were observed to have difficulty moving back and forth between the left and right columns.

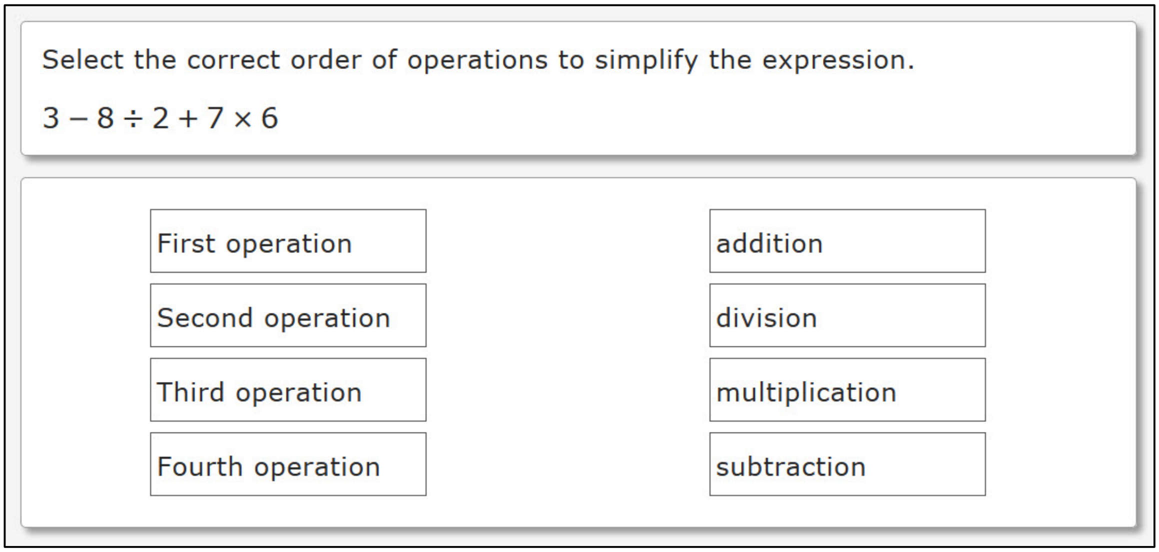


Figure 5.3.2.10. Elementary math matching item.

One student had difficulty understanding the matching function. This student requested clearer instructions for responding to the task (i.e., “Click a selection on the left, and then click a selection on the right.”). Two students completed the item in braille. One student did not give feedback on the item. The other student expressed confusion about how to complete the task.

The second elementary school matching item (Figure 5.3.2.11) posed issues due to scrolling. The students could only see one answer choice at a time; other answer choices were below the visible screen. Students suggested that drop-down menus would make the item more accessible.

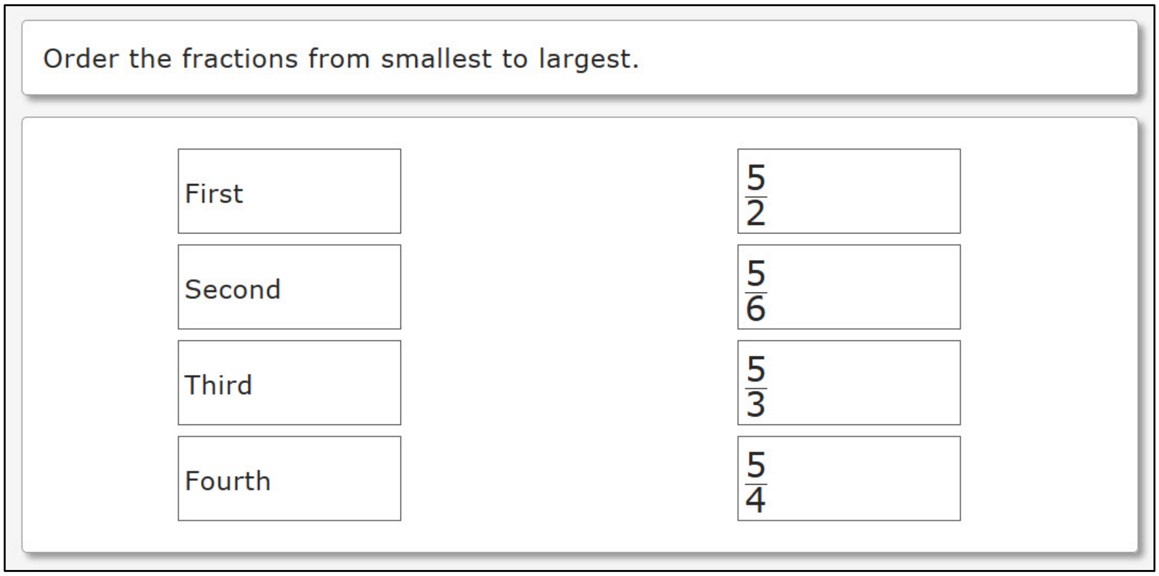


Figure 5.3.2.11. Elementary math matching item.

Students completed the item shown in Figure 5.3.2.12 either in the online format or in braille. The online version of this item was a matching item, as shown. One student who completed the item online had to enlarge the font to 42 points. This required the student to scroll from left to right to view all of the content, which increased the time necessary for task completion. The students suggested that the answer choices be placed closer together or that the item be changed to a drop-down menu format. Another suggestion for improving the item was to make the instructions clearer. One student attempted to click within the diagram to answer the question. The student suggested that instructions should clearly state where to locate the answer choices.

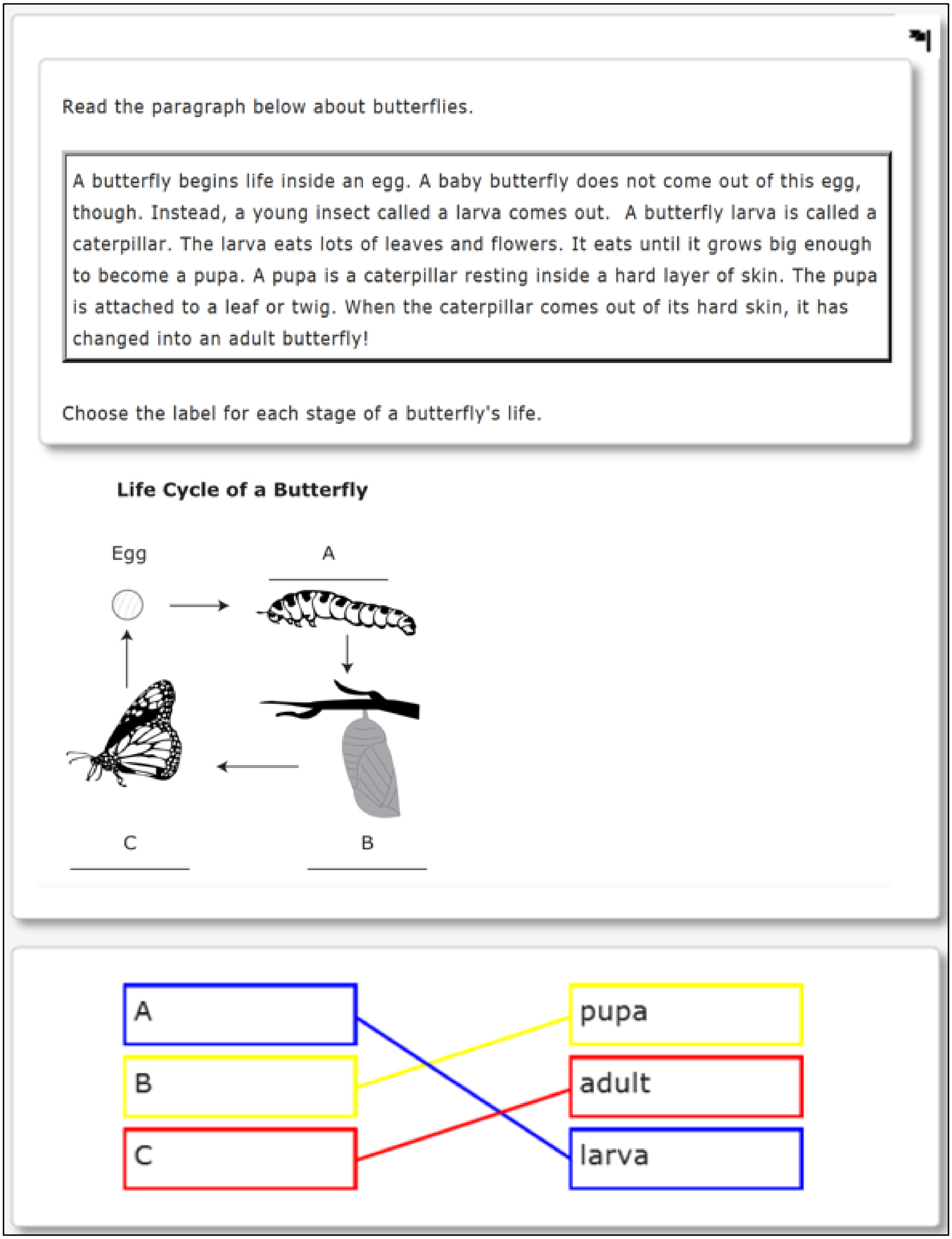


Figure 5.3.2.12. Elementary ELA matching item with background graphic.

Students who completed this item in braille noted that the braille graphic was confusing. Students were unable to determine which parts of the item represented answer choices. To improve the accessibility of the braille form, the instructions should be clarified so that students will know how to interact with the item. Students also suggested that the diagram be removed from the item. Instead of the diagram, the item could include a list of prompts that students could match to answer choices.

Middle school students completed a cognitive lab with a number line item in a matching format (see Figure 5.3.2.13). One student commented that the item required too much scrolling in order to view the item with screen magnification. Several students were unfamiliar with the content and had difficulty understanding how to complete the task. To improve the item, the instructions could be clarified to direct students to first solve the equations and then match them to the corresponding letters on the number line.

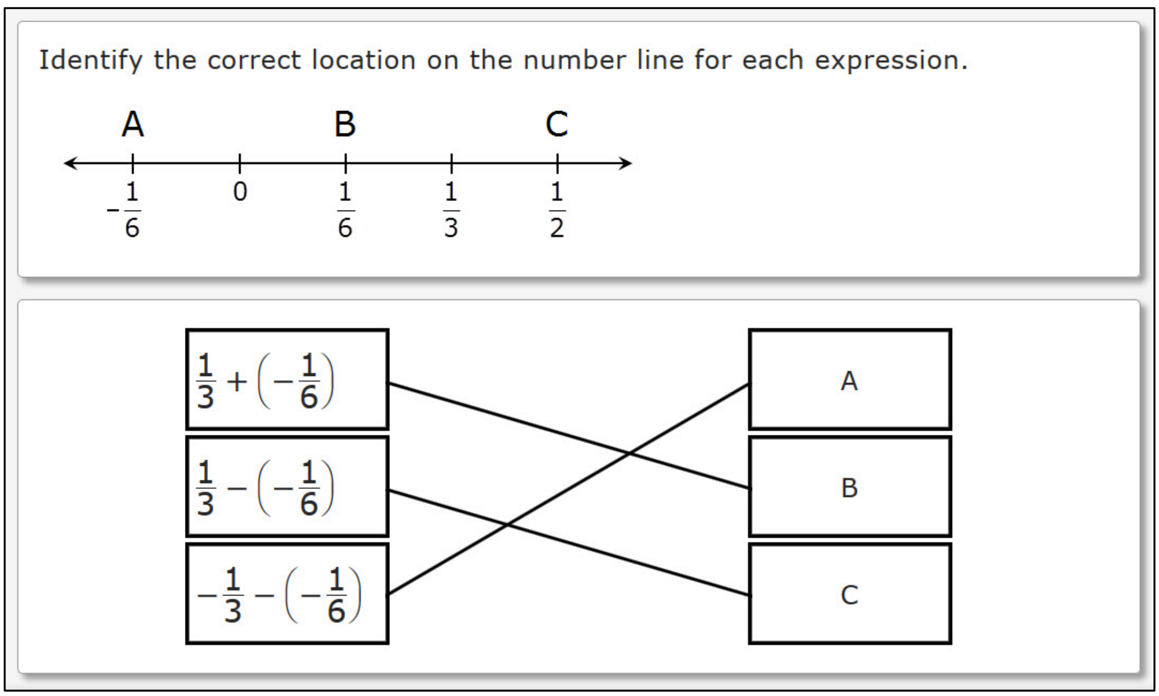


Figure 5.3.2.13. Middle school number line item.

Middle school students who completed the braille version of the item in Figure 5.3.2.13 did not understand what the item asked or how to respond. These students recommended clarifying the instructions. Students who completed the same item in the print form also expressed confusion about how to answer the item. On the print version, one student reported that he/she could not see the numbers on the number line because the print was too small.

One middle school math item required students to match shapes (see Figure 5.3.2.14). Middle school students who answered the item online reported three main issues. Three students reported that the directions were unclear; one student stated that they had to scroll up and down a lot in order to respond to the item; and another student was confused by the different colored lines once he selected his choices. No other issues or suggestions for improvement were made regarding the online form of this item.

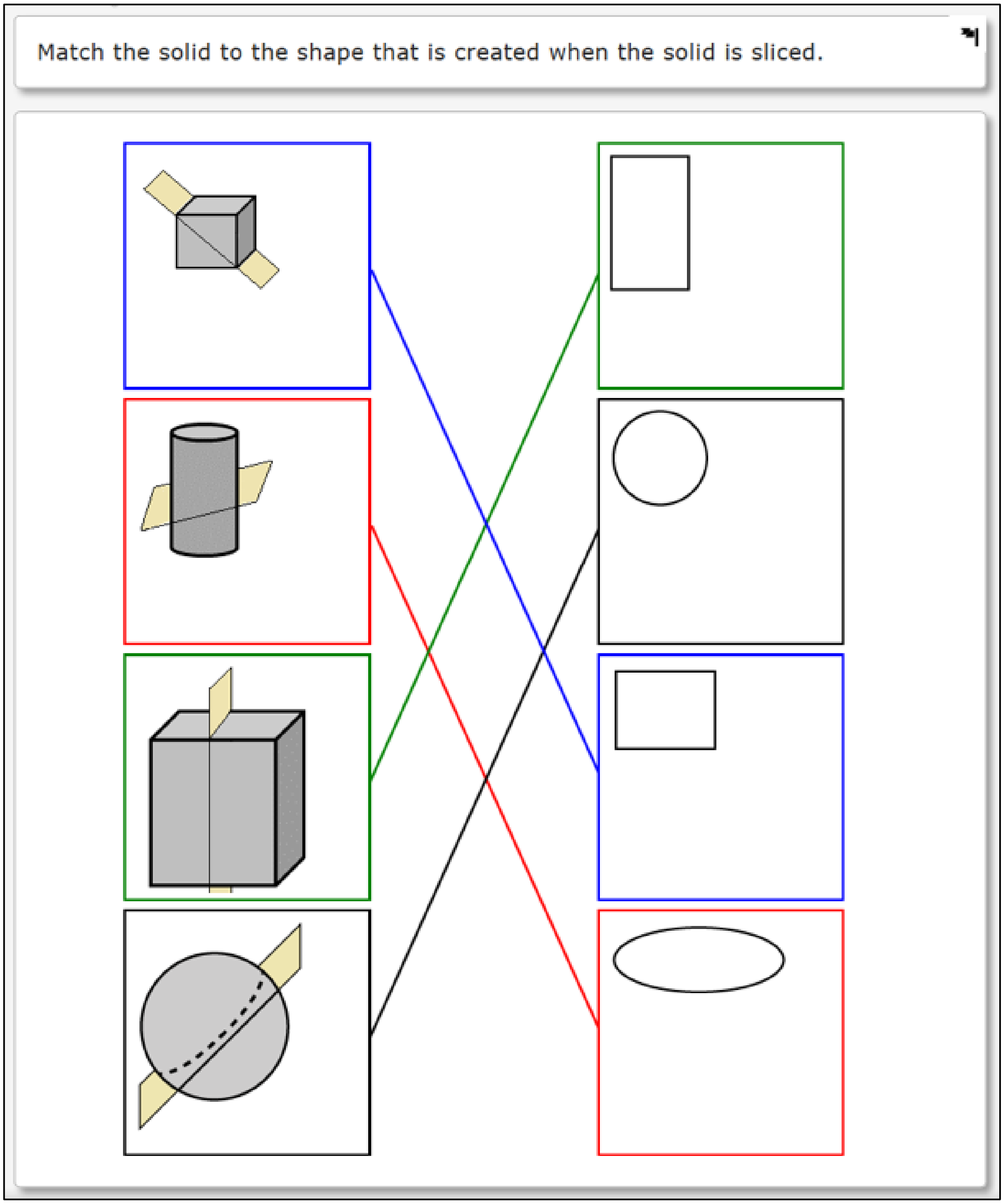


Figure 5.3.2.14. Middle school math matching item.

The middle school student who completed the item on the braille form had trouble understanding the solids. The student was not sure how to respond to the question and simply pointed to the answer choices. The student suggested that letters and numbers on the shapes in each column would improve the task. This student’s suggestion led to that revision, among others, for matching items presented in braille.

Middle school students who completed a print version of the item reported having some trouble understanding what the item asked them to do. One student stated that the pictures did not make sense, and another student stated that he did not understand the question. One student suggested that the instructions should be clearer with regard to how the figures in the left column relate to the shapes in the right column.

At the middle school level, students viewed the item shown in Figure 5.3.2.15.

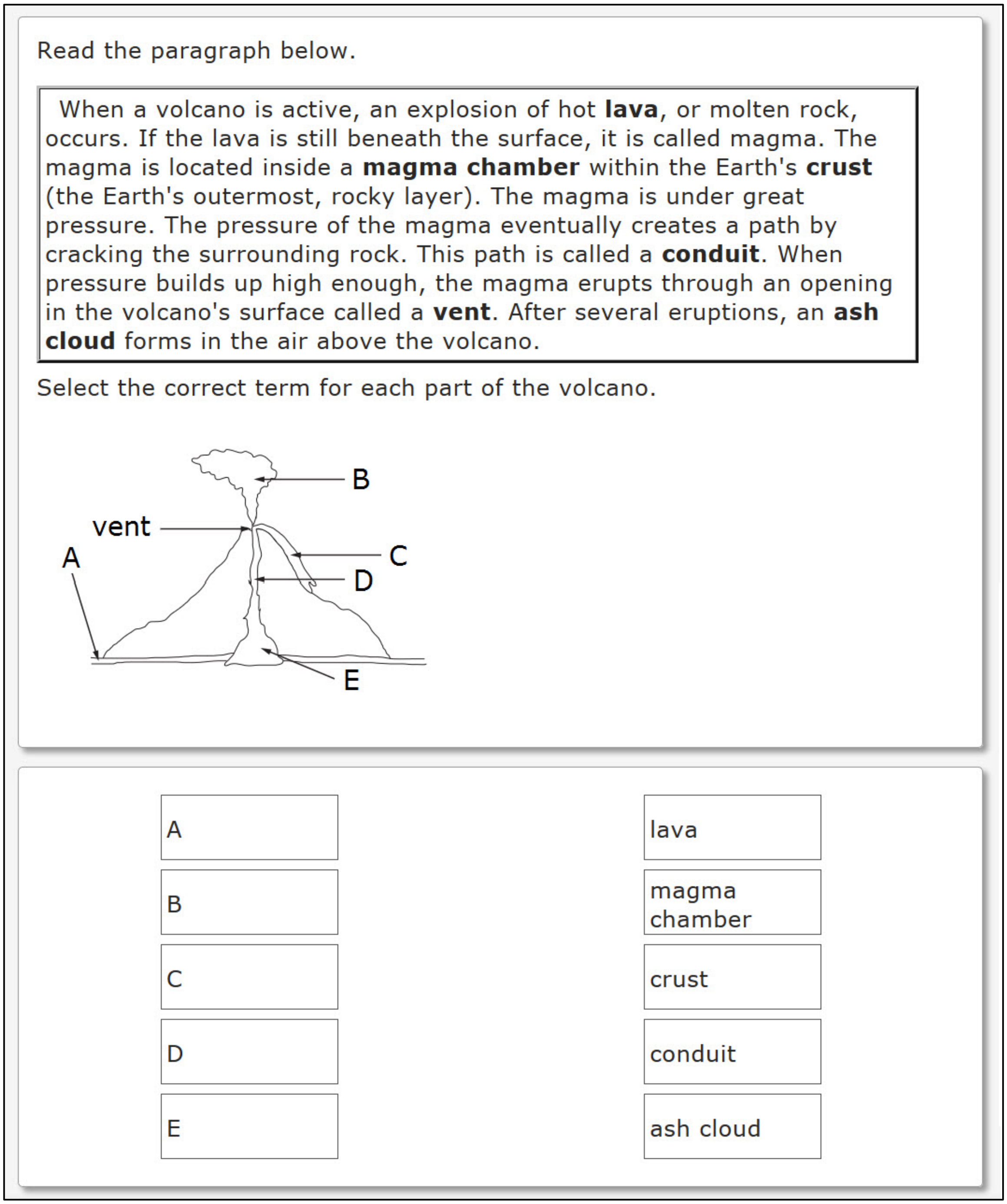


Figure 5.3.2.15. Middle school ELA matching item.

Middle school students who completed this item online commented that the information on the screen appeared stretched. Scrolling through the item presented a barrier to responding to the task, as well. For these reasons, students liked having the print form as a backup. Additionally, students were confused about the colors of the lines once matches were made. One student thought that a red line indicated a wrong answer and a green line indicated a correct answer. Some suggestions for improving the task were to have the letter from the answer choices turn blue when the letter was selected, clarify the instructions, provide a sample item that shows the student how to complete the task, and change the task to a drop-down-menu item type. Students also reported that the matching function did not work properly with screen enlargement, a defect that was reported.

Students who completed this item in the braille format stated that they did not understand what to do and thought they had incomplete information. The tactile graphic had erroneously omitted the letters marking the parts of the diagram, and the list of words to describe the parts of the volcano was missing. Therefore, students were understandably unsure how to respond to the task.

High school students also completed a matching item in math (see Figure 5.3.2.16). These students expressed many accessibility issues with the item. First, the screen reader did not read the expressions correctly. Students were also confused about how to respond to the task. One student commented that she did not understand how to match elements in the left column to expressions in the right column. Another student found it hard to delete a line in order to change his response.

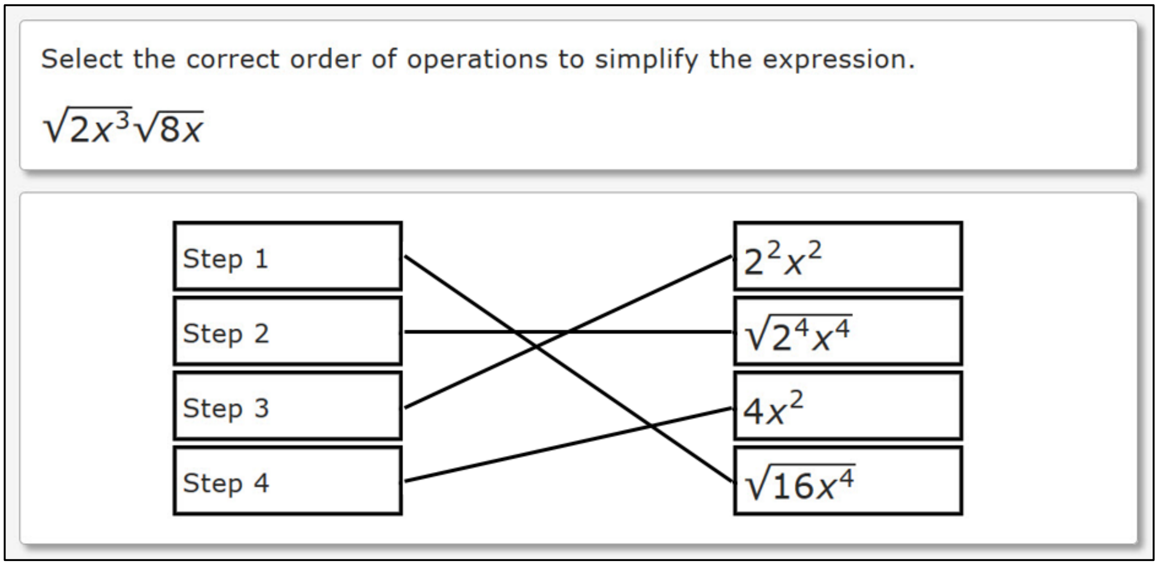


Figure 5.3.2.16. High school math matching item.

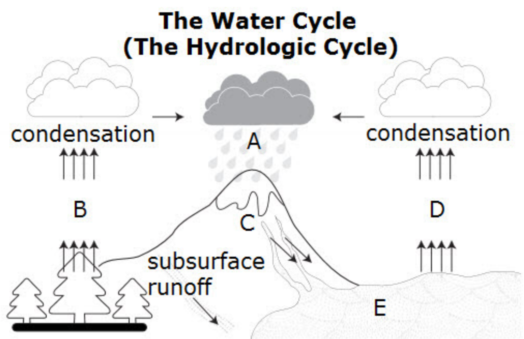
Students had different ideas about how to improve the online format of the item in Figure 5.3.2.16. One student suggested that the item should be presented in a multiple-choice format. Another student suggested a matrix of radio buttons. A third student preferred the matching format because it was "easier to use." Students suggested that, to improve the online version of this item, item instructions should be clarified and the process of changing an answer by deleting the lines should be made easier. Students who completed the item online noted a defect in that the test taker could alter or delete answer choices within the item. Identification of defects such as these helped developers improve the test delivery platform.

Teachers of high school students who completed the print and braille versions of the item in Figure 5.3.2.16 noted that their students were often unsure how to respond to matching items in braille, perhaps because of the unfamiliarity of the column layout and the lack of numbers and letters on the elements to be matched. Drop-down menus would be preferable.

High school students received the background graphic matching item shown in Figure 5.3.2.17. Students using the online version of the test said the graphic was difficult to see because it was small and dense with information, which created a barrier to accessing the content of the task. Test takers also reported that scrolling, which was required in order to see the full item, was labor intensive for students with motor impairments. When using the zoom feature, matching options were spaced far apart and the students had to move the screen from left to right. Students also reported that the online directions were not clear. Online test takers suggested using drop-down menus instead of letters or moving the matching options closer together.

Read the water cycle terms and definitions below.

Accumulation: a process in which water pools in large bodies such as oceans, seas, and lakes
Condensation: a process in which water vapor cools and forms the tiny drops of water that make up clouds
Evaporation: a process in which water from the surface of an ocean, lake, or land turns into vapor
Precipitation: a process in which water falls from the sky in the form of rain, snow, sleet, or hail
Subsurface Runoff: rain, melted snow, or any other water that flows in underground streams, drains or sewers
Transpiration: a process in which some of the water contained in plants evaporates into the atmosphere



Match the labels to the parts of the water cycle.

A	accumulation
B	evaporation
C	precipitation
D	surface runoff
E	transpiration

Figure 5.3.2.17. High school ELA matching item.

High school students who used the braille form also found the graphic cluttered and thought the diagram and answer choices were confusing. These students assumed that the letters in the diagram matched the letters of the answer choices. For instance, the students thought the letter E on the diagram (lake) was supposed to match letter E in the answers (transpiration). This confusion between the left and right columns in matching tasks was apparent to several braille users, who suggested an alternate vertical format for the letters and answer options.

High school students who used the print version of the test had issues accessing the content of the task. Students commented that the diagram was confusing, and they said the correct choice for letter A was not present. Print form respondents requested clearer labels on the diagram. Students did not like the way the diagram illustrated the water cycle, and students thought the letters were hard to find in the graphic.

Another matching item included in the high school cognitive labs is shown in Figure 5.3.2.18. Online test takers had to magnify the screen in order to see the item content. However, when the item was magnified, students then had to scroll from one part of the item to another, which was difficult. Some students were not able to select and match parts of the item. Online test takers requested making the item font larger to reduce the need for screen magnification. For the most part, however, online test takers had no trouble with the matching format, perhaps because they had already experienced several matching items.

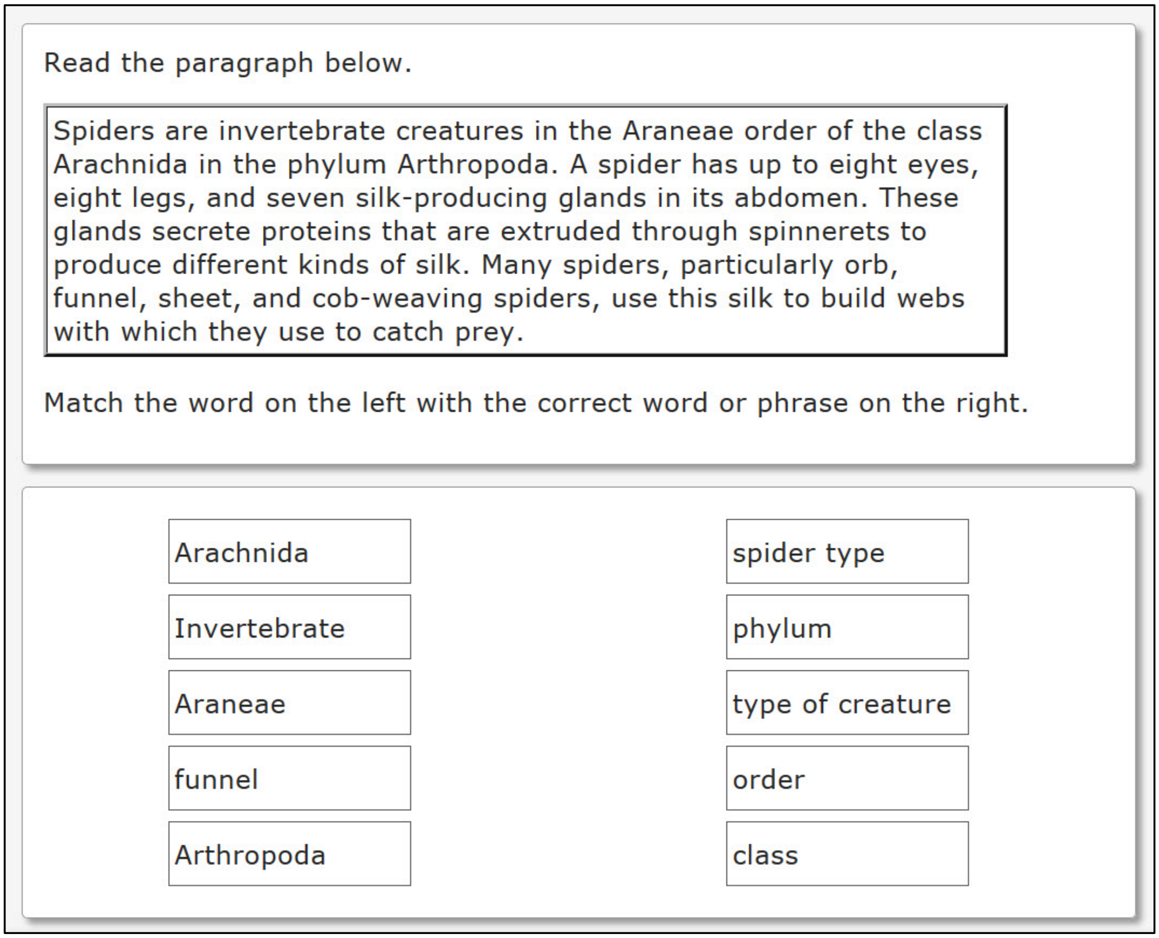


Figure 5.3.2.18. High school ELA matching item.

Braille users of the item in Figure 5.3.2.18 said that the screen reader worked accurately and that the numbers and letters on the elements in the left and right columns, respectively, enabled students to respond to the task, even though the item content was too difficult for some students. Teachers suggested using drop-down menus, radio buttons, or check boxes instead of the matching format.

Select Text. Select text items require the student to select information within the context of a paragraph or longer passage. The student may select one or more words, phrases, or sentences within a passage. When the cursor hovers over a word, phrase, or sentence, there is visual feedback to indicate that the element is a selectable response. At the time of the teacher panels and cognitive labs, a selectable response appeared outlined in a red box when the user hovered over it in the passage (see Figure 5.3.2.19). After the user selected a response, the selection was shown highlighted in yellow (see Figure 5.3.2.20).

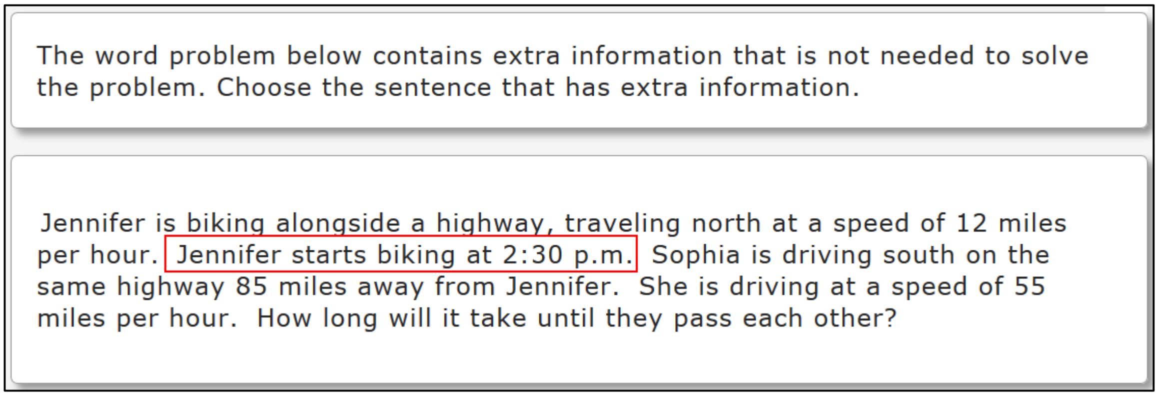


Figure 5.3.2.19. Select text item viewed as user hovers over a selectable response with the cursor.

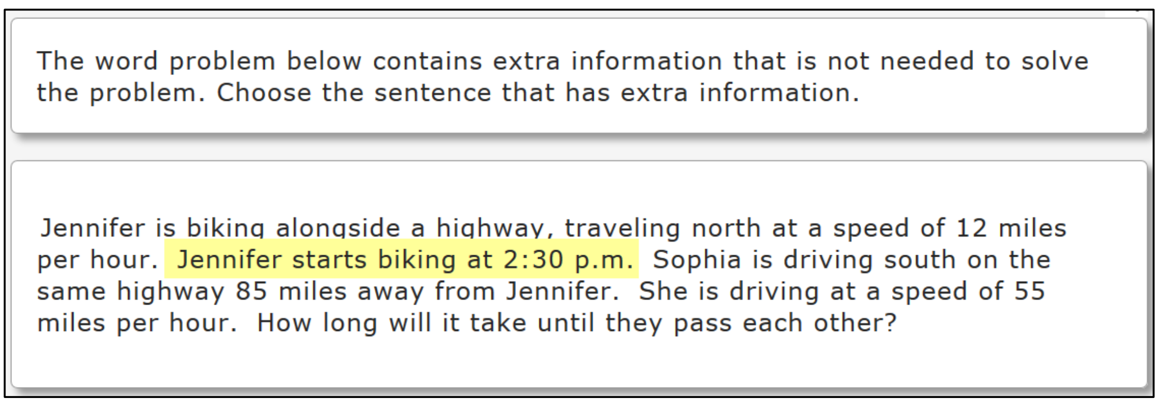


Figure 5.3.2.20. Select text item viewed after user has selected a sentence.

The presentation of select text items in KITE has changed since the teacher panels and cognitive labs. When the answer options are entire sentences and the student hovers over a sentence, a red box outlines the selection and the text is highlighted to indicate that it is a selectable response. After the student selects the text, the red box remains, outlining the selection (see Figures 5.3.2.21 and 5.3.2.22).

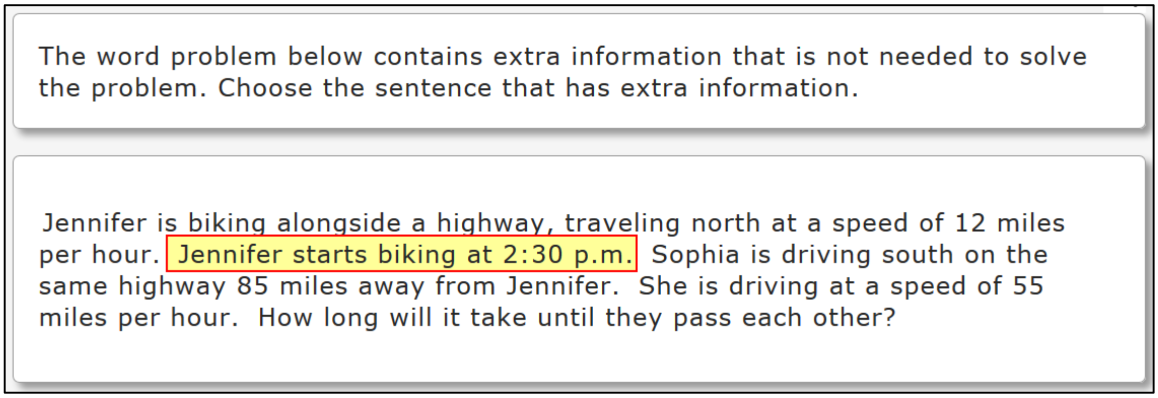


Figure 5.3.2.21. Revised select text item viewed as student hovers over a selectable response with the cursor.

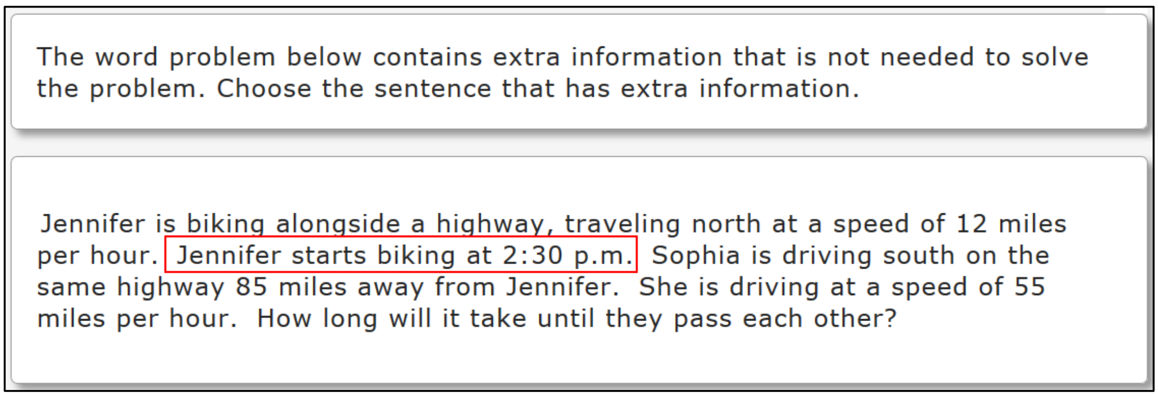


Figure 5.3.2.22. Revised select text item viewed after student has selected a sentence.

In a variant of the select text format where the answer options are words or phrases rather than complete sentences, the answer options appear in bold font so the student knows which words are available for selection (see Figure 5.3.2.23). Multiple responses are possible in both types of select text items.

Read the excerpt below from the conclusion of the *Personal Memoirs of General U.S. Grant*. Select the word that means "disgrace."

[The Southern States] saw their power waning, and this led them to encroach upon the prerogatives and independence of the Northern States by enacting such laws as the Fugitive Slave Law. By this law every Northern man was obliged, when properly summoned, to turn out and help apprehend the runaway slave of a Southern man. Northern marshals became slave-catchers, and Northern courts had to contribute to the support and protection of the institution.

This was a **degradation** which the North would not permit any longer than until they could get the power to **expunge** such laws from the **statute** books. Prior to the time of these **encroachments** the great majority of the people of the North had no particular **quarrel** with slavery, so long as they were not forced to have it themselves. But they were not willing to play the role of police for the South in the **protection** of this particular institution.

Figure 5.3.2.23. Select text item in which answer options are bolded.

Teacher panels with select text items. The panelists generally liked the red box that appeared around the answer choices when hovering with the cursor; however, some panel teams expressed concern that the red lines would be problematic for students with colorblindness. Panelists suggested that once a student selects a word, phrase, or sentence within the text, the answer choice should be distinguished from the surrounding text in several ways (e.g., highlighted, bolded, different font, or border). Several panel teams expressed concern that visually impaired students might require a scribe in order to answer these items.

In order to accommodate blind or visually impaired students, the panel recommended that a screen reader should be available. The panel also stated that answer choices be listed either before or after the passage so the sighted reader could easily view the answer choices. The screen reader should tell the student what he/she has selected. Additionally, the panel recommended the addition of a text box to allow students to enter their responses using a keyboard or braillewriter. The panelists also recommended increasing the font size, making the highlighting in selected choices brighter, and double-spacing between lines for ease of reading.

For students with motor disabilities, the panel recommended the use of an assistive technology device such as switch technology. The student could use the switch or tab key to cycle through words or sentences. In this manner, the student could still access the materials utilizing the keyboard and make a selection with the second switch or enter key.

In addition, the panelists made general recommendations to improve the accessibility of select text items. For items in which the possible answer choices were bolded, the panel suggested clarifying the directions (e.g., "Select from the words in bold.") The panelists also suggested that the directions should be more specific about how to change an answer once the student has made a selection.

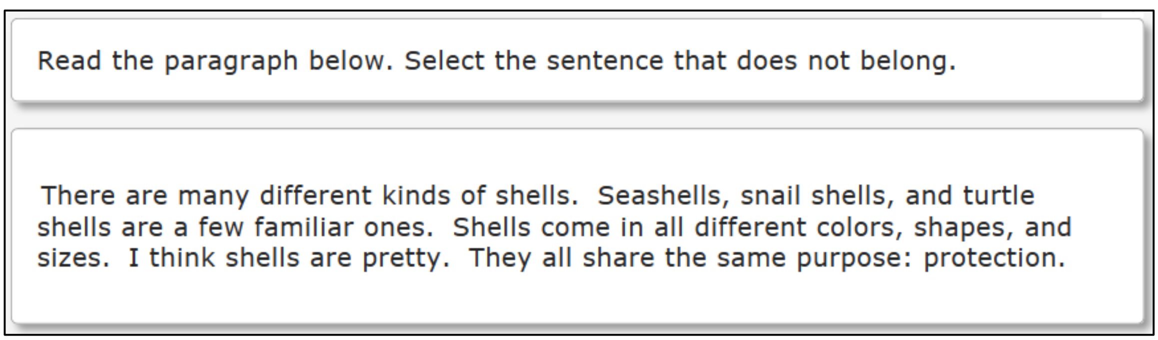
Cognitive labs with select text items. At the elementary level, the cognitive labs included two select text items. The first item (Figure 5.3.2.24) posed few barriers to accessibility for students. Students who completed the item online only expressed concern about scrolling through a magnified item. Students who completed the item in braille did not express any accessibility concerns.

> The word problem below has extra information that is not needed to solve the problem. Select the sentence that has the extra information.
>
> Amy bought a bookshelf with 3 shelves. She can put a maximum of 25 books on each shelf. The maximum weight each shelf can hold is 50 pounds. What is the greatest number of books Amy can put on the bookshelf?

Figure 5.3.2.24. Elementary math select text item.

The second select text item (Figure 5.3.2.25) posed accessibility concerns similar to the previous item for online users. Students who completed this item online with screen magnification had to scroll across the screen to read the passage. A way to improve this task would be to allow text magnification with text wrapping without forcing students to scroll across the screen. One student also had difficulty understanding how to respond to the task. This student requested clearer instructions (i.e., the instructions should tell the student to click on a sentence in the text).

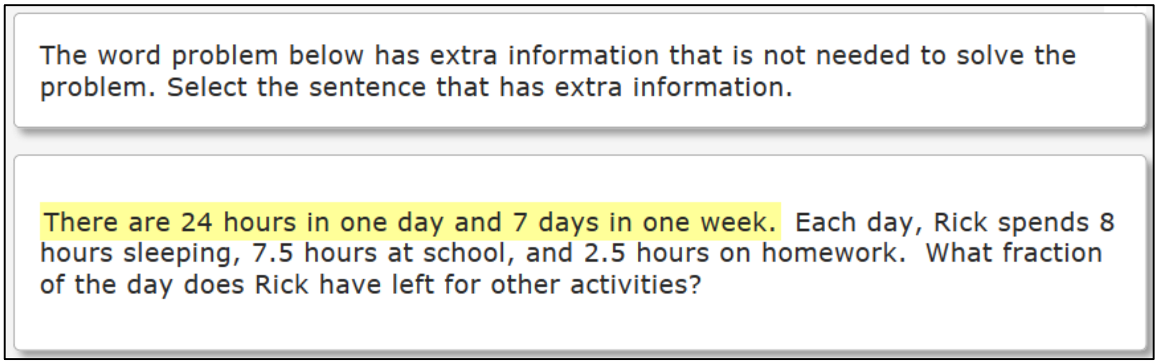
Braille users were unsure how to interact with the item and had difficulty indicating the correct answer because marking in the book is not a typical response for a braille reader. Instead, there should be letters for the answer options so that students can easily indicate the correct response with a braillewriter. Students using the braille format suggested the insertion of a blank line between the instructions and the paragraph in the braille booklet, which would be consistent with the onscreen appearance of the item and would better mark the beginning of the text.


Read the paragraph below. Select the sentence that does not belong.

There are many different kinds of shells. Seashells, snail shells, and turtle shells are a few familiar ones. Shells come in all different colors, shapes, and sizes. I think shells are pretty. They all share the same purpose: protection.

Figure 5.3.2.25. Elementary ELA select text item.

At the middle school level, the cognitive labs included two select text items. The first item (see Figure 5.3.2.26) did not pose any barriers to accessibility for print or braille users, though labeling of the answer options was requested to simplify responding with a braillewriter. One student requested increased text size in the online version.



The word problem below has extra information that is not needed to solve the problem. Select the sentence that has extra information.

There are 24 hours in one day and 7 days in one week. Each day, Rick spends 8 hours sleeping, 7.5 hours at school, and 2.5 hours on homework. What fraction of the day does Rick have left for other activities?

Figure 5.3.2.26. Middle school math select text item.

The second select text item at the middle school level presented more barriers to accessibility. The online version of the item (Figure 5.3.2.27) was confusing for students. Visually impaired students could not distinguish bolded text from other text. Some students were unsure how to respond to the item because they did not know where to click in the text or which pairs of lines were answer choices. One student was confused about how many pairs of lines he needed to choose. This student was also confused about how to change his answer. Another student had difficulty scrolling and referred to the print version of the item when selecting her answers. Online test takers suggested clarifying the instructions and providing a print version of the item with the online version.

Read the poem below by Robert Frost. Select the pair of lines that **best** supports the theme of realizing the importance a decision can have.

The Road Not Taken

by Robert Frost

Two roads diverged in a yellow wood,
And sorry I could not travel both
And be one traveler, long I stood
And looked down one as far as I could
To where it bent in the undergrowth;

Then took the other, as just as fair,
And having perhaps the better claim,
Because it was grassy and wanted wear;
Though as for that the passing there
Had worn them really about the same,

And both that morning equally lay
In leaves no step had trodden black.
Oh, I kept the first for another day!
Yet knowing how way leads on to way,
I doubted if I should ever come back.

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I --
I took the one less traveled by,
And that has made all the difference.

END OF TEXT

Public domain.

Figure 5.3.2.27. Middle school ELA select text item.

A student using the braille form of the item in Figure 5.3.2.27 had difficulty understanding which lines in the poem were answer choices. This student suggested rewriting the instructions so that they state, “The answer choices are in italics.” In braille, markers for emphasis take the place of bold font or italics, so this student was simply recommending the inclusion of a cue about how to locate the answer options in the text.

At the high school level, cognitive labs included two select text items. The first item posed no barriers to students in the online or print formats (Figure 5.3.2.28). Students who completed this item in braille had difficulty responding to the task because they were unable to select or highlight their answer choice in the braille booklet. Half of the students suggested that each sentence should have an accompanying letter, check box, or radio button. Braille readers could then “check” the unnecessary sentence instead of selecting it within the text.

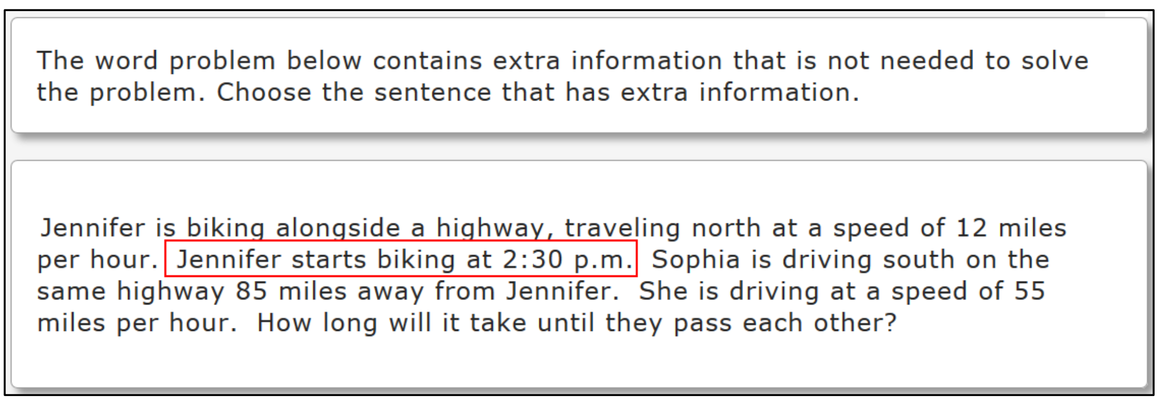


Figure 5.3.2.28. High school math select text item.

The second high school select text item is shown in Figure 5.3.2.29. Students who completed the item online had difficulty reading the content after magnifying the screen suggested that item instructions should say, “Select the bolded word” instead of “Select the word,” but otherwise reported no barriers to access. Print users had no difficulty.

Read the excerpt below from the conclusion of the *Personal Memoirs of General U.S. Grant*. Select the bolded word that means "disgrace."

[The Southern States] saw their power waning, and this led them to encroach upon the prerogatives and independence of the Northern States by enacting such laws as the Fugitive Slave Law. By this law every Northern man was obliged, when properly summoned, to turn out and help apprehend the runaway slave of a Southern man. Northern marshals became slave-catchers, and Northern courts had to contribute to the support and protection of the institution.

This was a **degradation** which the North would not permit any longer than until they could get the power to **expunge** such laws from the **statute** books. Prior to the time of these **encroachments** the great majority of the people of the North had no particular **quarrel** with slavery, so long as they were not forced to have it themselves. But they were not willing to play the role of police for the South in the **protection** of this particular institution.

Figure 5.3.2.29. High school ELA select text item.

For students who completed the item in braille, there was no way for students to select words in the text as instructed in the directions, so it was not possible to answer the item without writing in the booklet or dictating to a scribe. Braille students are not accustomed to writing in a test booklet and typically either dictate their answers or enter them using a braillewriter. As with the middle school select text item above, one student pointed out the difficulty of locating answer options that are marked with emphasis in braille text. After realizing that some words in the second paragraph were emphasized, he went back to the first paragraph to see if he had missed any answer options there. Several students suggested that answer choices could be listed below the passage with letters as in a multiple choice item. This way, students could use easily identify the available options and use letters to select their answers.

Matrix. Matrix items include a grid of columns and rows and serve as alternate representations of categorization tasks. In the matrix format, students review test elements in rows and answer options or labels as column headers, and they select their answers by clicking on radio buttons at the intersections of the rows and columns. When the student selects an answer, the radio button turns blue (see Figure 5.3.2.30). Following feedback from teachers and students, larger radio buttons appeared in matrix items and a green check mark was shown when students selected a response (Figure 5.3.2.31).

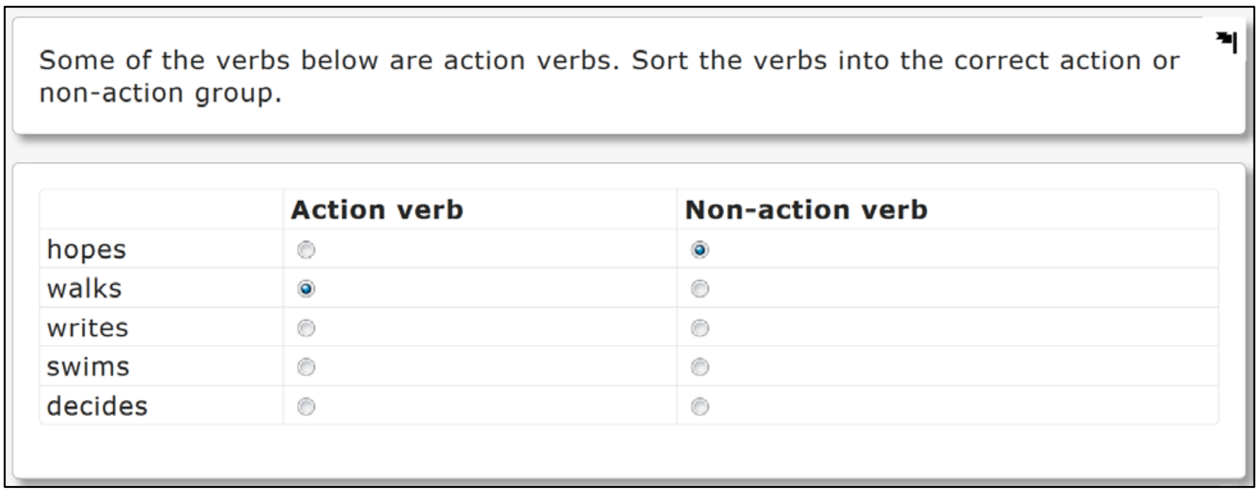


Figure 5.3.2.30. Matrix item with radio buttons in original format.

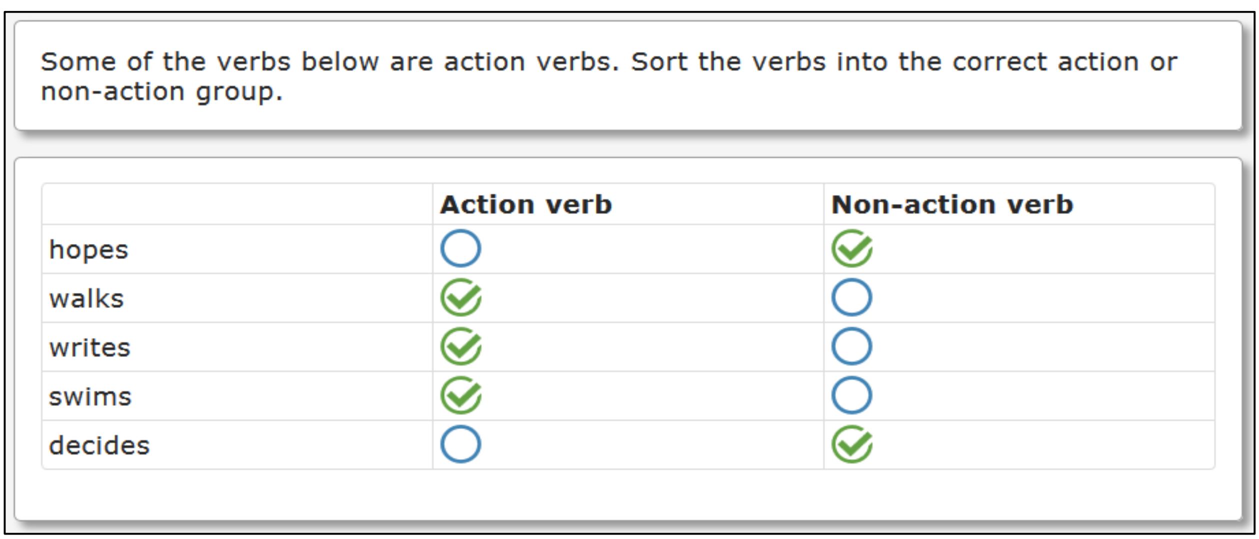


Figure 5.3.2.31. Matrix item with revised radio buttons.

Teacher panels with matrix items. Panelists were concerned about the accessibility of this item type for students with motor and visual impairments. The radio buttons were quite small, and panelists thought that the items required very fine motor skills. Furthermore, teachers thought students would have difficulty distinguishing visually between columns and rows on the matrix.

However, panelists believed that the matrix items (and other item types that contain radio buttons, such as multiple-choice and multiple-select items) could provide wide accessibility for students with a few formatting changes:

- Increase the size of the radio buttons or allow the student to select an entire box, not just a button inside of a box.
- Increase the spacing between rows.
- Increase the contrast between the buttons and the background.
- Make rows and columns more distinct (e.g., shade every other row/column; provide darker borders between rows/columns).
- Provide answer choices next to each radio button rather than just at the top of each column.
- Present the items in braille format.

Cognitive labs with matrix items. Elementary students completed two matrix items. The first task consisted of a passage with the following matrix item below it (Figure 5.3.2.32):

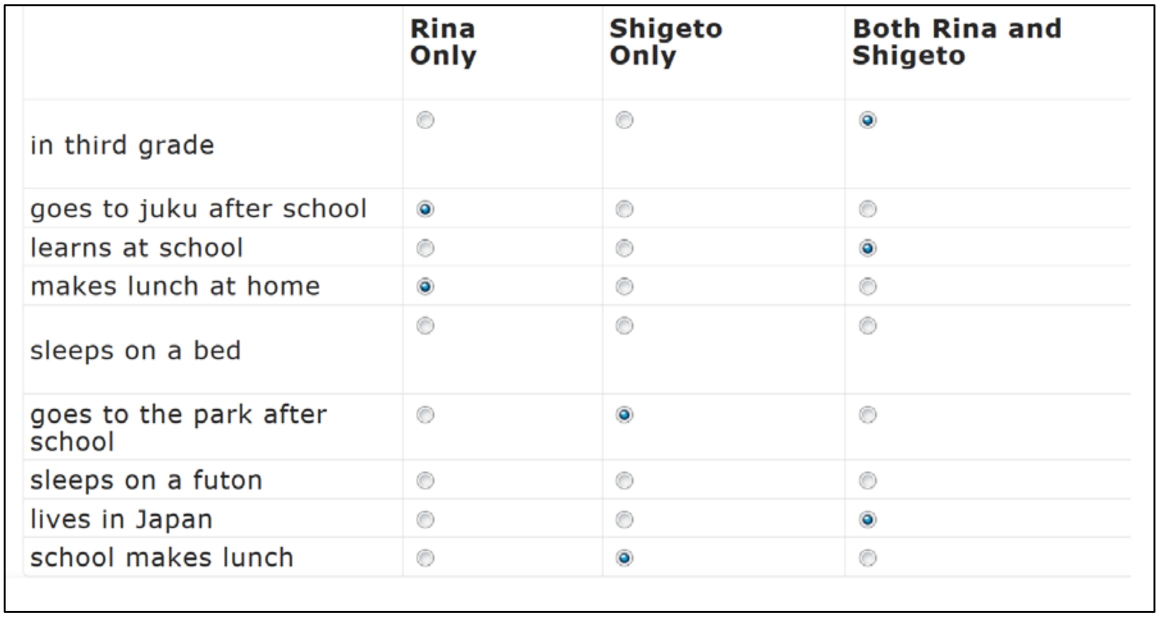


Figure 5.3.2.32. Elementary ELA matrix item used in cognitive labs.

This item posed several accessibility barriers for students who worked with the online version. Scrolling through the lengthy passage presented a barrier to accessing content. One student enlarged the text to 42-point font, and, since the words spilled off the screen, he had to scroll from left to right. The student lost his place a few times, and his reading speed was slower than if he had read the item on an enlarged font paper and pencil test form. The student suggested displaying the answer choices closer together and darkening the dividing lines. One student who completed the item in braille reported no issues but another suggested a revised task layout. The student suggested placing the list of answer choices under the question, as opposed to the right of the question, to improve the item for braille users.

The second matrix item that elementary students completed is shown in Figure 5.3.2.33. Students who completed the item online did not know where to provide their answers because the answer choices were on the far right side of the screen and outside of students' view with screen magnification. The students struggled to respond to the task because they had to scroll from left to right to view answer choices. Students suggested that the noun and verb columns should appear closer to each other. Alternatively, the noun and verb answer choices could be displayed in a drop-down menu format. However, the large radio buttons were noted to be an improvement. Students who completed this item in braille easily access the item but suggested that contents be displayed in lists instead of columns.

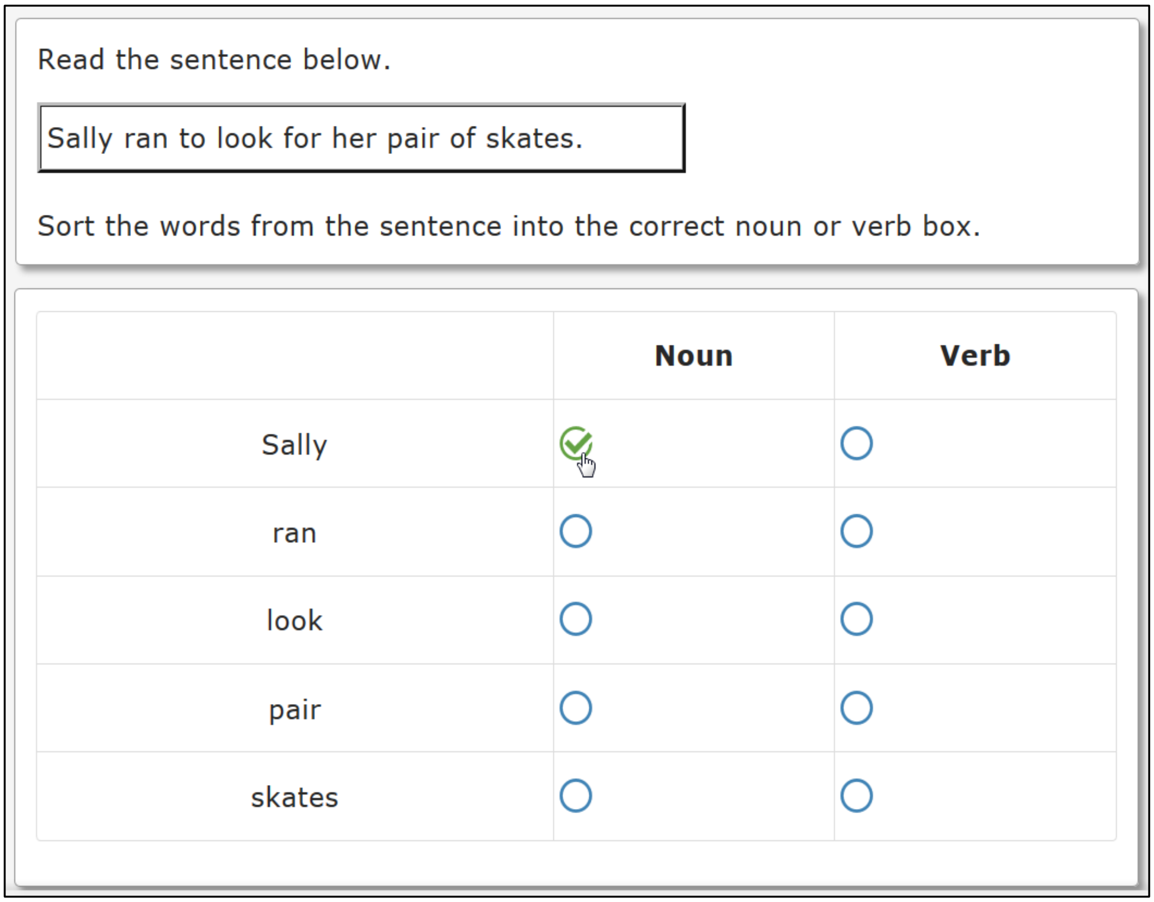


Figure 5.3.2.33. Elementary ELA matrix item used in cognitive labs.

Middle school students completed two matrix items. Students who completed the online version of the first item (Figure 5.3.2.34) had difficulty getting the item to function. A defect or technical glitch prevented some students from accessing the item. Students were also confused about how to respond to the item and had difficulty with the meaning of the word “square” (i.e., “square” as a geometric shape, versus “square” as a mathematical operation). One suggestion for improving this item was to allow students to make only one selection per column.

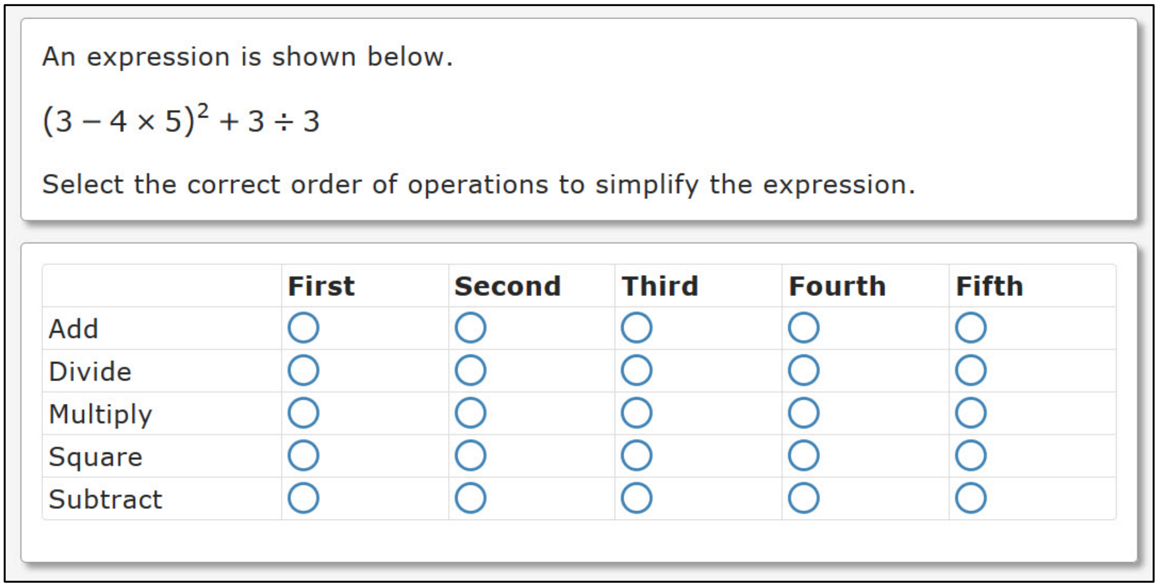


Figure 5.3.2.34. Middle school math matrix item with enlarged radio buttons.

The middle school student who completed this item in braille in a matching layout also expressed confusion about the instructions. The student suggested clarifying the item instructions. (E.g., "Match the operation with the correct step.") The students who completed the print form of the item did not have any suggestions for improvement.

The second matrix item that middle school students examined included a reading passage with prompts and answer choices below the passage. The prompts and answer choices are shown in Figure 5.3.2.35. Online users were confused about the item instructions and thought that they needed more directions. For example, some students did not know how to complete the task because they did not understand why some text within the passage, such as headings, was bolded. In this case, headings in bold were simply an accurate reproduction of the passage as written by the author.

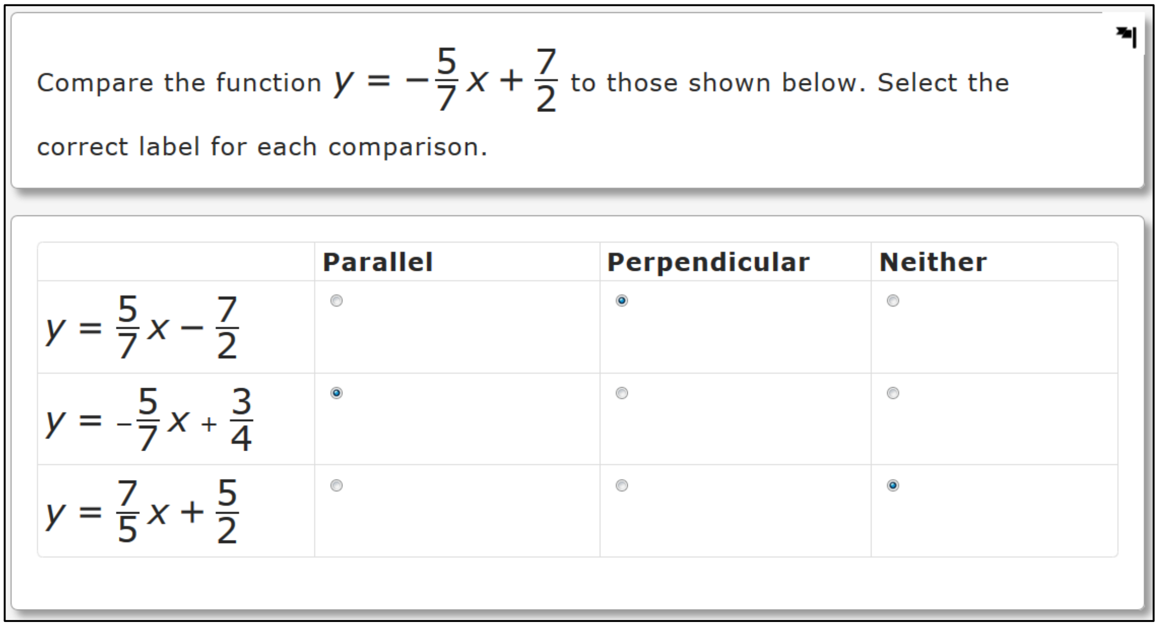
	Red Sky at Night	Red Sky in Morning	Both
caused by water vapor and dust particles in the sky	◯	◯	◯
weather lore	◯	◯	◯
stable air is coming from the west	◯	◯	◯
predicts good sailing weather	◯	◯	◯
higher water content in the atmosphere	◯	◯	◯
predicts bad sailing weather	◯	◯	◯

Figure 5.3.2.35. Middle school ELA matrix item.

One student who completed the item online tried to click on answer choices rather than radio buttons and clicked twice in the wrong row. One student had difficulty scrolling from the text to the answer options, so she used a print version of the item in addition to the online version. One suggested improvement was to organize the online format like the print version where the radio buttons were displayed directly below the answer choices instead of off-center. Students also suggested making the lines darker around the answer choice boxes.

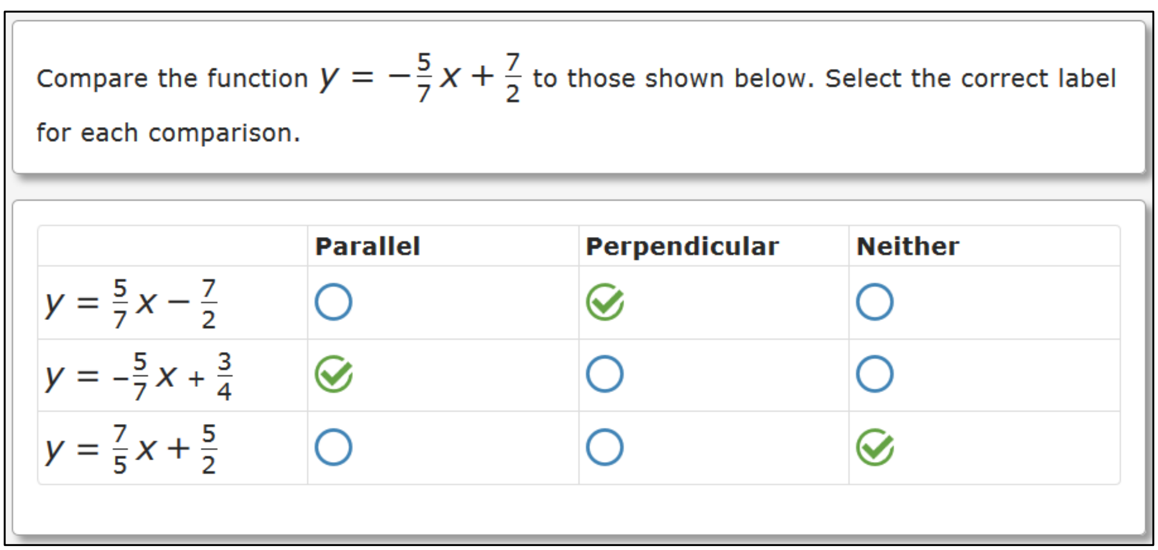
For the middle school students who completed the item in Figure 5.3.2.35 in braille, teachers read the selection aloud to speed up the testing process and the students followed along in the braille form. One student expressed concern that the passage and its title ("Red Sky at Night, Sailor's Delight. Red Sky in Morning, Sailor's Warning") were too long. In addition, the presence of "END OF TEXT" and "Public domain" at the end of the passage was distracting to students. Braille users thought this item could be improved by listing the answer choices under each statement or by changing the item to a multiple-choice format.

High school students completed two matrix items. Students reported that the screen reader did not read the equations correctly for the first item (Figure 5.3.2.36). Additionally, one student could check the buttons, but commented that he did not understand what he was selecting. Students also expressed concern that the item did not fit on one screen after using the CTRL + Zoom function to enlarge the screen. Students suggested making the radio buttons larger to improve the online form. An example of this item with revised radio buttons is shown in Figure 5.3.2.37.


Compare the function $y = -\frac{5}{7}x + \frac{7}{2}$ to those shown below. Select the correct label for each comparison.

	Parallel	Perpendicular	Neither
$y = \frac{5}{7}x - \frac{7}{2}$			
$y = -\frac{5}{7}x + \frac{3}{4}$			
$y = \frac{7}{5}x + \frac{5}{2}$			

Figure 5.3.2.36. High school math matrix item with original radio buttons.


Compare the function $y = -\frac{5}{7}x + \frac{7}{2}$ to those shown below. Select the correct label for each comparison.

	Parallel	Perpendicular	Neither
$y = \frac{5}{7}x - \frac{7}{2}$			
$y = -\frac{5}{7}x + \frac{3}{4}$			
$y = \frac{7}{5}x + \frac{5}{2}$			

Figure 5.3.2.37. Revised high school math matrix item with enlarged radio buttons.

Students who completed the item online also noted a technical defect, just as in the item shown in Figure 5.3.2.16. The item allowed test takers to type characters into the answer choices. In other words, test takers were able to alter or even delete the answer choices in the left column. This defect was reported to test developers.

High school students who completed the item in Figure 5.3.2.36 in braille in a matching format had difficulty understanding what to do in the item. Students did not understand the need to match labels in the first column with equations in the second column. Instead, they thought that the labels in each row were intended to go with the equations in the same row. These students suggested that the directions should be clarified. Additionally, they recommended numbering the rows in each column and adding column headings. The student who completed the item in print did not report any difficulties accessing the item, understanding the task, or responding to the item.

The second matrix item that high school students completed is shown in Figure 5.3.2.38. Students using the online version of the test reported that they had to zoom frequently in order to see the test (which then required a lot of scrolling), and that the vocabulary was difficult. The students suggested moving the answer choices closer together to avoid unnecessary scrolling or making the entire box a hot spot for responding instead of just the radio button. Several other online users reported no problems.

Read the Amendments V and VI to the United States Constitution below.

Amendment V
No person shall be held to answer for a capital, or otherwise infamous crime, unless on a presentment or indictment of a Grand Jury, except in cases arising in the land or naval forces, or in the Militia, when in actual service in time of War or public danger; nor shall any person be subject for the same offence to be twice put in jeopardy of life or limb; nor shall be compelled in any criminal case to be a witness against himself, nor be deprived of life, liberty, or property, without due process of law; nor shall private property be taken for public use, without just compensation.

Amendment VI
In all criminal prosecutions, the accused shall enjoy the right to a speedy and public trial, by an impartial jury of the State and district wherein the crime shall have been committed, which district shall have been previously ascertained by law, and to be informed of the nature and cause of the accusation; to be confronted with the witnesses against him; to have compulsory process for obtaining witnesses in his favor, and to have the Assistance of Counsel for his defence.

Select the appropriate category for each phrase.

	Amendment V	**Amendment VI**	**Both**
protects landowners	○	○	○
protects the accused	○	○	○
gives the right to an attorney	○	○	○
gives the right to a jury	○	○	○

Figure 5.3.2.38. High school ELA matrix item with enlarged radio buttons.

High school students who used the braille form of the item were not able to understand the table arrangement. Students could read the table but did not know how to choose the correct responses on the braille test. Respondents using the braille form suggested correcting the functionality of the test by placing radio buttons under each phrase to select Amendment V, Amendment VI, or Both, and creating a table like the one in the online version with the phrases listed down the left side. These suggestions led to a revised matrix layout in braille in which each row is numbered and each answer option is lettered.

Additionally, the letters of each answer option are repeated in the row and column intersections to facilitate entry of the correct response for each item with a braillewriter.

Drop-Down. Drop-down items require students to select their answer choices from one or more drop-down menus. The drop-down menus can serve several purposes in these items:

- *Ordering.* Students use drop-down menus to indicate the sequence number in a series of steps, operations, or sentences.
- *Matching.* Students use the drop-down menus to select the appropriate category for each element. For example, students read a sentence, and words from the sentence appear at the bottom of the item. Each word has a drop-down menu next to it. Using the drop-down menus, students select the correct part of speech (e.g., noun, verb, or adjective) for each word.
- *Graphic Labeling.* These items present a graphic, such as a drawing of an object or a number line. Parts of the graphic are lettered (e.g., A, B, C, etc.). A list of the letters appears below the graphic, and each letter has a drop-down menu next to it. Students select the label for that part of the graphic from each drop-down menu. Figure 5.3.2.39 displays an example drop-down item that features a graphic with labeled parts.
- *Multiple choice.* Drop-down menus appear embedded in text with a default word in each location. Each drop-down menu contains a list of options, and the student selects the appropriate option from each menu.

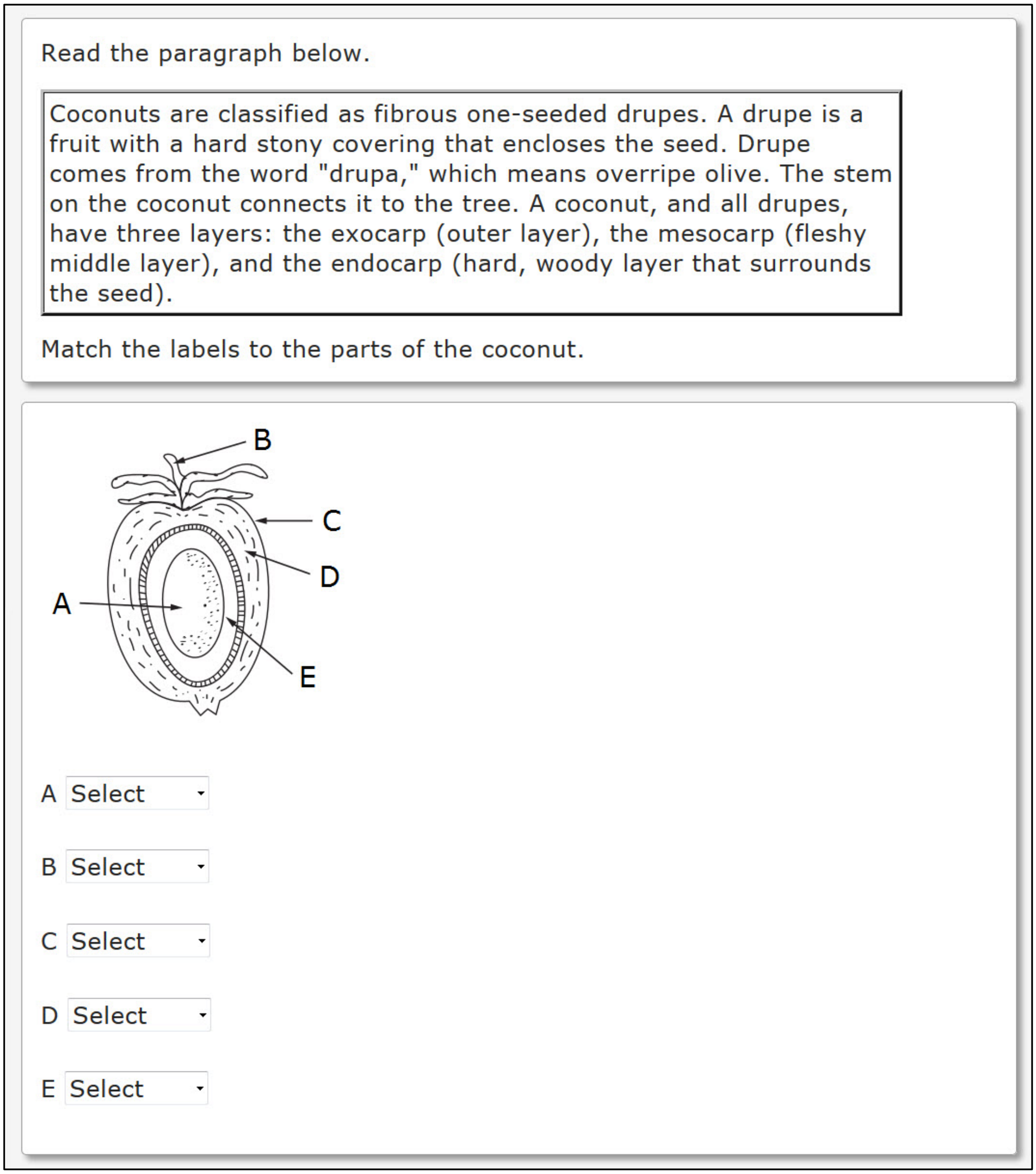


Figure 5.3.2.39. Drop-down item with graphic.

Teacher panels with drop-down items. For drop-down items with graphics, panelists recommended that the drop-down menus be placed within or beside the graphic to reduce scrolling. For some of the drop-down item types with graphics, panelists were concerned that the graphic was too small. Additionally, panelists thought that some drop-down items with graphics included too many labels and answer options. As an example, see Figure 5.3.2.40. This item requires students to select from seven answer options.

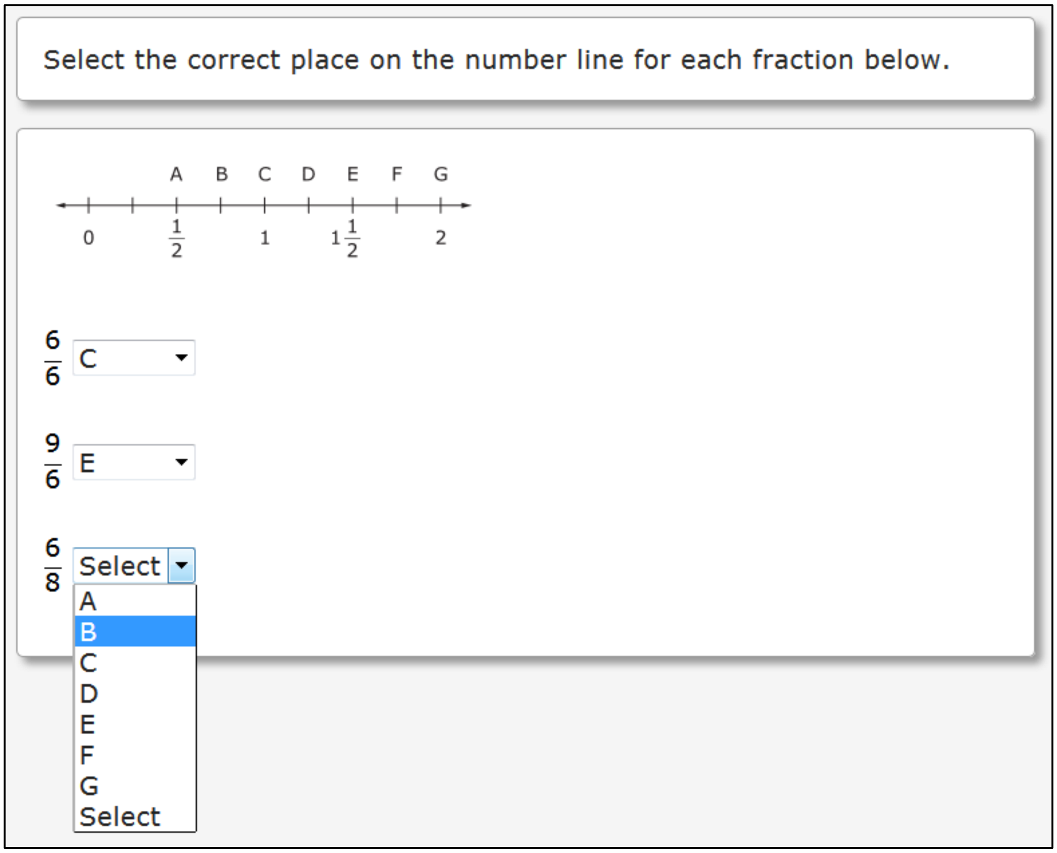


Figure 5.3.2.40. Drop-down item with number line and seven answer options.

Panelists generally believed that drop-down items would be accessible for students with visual and motor disabilities. The panelists often preferred drop-down items to other item formats, such as matching, ordering, and graphic labeling item types. There were a few exceptions, however. The panelists thought that drop-down menus unnecessarily increased cognitive load when the task involved ordering. One panelist team contended that the drop-down item type would be "difficult for kids who struggle with ordinals because [students] do not see the visual of the finished product." Panelist teams also thought that drop-down items with number lines (e.g., Figure 5.3.2.40) included unnecessary cognitive load.

Panelists generally agreed that the drop-down items were better than other item types for JAWS (Job Access with Speech screen reader) users. Some panelists thought that drop-down menus would work well for students using switches to scan the answer choices. However, other panelists thought that these items would require too many switch hits.

Panelists made several suggestions to make the drop-down items more accessible. Panelists suggested that drop-down menus should be larger with larger font size. The panelists also commented that the word "select" should not be used as a default word for the drop-down menus because it is not a valid answer choice and may cause confusion for students. In addition, panelists said that more space should be inserted between the drop-down menus. Alternatively, the answer choices could appear horizontally rather than vertically to eliminate the need for scrolling and to prevent the drop-down menus from overlapping other answer choices. For blind students, panelists suggested that answer choices should be provided in braille and that audio should be added to confirm the students' answer choices.

Cognitive labs with drop-down items. Elementary students completed three drop-down items during the cognitive labs. Students who completed the online version of the first item (see Figure 5.3.2.41) reported that scrolling posed a barrier to accessing the content of the task. One student started to select responses at the bottom of the item first because he did not understand how to complete the task.

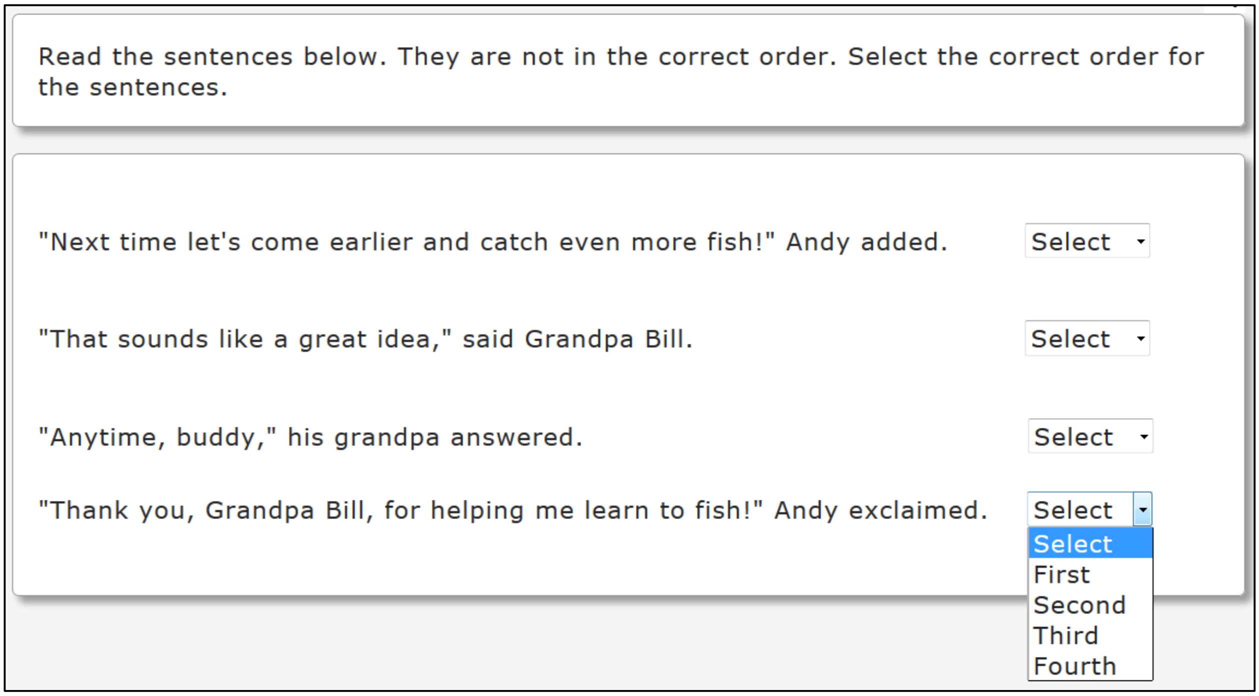


Figure 5.3.2.41. Elementary ELA drop-down menu item.

Elementary students who completed the braille version of the item in Figure 5.3.2.41 saw the item formatted as a matching item with the sequence numbers in the left-hand column (1, 2, 3, 4) and the sentences in the right-hand column. They reported that placing answer choices in columns should be avoided in braille if possible. The sentences were lettered in braille but were not lettered on the computer; students said the braille format and the online format should match. However, the drop-down format could not be presented in braille, so a matching version was used instead. Additionally, students suggested that the answer choices A, B, C, and D should appear on the left, with the item asking which sentence is first, second, third, and fourth, as opposed to having the answer choices on the right with "First," "Second," "Third," and "Fourth" on the left.

The second drop-down item that elementary students examined is shown in Figure 5.3.2.42. Students who completed the item online expressed difficulty scrolling across the screen to read the question. However, the drop-down format of the question seemed to work well for most students with visual impairments because the vertical presentation of options allowed students to see all of the options on the screen. One student, however, was confused about how to respond using the drop-down menu format. To prevent this barrier, the item instructions could better explain how to use the drop-down function.

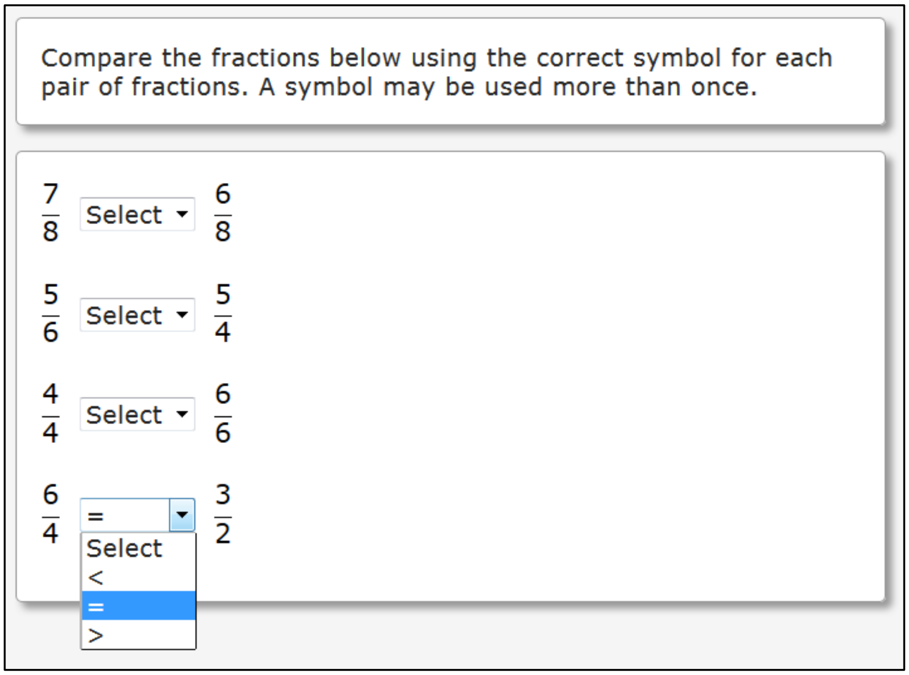


Figure 5.3.2.42. Elementary math drop-down menu item.

Two elementary school students completed the braille version of the item in Figure 5.3.2.42, which was formatted as a multiple choice item. One student did not report any barriers or suggestions regarding this item. The other student stated that the content was "mushed" together, which made the item confusing. This student suggested providing a legend for answer choices so that students could answer with letters (e.g., A, B, or C).

Elementary students completed a number line item in a drop-down format online and a matching format in braille and print (see 5.3.2.43). One student who completed the online version had difficulty navigating the item's layout. This student did not understand what the hash marks referred to, and he was uncertain how to provide his response. Another student did not understand how to respond using the drop-down menus. Suggestions for improving this item included placing the answer choices closer together and providing clearer instructions.

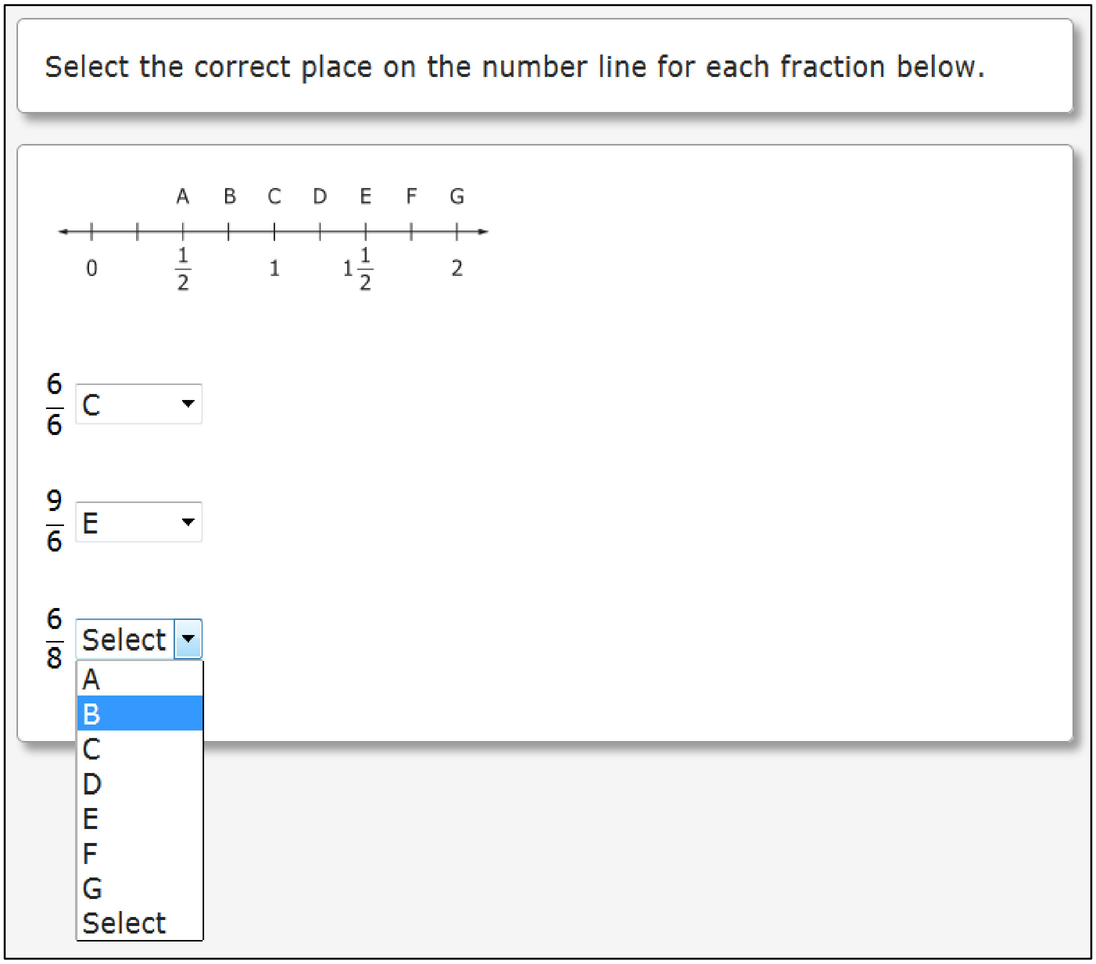


Figure 5.3.2.43. Elementary number line item.

The students who completed the braille version of the item in Figure 5.3.2.43 stated that the letters were missing from the number line. As a result, these students were unable to interact with the graph.

Middle school students completed two drop-down menu items during the cognitive labs. The first item provided a graphic with labeled parts. The students viewed a series of prompts and selected corresponding answer options from drop-down menus (see Figure 5.3.2.44).

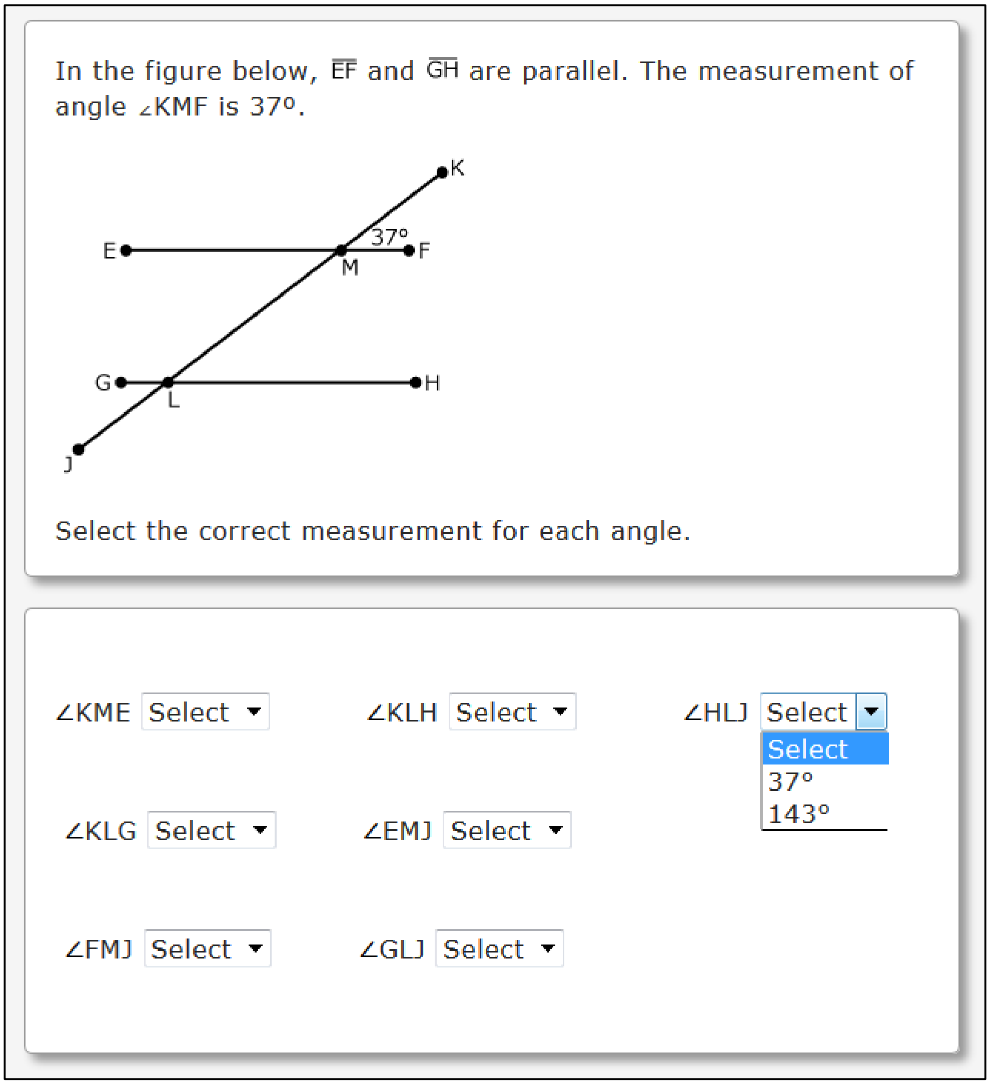


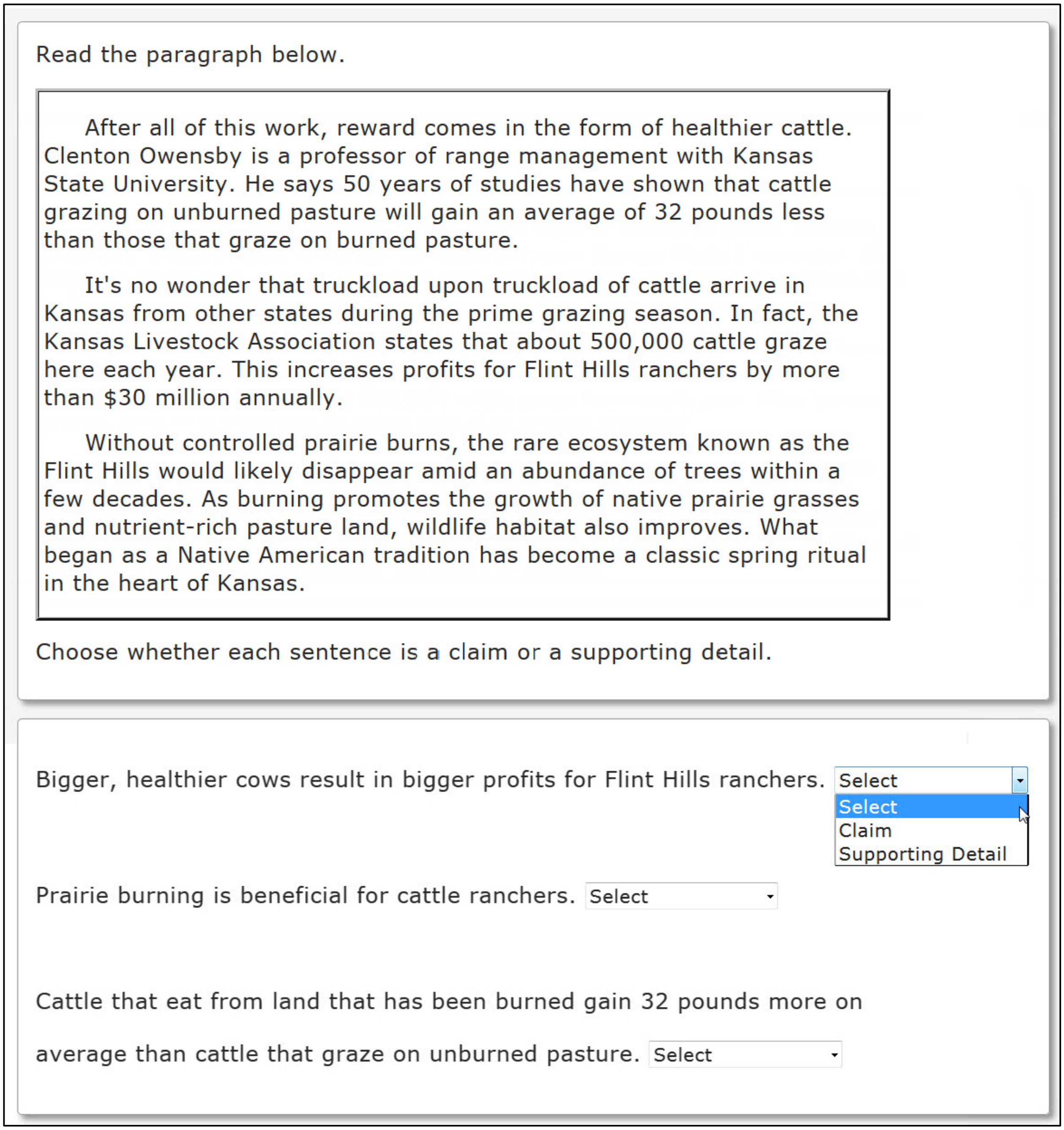
Figure 5.3.2.44. Middle school math drop-down menu item.

One middle school student who completed the online version of this item reported having to use the Return key to answer the question, as opposed to clicking with the mouse, and stated this process should be explained in the item directions. This was a defect that was reported to the developers. Another student who completed the online version of the item required magnification to access the content of the task. This student also had difficulty understanding how to complete the task. The student suggested reducing the number of angles presented in the item (to reduce scrolling) and clarifying the instructions. A fourth student, who used both the online and print versions, had difficulty scrolling with the track pad and preferred a mouse. This student referred to the print copy of the item when she had difficulty scrolling.

One student who used the print form with the online form liked the setup of this question. The teacher for this student added a note that the student had to magnify her screen three times which meant the student could not see the entire question and all answer choices at the same time. The student who completed this item in multiple choice format on the braille form did not report any issues with this item and stated no improvement was necessary.

Another student utilizing only the print form of the item in Figure 5.3.2.44 reported the image was not large enough, and he could not see the letters or angle measurements in the graphic. The student had trouble figuring out what to do, but the teacher added a note saying this may have had to do with not being able to see the graphic. The student suggested making the graphic larger.

The second drop-down menu item that middle school students completed is shown in Figure 5.3.2.45. For students who completed the item online, the item appeared to be "stretched." The teacher commented that it might help to narrow the screen and make the text bigger. Students had difficulty accessing the content because of the difficulty of the passage; additionally, the drop-down menu only worked with the Return key and not the mouse. This defect was reported to the developers. Test takers suggested three ways to improve the task for the online format: use a print form in addition to the online format, access the tasks using a mouse as well as a keyboard, and provide instructions that clearly tell the student how to navigate the drop-down menu function.



Read the paragraph below.

After all of this work, reward comes in the form of healthier cattle. Clenton Owensby is a professor of range management with Kansas State University. He says 50 years of studies have shown that cattle grazing on unburned pasture will gain an average of 32 pounds less than those that graze on burned pasture.

It's no wonder that truckload upon truckload of cattle arrive in Kansas from other states during the prime grazing season. In fact, the Kansas Livestock Association states that about 500,000 cattle graze here each year. This increases profits for Flint Hills ranchers by more than $30 million annually.

Without controlled prairie burns, the rare ecosystem known as the Flint Hills would likely disappear amid an abundance of trees within a few decades. As burning promotes the growth of native prairie grasses and nutrient-rich pasture land, wildlife habitat also improves. What began as a Native American tradition has become a classic spring ritual in the heart of Kansas.

Choose whether each sentence is a claim or a supporting detail.

Bigger, healthier cows result in bigger profits for Flint Hills ranchers. Select

Select
Claim
Supporting Detail

Prairie burning is beneficial for cattle ranchers. Select

Cattle that eat from land that has been burned gain 32 pounds more on average than cattle that graze on unburned pasture. Select

Figure 5.3.2.45. Middle school ELA drop-down menu item.

Middle school students who completed the item in Figure 5.3.2.45 in braille or print in matrix format reported little difficulty with the item. Beyond improving the instructions to explain how to respond to the matrix, students did not provide other suggestions for improving the item in braille.

High school students completed two drop-down items. The first is shown in Figure 5.3.2.46. Some online users had difficulty seeing the graphic for the first item, so they used the print version with the online version. Students needed to magnify the content so much that the item required a great deal of scrolling and moving laterally to see the entire item. Additionally, some of the vocabulary was difficult for the online test takers. Students commented that the online format could be improved by placing the answer choices to the right of the graphic or providing drop-down menus for the letters in the graphic instead of below it. Students also stated that a bigger screen was needed because they could not see the full image when they enlarged the graphic. One student used JAWS, but he was not extremely proficient at using a screen reader and had some difficulty with the screen reader commands.

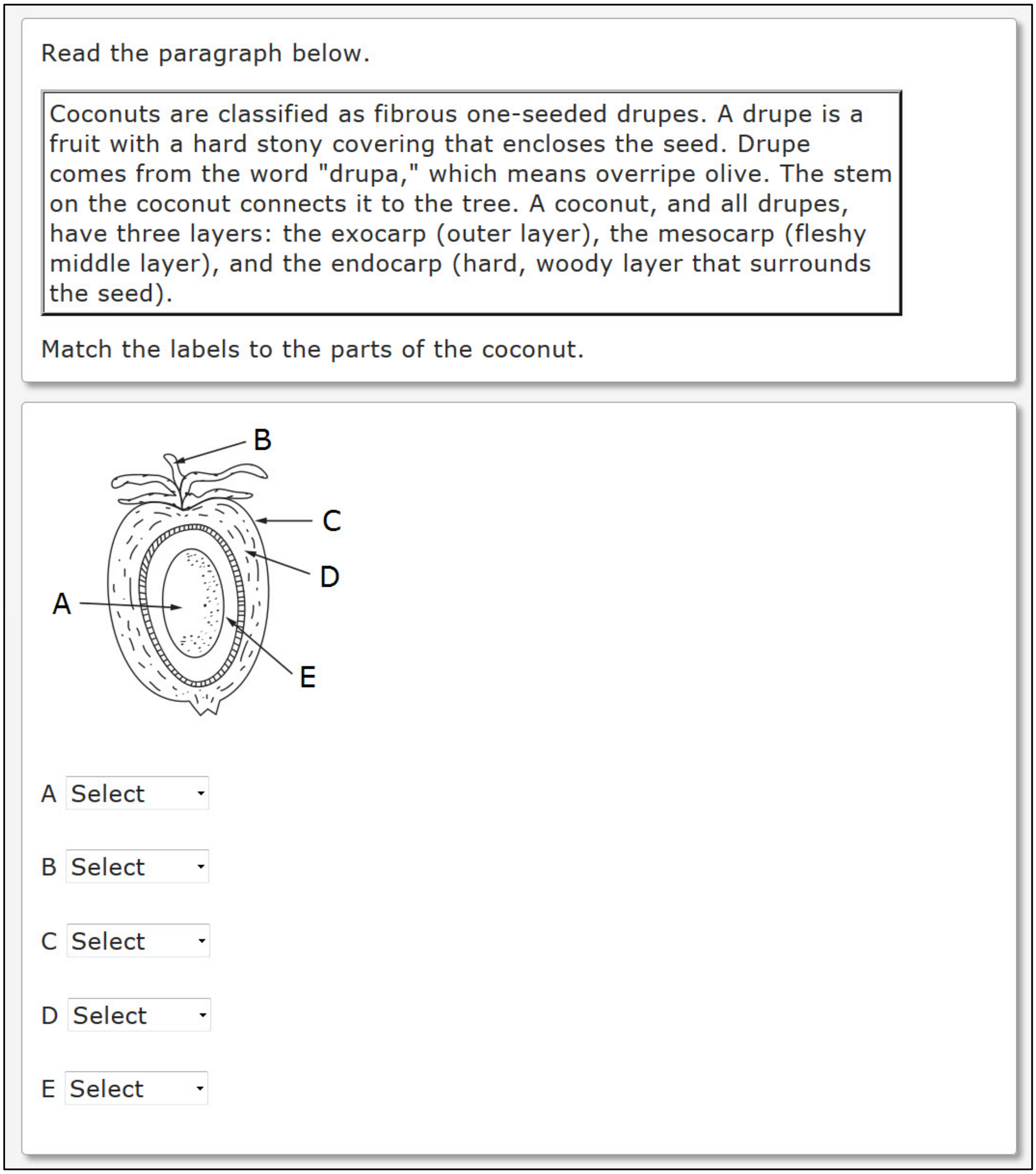


Figure 5.3.2.46. High school drop-down item in ELA.

High school students who completed the item shown in Figure 5.3.2.46 in braille in a matching format commented that answer choices A and E appeared to point to similar areas on the diagram. Barriers to responding to the task included unclear directions and difficulty in locating to what part of the diagram answer choice C pointed. Braille test takers suggested that using a new format, enlarging the tactile diagram, including a written description about the graphic with the diagram, and numbering the items would improve the task. High school students who took the item in print suggested making the diagram bigger. Students thought that the arrows for answer choices C and D pointed to the same thing.

High school students who participated in cognitive labs completed a number line item either online or in braille. This item contained the same content as the middle school item in Figure 5.3.2.34 above but formatted as a drop-down-menu item type. This item required students to simplify several expressions and then find the correct location for each expression on a number line (see Figure 5.3.2.47).

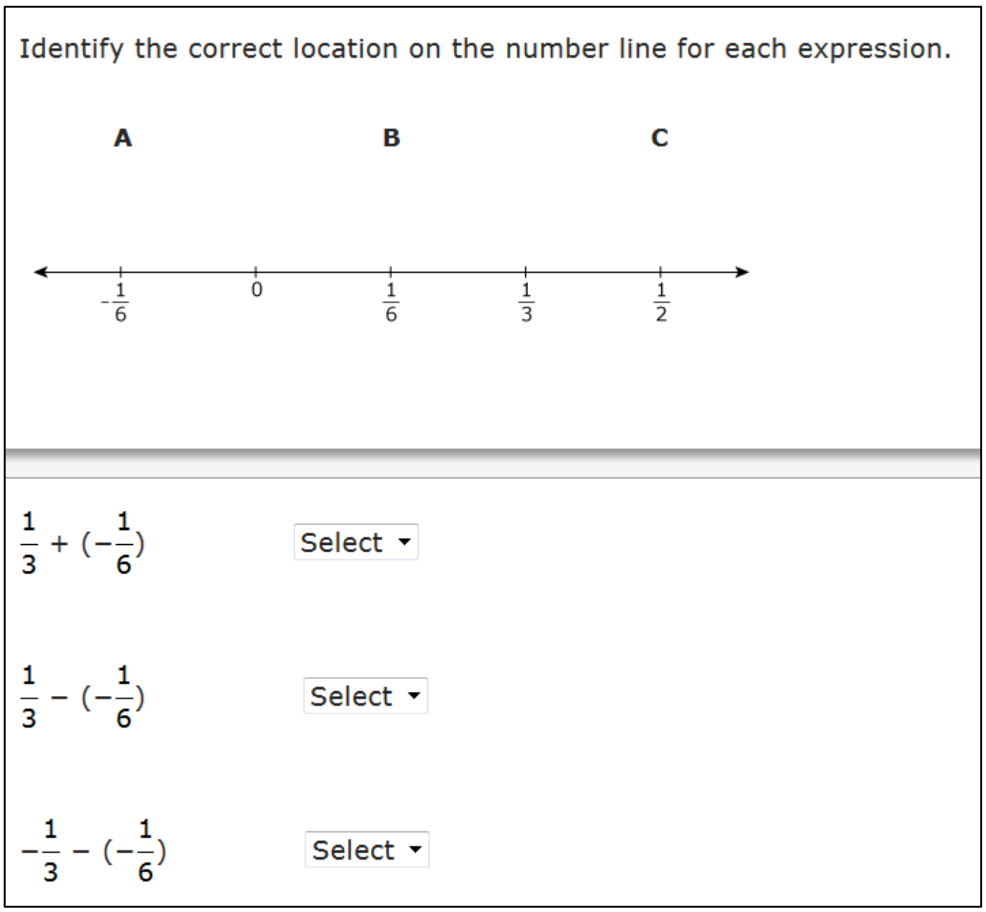


Figure 5.3.2.47. *High school number line item.*

High school students who completed the item online suggested reducing the amount of white space between the letter labels and the number line and making the number line larger. High school students who completed the item in Figure 5.3.2.47 in braille noted a few barriers to accessing the item. The screen reader could not read the number line correctly, so it was inaccessible online, but the braille version was reported to be perfect. One student commented that having the letters lined up with the fractions was confusing. A few students were also confused about how to respond to the task because they did not understand what the answer options were. One braille reader suggested numbering the parts of the item in braille to simplify responding with a braillewriter. Finally, the lack of congruence between the braille and online versions was bothersome to one student.

The student who completed the item in Figure 5.3.2.47 in print did not report any barriers to accessing the content, understanding how to complete the task, or responding to the task. However, the student did state that the number line should be larger.

5.3.3 Constructed Response

Text Entry. Text-entry items in math and ELA require that a student use the keyboard to type an answer into a provided text box. The complexity of this item type can range from recall to extending student thinking. Figure 5.3.3.1 is an example of a text-entry item. Text entry items, whether short or long responses are required, are accessible to students with vision and motor disabilities because they are equivalent to the tasks that students complete as part of their daily instruction. Therefore, text entry items were not included in sample tests presented to teachers and students. Teachers and students were asked to review another type of constructed response item, requiring students to plot points or graph lines, because of the apparent barriers to accessibility online.

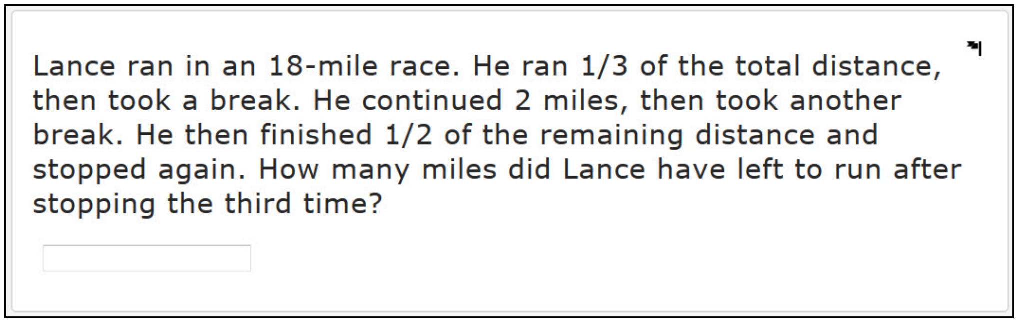


Figure 5.3.3.1. Text-entry item.

Graphing. Graphing items require students to click within the area of a graph to plot a single point. For line graph items, after a student has plotted two points, the computer generates a line that connects the points. From the student's perspective, graphing items are constructed response items even though they can be scored electronically. Figure 5.3.3.2 is an example of a line graph item shown after the student has plotted three points.

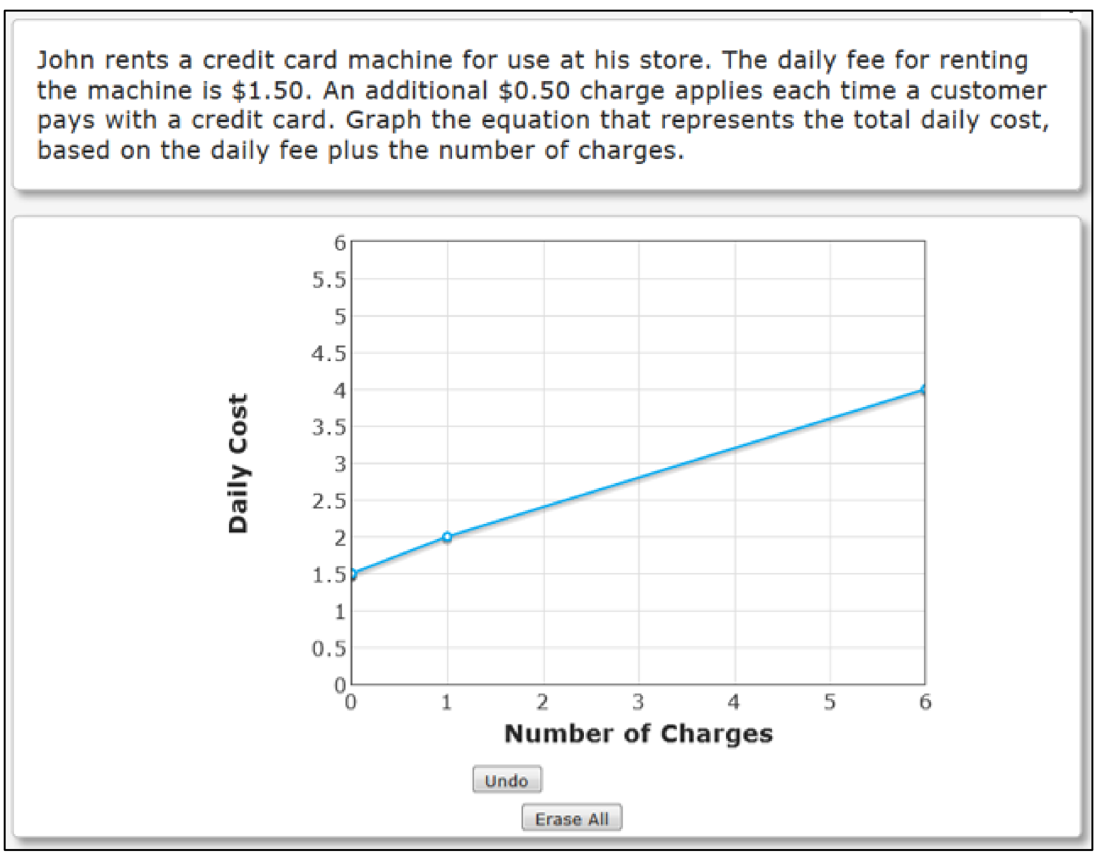


Figure 5.3.3.2. Graphing item viewed after student has plotted three points on the graph.

Teacher panels with graphing items. The panelists had specific accessibility concerns with line graph items (see Figure 5.3.3.2). The panelists thought the graph lines would be too light for low-vision readers. The panelists also though this item could present difficulties for students with motor impairments; plotting points on the graph requires very fine motor skills because hot spots on the graph can be sensitive. When a student clicks inside a box, rather than at the intersection of two graph lines, a point is plotted on the graph. As such, it is easy to accidentally plot a point on the graph.

Cognitive labs with graphing items. Cognitive labs for middle school and high school students included line graph items. The middle school line graph item (the online version) required students to create a line that represented a proportional relationship (see Figure 5.3.3.3). The students had problems understanding the terminology, understanding the question because they did not know the math content, and knowing how to create a point and a line. Two students suggested providing improved directions about how to interact with the item. One student had difficulty scrolling to see the full graph. This student was using a computer track pad rather than a mouse because a mouse was not available.

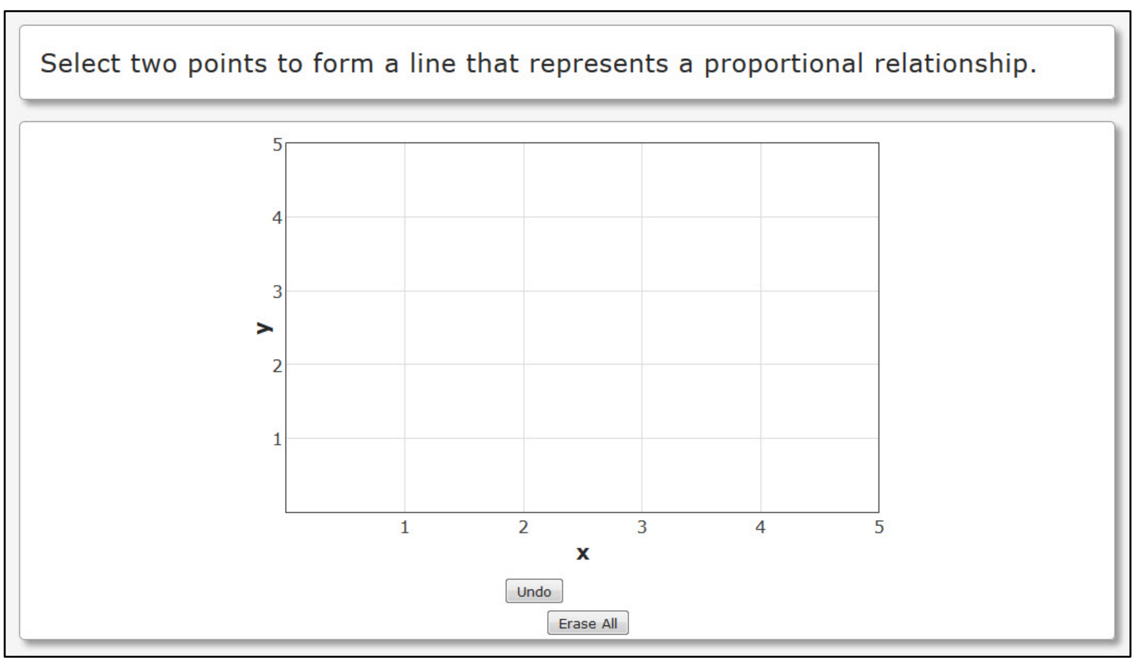


Figure 5.3.3.3. Middle school math graphing item.

Students who completed the item in Figure 5.3.3.3 in braille and print approached it as a constructed response task. The braille user did not report any issues with the item but suggested labeling zero to provide a reference point. This was an oversight in the creation of the sample item. The student who completed this item in print could not provide a response because he/she did not understand the instructions. This student also suggested using larger print on the graph.

High school students completed the line graph item shown in Figure 5.3.3.4. Students who completed the item online identified accessibility barriers. Several students were unable to respond to the task by graphing the lines. Some students did not understand how to complete the task (i.e., students were unsure how to create the graph lines).

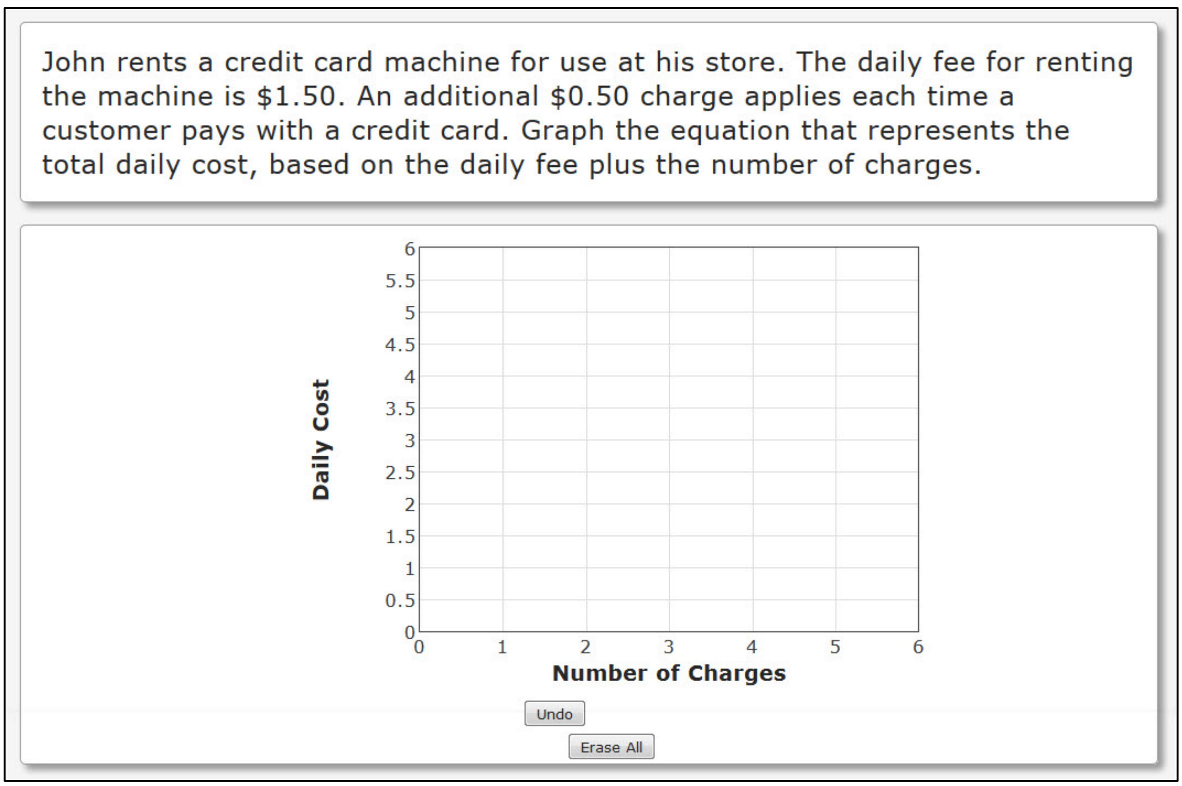


Figure 5.3.3.4. High school math graphing item.

For the braille version of this item, the high school students had many suggestions. Students stated that the instructions should be clarified, that more explanation of the graph should be added to the braille form, and that the graph should include a title. Braille users were not sure how to respond to the item using the tactile graphic. Several students suggested manipulatives for this item, such as tactile shapes or foam dots that would adhere to the graph. Another suggestion was to provide a paper and a stylus; the student could poke holes in the paper to indicate the graph plot points.

The high school student who completed the item in print suggested that the online interactive format was preferable to the paper-and-pencil version. This student also expressed confusion with the task's instructions.

5.4 Accessibility Across Item Types

In the accessibility reviews, the panelists made many accessibility suggestions that apply across TE item types. Some of these suggestions applied to all students, and other suggestions applied to students with specific types of disabilities.

Item instructions. One of the most pervasive criticisms of the TE items related to instructions. Unclear instructions can have an adverse effect on all students. This effect may be amplified for students who also struggle with visual or motor impairments. Panelists expressed confusion about the functionality of several item types due to vague instructions. For example, the line graph items did not explicitly tell students how to interact with the graph. As another example, some ELA items included text with words or phrases in bold or italics, but the instructions did not indicate the purpose of the bolded or italicized words. Additionally, drag-and-drop items should explicitly tell the student to click, hold, and drag an element into place.

One item type that was particularly difficult was click-to-select categorization. Panelists overwhelmingly agreed that the directions for this item type were unclear. Since click-to-select categorization items look like drag-and-drop categorization items, a test taker's instinct might be to click, hold, and attempt to drag an answer choice into place. If a student attempts a drag-and-drop motion with a click-to-select categorization item, the answer choice snaps back into its original position. The instructions do not tell students how to manipulate elements in the item (i.e., click left and then click right, rather than drag and drop).

This feedback was instrumental in driving student preparation for TE items in KITE. Before testing with KITE, students have access to practice tests that allow them to become familiar with all of the item types and how to manipulate them. These practice tests are also available with online accommodations activated, including audio and switch functionality. Furthermore, a help "?" was added to the navigation bar for students to use during testing. The help menu provides brief tips in case the student is confused about an item type. In braille tests, sample items of each type and instructions for responding are in the front of each booklet for teachers and students to review.

Graphics and fonts. Across item types, panelists suggested that items be made more accessible with regard to font size and graphics. Additionally, panelists suggested increasing the accessibility of item graphics by simplifying graphics, making graphics larger, making graphics less "cluttered," and increasing color contrast. Panelists also suggested avoiding the use of colors and light greys to distinguish between parts of an item in order to increase accessibility for students with colorblindness and low vision. If an item requires color, the colored element should be distinguished in a variety of other ways (e.g., bolded, outlined in a box, etc.).

Scrolling. Another major barrier to accessibility occurs when students are required to scroll in order to view an entire item. Scrolling can present barriers for students with visual impairments, motor impairments, and students with special cognitive needs. Scrolling between a graphic or passage and answer choices adds to the short-term-memory and motor-skill requirements of the item. One panelist even added, "Scrolling through the passage [and answer choices] on a computer screen may make a student sick who has vestibular issues."

5.5 Summary

Expert panelists provided many suggestions for increasing the accessibility of TE items. The TE items reviewed served several assessment purposes within math and ELA. (see Table 5.5.1).

Table 5.5.1

Technology-Enhanced Item Types and Tasks

	Task Demand				
Item type	Ordering	Matching/ labeling	Categorizing	Plotting lines	Selecting information
Select text					X
Ordering	X				X
Labeling		X			
Categorization			X		
Venn diagram			X		
Number line		X			
Line graph				X	
Matching		X			
Drop-down	X	X	X		X
Radio buttons		X	X		

Items required students to place elements in order, match elements, categorize, label, plot lines on a graph, and identify main ideas or superfluous information in text. An advantage of TE items is that they can provide students with a more authentic assessment experience. TE items are also quite flexible; most item types can be used for multiple tasks.

Preferred item type for each task. It is clear from the panelists' reviews that certain item types are more universally designed and thus more accessible for a wider range of students. Drag-and-drop items, while versatile and intuitive for many students, can require fine motor skills. This presents an accessibility barrier to students with limited motor abilities. When the assessment purpose is ordering, matching/labeling, or categorization, panelists preferred click-to-select-categorization and drop-down-menu item types to the drag-and-drop format. With some revisions (e.g., larger buttons, larger hot-spot area), radio button items might also be accessible to students with motor impairments.

For matching tasks, the review revealed drop-down-menu and click-to-select-categorization item types to be the most accessible. The matching item type was described as cluttered and confusing for students with visual impairments.

Item accessibility can be improved for all students on all item types by providing access to assistive technologies such as switches and screen readers. Panelists also recommended professional training on specific pieces of technology (e.g., Window eyes, JAWS, embosser programs, braille transcribing programs).

Additionally, as mentioned in Section 4.0, item instructions should clearly convey what students are expected to do, graphics and font should be simple and sufficiently large, colors should be avoided, and scrolling should be minimized in the items. These recommendations are in agreement with previous research on item accessibility (Kopriva, 2000).

Students with vision disabilities. Item accessibility can be enhanced for visually impaired students by improving alignment between computer-based and braille forms of items. There were some discrepancies between computer-based and braille forms. Braille forms should also be designed so that students have a clear understanding of the answer choices. This was problematic on some select text items (e.g., bolding of answer choices did not translate into braille).

Other steps to improve item accessibility for visually impaired students include:

- Provide onscreen magnification
- Improve screen reader functionality
    - Make sure the screen reader correctly reads all fractions, numbers on a number line, etc.
    - Make sure the screen reader reads text at a reasonable rate

  - Once a student has completed a question, provide auditory support so the student knows when to move to the next question.
  - Make sure students can review their answers by having answers read to them.
- Make the cursor larger.
- If the student has keyboarding knowledge, provide a text box for the student to provide his/her responses.
- When a student selects an answer choice, a "click" sound should confirm that the student has provided an answer.
- When magnification is used, make sure the image on the screen is still clear and the text is legible.
- On braille forms, provide a number for each part of an item that the student should answer. This way, students will be less likely to accidentally skip part of the item.
- If graphics must be used in the item:
  - Provide tactile versions of the graphics (braille) that are sufficiently large.
  - Provide a description for each part of a graphic (e.g., coconut item graphic)

Students with motor impairments. Students with motor disabilities will likely have difficulty with items that require the drag-and-drop function with a mouse, a conventional mouse pad, or a touch screen. In many cases, accommodations can be made to these items by changing the item format in a way that will not affect the complexity level. Sighted students with motor disabilities may benefit from a scribe to enter responses, which eliminates differences in presentation.

To alleviate motor demands, the panelists made the following suggestions:

- Present students with items types that do not require the drag-and-drop function, such as click-to-select categorization, drop-down menu, or radio buttons.

- Make items accessible via keyboard (e.g., arrow keys, Tab, Enter, etc.).
- Increase the size of the cursor, hot spots, etc.
- Items with graphics/diagrams (if drag and drop is necessary):
    - Make graphics sufficiently large for students to manipulate elements in the graphic.
    - Reduce the amount of space between the graphic and the answer choices.

5.6 Conclusions

Item writers need to be thoughtful about students with disabilities when preparing assessment tasks. The goal in any item-writing endeavor is to create universally designed items that are accessible to the widest range of students. When items pose accessibility concerns, the necessary accommodations come with tradeoffs.

For example, onscreen magnification may help visually impaired students see elements in an item. However, the magnification may make the item more complex because the student has to scroll through several screens and retain parts of the item in working memory in order to provide a response. Similarly, audio presentation can present a memory burden that disadvantages visually impaired students.

Additionally, providing tactile graphics of pictures to blind students may not be a simple solution. Students need multiple occasions to interact with similar graphics in order to become familiar with them and to be prepared to use them on an assessment. In addition, tactile graphics might be simplified from their visual versions, and this can lead to the elimination of useful contextual material.

As evidenced in these item reviews, item cognitive complexity is not fixed. For instance, the DOK level of an item may be different for a student without a documented disability, a student with motor disabilities, and a student with visual impairment. This is true for math items, for example, on which a blind student is at a disadvantage due to reduced experience with visual relationships. The panel noted that a blind person would typically learn synthetically in a sequence from parts to wholes. However, on one of the sample items, the question was approached from the analytic perspective of the whole to a part. This could increase the cognitive complexity for this population of users.

Furthermore, translating the computer-based, interactive TE items into hard-copy braille or print versions may not provide an experience with identical cognitive demands. Therefore, further research must focus on the preparation of assessment tasks that are accessible and contain features similar to these TE items. Reducing assessments to traditional multiple-choice questions for students with vision disabilities would deny these students access to new technologies and the engaging tasks that are now being developed.

Despite these obstacles, panelists generally agreed that several TE item types were quite accessible for students with a wide range of disabilities. Some item types appeared to be universally designed, including click-to-select-categorization, drop-down-menu, and radio button items. The panelists had suggestions to further improve the accessibility of these item types (e.g., make the radio buttons larger). Many of their suggestions were implemented without altering the item content or difficulty.

6.0 Characteristics of Students with Vision and Motor Disabilities

This section presents the results of a detailed survey completed by teachers of students with vision and motor disabilities who participated in the 2014 field tests and item tryouts. The characteristics of this population have not previously been described at this level of specificity. Appreciating the vision and motor needs of these students will enable the reader to better understand the barriers presented by online test items and tasks, particularly those that require hand-eye coordination for response.

Thirty-five teachers provided detailed information about 66 students with vision or motor disabilities who participated in the Kansas field tests or in the item tryouts in other states. Thirty-seven (56%) of the students were male and five (7.6%) were of Hispanic origin. Forty-eight (73%) students were white, seventeen (25.6%) were black, two (3%) were Native American, one (1.5%) was Asian, and one student (1.5%) was marked as Other. These 66 students represented 11 states as shown in table 6.0.1.

Table 6.0.1

Participating Students by State

State	Number	Percent
Colorado	1	1.5
Illinois	1	1.5
Kansas	10	15.2
Maryland	3	4.5
Michigan	13	19.7
Mississippi	9	13.6
Missouri	7	10.6
Nebraska	1	1.5
Ohio	7	10.6
Oklahoma	6	9.1
West Virginia	8	12.1
Total	66	99.9

Students represented grades 3 through 12, as shown in Table 6.0.2. Students came from predominantly large towns and urban environments, as shown in Table 6.0.3.

Table 6.0.2

Participating Students by Grade

Grade	Number	Percent
3	11	16.7
4	6	9.1
5	4	6.1
6	9	13.6
7	7	10.6
8	10	15.2
9	3	4.5
10	5	7.6
11	1	1.5
12	4	6.1
No grade reported	6	9.1
Total	66	100

Table 6.0.3

Participating Students by Population Density

Population Density	Number	Percent
Rural (population of less than 2,500 and not within a larger metropolitan area)	12	18.2
Small town (population of 2,500 to 25,000 and not within a larger metropolitan area)	13	19.7
Large town (population of 25,000 to 250,000 and not within a larger metropolitan area)	20	30.3
Urban (within a metropolitan area with a population of over 250,000)	21	31.8
Total	66	100

Teachers selected one or more primary and secondary special education categories for their students, as shown in Table 6.0.4. The majority of students (89.4%) experienced blindness or low vision with additional primary and secondary disabilities distributed among 11 categories. Students received instruction in several types of classroom settings, as shown in Table 6.0.5. The majority of students attended regular classes, and the second largest percentage attended state residential schools for the blind.

Table 6.0.4

Primary and Secondary Special Education Categories

Category	Primary	Percent	Additional	Percent
Autism			4	6.1
Blind or low vision (vision impairment)	59	89.4		
Deaf or hard of hearing (hearing impairment)	1	1.5	1	1.5
Deaf-blindness	2	3.0		
Emotional disturbance			1	1.5
Intellectual disability (mental retardation)			1	1.5
Multiple disabilities	1	1.5	1	1.5
Orthopedic impairment			3	4.5
Other health impairment	3	4.5	4	6.0
Specific learning disability	2	3.0	2	3.0
Speech/language impairment			3	4.5
Traumatic brain injury			1	1.5

Table 6.0.5

Classroom Placement of Participating Students

Category	Number	Percent
Regular class	39	59.1
Resource room	6	9.1
Separate class	2	3.0
Separate school	5	7.6
Residential facility	14	21.2
Total	66	100

Sixty-three (95.5%) students were reported to have vision loss. Of these students with blindness or low vision, 38 (57.6%) wore glasses or contact lenses, 34 (51.5%) required enlarged print, 14 (21.2%) required tactile graphics and symbols, and 25 (37.9%) required Braille.

Teachers were asked to select the technology used by these students. Multiple selections were possible for each student. Table 6.0.6 displays the technology used by blind and low-vision students.

Table 6.0.6

Technology Used by Blind and Low-Vision Students

Technology	Number	Percent
Magnifier	25	37.9
Computer screen magnifier	11	16.7
Screen magnification software	26	39.4
CCTV	21	31.8
Screen reading software	21	31.8
Scanner with talking word processor	1	1.5
Manual braillewriter	23	34.8
Electronic braillewriter	11	16.7
Refreshable braille display	18	27.3
Light box	2	3.0

Note. Closed-circuit TV (CCTV) is an electronic magnification system.

Three students were reported to have hearing loss in addition to blindness or low vision. Of the three, one student (1.5%) used a unilateral hearing aid and two students (3.0%) used bilateral hearing aids. One of the students (1.5%) used a personal or classroom amplification device. No students were reported to use sign language in place of or in addition to speech to meet expressive communication needs.

Six students (9.1%) were reported as having a physical or orthopedic disability or motor skill problems. Of these six, two also had vision disabilities and a third student was categorized as deaf-blind. One of these six students (1.5%) walked with a cane. One student (1.5%) used a wheelchair without assistance, while another three students (4.5%) used a wheelchair with assistance. Four students (6.1%) required specialized positioning equipment, such as a standing frame.

In terms of arm and hand control, which may affect computer access, one student (1.5%) used only one hand to perform tasks and two students (3.0%) required assistance to perform tasks with their hands. These three students (4.5%) used a standard computer keyboard with their fingers and one of these students (1.5%) also used a touch screen and voice-recognition software. Three students (4.5%) could not use their hands to complete tasks. Of these three students, one student (1.5%) could not access a computer, while the other two students (3.0%) used switch systems with eye gaze, head, knee, foot, or leg access. Two of the six students with physical impairments (3.0%) had restricted range of head motion and four students (6.1%) required consistent head support throughout the day.

Finally, the teachers indicated the students' communication preferences. No students were marked as using sign language or augmentative or alternative communication systems. Though 14 students were not marked as using speech, in the absence of any other marked choice, it must be assumed that all students used speech as their primary mode of communication.

7.0 Field Tests

Large-scale field tests of the new Kansas achievement tests were conducted in spring 2014. These field tests served several purposes. First, the field tests utilized Kansas' proprietary technology platform, KITE. KITE includes tools for all aspects of assessment delivery: item entry; external and internal item review; media management and review; building and publication of test forms; delivery of formative, practice, and live assessments; management of testing accommodations; management of student rosters and individual special needs by teachers and test administrators; delivery of professional development; and reporting of results to stakeholders. KITE functionality was utilized by Kansas teachers, administrators, and students for the first time during spring 2014 field tests.

Second, the field tests allowed CETE to deliver a large number of new passages, items, and tasks for the development of future test forms. These new items and tasks included innovative, TE tasks along with constructed response and selected response items. The KITE platform was explicitly built to support the development of new kinds of assessment tasks, such as interactive test items with moving pieces and constructed response items such as graphing and partitioning, with multimedia delivery to students.

Third, and most important for the ATEA project, the timing of these field tests permitted the placement of identical item content in different formats onto matched test forms, referred to internally as form A and form B. At each grade and for each subject, forms A and B were precisely the same in length and content. Matched forms A and B were prepared for both ELA and math at grades 3-8 and high school. The length of the forms varied by grade and content, and the forms comprised 52-58 items in ELA and 60 items in math. ELA tests contained reading passages with item sets and stand-alone writing items. Math tests contained only stand-alone items.

The selected response items on forms A and B were identical. However, the TE items that were believed to be inaccessible to students with vision and motor disabilities were reformatted into accessible alternative formats for form B. Project staff based the identification of TE items that were likely to create barriers for students with disabilities on the qualitative feedback obtained during the first year of the ATEA project.

Form A contained all of the original test items that were written and selected to meet state-mandated test specifications. These items passed through the standard test-development process, including multiple internal reviews, editorial reviews, internal accessibility reviews, external teacher reviews, and external bias and sensitivity reviews. After items had completed that process, the inaccessible TE items on form A were replaced with accessible alternatives on form B. The alternative items maintained the same wording as the original items, but appeared in formats that provided for different item presentation and response modalities. Form B became the base form for all special forms and online accommodations. Form B was delivered with text-to-speech audio, in braille, in print, and in large print. Form B was also available to students in the general population as an unaccommodated online test. Form A, along with other forms used to field test additional items at each grade and subject, was delivered only online and without accommodations.

7.1 Procedures

Test forms were assembled and built in KITE for online delivery to Kansas students. When the testing window opened, online tests were assigned to students randomly, with some exceptions, as they logged into KITE for online testing. Form assignment was managed electronically. All test forms, including the accessible form B, were available in online, unaccommodated formats. In addition, students with accessibility requirements were enrolled in form B as a result of auto-enrollment rules that matched students with accessibility needs to test forms that supported those needs. Therefore, form B was the single form assigned to all students with accessibility needs. Form B was also assigned to students without accessibility needs on a random basis.

Special forms for offline testing (braille, print, and large print), all of which matched forms A and B in content, were delivered to students individually as requested by their IEP teams. Students who received offline tests also received KITE logins. Their teachers transcribed the students' responses into KITE after the students completed testing by marking responses onto print forms, recording responses with a braillewriter, or dictating responses to a teacher or other test administrator.

Teachers requested Spanish print tests in math and science, another type of special form, for some students. Teachers entered student responses to the Spanish print tests into KITE using the students’ logins to the corresponding online English-language tests. Spanish test items were not entered into KITE, hence all testing in Spanish utilized print test forms and required a transcriber.

One of the accommodations delivered via KITE is text-to-speech audio, which is available to individual students upon the request of their IEP teams. Students who required audio were administered the online test in form B and were permitted to play the audio for admissible portions of the tests. Only form B contained audio files for on-demand use. However, there is no way to know if students used the audio option, or how often they used it.

Smooth auto-enrollment of students to test forms on a random basis (other than students with accessibility needs) was interrupted by several minor technological problems that surfaced in the new KITE system. For example, test takers who logged in early in the testing window were not randomly assigned to form B in the same numbers as students assigned to other forms. That problem was corrected as soon as it was identified, but one outcome was unequal group sizes for forms A and B. In addition, a distributed denial-of-service attack from an external source interrupted testing for a significant portion of Kansas test takers a few weeks into the testing window. That external attack caused a closure of testing for about three days while CETE contracted for a solution to protect against the ongoing attack and prevent similar future attacks. As a result, the validity of testing for schools and districts who had elected to administer tests during that time was questioned. Schools and districts who had scheduled testing to be completed by a particular date were unable to meet that deadline and chose to return to instruction rather than return to testing following the delay. Some entire districts tested either early or late in the testing window and did not experience difficulties; therefore, the students affected by these problems were not random. The state of Kansas made the decision not to report test scores or results for any students because valid test scores were not available for some students. Fortunately, because it was a field test year, test scores did not have consequences for students, schools, or teachers.

7.2 Subjects

The subjects of the analyses for the spring 2014 field tests were students who had been assigned to online test forms A and B *without* accommodations. This comparison served as a preliminary step to investigating student performance with accommodations, which is the ultimate question of the ATEA project. The spring 2014 field tests were an excellent opportunity to evaluate the performance of large groups of students on alternate item formats as a prerequisite to concluding that those alternate formats are fair and equitable for students with accessibility needs.

The subjects of field test analyses are large samples of students in the general testing population, including most students with disabilities who do not require testing accommodations or special forms. Students with disabilities in these samples include students with learning disabilities, emotional or behavior problems, speech-language disorders, orthopedic disabilities, other health impairments, and intellectual disabilities, among other disabilities. Students typically not included in this group include students who are deaf or hard of hearing, students who are blind or have low vision, students with deaf-blindness, and students with significant cognitive disabilities who are eligible for an alternate assessment. Students who used any accommodation, including online audio, were not included in these analyses.

Because of the lack of fully random assignment of students to unaccommodated online test forms, the first priority was the investigation of the equivalence of the student groups assigned to forms A and B. Three methods were used to evaluate equivalence. First, student demographic characteristics—the percentage of students with disabilities (SWD) and the percentage of students eligible for English-for-Speakers-of-Other-Languages (ESOL) programs—were compared using chi-square analysis with the phi coefficient as a measure of effect size. Second, the percent correct for the 2014 field tests was computed for these students and compared using univariate analysis of variance with eta squared as a measure of effect size. Third, the scaled scores from 2013 Kansas tests were compared using univariate analysis of variance and eta squared. The 2013 tests were part of the previous test cycle and were unrelated to the field tests administered in 2014. Even the testing platform and accommodations were different in 2013. These results were available for the major part of each student group other than third grade. Results of these analyses are shown in detail in Appendix C and are summarized here.

Chi-square analyses were non-significant for the percentage of SWD in each group in all cases except grade 8 math. Even then, the chi-square value of .049 barely achieved significance. Chi-square, like other significance tests, is sensitive to sample size. Large sample sizes may produce significant results even with tiny differences between groups. The phi coefficient is an appropriate measure of effect size for two x two contingency tables, such as the form x group contingency tables used for these analyses. Phi eliminates the effects of large sample sizes by dividing the chi-square result by the sample size and then taking the square root, resulting in a range of values from 0 to 1, where 0 indicates no relationship between membership in one group and membership in the other group. The goal for this analysis was to verify that membership in the student group was not related to the test form that was administered. The phi coefficients were tiny in all cases, showing that the relationship between students grouped by disability status and the test form to which they were assigned was negligibly small.

In contrast, about half of the groups differed significantly in the percentage of students eligible for participation in ESOL programs. English-Language Learners (ELLs) make up a larger proportion of each student group than do students with disabilities, particularly at the lower grades. Some ELLs were eligible for Spanish test forms in math, resulting in slightly lower participation in the online English-language tests in math than in ELA. Nonetheless, small phi coefficients in all cases suggest that any chance relationship between ELL status and assignment to test form is extremely weak.

Next, comparison of percent correct on the 2014 field tests revealed significant differences in grades 3, 4, and 6 in ELA and grades 5 and 11 in math. For univariate ANOVA, partial eta squared is a measure of effect size that can be used to clarify the magnitude of significant results for large sample sizes. Partial eta squared can be interpreted as the amount of variance in scores attributable to test form assignment. Since all of the partial eta squared values are less than 0.01, less than 1% of the difference in scores is attributable to test form assignment.

For achievement test results on the 2013 assessments, scaled scores were used to compare student groups. The 2013 assessments are not comparable to the 2014 field tests. Furthermore, 2013 scores are scaled scores, rather than percent correct, and carry a different meaning. Nonetheless, these different tests serve as another useful check on group equivalence.

Significant results were obtained for grades 4, 6, and 11 in ELA and grade 8 in math. The same tiny effect sizes are evident for 2013 scores as for the 2014 percent correct scores with the exception of grade 11 ELA scores. At that grade and subject, test form assignment in 2014 was associated with just over 1% of test score variance from 2013. Third graders have no previous test scores and therefore groups were not compared on this variable.

Based on disability status, ELL status, 2014 percent correct scores, and 2013 scaled scores (where available), the groups assigned to forms A and B appear to be justifiably equivalent. This outcome provides a sound basis for continuing to evaluate student performance on TE items in different formats, which were randomly assigned to these groups.

7.3 Technology-Enhanced Items

Form A at each grade and subject consisted of the original TE items of all types, selected according to test specifications. Upon review of the TE items in form A, items that were likely to be inaccessible to students with vision or motor disabilities were adapted into more accessible alternative formats on form B for the purpose of testing these alternatives on the general population of students. Adapted items were reformatted to allow different presentation and response modalities (e.g., audio presentation or switch response) while retaining the same wording as the original items. All other items were unchanged.

In addition to these reformatted items on form B, there were additional opportunities to test alternative formats on forms C and D, which were otherwise different test forms. For example, an ordering item on form A may have been reformatted into a matrix item on form B and a matching item on form C. Other than these additional accessible alternatives and some linking items used for item parameter estimation across forms, forms C and D contained different items than forms A and B. For that reason, test scores on forms C and D are not comparable to those of forms A and B. However, the following sections, which describe format comparisons at the item level, include data on adapted items presented on forms C and D. Table 7.3 shows the total number of test items, the number of TE items (TEIs), and the number of TE items that were altered for accessibility on form B and again on forms C and D, by grade and subject.

Table 7.3

Total Items, Technology-Enhanced Items, and Altered Technology-Enhanced Items on 2014 Field Tests

Subject and grade	Total items	TEIs	Altered TEIs on form B	Additional TEIs on forms C and D
ELA Grade 3	52	10	6	2
ELA Grade 4	53	18	5	8
ELA Grade 5	54	9	3	4
ELA Grade 6	57	11	3	3
ELA Grade 7	55	10	5	5
ELA Grade 8	57	12	4	2
ELA Grade 11	58	14	7	9
Math Grade 3	60	14	5	4
Math Grade 4	60	14	7	4
Math Grade 5	60	12	6	5
Math Grade 6	60	10	3	1
Math Grade 7	60	9	3	1
Math Grade 8	60	8	3	3
Math Grade 10	60	8	1	0
Totals		159	61	51

Not all of the TE items on form A were adapted or altered for form B. There are three reasons why some items were not reformatted. First, some of the TE items on form A were already in accessible formats, such as matrix or matching, and did not require revision. Second, some of the TE items on form A, such as plotting points and graphing items, had no accessible options and were left the same. Since this part of the research project was to evaluate performance on paired items by students who did not require accommodations, these items did not present barriers for the students responding to the online tests. However, these items were reformatted for print and braille test forms. Those changes and their effects will be discussed in the Item Tryouts section of this report.

Finally, some TE items were initially believed to be accessible online, particularly drop-down items. Because of this belief, some original items were not altered. In addition, some original items of other types were reformatted into drop-down items as the accessible alternative. Drop-down items were reformatted for print and braille forms, and outcomes from those alternatives are described later. After the field tests, the project staff discovered that drop-down items are not accessible to students who require accommodations. Therefore, some of the items listed in the table above as tested accessible alternatives were not truly accessible and were omitted from the item comparisons.

It should be noted that item and task types selected for form B were based on the item types and cognitive requirements of the original items in form A. These items were written to meet specific targets and claims for the assessed standards, rather than to test a variety of innovative tasks. Some targets and claims lend themselves to particular cognitive tasks, such as ordering, sorting, or labeling. There was no attempt to vary the TE item types within any one test form or to include the most task types possible. Furthermore, several different types of inaccessible TE items were translated into just one or two accessible alternatives. As a result, the variety of innovative task types was reduced in form B as compared to form A. In fact, for some of the B forms, all of the altered items were the same item type.

In sum, there were only a small number of paired items in different formats at each grade and subject, and these pairs often exhibited little variety of tasks. Grouping items from different tests, grades, and subjects into pairs that share both the original and altered formats allows a broader look at how those two formats compare than would be possible by evaluating only within each grade and subject. The next section describes analyses that use the item pair as the unit of measurement. Following those analyses, are the results of DIF studies conducted on all of the changed item pairs for SWD and ELLs.

7.4 Item Results

Scatterplots of item pairs. Item pairs were selected for each scatterplot based on their shared original and altered formats. For example, 26 items whose original format was ordering were transformed into matching items for the accessible forms. These 26 items covered a range of difficulty, and they provide a sufficient set of data from which to draw preliminary conclusions about the efficacy of the format change. When there were not enough item pairs with a single original format that also shared an altered accessible format, the items were grouped together for analysis based on the altered format. For example, only a few items used labeling, background graphics, drop-down menus, and categorization on form A. These items were converted to matching items on form B. These items were grouped to more broadly evaluate how the matching format functions as an accessible alternative item type. This occurred because there were more original item formats than altered item formats, as described above.

As noted, drop-down items turned out not to be viable accessible alternatives; therefore, they are excluded from the scatterplots except as original item types. Furthermore, only unique item pairs are included in any single scatterplot. When an item on form A was altered into an accessible version on form B, and a different accessible version on form C, that item appears on both scatterplots for the different accessible versions. For example, a labeling item that was presented in matching format on form B and matrix format on form C is shown on both the matching and matrix scatterplots.

Some items that appeared on three forms had only one accessible option, and in that case, the item is included on the accessible alternative scatterplot only once. An example is an item that appeared as labeling on form A, matching on form B, and drop-down on form C. Of those three alternatives, matching is the only truly accessible version. Drop-down is not fully accessible and is not used in these analyses as an accessible option. Because the original version of this item was labeling, the item appears only once on the scatterplot showing matching as the accessible item type. Finally, items that had extreme IRT values (below -4 or above 4) were excluded from all scatterplots. In conclusion, the number of items shown on these scatterplots does not match the total number of TE items or of adapted and accessible TE items.

The scatterplots are in three groups to demonstrate the paired values of three different item characteristics. Appendix D contains the first set of scatterplots referencing classical item difficulty. In each case, the 1−p value of the original format is plotted on the x-axis and the 1−p value of the accessible format is plotted on the y-axis. Each scatterplot shows the diagonal along which identically performing items would fall. The regression line is the regression of y on x, or the prediction of performance of the altered items (y) by their originals (x). The standard major axis regression line is an alternative that removes the correlation between the item pairs from the regression equation. Finally, each plot shows the R^2 value from the regression. This value is a measure of the stability between the paired items, in spite of their format change. Each pair of items plotted on the graph is unique, and the set of items includes different grades and subjects; therefore, the relationship shown R^2 may also be interpreted as a measure of construct consistency.

Item difficulty: P values. The first group of scatterplots shows item difficulty as measured by 1 minus the p value. P values show item difficulty in terms of the percentage of students who answered the item correctly. Low p values indicate difficult items because few students answered correctly. High p values designate less difficult items that a high percentage of students responded to correctly. The p-value scale was inverted so that the easiest items would be plotted closest to the origin; item difficulty increases as the values extend along the axes. This intuitive scale of easy to difficult is consistent with the direction of the scale used in the second group of scatterplots, which shows IRT b values, another measure of item difficulty. With the p-value scale inverted to 1−p, the two plots of item difficulty can be compared side by side.

A sample scatterplot is shown in Figure 7.4.1. This plot will be described in detail to facilitate the interpretation of the additional plots in Appendix D.

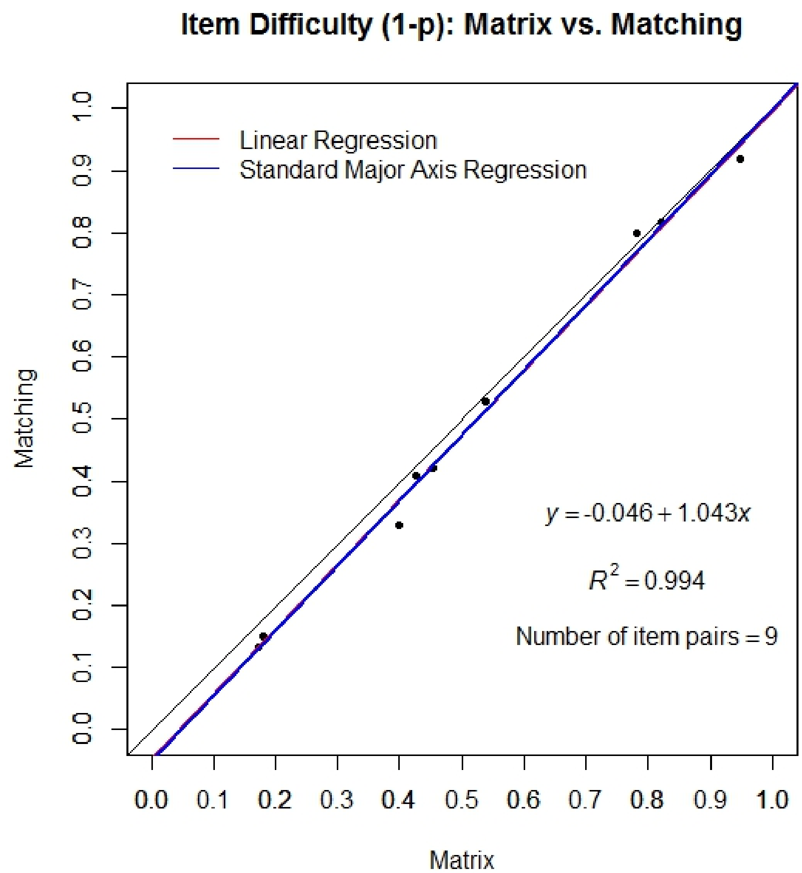


Figure 7.4.1. P values for matrix and matching items plotted as 1−P.

This particular scatterplot does not show items transformed from an inaccessible original format to a more accessible alternative. Instead, it represents items that appeared in two different accessible formats, matrix and matching. In some cases, the matrix or matching format may have been the original format. In other cases, an original item of a different format was converted to these formats and administered twice, once as matrix and once as matching.

This plot demonstrates the extremely close relationship between these two accessible item formats. These formats are virtually interchangeable with one another as accessible alternatives. This is evident in several ways. First, there are nine items that span a range of difficulty, providing a sufficient basis for preliminary conclusions about the relationship between these formats. All of the items are on or very close to the diagonal. The intercept of the regression equation is near zero, indicating no systematic difference in difficulty between the formats. Next, the slopes of the regression lines are extremely close to 1, indicating equal changes in difficulty for both formats across the difficulty range. (Note that the two types of regression lines overlap.) Finally, the R^2 value is greater than .99, indicating extreme consistency between the item pairs across the entire set of items.

The other 1−p-value scatterplots also show high concordance between the different item types in terms of item difficulty, and they can be interpreted in the same way as the figure above. There are six of these scatterplots. The first, ordering format versus matching format, shows that matching items appear to be slightly more difficult than ordering items at the easier end of the difficulty range because the 1−p value is slightly higher, meaning that the p value is slightly lower, than corresponding ordering items. There appears to be greater variability among easier items as well. The second plot, showing a combination of original formats with matching as the accessible alternative, also shows the pattern of greater variability at the easy end of the difficulty range and slightly more high-difficulty matching items. For both of these matching scatterplots, however, the intercept, slope, and R^2 values show high consistency between formats.

The third scatterplot displays matrix format items as the accessible alternative for a combination of original formats. This plot also shows high concordance between formats, with the same pattern of greater variability among easier items as was shown in the matching format scatterplots. The fourth scatterplot is the matrix to matching scatterplot illustrated above.

The fifth scatterplot shows drop-down items as the original format with selected response as the accessible alternative. Formatting a drop-down item into a selected response item was possible if there was only one drop-down menu in the item. Because selected response items are fully accessible, they are always a good alternative. This plot shows good consistency except for two items that were slightly more difficult in the selected response versions. While the variability among item pairs is greater, the consistency of the formats overall, as indicated by the regression equation, is still very high.

The sixth scatterplot is a “baseline” showing identical selected response items from the A and B forms. This scatterplot is included for the 1−p-value plots as a comparison of the differences that might occur as a function of the two different samples for each item pair. Fortunately, because these items were identical on both forms, they are highly consistent. Nevertheless, the unevenness around the diagonal demonstrates that there is some variability simply because two different samples were the basis of the data for the two p values for each item pair. These data provide a sense of what kinds of nonzero values the intercept and slope might have solely by chance.

Item difficulty in IRT b values. The b value in IRT is a measure of item difficulty obtained by jointly estimating the parameters of the original and adapted TE items using the unchanged SR and TE items on forms A and B as the linking block. A two-parameter logistic model was used to obtain the parameter estimates shown in the second and third sets of scatterplots. IRT b-value scatterplots use the standard IRT z-score scale of -4 to 4, where 0 is the mean of item difficulty and each integer shows a difference of one standard deviation in difficulty. Items with the lowest values have the lowest difficulty (b value).

The second set of scatterplots comprises six comparisons using the same sets of items as the 1−p scatterplots. Corresponding scatterplots that show the two item-difficulty estimates can be compared side by side. These scatterplots include ordering items transformed into matching items, combined original item types changed to matching format, combined original item types adapted to matrix format, matrix format compared to matching format, and drop-down items formatted as selected response items. A final sixth scatterplot again shows a selection of identical items from form A and form B to illustrate the amount of random variation that occurs between student groups even when the items are identical. Appendix E comprises these scatterplots, which are summarized here.

The first scatterplot, ordering format versus matching format, shows that matching items are somewhat more difficult than ordering items at the easier end of the item difficulty range, similar to the results shown in the 1−p-value scatterplot. There appears to be less variability between item types for the more difficult items. The scatterplot that shows a combination of other-format items that were transformed into matching items is also similar to the 1−p scatterplot, with matching items having slightly more difficulty than the other item types, particularly at the easier and middle portions of the item difficulty range. The third plot, combined item types compared with matrix items, is again extremely similar to the 1−p scatterplot. In this case, however, there is a somewhat higher correspondence between the difficulties of matrix items with their original versions as demonstrated by the R^2 values. The matrix versus matching scatterplot, that compares these two accessible versions with each other rather than with an inaccessible item type, again shows that these two formats are virtually interchangeable in terms of item difficulty.

In contrast to these first four similar scatterplots, the drop-down format versus selected response format plot shows greater variability among b values than among 1−p values. While the R^2 values are extremely similar, the slope of the b-value plot tips the balance in favor of drop-down items over selected response items. Most of these item pairs were relatively easy items, leaving the question open of how students would have performed on difficult tasks. Finally, the sixth plot shows almost equal values for a set of identical items pairs from each form. These IRT values were drawn from a separate calibration of parameter estimates, rather than the joint calibration in which these items were part of the linking block.

Item discrimination: IRT a values. The third set of scatterplots shows IRT a values, which are estimates of item discrimination, or the efficiency of the item in distinguishing between students of similar ability levels. For IRT a values, the scale of 0 to 2 shows the slope of the item discrimination curve, or the power of the item to discriminate between students whose ability is similar. Items with high values have good precision in differentiating between students with similar theta values. Items with values close to 0 are poor at discriminating because they do not distinguish well even between students whose abilities on the theta scale are farther apart. These plots cannot be directly compared to the first two sets of plots because these parameters measure a different aspect of item functioning.

Unlike item difficulty, for which test developers seek to prepare items across much of the range, high item discrimination values are always desirable. The values on this set of scatterplots are likely to fall within a narrower range, the correlation of the pairs of values will necessarily be lower, and therefore lower R^2 values are to be expected. The plot of item pairs tends to look more like a cloud in the middle of the scatterplot than a line of items roughly along a diagonal. The intercept, slope, and R^2 values do not have clear interpretations for these plots as they do for the difficulty plots. Instead, these values reveal the restricted range of the item parameters. By visual inspection, most values should be in the middle to upper ranges of the plot and the distances away from the diagonal should be minimal. When an item pair falls much above or below the diagonal, that item shows much better discrimination in one format or the other. All plots are included in Appendix F and are summarized here.

The first plot includes the same item pairs as the first plot in the other two sets. Because the a value shows a different feature of the item pairs, this plot does not look like the others. The first conclusion is that matching items tend to offer higher levels of discrimination than do the original drag-and-drop ordering items, because the matching values tend to be higher than the ordering values. However, there is more variability around the diagonal than was evident for item difficulty. The lower R^2 value is doubtless affected not only by that variability, but by the restriction of range for the item discrimination values.

The second and third plots continue the pattern of the first, and show that the accessible item types of matching and matrix tend to have higher discrimination values than the original item versions. The fourth plot, which contrasts the two accessible types of matching and matrix, shows high concordance according to intercept and slope. The R^2 value is probably affected more by restriction of range than by lack of consistency between the two item types. The fifth plot of drop-down items versus selected response items shows high consistency, though selected response items are somewhat better discriminators than their drop-down counterparts. Finally, the sixth plot shows extreme similarity between identical items whose discrimination parameters were estimated on separate groups.

Differential item functioning. DIF analyses were conducted on all changed TE items, contrasting performance of students who took form A with those who took form B. These analyses were conducted on separate groups of students with and without disabilities (SWDs and SWODs), and again on separate groups of students who did and did not qualify for ESOL services. The purpose of conducting the analyses in this nested way was to identify an interaction between item format and student group if DIF was detected. For example, if uniform DIF favored form A for students without disabilities but favored form B for SWD, it would suggest an unwanted interaction between item type and disability status.

Because of their length, Appendix G contains the tables of DIF results. These results are summarized here. Two guidelines for the magnitude of DIF results can be considered in the interpretation of the outcomes. Jodoin and Gierl (2001) suggest that ΔR^2 values greater than 0.035 and less than or equal to 0.070 constitute moderate DIF while values greater than 0.070 indicate large DIF between the contrasted groups. DIF values lower than 0.035 are negligible to small. None of the TE items showed large DIF. The number of TE items (TEIs) that demonstrated moderate DIF are shown by grade and subject in Table 7.4.2.

Table 7.4.2

Number of Items Showing Moderate DIF

Subject and grade	Number of changed TEIs	Number of TEIs with DIF by group			
		SWOD	SWD	Non-ELL	ELL
ELA Grade 3	6				2A, MC
ELA Grade 4	5				
ELA Grade 5	3	†1A, match		†1A, match	†1A, match
ELA Grade 6	3				
ELA Grade 7	5		1B, matrix		1A, match
ELA Grade 8	4				
ELA Grade 11	7				
Math Grade 3	5				
Math Grade 4	7	1N, match			
Math Grade 5	6		1A, match		
Math Grade 6	3				
Math Grade 7	3				
Math Grade 8	3				
Math Grade 10	1				
Totals	61	3	2	1	4

Note. † indicates one item that showed DIF for multiple student groups. Uniform DIF is shown by the number of items and the letter of the favored form. Non-uniform DIF is shown by the number of items and the letter N.

Sixty-one changed TE items were evaluated for four student groups: students without disabilities, students with disabilities who did not require accommodations, students who were not eligible for ESOL services, and students who were eligible for ESOL services. Nine instances of moderate DIF were identified for seven unique items, and no occurrences of large DIF (Jodoin & Gierl, 2001) were found. Of those seven unique items, six showed uniform DIF, with five items favoring the A form and one item favoring the B form. DIF occurred within each student group. Except in the case of grade 5 ELA, an item that showed DIF within one student group did not show DIF in any other student group. There were no interactions in which the same item favored one form for one group but the other form for the corresponding group. For SWD, a particularly crucial group for this research, there were only two items that demonstrated DIF, and those items favored different forms. However, the A form was favored more often than the B form overall.

One grade 5 ELA item showed DIF and favored the A form for three of the four student groups. This item was an easy ordering item with five elements, and it was transformed into a matching item of medium difficulty for the accessible form B. Examining the 1−p-value scatterplot that compares ordering format to matching format, this item was the relatively easy item farthest from the diagonal. The difference between the p values for the two formats of this item was the largest of any item pair. As a result, it favored form A for three student groups, though not for SWD.

Overall, these DIF results are encouraging in terms of demonstrating that the different item types on forms A and B did not disadvantage any particular student group. Only 7 out of 61 TE items showed moderate DIF based on test form, while no items showed large DIF. DIF results were scattered among the student groups rather than predominating in one student group, which would have called into question the fairness of those item types for those students. Except for one item that was clearly easier in its original format and showed DIF for three groups, each instance of DIF was evident for only one group. It should be noted, however, that most of the DIF favored the A or original form rather than the B form that contained reformatted test items.

7.5 Assessment Results

Differences in scores on TE items. The analyses of item parameters, as displayed on the scatterplots, used each individual pair of matched (but differently formatted) items as the unit of analysis, with item pairs aggregated across grades and subjects. These analyses included matched and adapted items from additional test forms that were not identical in content to form A. Because the item pair, rather than the test form, was the focus, a broader net could be cast across all accessible alternative items tested in spring 2014.

The second set of analyses evaluated DIF, which is an implicit contrast of student performance on the matched items within a test form for key student groups for each grade and subject. DIF was conducted for the A and B forms; therefore, DIF results did not address performance on some of the additional items that were sprinkled throughout the C and D forms.

These two initial methods of analysis investigated performance at the finest grain—individual TE items. The next set of analyses changes the focus from individual items to testlets, or groups of items. These analyses contrast scores on the sum of the matched and reformatted TE items for each form, given that there are different numbers and types of TE items for each grade and subject.

To understand these analyses, refer to the numbers of changed TE items for each grade and subject shown in Table 7.4.2. First, the total score for the changed TE items was computed. That total score is the sum of the same number of matched items in different formats for forms A and B at any one grade and subject. Scores for TE items that were identical on forms A and B were not included so that only the different TE formats would be contrasted. Those scores were compared using univariate ANCOVA with the changed TE item score as the dependent variable. Form type was entered as a fixed factor, and disability or ELL status was entered as a random factor for separate ANCOVAs. The total score on the identical SR items on each form was used as a covariate to control for achievement in that subject. Because all of the SR items were identical between forms A and B, the total SR score has the same meaning for all students in each grade and subject.

An interaction would suggest that the contrasted student groups performed differently on the different TE items; for example, if SWD performed better on the adapted TE items on form B while SWOD performed better on the original TE items on form A. The research hypothesis was that there would be no interaction between form and student group.

Main effects indicate the relative difficulty of the changed TE items on the two forms and the relative performance of the two student groups on those items. Students with disabilities often perform worse overall than SWOD because of the impact of their disabilities on instruction and educational attainment. However, since overall subject-area achievement was controlled by using the total SR score as a covariate, the a priori research hypothesis was that student groups would not differ on the TE items. The research hypothesis for TE item difficulty was that the TE item summed score will be equivalent regardless of form. Table 7.5.1 shows the results of these ANCOVAs for each grade and subject. Cohen’s d, which is an effect size measure of difference between two means in terms of their pooled standard deviations, was computed and is shown whenever a comparison of two means was significant. A Cohen’s d value indicating a small effect is generally about .20, a medium effect is .50, and a large effect is around .80. The interpretation of the value also depends on the two means being compared and the consequences of effects of that size to the groups being compared. For example, a medium effect of half of a standard deviation difference in scores may be an unacceptable outcome if it represents worse performance by a vulnerable or disadvantaged population of students.

Table 7.5.1

ANCOVAs with Summed TE Item Score as Dependent Variable and Summed SR Item Score as Covariate

	SWD					ELL				
Grade and subject	Form x Group	Form	Cohen's d	Group	Cohen's d	Form x Group	Form	Cohen's d	Group	Cohen's d
ELA Grade 3		.039	0.357			.002				
ELA Grade 4		.027	0.679	.022	0.169					
ELA Grade 5										
ELA Grade 6				.005	0.189					
ELA Grade 7							.034	0.143	.000	0.061
ELA Grade 8				.040	0.134					
ELA Grade 11										
Math Grade 3		.039	0.119							
Math Grade 4		.018	0.057	.002	0.016					
Math Grade 5		.044	0.162							
Math Grade 6										
Math Grade 7										
Math Grade 8				.039	0.159					
Math Grade 10									.024	-0.116

Note: p values are shown where significant effects were found; empty cells indicate no significant effects. Both groups performed worse on Form B when main effects were significant; SWD or ELL groups performed worse when group main effects were significant except in Math Grade 10. Cohen's d is shown for significant differences between two means.

In grade 3 ELA, there were six TE items from form A that were adapted for form B. Four of the items were changed to the matching item type and the other two were changed to SR items. The TE items on form B, as a group, were more difficult for all students. In grade 4 ELA, all five inaccessible TE items were adapted to matching items on form B. Performance on the TE items on form B was again significantly lower than for the corresponding items on form A. The medium to large Cohen's d values for these contrasts demonstrate the magnitude of the difficulty these matching items caused at the younger grades.

At grades 3, 4, and 5 in math, as at grades 3 and 4 in ELA, the TE items on form B were more difficult than the corresponding items on form A, though the sizes of the effects were much smaller than in ELA. Over all of these B forms, there were several item types utilized as accessible alternatives, although the matching item type predominated. The pattern of lower performance on the TE items on the B forms suggests that closer attention should be paid to how students respond to matching items and whether the matching type is the best alternative for drag-and-drop items that must be reformatted for SWD. The results at these grades also imply that matching items may be causing more of a problem for younger students than the original drag-and-drop items on these tests. In summary, the TE items on form B were more difficult than the TE items on form A in several lower grades and both subjects. This unexpected outcome warrants concern about specific TE item types and why they tend to induce lower performance for general-education students and for SWD.

SWD performed significantly more poorly than SWOD on the altered TE items on five assessments: ELA grades 4, 6, and 8 and math grades 4 and 8. These five assessments compose almost one third of the tests that were evaluated, signifying that further investigation is necessary into why SWD had more trouble than SWOD on these adapted item types. Even though the Cohen's d values were small, the trend was consistent.

In terms of ELLs, there was an interaction with ELL status on grade 3 ELA. ELL students performed more poorly than non-ELL students on form B but not on form A. At grade 7 ELA, ELL students performed slightly lower than non-ELLs while all students performed worse on the form B TE items. Finally, ELL students performed slightly better than non-ELL students on the TE items in grade 10 math. ELL students did not demonstrate as many difficulties on the adapted TE items on form B as did SWD. Adapted TE item scores produced significant effects in only three grade levels, and effect sizes were all quite small. The low incidence of lower performance on adapted TE items based on form or ELL status is welcome news.

Differences in scores on selected response (SR) items. The next set of investigations concerned the scores on the SR items on each test. A total SR score was computed for each test, and then test forms A and B were compared. The purpose for this investigation was to determine whether the different versions of the TE items on forms A and B had any effect on SR scores, which consisted of identical items. These analyses do not look directly at TE items and scores, but rather on whether the SR items on each test were affected by the format of the TE items on those tests. The research hypothesis was that total SR scores should be identical for each pair of forms because they consist of identical item sets.

These analyses used univariate ANOVA with the total SR score as the dependent variable. Test form type was entered as a fixed factor and disability or ELL status as random factors. TE item performance was not used as a covariate for these analyses because the purpose was to tease out any effect of TE items on SR item scores based on form type. Table 7.5.2 shows the results.

Table 7.5.2

ANOVAs with Summed SR Item Score as Dependent Variable

	SWD					ELL				
Grade and subject	Inter-action	Form	Cohen's d	Group	Cohen's d	Inter-action	Form	Cohen's d	Group	Cohen's d
ELA Grade 3									.034	0.579
ELA Grade 4				.036	0.698	.002				
ELA Grade 5				.011	0.818				.024	0.508
ELA Grade 6		.039	0.078	.003	1.010				.028	0.528
ELA Grade 7				.036	1.107				.032	0.690
ELA Grade 8				.048	1.039				.017	0.601
ELA Grade 11				.036	1.108		.010	0.032	.000	1.022
Math Grade 3	.049					.032				
Math Grade 4				.006	0.631					
Math Grade 5				.032	0.678				.014	0.518
Math Grade 6				.015	0.769				.009	1.083
Math Grade 7		.039	-0.027	.001	0.816				.008	0.464
Math Grade 8				.039	0.831	.023				
Math Grade 10										

Note: p values are shown where significant effects were found; empty cells indicate no significant effects. Both groups performed worse on Form B when main effects were significant except on Math Grade 7; SWD or ELL groups performed worse when group main effects were significant. Cohen's d is shown for significant differences between two means.

Project staff anticipated that historically low-performing students, such as SWD or ELLs, might perform worse than SWOD or non-ELLs and that this outcome would be evident in the main effect for group. This hypothesis proved true. SWD performed more poorly on the total SR item score, which consisted of the majority of items on each test, at all grades and subjects except grade 3 in both math and ELA and grade 10 in math. ELLs similarly performed worse than non-ELLs on all ELA tests except grade 4. They also showed significantly lower performance at grades 5, 6, and 7 math. The fact that ELLS performed worse on more ELA tests than math tests was not unexpected; the barrier of English-language proficiency is likely to be more pronounced on ELA tests. The magnitude of the effects of group differences was medium to large in each case.

For SWD, there were some other unanticipated outcomes. An interaction between form type and student group was evident at grade 3 math, where SWD performed slightly worse on form B than on form A, while scores of SWOD did not differ. In two instances, at grade 6 ELA and grade 7 math, total SR scores were significantly different for both SWD and SWOD. Form A was easier at grade 6 ELA, while form A was more difficult at grade 7 math. These outcomes were very slight, as shown by the effect sizes, and may simply have been due to random group differences.

Three instances of interaction effects occurred for ELLs. On grade 4 ELA and grade 8 math, ELLs performed worse on the SR items on form B than on form A, whereas non-ELLs performed the same on both forms. For grade 3 math, ELLs performed better on the SR items on form B than on form A, but scores for non-ELLs did not differ between forms. A main effect for form occurred for grade 11 ELA, where SR scores on form A were slightly higher than those on form B.

In summary, other than expected group main effects, the project found only three other significant outcomes in the comparison between SWD and SWOD. Four additional outcomes were evident when the project staff contrasted ELL and non-ELL students. Fortunately, none of these effects were large, and the directions of the effects were mixed.

7.6 Conclusions

These results show considerable support for the equivalence of different item layouts for students with and without disabilities taking computerized assessments without accommodations. Some areas of concern were revealed in the trend of matching items being slightly more difficult than their drag-and-drop original versions, particularly for elementary children. Significant differences between the scores on the TE items on the paired tests were evident at grades 3-5 on the ANCOVA analyses, where matching items were the predominant alternate version. Furthermore, item difficulty scatterplots show a consistent trend toward the greater difficulty of matching items.

On the scatterplots, items formatted as matrix tasks appeared to be somewhat closer to their originals in difficulty and discrimination. However, items that were administered in both matching and matrix formats were essentially equivalent. Mixed results were apparent in that more difficult item formats also tended to show better discrimination between students of different ability levels.

Null hypotheses were rejected for some of the ANOVA analyses. One significant interaction between form and disability status, plus two instances of significant main effects for form were revealed by the ANOVAs, which compared scores on the identical SR items on the paired A and B forms. It is impossible to know whether the TE items on those three tests affected performance on the SR items, or if these are cases of random differences between groups.

Overall, there were fewer significant results in the ANCOVAs and ANOVAs for students eligible for ESOL programs. These student groups were not the target of the ATEA project, yet these field tests provided an interesting opportunity to contrast ELL and non-ELL students on the same measures. These outcomes were not as consistent as those for SWD, suggesting that they are more likely to reflect sample differences than systematic effects of the TE item formats.

On the DIF analyses, 7 of 61 TE items showed moderate DIF between the contrasting versions. Four of the seven were matching items, one was a matrix item, and two were multiple-choice items. These DIF outcomes were scattered among the four student groups based on disability or ELL status. Only one item showed DIF for multiple groups. Fortunately, no items showed large DIF. Clearly, identifying the pros and cons of the different item types for various ages, grades, and subjects will require more investigation as these new types are adopted by test developers and become widely used.

The field tests conducted in Kansas during spring 2014 were the largest tests of contrasting item formats undertaken for the ATEA project. Embedding different item layouts and formats in otherwise matching tests allowed quantitative comparison of item parameters and DIF studies based on large and equivalent samples of students, including students with disabilities. These studies permit preliminary conclusions about the effectiveness and fairness of alternative item presentations. With these outcomes as a starting point, further investigation can proceed to evaluate how SWD respond to items when presented with accommodations and on special forms.

8.0 Item Tryouts

8.1 Procedures

The purpose of the item tryouts was to administer TE items in accommodated formats to sufficient numbers of students with vision and motor disabilities to permit quantitative analysis of performance compared to students who responded to non-accommodated online TE items. To that end, the TE items from the Kansas accessible forms were grouped into short tests in ELA and math for administration in ATEA consortium states other than Kansas. These tests were administered via CETE's KITE system to students whose teachers had obtained signed parental consent.

Originally, field tests were intended to occur simultaneously with the Kansas field tests during the spring of 2014. However, representatives of the member states' departments of education expressed concern about other field tests occurring at the same time, such as those conducted by the general assessment consortia SBAC and PARCC, and the alternate assessment field tests conducted by the NCSC and DLM consortia. After polling ATEA member states and requesting an extension of ATEA project activities for an additional year, the field tests were reformatted into tryouts of only the TE items and scheduled for October 2014. Ultimately, testing extended into December 2014 in order to obtain as many participating students as possible.

TE items were administered one grade level higher than in Kansas, i.e., grade 4 items were administered to fifth graders. This was done because the items would probably have been too difficult for fourth graders who had received only two to three months of instruction by the time of the item tryouts. However, item performance was combined with that of Kansas students who had responded to the same items on special forms the previous spring during the testing window from March through April.

8.2 Subjects

Table 8.2.1 shows the numbers of participants in the Kansas field tests and the ATEA consortium item tryouts, by grade and subject. While most students took both ELA and math tests, some individual students, in Kansas as well as in other states, responded to only one of the tests. For this reason, the numbers of participants do not match at every grade.

Table 8.2.1

Participation in Special Forms Field Tests (Kansas) and TE Item Tryouts (Other States)

Grade and subject	Participants		
	Kansas	Other states	Total
Grade 3 ELA	12	7	19
Grade 4 ELA	7	3	10
Grade 5 ELA	8	6	14
Grade 6 ELA	6	4	10
Grade 7 ELA	8	5	13
Grade 8 ELA	15	4	19
High School ELA	4	10	14
Grade 3 Math	12	7	19
Grade 4 Math	9	3	12
Grade 5 Math	9	6	15
Grade 6 Math	5	5	10
Grade 7 Math	8	5	13
Grade 8 Math	17	4	21
High School Math	7	10	17

As Table 8.2.1 demonstrates, the numbers of students responding to TE items at each grade and subject was tiny. The entire Kansas population of students who used accommodated forms was tested. However, participation in other ATEA-consortium states varied based on the state’s commitment to the project, the dissemination of information to teachers and recruitment of subjects, and the effect of competing priorities on the time and attention of state department of education staff, teachers, and parents. Furthermore, no students with motor disabilities enrolled in field tests or in item tryouts.

ATEA member states noted that other field tests that occurred at the same time were of higher priority, especially since other field tests were not optional. For this reason, two ATEA member states declined to participate at all. Several other states disseminated information to teachers in their states, but did not actively recruit students.

When low enrollment became evident, ATEA project staff undertook additional measures to recruit students. Project staff encouraged personnel at ATEA-consortium state departments of education to solicit teachers in their states. Staff sent e-mails to state schools for the blind when contact information could be obtained. In several cases, this resulted in enthusiastic participation of teachers and students. Table 8.2.2 shows the breakdown of participation by state.

Table 8.2.2

Participants by State

State	Participants
CO	1
IL	2
KS	68
MD	3
MI	8
MO	6
MS	9
NE	1
OH	5
OK	5
WV	1
Total	109

8.3 Results and Conclusions

The population of students who used special forms turned out to be too small for quantitative comparisons. The challenge that appeared at the conclusion of the item tryouts was how to evaluate the performance of students with accommodations when the planned propensity score matching and DIF were not possible. Evaluation of the accessibility of items on special forms had to take another direction. The investigation now attempted to determine whether items on the special forms were accessible by qualitatively comparing items on special forms with those on online forms. The first step in that investigation was to determine how many students on special forms were able to answer an item correctly, which would imply that the item content was sufficiently accessible. There are several possible reasons why an item was not answered correctly by most or all of these students. The item may have been inaccessible even with the print or braille accommodation; the item content may have been too difficult for students with vision disabilities; or the item may have been too difficult overall, which would be confirmed by the difficulty of the item on the computerized form.

Because of the small number of students, the performance of students on the special forms (print, large print, and braille) was combined in order to compute each item's p value. While this was not an ideal solution, it did provide 10 or more student responses for each test item. This procedure was defensible for most items on the special forms. First, the layout of items on print and large print was identical. In most cases, the layout in braille was also identical to that on print forms. Only for matching items was the layout modified for braille. In braille, matching items are laid out with the columns presented vertically, top to bottom, rather than horizontally, left to right. After the item stem, the column containing answer choices or letters is presented first. The column with item numbers is presented second. This layout is introduced in the sample items at the beginning of each test (as are all item layouts on all special forms) so that the braille reader is prepared to review the answer choices before selecting an answer choice for each numbered item. This layout was suggested by students in cognitive labs who were confused by the standard matching format. Furthermore, CETE received no complaints from teachers or students as a result of introducing this layout for Kansas and ATEA tests in 2014. Figures 8.3.1 and 8.3.2 show the directions and layouts of the sample matching items for print and braille.

Sample matching question for print.

The first column on the left shows four entries numbered 1, 2, 3, and 4. The second column on the right has five answer choices marked A, B, C, D, and E. Match each entry in the first column with the correct answer choice in the second column. Draw lines in your test booklet between the columns or write the letter for the correct answer choice next to the number in the first column. Some entries or answer choices will not be matched.

S3. Match each expression with its value.

1.	10 – 4	A)	5
2.	7 – 0	B)	6
3.	5 + 4	C)	7
4.	3 + 2	D)	8
		E)	9

Figure 8.3.1. Sample matching item for print forms.

Sample matching question for braille.

There are five answer choices marked A, B, C, D, and E in the first list. The second list shows four entries numbered 1, 2, 3, and 4. Match the letter of the correct answer choice in the first list with the corresponding entry in the second list. Some entries or answer choices will not be matched.

S3. Match each expression with its value.

A) 5

B) 6

C) 7

D) 8

E) 9

1. 10 – 4
2. 7 – 0
3. 5 + 4
4. 3 + 2

Figure 8.3.2. Sample matching item for braille forms.

The majority of the 98 ATEA TE items presented on special forms (54%) had p values greater than .5. These items were evidently accessible to most or all of the students who used these forms. P values for the special forms TE items tallied by deciles are shown in Table 8.3.1.

Table 8.3.1.

ATEA TE Item P Values

P value	Number of items	Cumulative %
.80 or above	11	11
.70 to .79	12	23
.60 to .69	11	35
.50 to .59	19	54
.40 to .49	19	73
.30 to .39	7	81
.20 to .29	12	93
Less than .20	7	100
Total	98	

The seven items with p values less than 0.2 were obvious candidates for further investigation. The layouts of these items on the special forms were compared with the layouts on the accessible online forms to verify that item content was accurate and identical. The p values of the online forms were obtained to compare the magnitude of the differences for students on print and braille forms. These p values are shown in table 8.3.2.

Table 8.3.2.

Comparison of P Values of Difficult Items on Special Forms

Grade and subject	Item order	N	Item type	Special forms p value	Online forms p value
Grade 11 ELA	1.1.16	14	matching	0.07	0.22
Grade 3 Math	1.15	14	matrix	0.07	0.43
Grade 4 Math	1.13	12	matrix	0.08	0.29
Grade 4 Math	2.17	12	matching	0.17	0.25
Grade 8 Math	1.2.2	20	matching	0.10	0.21
Grade 8 Math	2.2.3	21	matrix	0.19	0.38
Grade 10 Math	2.1.6	17	matrix	0.12	0.26

Clearly several of these items were also difficult (p values were less than .3) for the general population of students. Some of the increased difficulty on the special forms versions of the items may be due to the impact of disability on educational achievement and performance, though without larger groups of students on special forms, it is impossible to test this hypothesis quantitatively.

Nonetheless, two of the difficult items stood out as significantly different from their online counterparts. The grade 3 math matrix item and the grade 8 math matrix item were not overly difficult for the students who took the online forms. Possible reasons for the greater difficulty include lack of visual or tactile clarity in the print or braille presentations or an unanticipated interaction of disability with item presentation. Figures 8.3.3 and 8.3.4 show the layouts for these two items (with secure content removed), and Figure 8.3.5 shows a screenshot of the online item layout (with secure content removed). Though the online version of the item appears small on the printed page, it would use the entire screen and would actually be larger and perhaps easier to read than the print version.

Use the symbols to compare the fractions. Not all symbols will be used.

	=	>	<
	○	○	○
	○	○	○

Figure 8.3.3. Grade 3 matrix item in math, print version.

Two numbers are expressed as powers of 10.

X =

Y =

Choose True or False for each statement about X and Y.

	True	False
X is times larger than Y.	○	○
Y is times larger than X.	○	○
X and Y are equal.	○	○

Figure 8.3.4. Grade 8 matrix item in math.

Use the symbols to compare the fractions. Not all symbols will be used.

	=	>	<
	○	○	○
	○	○	○

Figure 8.3.5. Grade 3 matrix item in math, online version.

Like the other ATEA items on special forms, these items are almost identical to the layout and format presented in the accessible online forms, with the exception of the shaded text box around the item stem. Neither item is a matching item, which would have caused a possible confound given that matching items are laid out differently in print and braille. The reason these items were much more difficult for students using special forms is impossible to determine, but extreme variability due to the small samples of students who used special forms cannot be ruled out. The fact that the online items are larger than the print items may also contribute to their higher p values for students who can access them online, but this does not explain why only a few special forms items were significantly more difficult than their online counterparts. In fact, among the items that were within an acceptable range of difficulty, the items on the accommodated forms sometimes had higher p values than the online originals. Again, however, this may be an artifact of the unstable results provided by small samples.

Some tentative conclusions can be drawn from these item tryouts, pending more research. First, most of the item layouts in print and braille appear to have been sufficiently accessible to blind and low-vision students; their p values fall within an acceptable range. Second, the different item types were distributed among the p values. No single item type dominated the group of difficult items, which confirms the findings of the general-population samples that the item types are defensibly interchangeable.

The overwhelming limitation to this part of the study is the small samples of students accessing special forms. The influence of individual students within these small samples may have resulted in extreme variability in the p values. Hence, these results do not permit any firm conclusions about the performance of these students. Further study with these populations of students will have to await several years of TE item presentation in various formats and with different accommodations.

9.0 Final Expert Review

The six experts who had evaluated the first prototype items were engaged two years later to review the final TE items and to provide feedback on the progress that had been made. A seventh expert reviewer, an administrator from a state school for the blind, was added to the group.

Reviewers found that the adapted item types were accessible in all of the formats they evaluated. Reviewer comments largely focused on online tools and accommodations as well as the quality of the hard copy braille test forms. These comments form the basis for continued improvement in both online and external accommodations.

Reviewers commented that the audio delivery and pace were very good. Reviewers noted, however, that audio delivery could be improved by having the audio respond to the student's choices, such as by having the selected sentence read aloud in the Select Text item type. They also requested audio for the online calculator. This suggestion will be implemented by providing a talking calculator to students who need it. Adjustable reading speed was also suggested, and this enhancement is being considered. Concerns were voiced about the forward and back audio buttons being too small. One reviewer noted that if audio and screen magnification were used together, the audio play box might cover part of the item.

Expert reviewers noted that screen magnification worked inconsistently. For example, in some item types the text wrapped so that horizontal scrolling was not required. In other situations, such as when text is placed in an image to conserve its layout as is required for poetry, the text did not wrap, forcing the student to scroll both vertically and horizontally. In some items, the text but not the graphics responded to screen magnification. This is another example of images reacting to onscreen tools differently than text. Issues involving inconsistent response to accommodations may be able to be addressed through technology improvements. Reviewers also noted that students with motor impairments may have difficulty navigating with the zoom feature, particularly when both vertical and horizontal scrolling are required. Unlike the technological issues mentioned above, this problem is inherent in using screen magnification and represents a situation in which the accommodation for the disability may introduce its own burdens.

Like screen magnification, switches were noted to work inconsistently across different item types. Reviewers also noted that the step scanning system can be confusing and suggested that students complete a tutorial before encountering it during a test. Some of the online tools were noted to be unavailable when step scanning was selected, with no obvious way to disable step scanning in order to access other functions. These comments referred to known issues in KITE, ATEA's testing platform, and may not apply to other testing platforms. Enhancements to audio, screen magnification, and switches in KITE are currently in development.

Experts who reviewed braille test forms remarked that having the hard copy in braille provided complete access to the text, eliminating the need to scroll and navigate with a refreshable braille display or synthesized speech. They also suggested additional descriptions for some graphics, such as tables and bar graphs. These descriptions would be crucial for students who have been blind from birth. In one item, one of the answers to a spelling question was a word that is usually contracted in braille. Reviewers noted that providing the complete spelling, either correct or incorrect, for a contracted word would give away the answer, unless the entire passage is presented in uncontracted braille. These suggestions warrant annual review and consultation with braille readers during the development of items that will be presented in braille.

Off-screen accommodations were recommended for students using braille, including an abacus, talking calculator, and braillewriter. For students utilizing on-screen delivery, reviewers commented, a scribe may be necessary to assist in locating tools and activating features as well as to record responses. This is a reminder that developing accessible alternatives to TE items does not eliminate the need for accommodations when students use those methods for instructional access.

10.0 Conclusions

The ATEA project set out to investigate the accessibility of innovative tasks and test items on computerized assessments for students with vision and motor disabilities. As with any ambitious research project, that goal was only partially attained. Nonetheless, the seven research activities that comprised the project over the course of three years yielded fascinating and novel information for the development of both computerized assessments and special forms tests for students with vision and motor impairments.

In terms of qualitative information obtained from experts, students, and teachers, many variations in item types and accessibility features were suggested and later tested. Some of those suggestions led to dead ends when anticipated features, such as audio presentation or switch response for drop-down menu items, did not work as planned. Initial attempts at formatting TE items in braille were wildly off base, for example, but subsequent guidance from braille readers, both teachers and students, enabled the adaptation of braille items much more effectively. Other recommendations were extremely successful right out of the box, as when the matrix item type was found to be adaptable for both online and offline testing, and, as a bonus, turned out to be nearly equivalent to original TE items in terms of item difficulty and construct measurement for students without vision or motor impairments.

Quantitatively, the project was both a success and a failure: a success in the opportunity to test matched pairs of items with very large samples of students taking online tests without accommodations; a failure in terms of the impossibility of enrolling sufficient samples of students with vision and motor impairments to afford statistical comparisons with matched groups of students without these disabilities. Testing adapted items with general population samples without accommodations, while not even conceptualized in the original proposal, turned out to be hugely valuable in terms of justifying that identical content presented in different layouts or item formats can measure the same construct with the same difficulty as TE items. Adapted items even showed some improvement in item discrimination in online testing. This outcome serves as the basis for claiming that adapted items are fair and equitable for students who must utilize either online accommodations or special forms outside of computerized testing. Without strong claims for this kind of item equivalence, which has been only partially achieved by this project, there cannot be equally strong claims that the assessment results and scores of students in small, unique populations mean the same thing and can be interpreted in the same way as the scores of students in the general population who do not require assessment adaptations or accommodations.

A less stable outcome was obtained from samples of students with vision and motor challenges, suggesting on the basis of reasonable *p* values that adapted items in special forms were indeed accessible. These tentative conclusions were confirmed by a final item and test review by the same experts who had initiated the project’s research trajectory. These experts critically reviewed both online and special forms, informing CETE once again about the successes and shortcomings of the work of the project thus far.

Limitations to this project warrant scrutiny for others wishing to evaluate or continue this research program. First, all test items, tasks, accommodations, and assessments were delivered via the Kansas Interactive Testing Engine, or KITE. This proprietary testing platform forms the basis for delivery of both the Kansas and Alaska state assessment programs. KITE is a complex and multi-faceted platform designed for use by test developers, item writers, teachers, administrators, and students themselves. Tasks and assessments delivered on KITE meet the specifications and capabilities required by CETE. These specifications and capabilities may be quite different from those of other test development and delivery programs. Differences will doubtless be evident in standard and TE item types, assessment features and functions, available accommodations and their functionality, and other accessibility tools. Conclusions reached with KITE may differ from those attainable in other settings. Before readers of this report conclude that the ATEA results are the final word on any aspect of accessibility, they would do well to confirm those results with other types of TE items and test delivery platforms. ATEA results should serve only as initial guidance toward development of greater access.

Furthermore, the samples of students who participated in the large-scale field tests were restricted to the state of Kansas. These students may or may not be similar in their characteristics and educational experiences to students in other states and regions. Students who participated in small-scale TE item tryouts spanned ten states. Even greater caution should accompany those results, however, due to the difficulty encountered in enrolling those students and obtaining data from them. The samples are much too small to draw anything more than very cautious conclusions. This particular area of research, as always, continues to demand much more effort.

Another crucial topic would address the struggle to assemble sufficient numbers of students in low-incidence populations. The major stated goal of the ATEA project was to accomplish that goal by enrolling multiple states in the research effort and pooling the results from their students with motor or vision challenges. The project ran into major roadblocks in this effort. First, due to pressures on ATEA member states to participate in other field tests prepared by major assessment consortia, both for general students and for students taking an alternate assessment based on alternate achievement standards, states either declined to participate in the ATEA field tests or were unable to exert sufficient pressure on the small number of teachers in the field who work with the target students. Another possible explanation is that state departments of education, who served as the nexus of communication by ATEA staff, viewed these small populations as less reachable or as a lower priority for their outreach efforts. Evidence that may substantiate that hypothesis includes the fact that state schools for the blind, when contacted directly, even though they were not in ATEA states and were unfamiliar with the project, often participated enthusiastically and ultimately contributed at least half of the special forms sample. This outcome speaks to the need to reach out to teachers and administrators for low-incidence populations directly rather than through intermediaries. These teachers and administrators have the needs of their special students at the forefront and will often jump at the chance to become involved in research activities that hold promise to benefit those students directly. They are often less involved in educational development activities and decisions, such as assessment development and review, that focus primarily on more typical students. Therefore, they may be eager to take on even significant challenges because the opportunity to do so is so rare.

A potentially subtle but pernicious issue precedes the difficulty of assembling large groups of students from low-incidence populations, however. That issue is the possible bias on the part of grant writers and reviewers toward overly optimistic plans for research that is extremely difficult to do. In the case of the ATEA proposal, research involving the sample sizes of students with vision or motor disabilities had never previously been carried out, and with good reason. It is simply not feasible to enroll that many students voluntarily with the protections appropriately required by university Institutional Review Boards. Obtaining larger groups of students with very specific characteristics is probably possible only if the data from those students is gathered as part of a state assessment program. That method is the hope for future replication and refinement of the initial findings from ATEA.

Project staff sincerely hope that the activities, results, and conclusions presented in this report will serve to initiate further experimentation and evaluation in other states and assessment programs. That is the means by which often-overlooked students, such as the target students of this grant, will be best and most equitably served by assessment results.

11.0 References

Abbott, D. (2005). Reasonable adjustments: The practicalities. *Occupational Health Review*, *115*, 26–28. Retrieved from http://enricosmog.com/sites/default/files/OHR%20115%20p26-28.indd_.pdf

Allman, C. (2006). *Position paper: Accommodations for testing students with visual impairments*. American Printing House for the Blind, Inc. Retrieved from http://www.aph.org/tests/accomodations.html

Amato, S., & Smith, D. W. (2012). Synthesis of available accommodations for students with visual impairments on standardized assessments. *Journal of Visual Impairment & Blindness*, *106*, 299-304.

American Evaluation Association (AEA). (2005). *Guiding principles for evaluators*. Retrieved from http://www.eval.org/Publications/GuidingPrinciplesPrintable.asp.

American Foundation for the Blind. (2005). *Building assessment initiatives for schools: An accommodations guide for parents and students with visual impairments*. New York, NY: Author. Retrieved from http://www.afb.org/section.aspx?SectionID=58&TopicID=264&DocumentID=2762

American Foundation for the Blind. (2012a). *Interpreting Bureau of Labor Statistics employment data*. New York, NY: Author. Retrieved from http://www.afb.org/section.aspx?SectionID=15&SubTopicID=177

American Foundation for the Blind. (2012b). *Refreshable braille displays*. New York, NY: Author. Retrieved from http://www.afb.org/ProdBrowseCatResults.asp?CatID=43

Arlene R. Gordon Research Institute. (2012). *Computer use and internet access.* New York, NY: Author. Retrieved from http://www.lighthouse.org/research/statistics-on-vision-impairment/computeruse-internet-access/

Armstrong, H. L., & Murray, I. D. (2010). Adapting advanced information technology network training for adults with visual impairments. *Journal of Visual Impairment & Blindness*, *104*, 504–509.

Assessment Services Supporting ELs through Technology Systems. (2011). *Enhanced Assessment Instrument Grant (EAG) for a new English language proficiency test*. Wisconsin Department of Public Instruction. Retrieved from http://assets.wceruw.org/resources/assets-grant-proposal.pdf

Assessment Services Supporting ELs through Technology Systems. (2013). *The WIDA consortium English language proficiency assessment framework: annual summative and on-demand screener.* Washington, D.C.: Center for Applied Linguistics. Retrieved from https://www.wida.us/get.aspx?id=832

Bechard, S., Sheinker, J., Abell, R., Barton, K., Burling, K., Camacho, C., . . . Tucker, B. (2010). Measuring cognition of students with disabilities using technology-enabled assessments: Recommendations for a national research agenda. *Journal of Technology, Learning, and Assessment*, *10*(4).

Beck-Winchatz, B., & Riccobono, M. A. (2008). Advancing participation of blind students in science, technology, engineering, and math. *Advances in Space Research*, *42*(11), 1855-1858. doi: http://dx.doi.org/10.1016/j.asr.2007.05.080

Brodin, J. (2010). Can ICT give children with disabilities equal opportunities in school? *Improving Schools*, *13*, 99–112. doi: 10.1177/1365480209353483

Capella-McDonnall, M. E. (2005). Predictors of competitive employment for blind and visually impaired consumers of vocational rehabilitation services. *Journal of Visual Impairment & Blindness*, *99*, 303–315. Retrieved from http://www.eric.ed.gov/PDFS/EJ687986.pdf

Center for Applied Special Technology (2011). *Universal design for learning guidelines version 2.0.* Wakefield, MA: Author. Retrieved from http://www.udlcenter.org/aboutudl/udlguidelines

Center for Universal Design. (1997). *Principles of universal design.* Raleigh, NC: North Carolina State University. Retrieved from http://www.ncsu.edu/ncsu/design/cud/about_ud/udprinciplestext.htm

Cho, H. J., Lee, J., & Kingston, N. M. (in press). Examining the effectiveness of test accommodation using DIF and a mixture IRT model. *Applied Measurement in Education*.

Christensen, L. L., Braam, M., Scullin, S., & Thurlow, M. L. (2011). *2009 state policies on assessment participation and accommodations for students with disabilities* (Synthesis Report 83). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from http://www.cehd.umn.edu/NCEO/onlinepubs/Synthesis83/default.htm

Common Core State Standards Initiative. (2012). Retrieved from http://www.corestandards.org/

Cormier, D. C., Altman, J. R., Shyyan, V., & Thurlow, M. L. (2010). *A summary of the research on the effects of test accommodations: 2007-2008* (Technical Report 56). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from http://www.cehd.umn.edu/NCEO/onlinepubs/Tech56/default.htm

Dolan, R. P., Burling, K. S., Rose, D., Beck, R. Murray, E., Strangman, N.,...Strain-Seymour, E. (2010). *Universal design for computer-based testing (UD-CBT) guidelines*. San Antonio: Pearson.

Dynamic Learning Maps Consortium (2014). *Dynamic Learning Maps Essential Elements for mathematics (version 4).* Lawrence, KS: University of Kansas. Retrieved from http://dynamiclearningmaps.org/sites/default/files/documents/Math_EEs/DLM%20Essential%20Elements%20Math%20%282013%29%20v4.pdf

Dynamic Learning Maps. (2015). *Test development.* Retrieved from http://dynamiclearningmaps.org/content/test-development

English Language Proficiency Assessment for the 21st Century. (2014). *Assessment system.* Retrieved from http://www.elpa21.org/assessment-system

English Language Proficiency Assessment for the 21st Century. (2015). *Features & benefits.* Retrieved from http://www.elpa21.org/assessment-system/features-benefits

Fuchs, L. S., & Fuchs, D. (2001). Helping teachers formulate sound test accommodation decisions for students with learning disabilities. *Learning Disabilities Research & Practice*, *16*, 174–181.

Fujiura, G. T., & Yamaki, K. (2000). Trends in demography of childhood poverty and disability. *Exceptional Children*, *66*, 187–199.

Hansen, E. G., Lee, M. J., & Forer, D. C. (2002). A 'self-voicing' test for individuals with visual impairments. *Journal of Visual Impairment & Blindness*, *96*(4), 273.

Hansen, E. G., Shute, V. J., & Landau, S. (2010). An assessment-for-learning system in mathematics for individuals with visual impairments. *Journal of Visual Impairment & Blindness*, *104*(5), 275-286.

Hasty, L. (n.d.). *Teaching tactile graphics* [Webcast]. Retrieved from http://support.perkins.org/site/PageServer?pagename=Webcasts_Teaching_Tactile_Graphics

IMS Global Learning Consortium. (2012). *Accessible portable item protocol*. Retrieved from http://www.imsglobal.org/apip/index.html

Individuals with Disabilities Education Improvement Act, 20 U.S.C. § 1400 (2004).

Jodoin, M. G. (2003). Measurement efficiency of innovative item formats in computer-based testing. *Journal of Educational Measurement*, *40*, pp. 1-15.

Joint Committee on Standards for Educational Evaluation. (2010). *The program evaluation standards* (3rd ed). Thousand Oaks, CA: Sage.

Kamei-Hannan, C. (2008). Examining the accessibility of a computerized adapted test using assistive technology. *Journal of Visual Impairment & Blindness*, *102*(5), 261–271.

Kelly, S. M. (2011). The use of assistive technology by high school students with visual impairments: A second look at the current problem. *Journal of Visual Impairment & Blindness*, *105*(4), 235-239.

Kim, J. S. (2012). The effect of "read-aloud" as a test accommodation for students with visual impairments in South Korea. *Journal of Visual Impairment & Blindness, 106*, 356-361.

Kopriva, R. (2000). *Ensuring accuracy in testing for English Language Learners*. Washington, DC: Council of Chief State School Officers.

Krufka, S. E., & Barner, K. E. (2006). A user study on tactile graphic generation methods. *Behaviour & Information Technology, 25*(4), 297-311. doi: 10.1080/01449290600636694

Landau, S., Russell, M., & Erin, J. N. (2006). The use of the Talking Tactile Tablet as a testing accommodation. *RE:View: Rehabilitation and Education for Blindness and Visual Impairment, 38*(1), 7–21.

Learning, Sight & Sound Made Easier. (2011). *Talking calculators*. Retrieved from http://www.lssproducts.com/category/calculators

Lebaz, S., & Picard, D. (2012). Identifying raised-line drawings by touch: a hard but not impossible task. *Journal of Visual Impairment & Blindness, 106*, 427-431.

Lenth, R. V. (2006-09). *Java applets for power and sample size* [Computer software]. Retrieved from http://www.stat.uiowa.edu/~rlenth/Power.

Leonard, R. (2002). *Statistics on vision impairment: A resource manual* (5th ed.). New York, NY: Arlene R. Gordon Research Institute of Lighthouse International. Retrieved from http://www.gesta.org/estudos/statistics0402.pdf

Minnesota Department of Education. (2009). *Minnesota manual of accommodations for students with disabilities in instruction and assessment*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from http://www.cehd.umn.edu/nceo/OnlinePubs/MNmanuals/MNmanual.pdf

Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice, 25*, 6–20. Retrieved from http://www.em2007.mpg.de/files/MislevyHaertel.pdf

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). *A brief introduction to evidence-centered design* (Research report 03-16). Princeton, NJ: Educational Testing Service.

National Center and State Collaborative. (2015). *Developing a large-scale assessment using components of evidence-centered design: Did it work?* Retrieved from http://www.ncscpartners.org/resources

National Center on Educational Outcomes. (2011). *Don't forget accommodations! Five questions to ask when moving to technology-based assessments* (NCEO brief #1). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from http://www.cehd.umn.edu/NCEO/onlinepubs/briefs/brief01/default.htm

National Science Foundation. (1996). *Women, minorities, and persons with disabilities in science and engineering* (NSF 96-311). Arlington, VA: Author. Retrieved from http://www.nsf.gov/statistics/nsf96311/2student.htm

New York State Education Department. (2006). *Test access & accommodations for students with disabilities*. Retrieved from http://www.p12.nysed.gov/specialed/publications/policy/testaccess/guidance.htm

Oregon Department of Education. (2012). *Online test administration.* Retrieved from http://www.ode.state.or.us/search/page/?id=3370

Parette, H. P., Peterson-Karlan, G. R., Smith, S., Gray, T., & Silver-Pacuilla, H. (2006). The state of assistive technology: Themes from an outcomes summit. *Assistive Technology Outcomes and Benefits, 3*(1), 15-33.

Parshall, C. G., Davey, T., & Pashley, P. J. (2000). Innovative item types for computerized testing. In W. J. van der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and Practice*, pp. 129-148.

Parshall, C. G.; Harmes, J. C. (2009). Improving the quality of innovative item types: Four tasks for design and development. *Journal of Applied Testing Technology*, *10*, 20-41.

Partnership for Assessment of Readiness for College and Careers. (2013). *Item guidelines for ELA/Literacy PARCC summative assessment*. Retrieved from http://www.parcconline.org/sites/parcc/files/Item%20Guidelines%20%2004%2025%202013%20Version%208_0.pdf

Partnership for Assessment of Readiness for College and Careers. (2014). *PARCC accessibility features and accommodations manual* (3rd ed.). Retrieved from http://www.parcconline.org/sites/parcc/files/parcc-accessibility-features-accommodations-manual-11-14_final.pdf

Partnership for Assessment of Readiness for College and Careers. (2015). *Accessibility accommodations and fairness.* Retrieved from http://www.parcconline.org/accessibility-accommodations-and-fairness

Rosano, A., Mancini, F., & Solipaca, A. (2009). Poverty in people with disabilities: Indicators from the capability approach. *Social Indicators Research*, *94*, 75–82.

Rosenblum, P. L., & Herzberg, T. (2011). Accuracy and techniques in the preparation of mathematics worksheets for tactile learners. *Journal of Visual Impairment & Blindness*, *105*(7), 402-413.

Russell, M., Mattson, D., Higgins, J., Hoffmann, T., Bebell, D., & Alcaya, C. (2011). *A primer to the Accessible Portable Item Profile (APIP) standards.* Minnesota Department of Education. Retrieved from http://apipstandard.org/archive/papers/APIP%20Primer%20-%20Final.pdf

Sireci, S. G., Scarpati, S. E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research*, *4*(75), 457–490.

Smarter Balanced Assessment Consortium. (2012). *Introduction to Smarter Balanced item and performance task development* [PowerPoint slides]. Retrieved from http://www.smarterbalanced.org/smarter-balanced-assessments/item-writing-and-review/

Smarter Balanced Assessment Consortium. (2012). *Smarter Balanced Assessment Consortium: General item specifications.* Measured Progress/ETS Collaborative. Retrieved from http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/05/TaskItemSpecifications/ItemSpecifications/GeneralItemSpecifications.pdf

Smarter Balanced Assessment Consortium. (2012). *Smarter Balanced Assessment Consortium: General accessibility guidelines.* Measured Progress/ETS Collaborative. Retrieved from http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/05/TaskItemSpecifications/Guidelines/AccessibilityandAccommodations/GeneralAccessibilityGuidelines.pdf

Smith, D. W., & Smothers, S. M. (2012). The role and characteristics of tactile graphics in secondary mathematics and science textbooks in braille. *Journal of Visual Impairment & Blindness*, *106*, 543-554.

Stone, E., Cook, L., Laitusis, C. C., & Cline, F. (2010). Using differential item functioning to investigate the impact of testing accommodations on an English-language arts assessment for students who are blind or visually impaired. *Applied Measurement in Education*, *23*, 132-152.

Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, *27*, 361–370.

Tennessee 3-5 Grade Band Training (ND). *Introduction to Webb's depth of knowledge levels*. Retrieved from http://archive.jc-schools.net/dynamic/math/webbs-depth.pdf

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.html

Thompson, S., Thurlow, M., & Moore, M. (2003). *Using computer-based tests with students with disabilities* (Policy Directions No. 15). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from http://education.umn.edu/NCEO/OnlinePubs/Policy15.htm

Thurlow, M. L., Laitusis, C. C., Dillon, D. R., Cook, L. L., Moen, R. E., Abedi, J., & O'Brien, D. G. (2009). *Accessibility principles for reading assessments*. Minneapolis, MN: National Accessible Reading Assessment Projects. Retrieved from http://www.narap.info/publications/reports/NARAPprinciples.pdf

Uslan, M. M. (1992). Barriers to acquiring assistive technology: Cost and lack of information. *Journal of Visual Impairment & Blindness, 86*, 402–407.

Watts, E. H., O'Brian, M., & Wojcik, B. W. (2004). Four models of assistive technology consideration: How do they compare to recommended educational assessment practices? *Journal of Special Education Technology, 19*, 43-56.

Wells-Moreaux, S., Bechard, S., & Karvonen, M. (2015). *Accessibility manual for the Dynamic Learning Maps Alternate Assessment 2014-15.* Lawrence, KS: University of Kansas Center for Educational Testing and Evaluation. Retrieved from http://dynamiclearningmaps.org/sites/default/files/accessiblity_manual_0.pdf

Wendt, A., Kenny, L. E., & Marks, C. (2007). Assessing critical thinking using a talk-aloud protocol. *CLEAR Exam Review, 18,* 18-27.

Wright, S. (2008). *Guide to designing tactile illustrations for children's books*. Louisville, KY: American Printing House for the Blind.

Zebehazy, K. T., Zigmond, N., & Zimmerman, G. J. (2012). Ability or access-ability: differential item functioning of items on alternate performance-based assessment tests for students with visual impairments. *Journal of Visual Impairment & Blindness, 106*, 325-338.

Zwern, A. L., & Goodrich, G. L. (1996). *Virtual computer monitor for visually-impaired users*. Proceedings of the 1st European Conference on Disability, Virtual Reality, and Associated Technologies. Maidenhead, UK. Retrieved from http://www.icdvrat.reading.ac.uk/1996/papers/1996_03.pdf

Appendix A

Expert Review Template

Item Number	ATEA_3E_BkgdGr_Butterfly_General
Grade Level	3
Subject	ELA
Item Description	The student has to drag each word to the corresponding line that identifies the stage that needs a label.
Targeted Construct	Standard: RI.3.7
Target or Focal KSAs	Use words in a text to label an illustration to demonstrate understanding of the text
Additional KSAs related to the target KSA	The student must have grade level reading comprehension.
Ancillary KSAs unrelated to the target KSA	
Depth of Knowledge (DOK 1-4)	
Barriers to Access	
Perceptual processing	
Linguistic processing	
Cognitive Processing	
Motoric Processing	
Executive Processing	
Affective Processing	
Access Recommendations	
Perceptual processing	
Linguistic processing	
Cognitive Processing	
Motoric Processing	
Executive Processing	
Affective Processing	
Links to Classroom Instruction	

General Directions for Task Tryouts

- The student should use the equipment and materials that he or she uses during instruction, including manipulatives and tools.
- You may read questions aloud to the student and provide print or braille copies as needed. Screen readers may be used for navigation, but they will not be able to read the onscreen questions accurately.
- Please video (preferred) or audio record your interaction with the student. You will be engaged in asking questions and guiding the student, and the video or audio will help me understand the issues that the student and you experienced during the sample tasks. I will provide a location for electronic upload of your video or audio files.
- Please complete this feedback sheet for each student who participates in the task tryouts. You may complete this sheet after the task tryouts.
- Please feel free to give me any other feedback in addition to the questions included here. You may use additional pages or email me your comments.

Student Assent Sample Script

Today we are going to do some math and reading on the computer. We will be working on things that are similar to what you do at school.

As a part of what we do, I will be asking you questions and we will be videotaping what we do together. This will take up to an hour.

If you don't want to do this, you don't have to. You can stop at any time, and that will be all right. Do you want to take part in this project?

Do you have any questions?

Let's get started…

Task Tryout Sample Script

Use the following script as a guide to the questions you ask the student.

Prompts and questions may be used at any time, as often as needed.

The goal of the task tryouts is to learn how the student accesses the content and what cognitive demands are made of the student for these new kinds of tasks. It does not matter if the student is able to answer correctly. The sample tests are not scored.

Sequence	Examples of possible prompts
Presentation	• *Have you taken tests on a computer before?* • *What tools do you usually use when you take a test? (Examples: Braille print, tactile graphics, physical tools like math manipulatives, assistance from another student or adult)*
	• *Do you have any questions before we begin?* • *You can take a break at any time.* • *Let me know if you have any questions.*
During the tasks	• Items may be read and repeated as needed. • Thinking aloud or talking through the test question is encouraged. • Accommodations, supports, and tools may be offered both before and during test items to assist students to answer the questions.
	• *Do you understand the question?* • *Are you stuck on this one?* • *What would help you answer the question?* • *Thank you for working so hard!*
Conclusion	• *Did you have any difficulty answering this question?* • *What were you thinking while you worked on this question?* • *Did that seem easy or hard to you?* • *Did any part of the question confuse you?*

Thank you for your participation in these task tryouts! Our goal is to improve accessibility of computerized assessments for students with blindness, low vision, and motor disabilities.

Appendix C

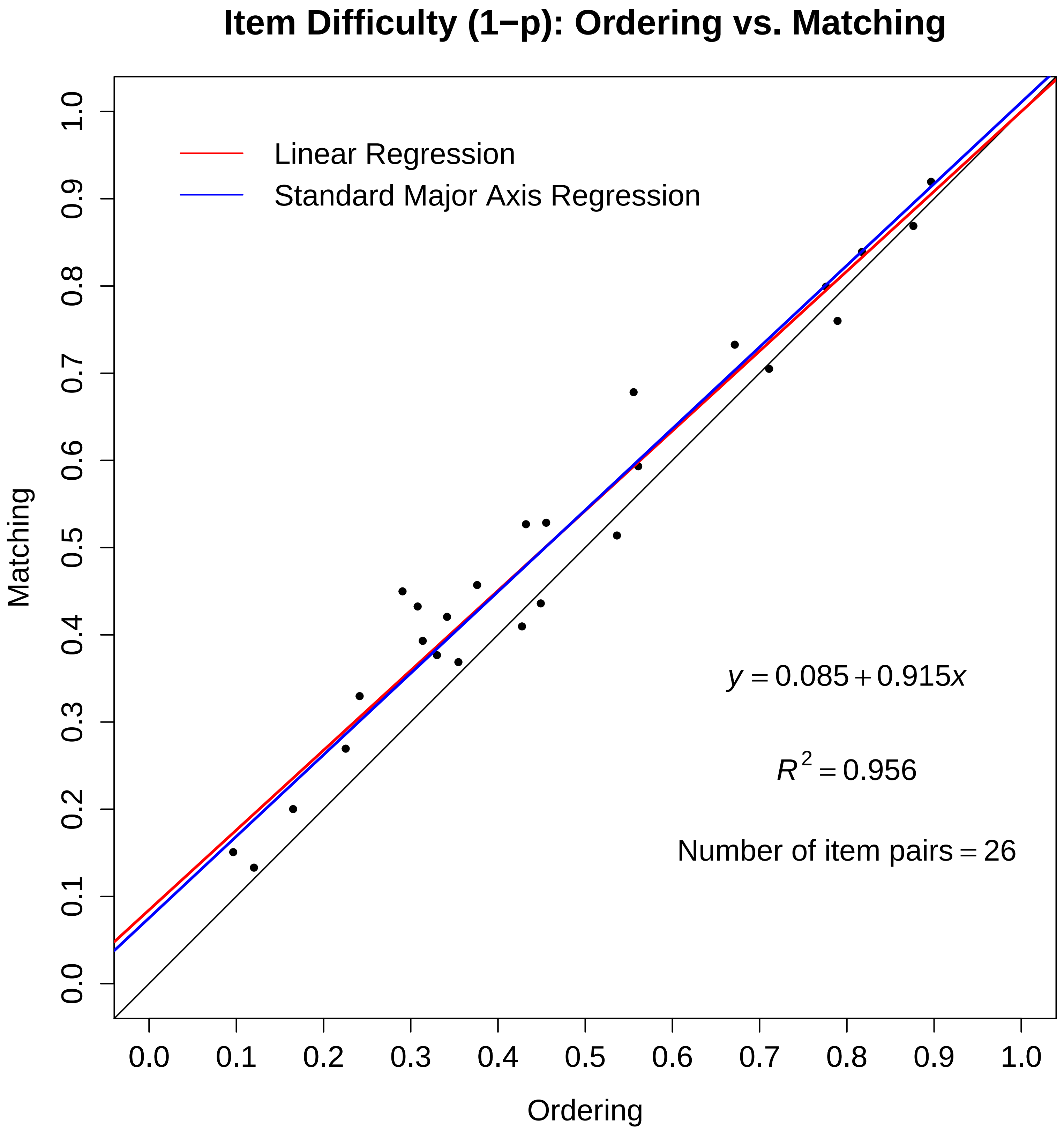
Results of Tests of Group Equivalence

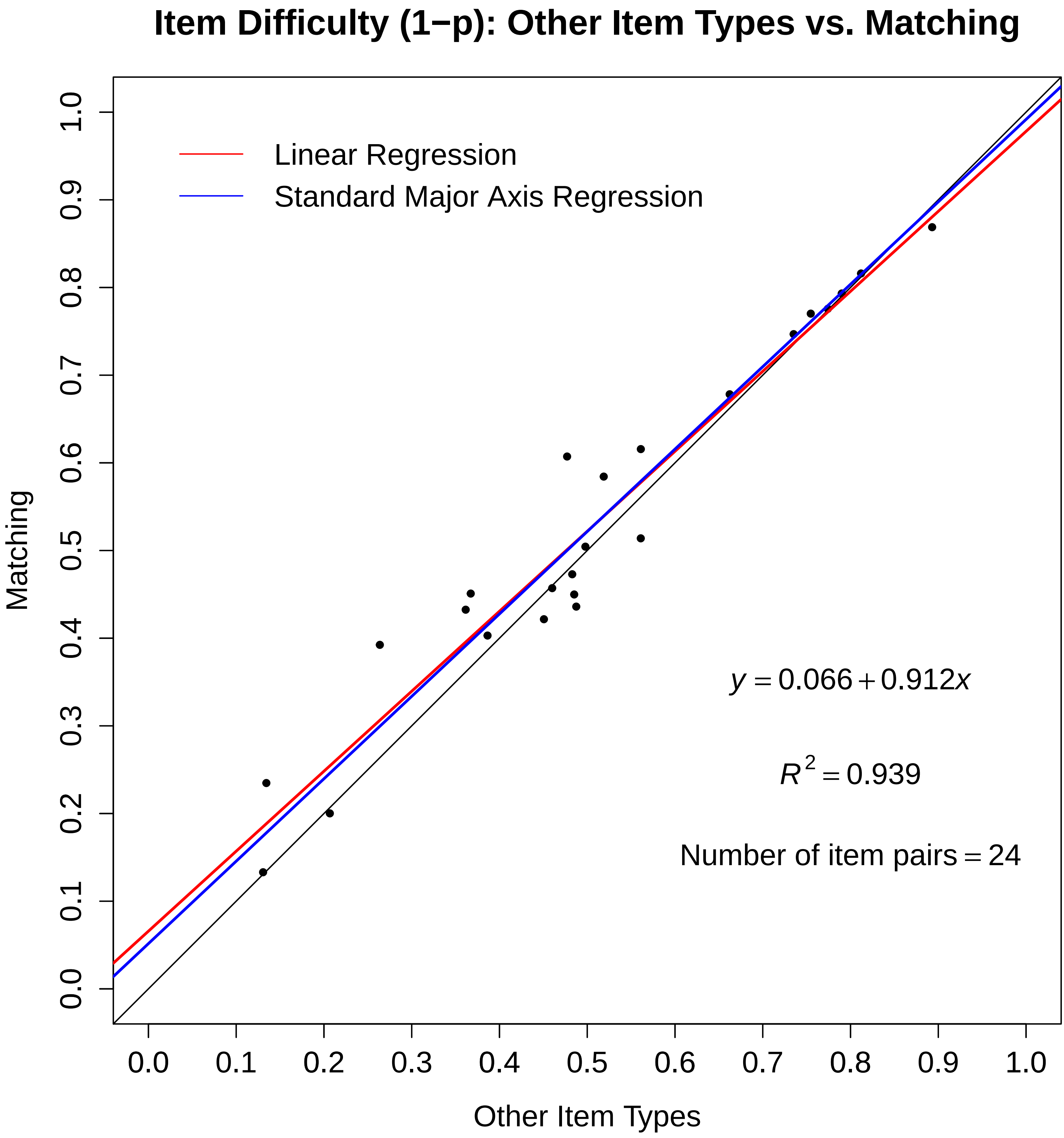
Grade/Subject Form	N	%	SWD	Chi-Square	Phi	ELL	Chi-Square	Phi	2014 total score	SD	Sig	Partial Eta squared	2013 N	2013 scaled score	SD	Sig	Partial Eta squared
Form A	3897	57.0	9.5			10.9			32.75	10.04							
Form B	2941	43.0	8.3			15.0			30.81	10.41							
Grade 3 ELA	6838		9.0	0.07	-0.02	12.6	0.00	0.06	31.92	10.24	0.00	0.01	NA	NA	NA	NA	NA
Form A	3714	57.8	8.3			14.1			34.86	9.00			3507	78.75	13.65		
Form B	2715	42.2	8.6			18.7			33.83	9.77			2580	77.73	14.40		
Grade 4 ELA	6429		8.4	0.62	0.01	16.1	0.00	0.06	34.43	9.34	0.00	0.00	6087	78.32	13.98	0.01	0.00
Form A	4074	64.0	7.8			15.6			42.64	8.72			3846	80.65	12.20		
Form B	2291	36.0	7.9			16.4			42.35	8.63			2197	80.67	12.07		
Grade 5 ELA	6365		7.8	0.91	0.00	15.9	0.43	0.01	42.53	8.69	0.20	0.00	6043	80.66	12.15	0.95	0.00
Form A	4018	55.3	7.3			10.5			38.70	9.47			3786	81.59	12.43		
Form B	3243	44.7	7.0			10.4			37.81	9.52			3085	80.66	12.65		
Grade 6 ELA	7261		7.2	0.54	-0.01	10.4	0.93	0.00	38.30	9.50	0.00	0.00	6871	81.17	12.54	0.00	0.00
Form A	3846	56.6	5.7			11.8			39.59	9.60			3631	78.40	13.61		
Form B	2945	43.4	6.2			13.8			39.19	9.46			2849	78.60	13.17		
Grade 7 ELA	6791		5.9	0.36	0.01	12.7	0.02	0.03	39.42	9.54	0.08	0.00	6480	78.48	13.42	0.36	0.00
Form A	3651	53.7	6.5			11.1			39.68	9.22			3458	79.56	12.31		
Form B	3146	46.3	5.9			13.3			39.43	9.47			3020	79.08	12.65		
Grade 8 ELA	6799		6.2	0.32	-0.01	12.1	0.01	0.03	39.57	9.34	0.27	0.00	6478	79.33	12.47	0.13	0.00
Form A	1938	42.5	7.1			5.2			47.33	11.85			360	70.55	15.02		
Form B	2620	57.5	7.4			5.1			47.08	11.96			418	66.99	16.68		
Grade 11 ELA	4558		7.3	0.68	0.01	5.1	0.95	0.00	47.18	11.91	0.47	0.00	778	68.64	16.02	0.00	0.01
Form A	7433	57.0	8.0			10.7			26.76	9.56							
Form B	5598	43.0	8.6			9.4			26.69	9.78							
Grade 3 Math	13031		8.3	0.16	0.01	10.1	0.01	-0.02	26.73	9.65	0.71	0.00	NA	NA	NA	NA	NA
Form A	6250	54.0	8.2			12.7			28.85	10.42			5906	82.64	14.41		
Form B	5334	46.0	8.3			14.7			28.79	10.51			5073	82.71	14.56		
Grade 4 Math	11584		8.3	0.90	0.00	13.6	0.00	0.03	28.82	10.46	0.77	0.00	10979	82.67	14.48	0.80	0.00
Form A	7744	54.9	7.1			9.9			27.92	11.21			7327	78.98	14.30		

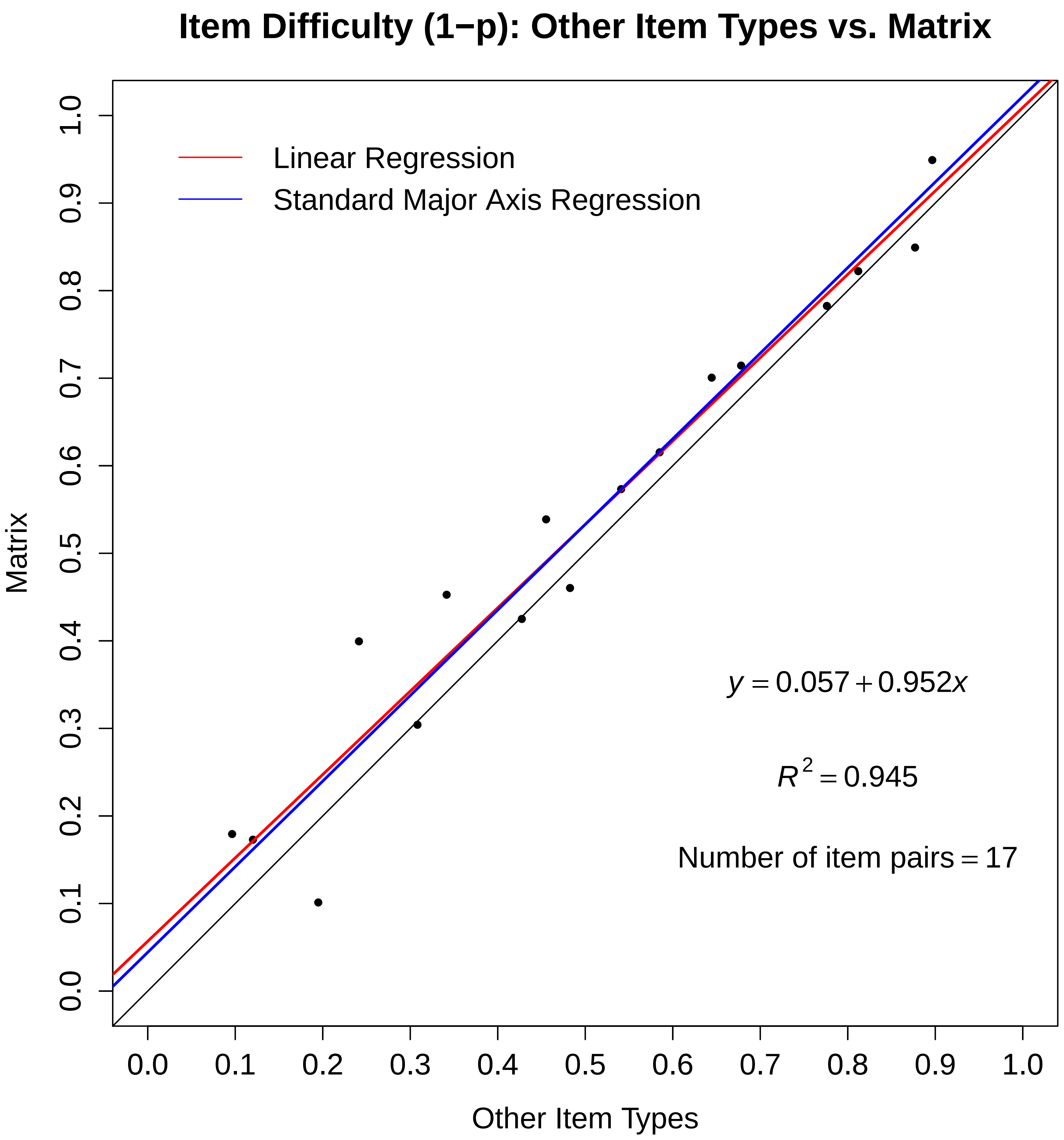
Grade/Subject Form	N	%	SWD	Chi-Square	Phi	ELL	Chi-Square	Phi	2014 total score	SD	Sig	Partial Eta squared	2013 N	2013 scaled score	SD	Sig	Partial Eta squared
Form B	6360	45.1	7.5			13.7			27.16	11.05			6018	78.66	14.17		
Grade 5 Math	14104		3.0	0.48	0.01	11.6	0.00	0.06	27.57	11.14	0.00	0.00	13345	78.84	14.24	0.20	0.00
Form A	6855	53.5	7.6			11.1			25.72	9.39			6464	76.87	14.72		
Form B	5966	46.5	6.9			10.8			25.75	9.27			5619	77.20	14.35		
Grade 6 Math	12821		7.3	0.11	-0.01	10.6	0.63	0.00	25.73	9.34	0.83	0.00	12083	77.02	14.55	0.22	0.00
Form A	5519	46.4	6.0			10.1			22.57	8.21			5267	77.30	15.78		
Form B	6363	53.6	5.7			8.2			22.81	8.36			6043	76.78	15.66		
Grade 7 Math	11882		5.8	0.48	-0.01	9.1	0.00	-0.03	22.70	8.29	0.11	0.00	11310	77.02	15.72	0.08	0.00
Form A	6160	51.2	5.1			8.5			21.69	8.19			5809	70.07	15.95		
Form B	5882	48.8	5.9			9.4			21.41	8.56			5557	68.05	16.35		
Grade 8 Math	12042		5.5	0.05	0.02	9.0	0.09	0.02	21.55	8.37	0.07	0.00	11366	69.08	16.18	0.00	0.00
Form A	4062	53.3	6.6			5.9			22.86	9.06			982	49.36	17.36		
Form B	3556	46.7	7.3			6.5			21.21	8.16			842	46.93	17.22		
Grade 11 Math	7618		6.9	0.30	0.01	6.2	0.31	0.01	22.09	8.69	0.00	0.01	1824	48.24	17.33	3.00	0.00

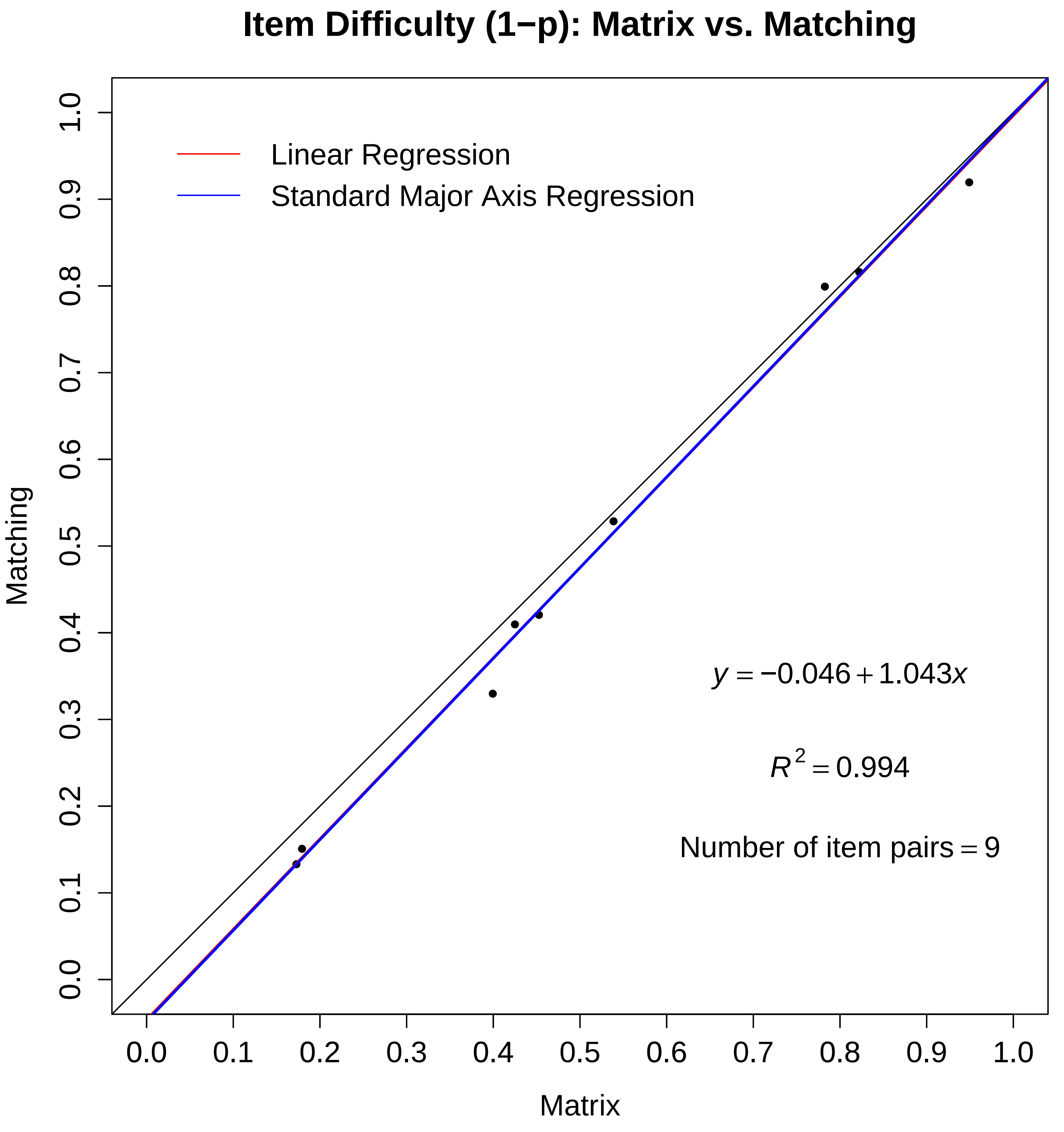
Appendix D

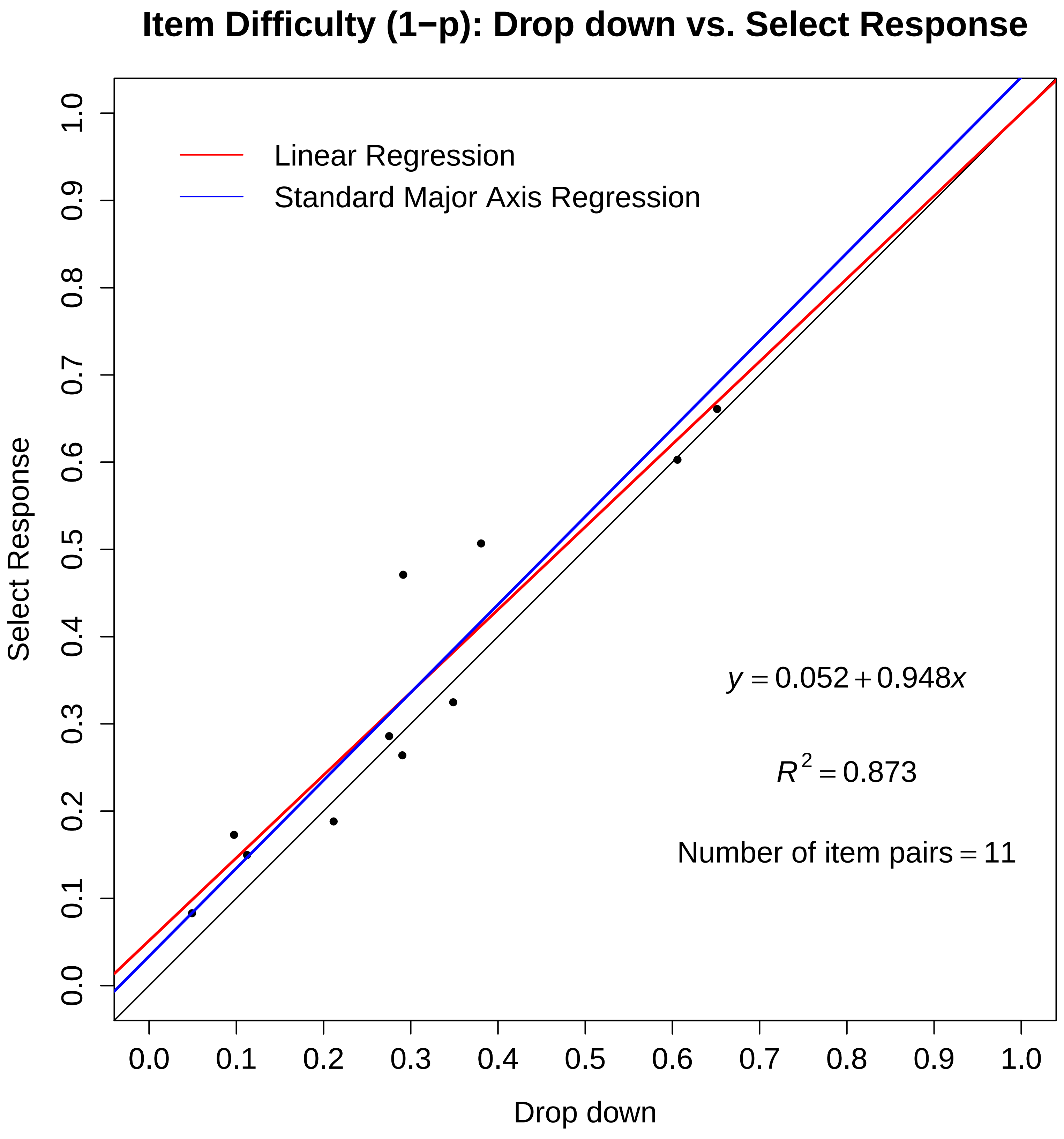
P-Value Item Difficulty

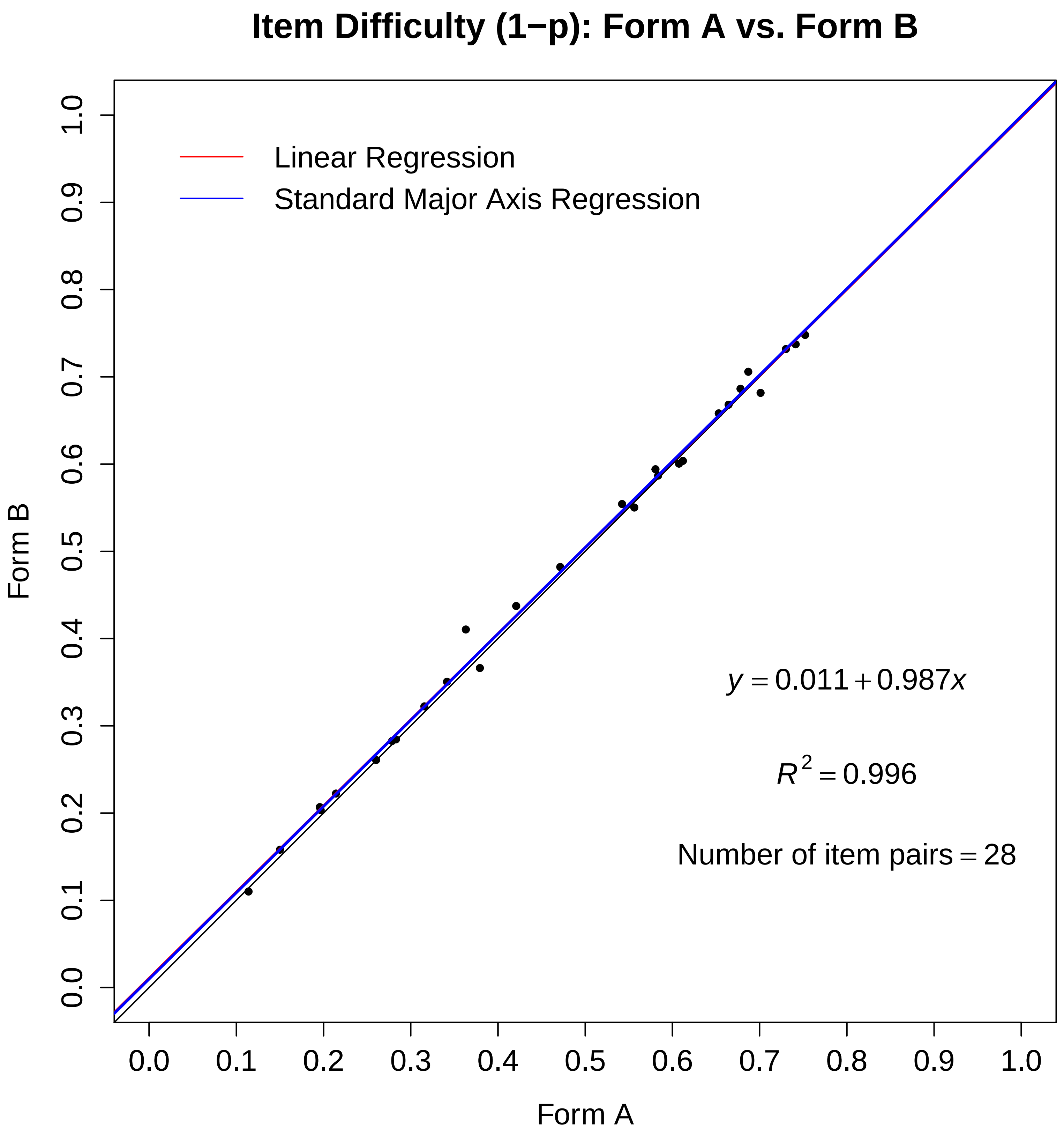



Item Difficulty (1−p): Other Item Types vs. Matching
Linear Regression
Standard Major Axis Regression
$y = 0.066 + 0.912x$
$R^2 = 0.939$
Number of item pairs = 24
Matching
Other Item Types
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0


Item Difficulty (1−p): Other Item Types vs. Matrix
Linear Regression
Standard Major Axis Regression
y = 0.057 + 0.952x
R2 = 0.945
Number of item pairs = 17
Matrix
Other Item Types
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0


Item Difficulty (1−p): Matrix vs. Matching
Linear Regression
Standard Major Axis Regression
y = −0.046 + 1.043x
R2 = 0.994
Number of item pairs = 9
Matching
Matrix
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

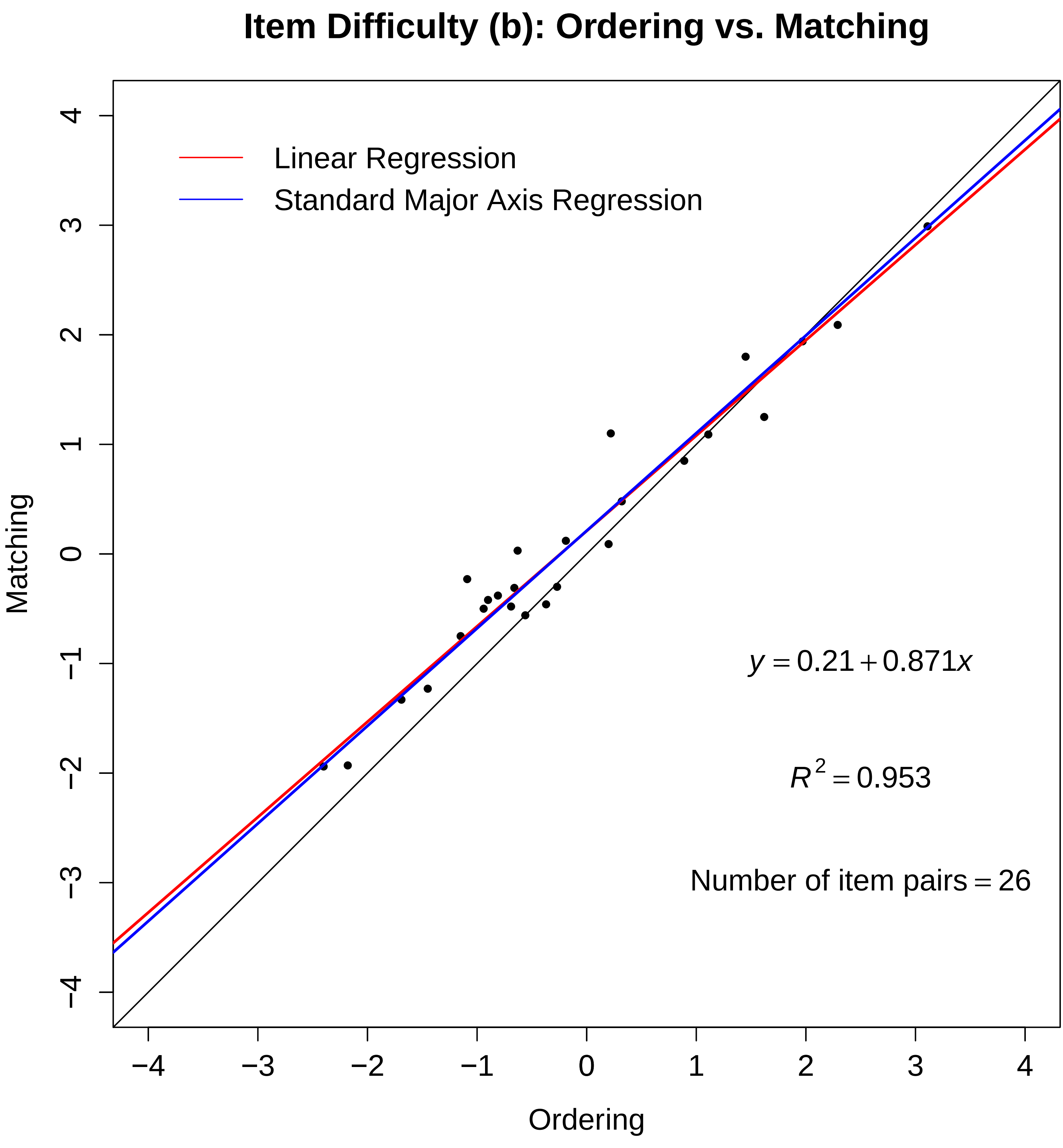

Item Difficulty (1−p): Drop down vs. Select Response
Linear Regression
Standard Major Axis Regression
y = 0.052 + 0.948x
R2 = 0.873
Number of item pairs = 11
Select Response
Drop down
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

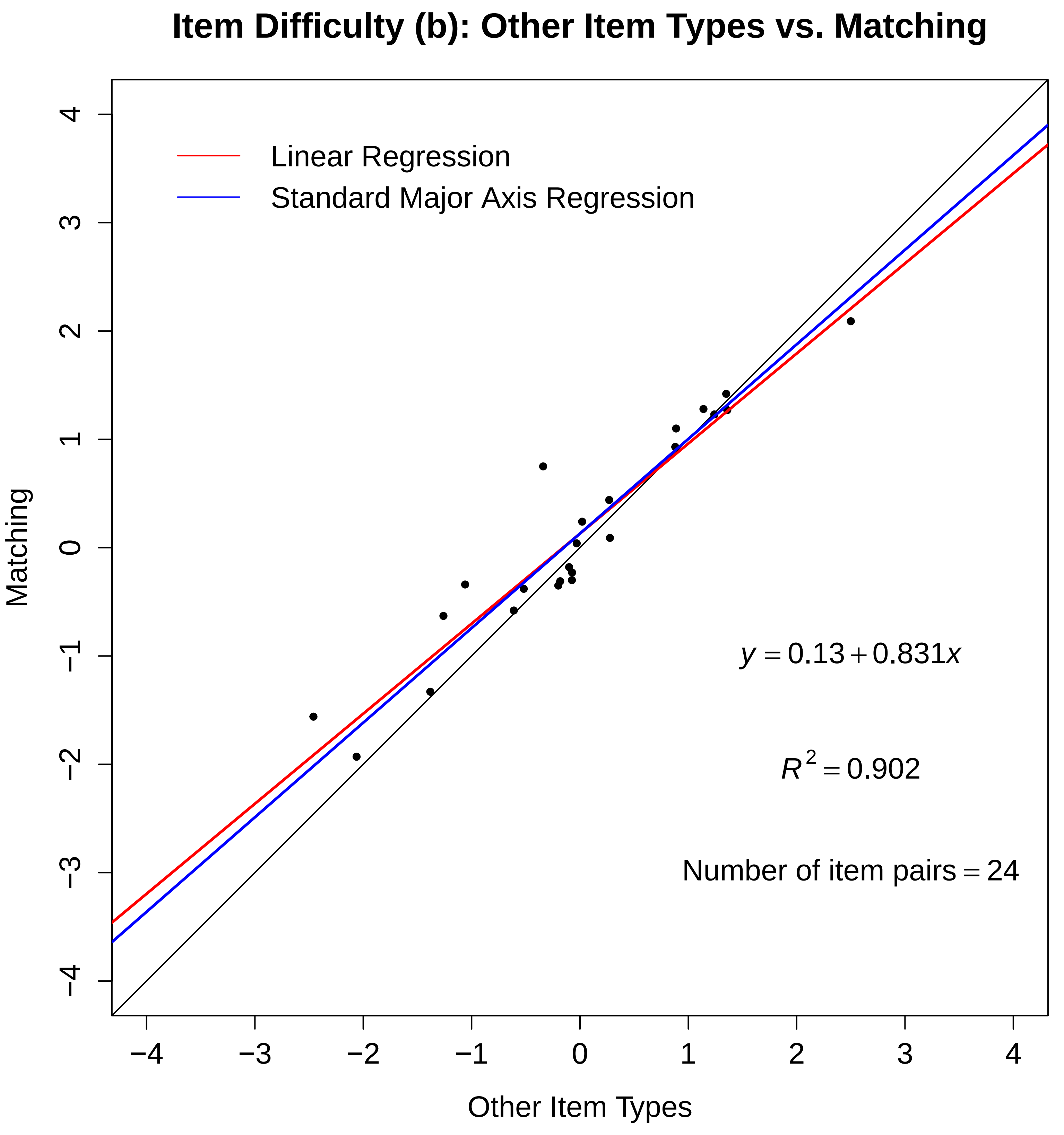


Item Difficulty (1−p): Form A vs. Form B
Linear Regression
Standard Major Axis Regression
$y = 0.011 + 0.987x$
$R^2 = 0.996$
Number of item pairs = 28
Form B
Form A
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

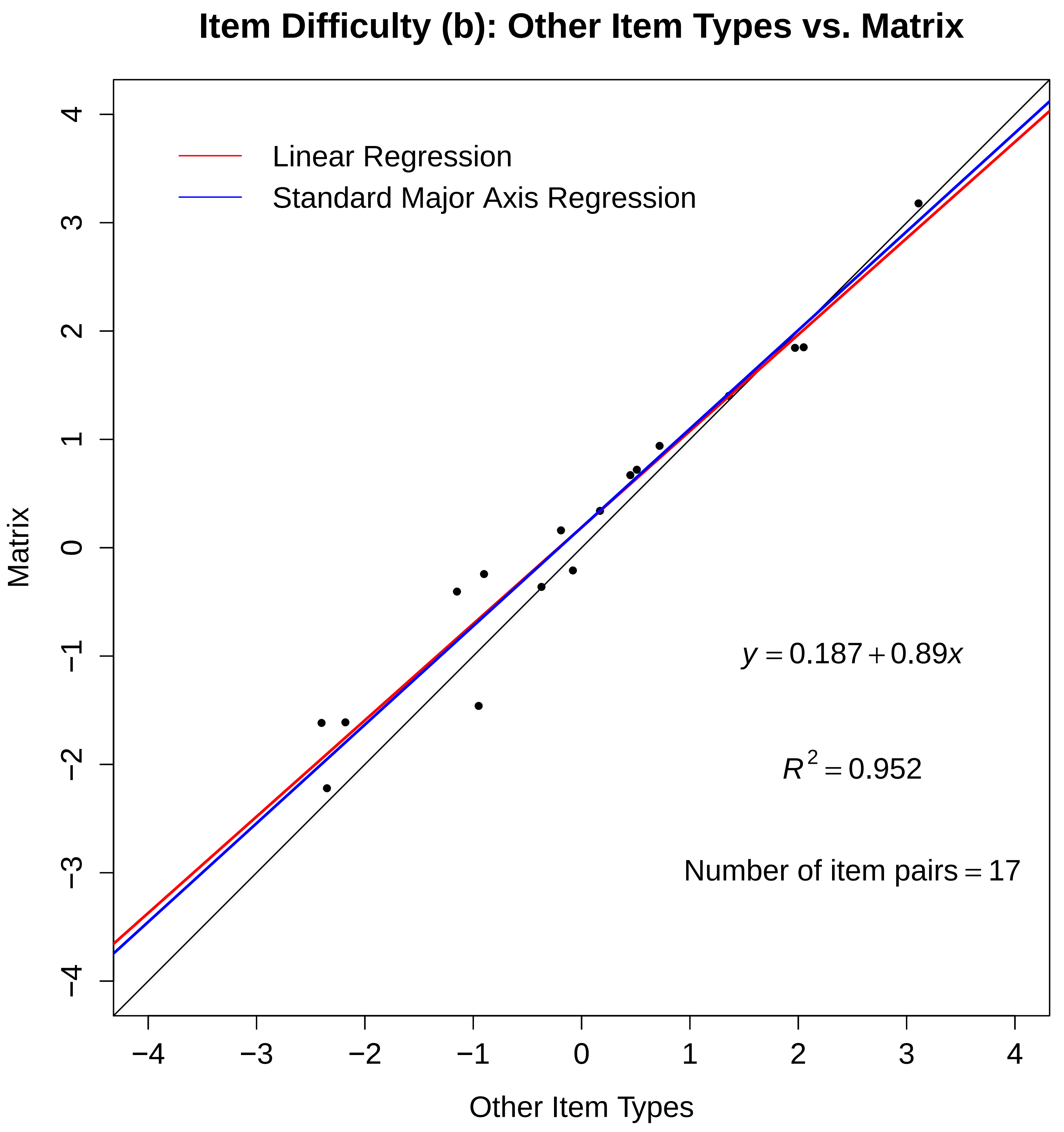
Appendix E

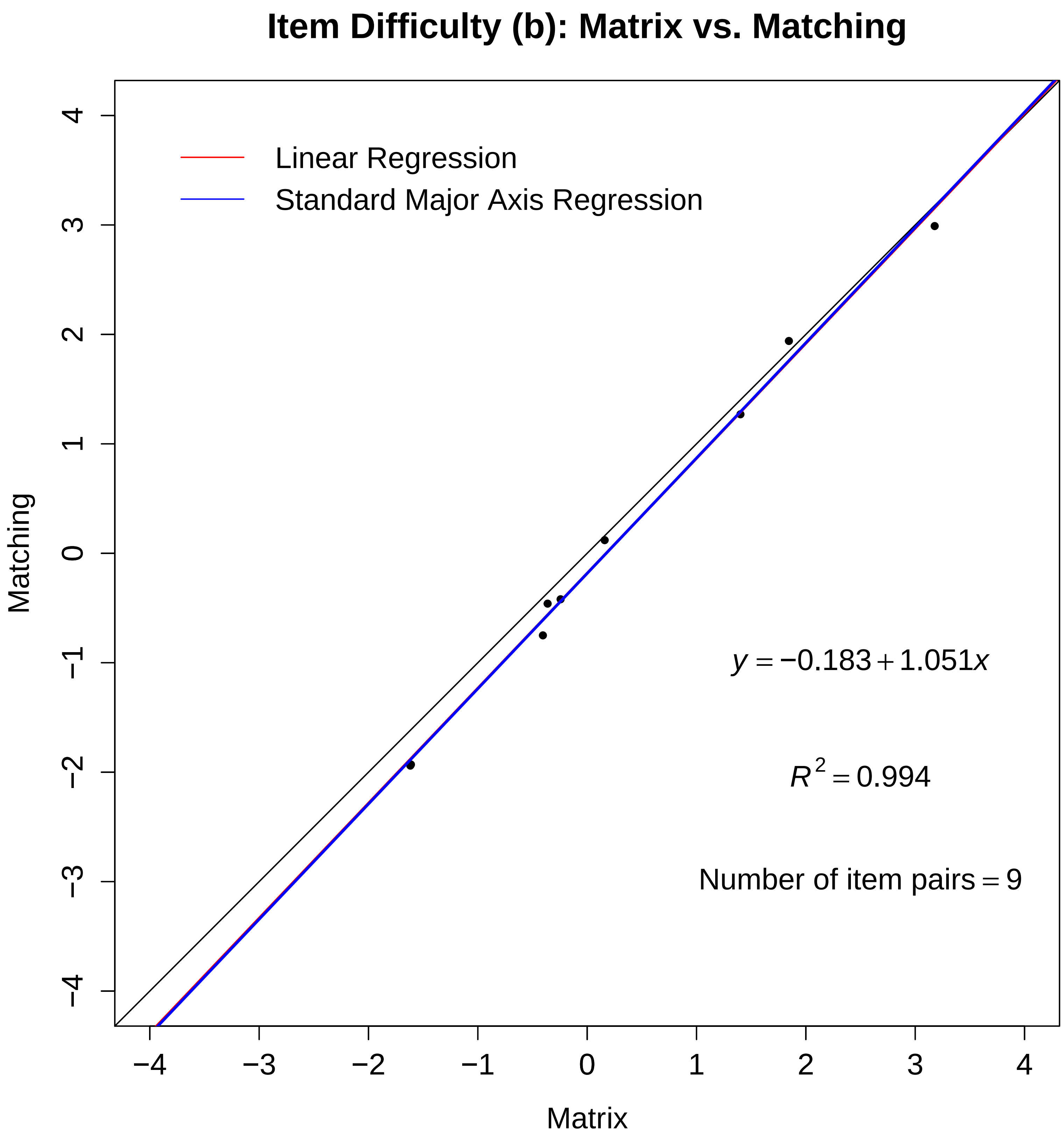
IRT Item Difficulty

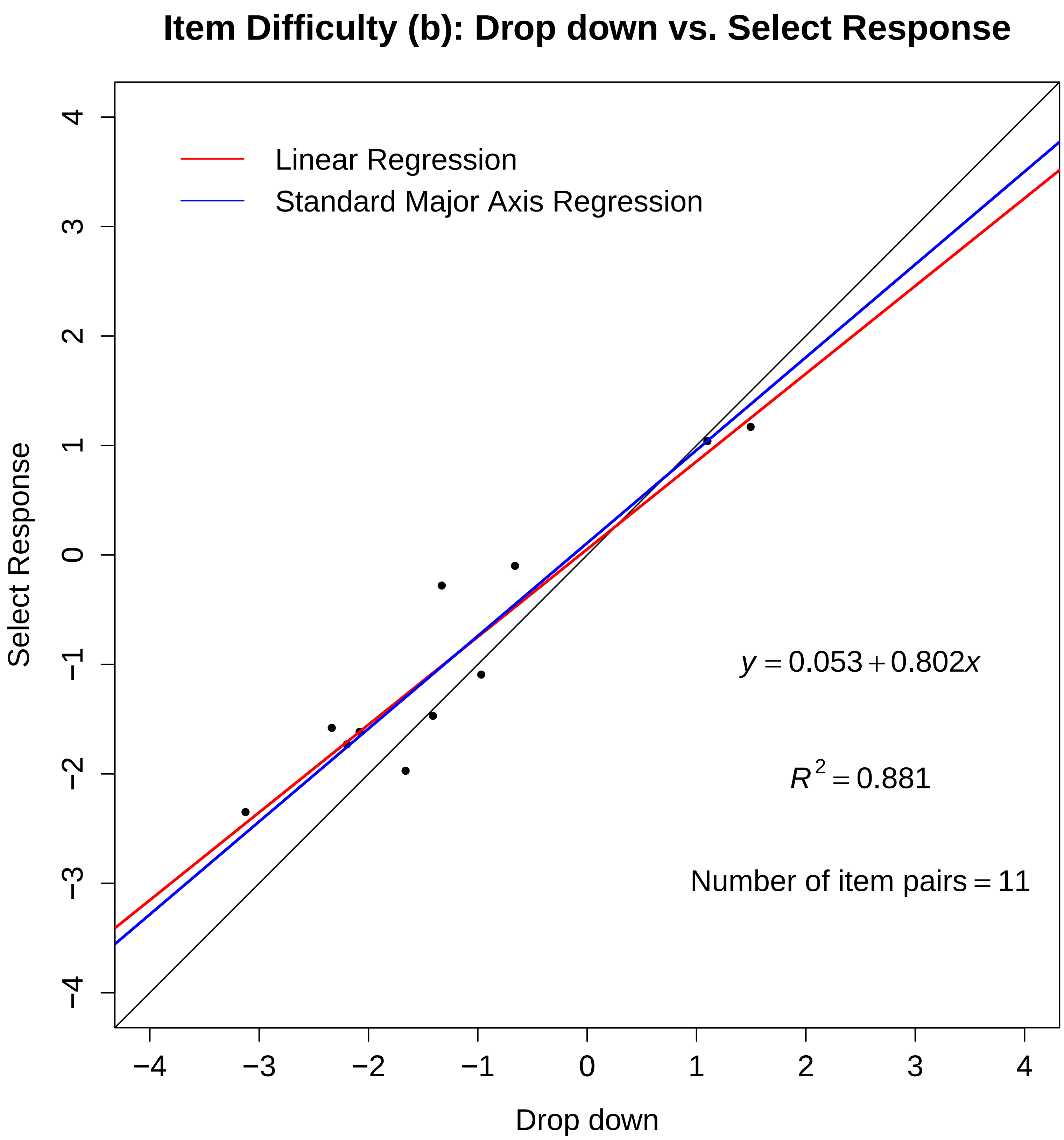




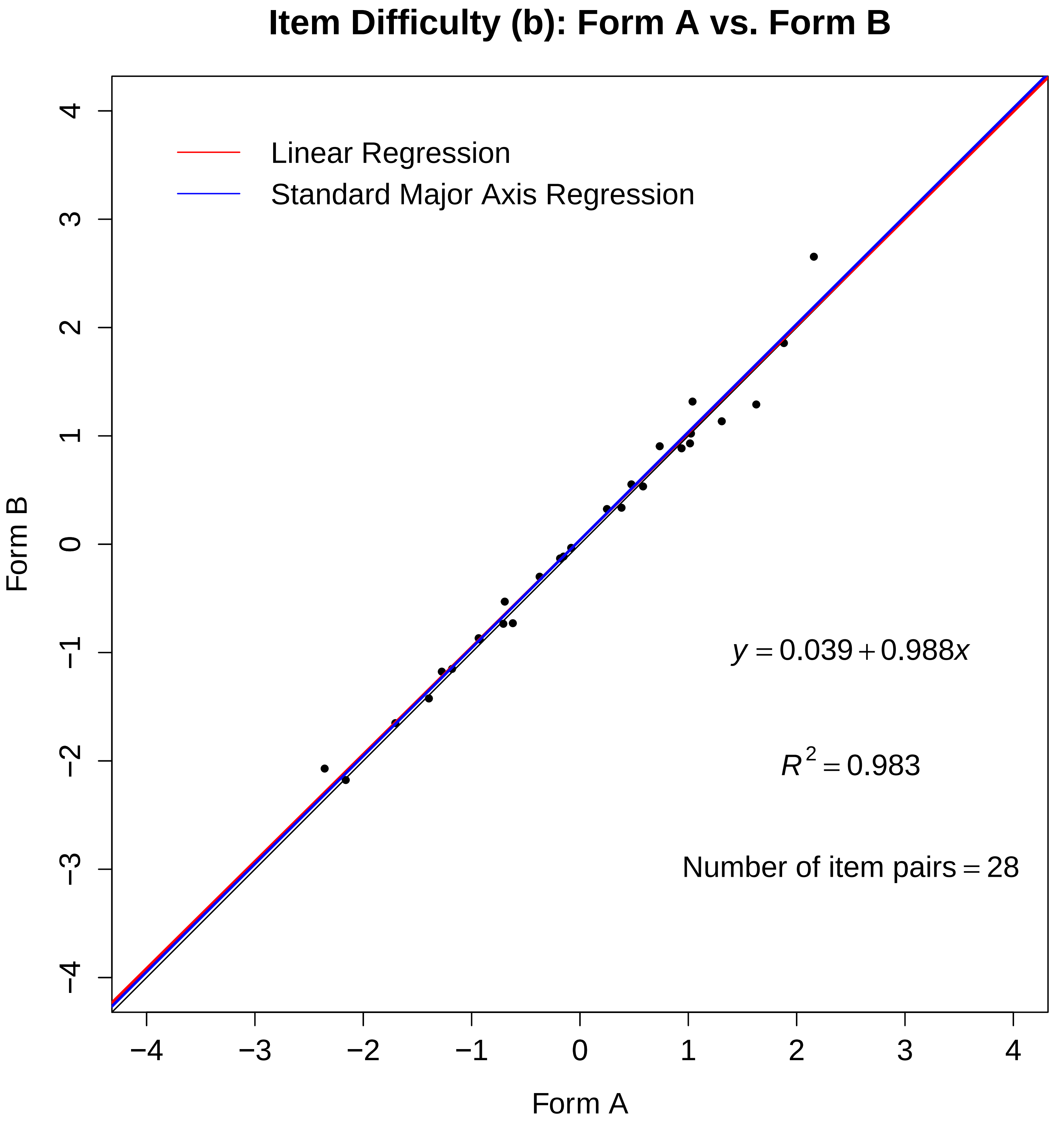
Item Difficulty (b): Other Item Types vs. Matching
Linear Regression
Standard Major Axis Regression
y = 0.13 + 0.831x
R^2 = 0.902
Number of item pairs = 24
Matching
Other Item Types
−4
−3
−2
−1
0
1
2
3
4


Item Difficulty (b): Other Item Types vs. Matrix
Linear Regression
Standard Major Axis Regression
y = 0.187 + 0.89x
R^2 = 0.952
Number of item pairs = 17
Matrix
Other Item Types
−4
−3
−2
−1
0
1
2
3
4


Item Difficulty (b): Matrix vs. Matching
Linear Regression
Standard Major Axis Regression
y = −0.183 + 1.051x
R² = 0.994
Number of item pairs = 9
Matching
Matrix
−4
−3
−2
−1
0
1
2
3
4



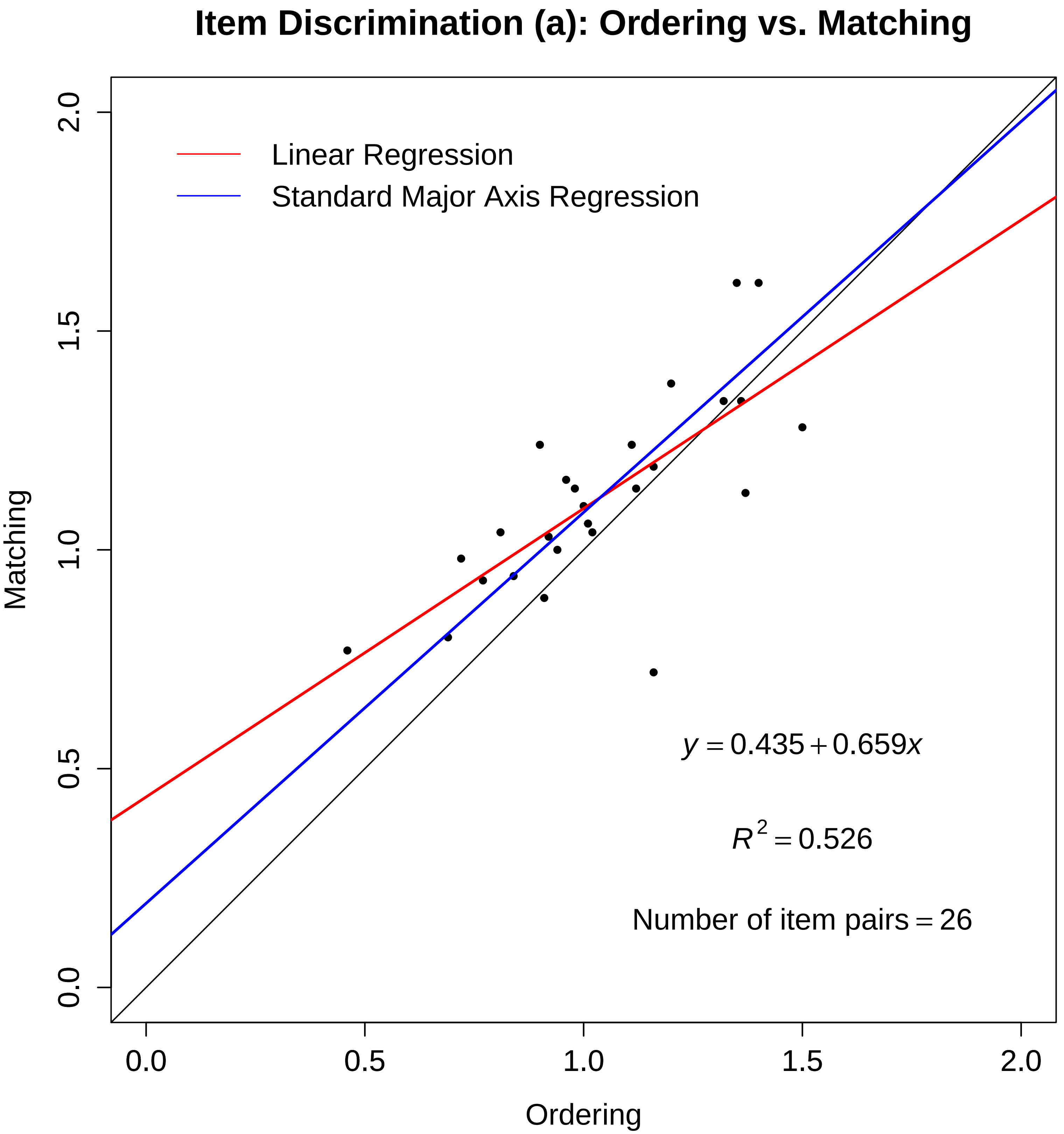
Item Difficulty (b): Drop down vs. Select Response
Linear Regression
Standard Major Axis Regression
$y = 0.053 + 0.802x$
$R^2 = 0.881$
Number of item pairs = 11
Select Response
Drop down
−4
−3
−2
−1
0
1
2
3
4

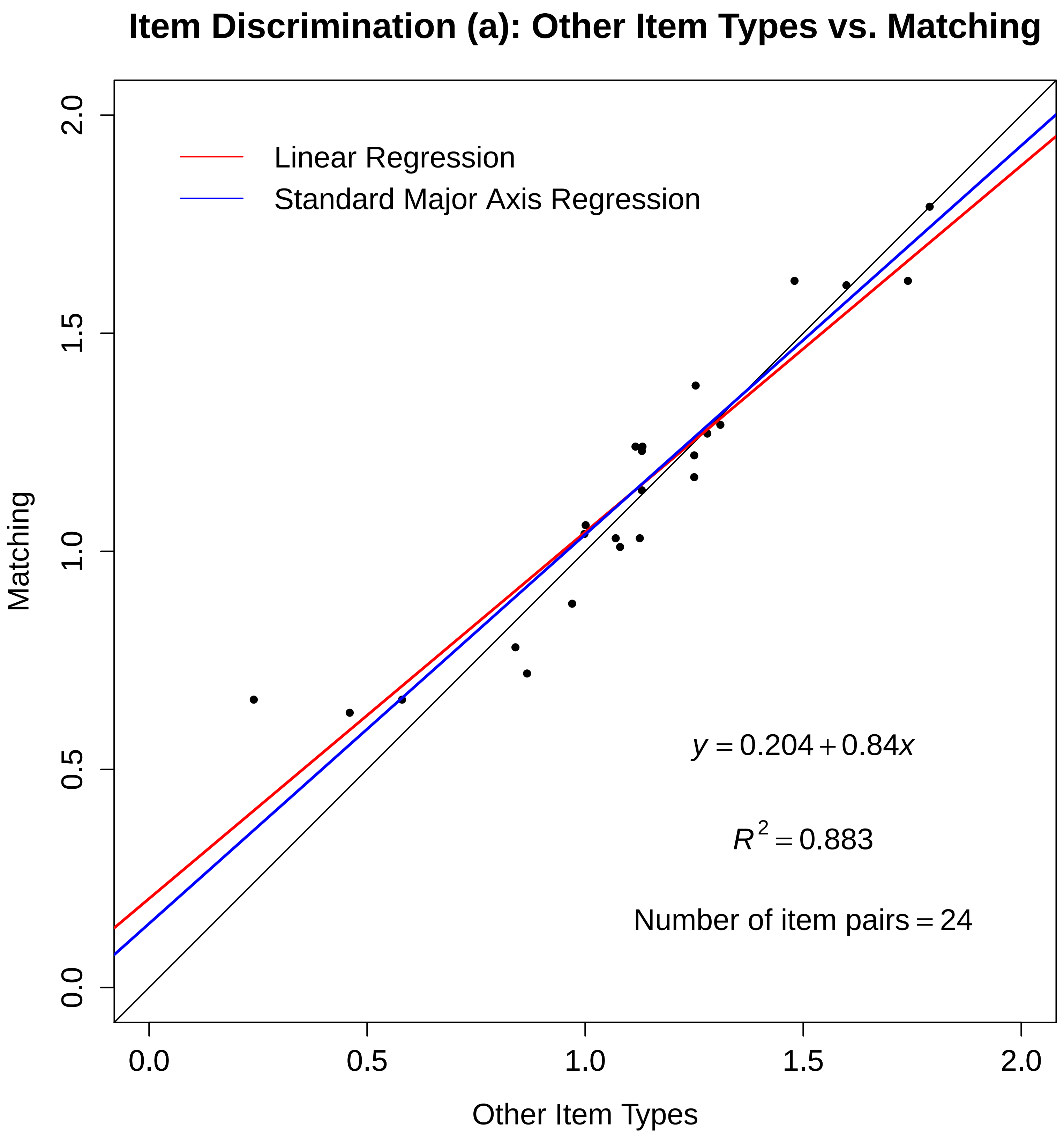


Item Difficulty (b): Form A vs. Form B
Linear Regression
Standard Major Axis Regression
$y = 0.039 + 0.988x$
$R^2 = 0.983$
Number of item pairs = 28
Form A
Form B
−4
−3
−2
−1
0
1
2
3
4

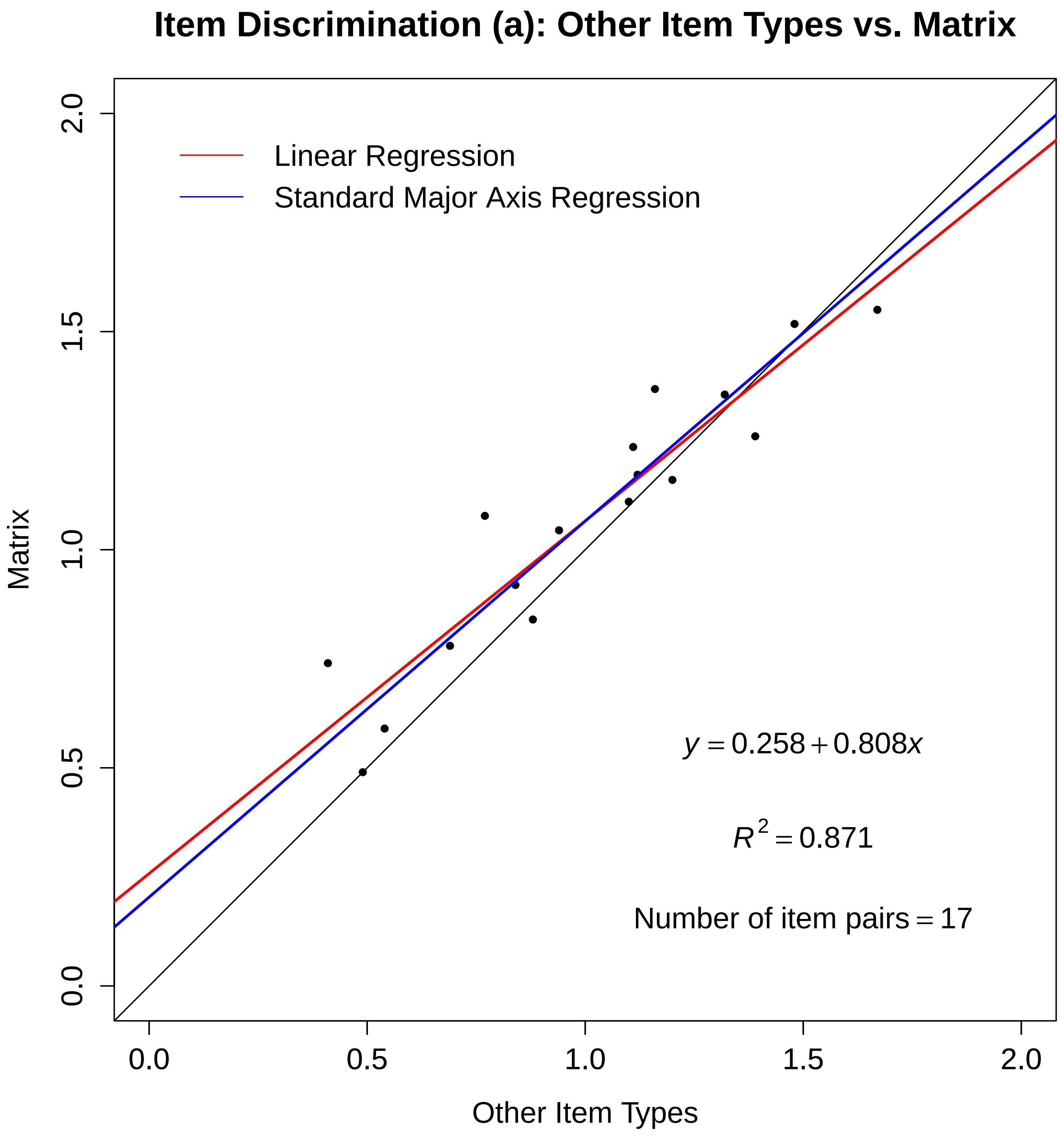
Appendix F

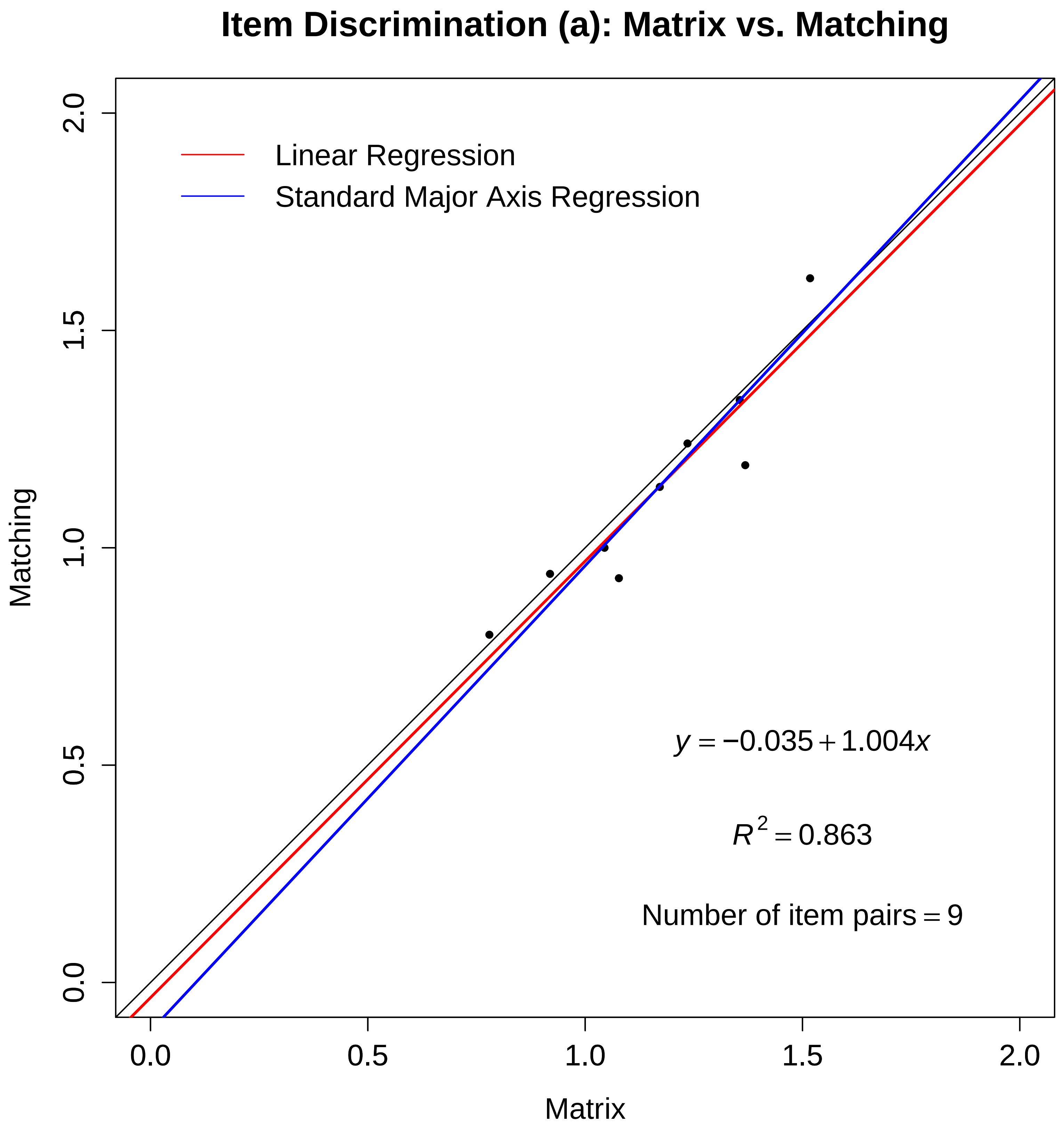
IRT Item Discrimination



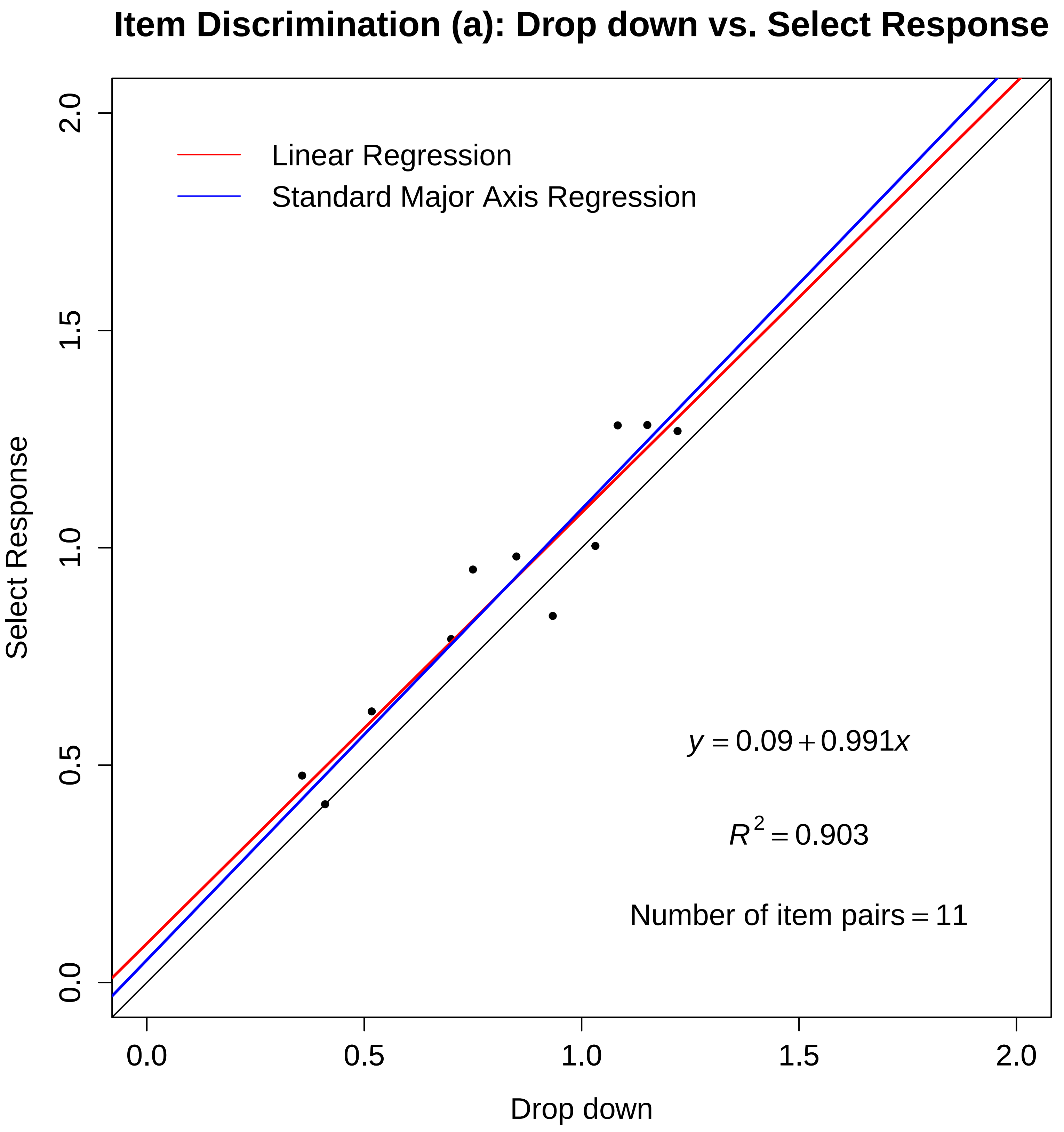


Item Discrimination (a): Other Item Types vs. Matching
Linear Regression
Standard Major Axis Regression
$y = 0.204 + 0.84x$
$R^2 = 0.883$
Number of item pairs = 24
Matching
Other Item Types
0.0
0.5
1.0
1.5
2.0

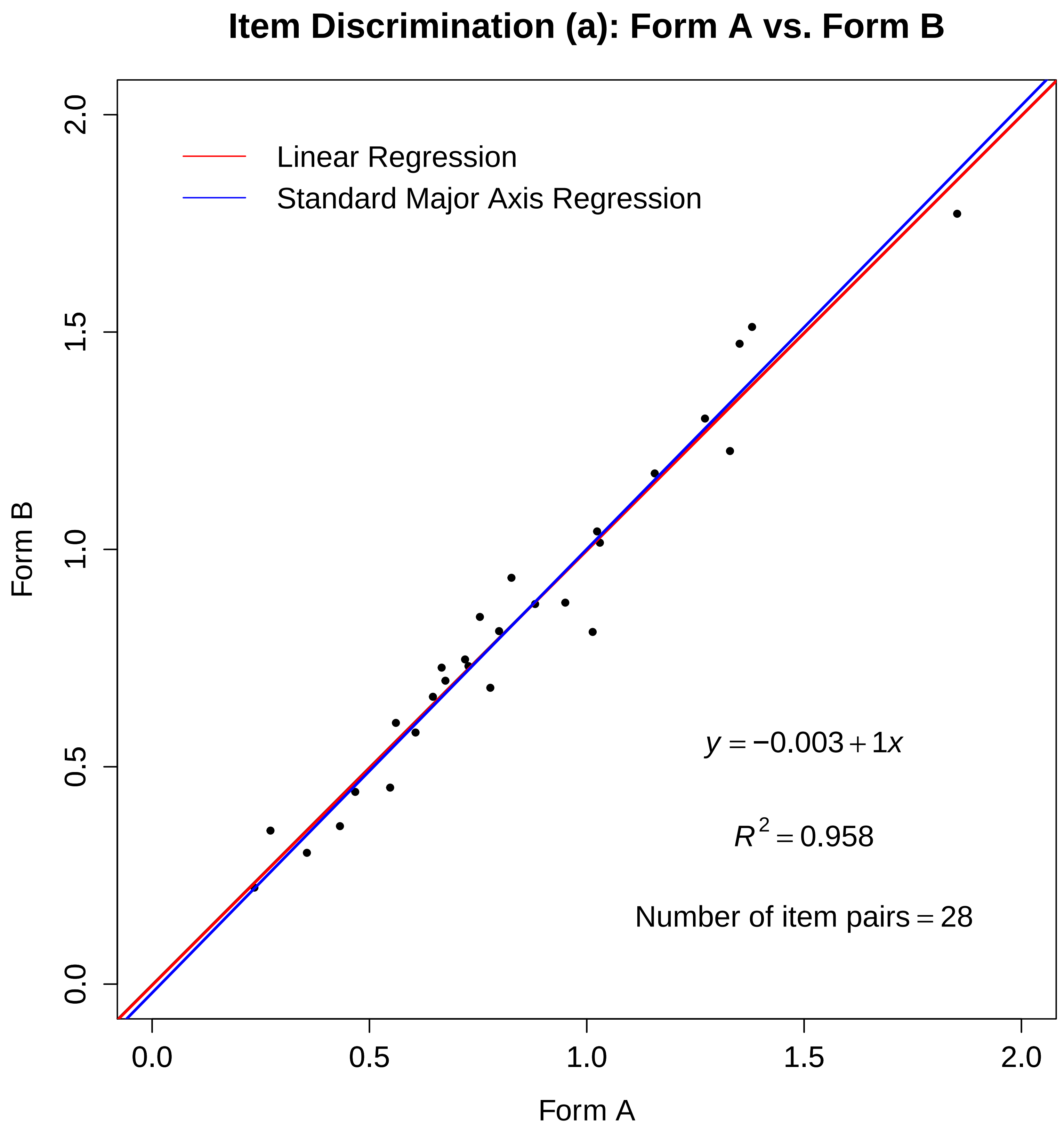

Item Discrimination (a): Other Item Types vs. Matrix
Linear Regression
Standard Major Axis Regression
$y = 0.258 + 0.808x$
$R^2 = 0.871$
Number of item pairs = 17
Matrix
0.0
0.5
1.0
1.5
2.0
Other Item Types
0.0
0.5
1.0
1.5
2.0



Item Discrimination (a): Matrix vs. Matching
Linear Regression
Standard Major Axis Regression
$y = -0.035 + 1.004x$
$R^2 = 0.863$
Number of item pairs = 9
Matching
Matrix
0.0
0.5
1.0
1.5
2.0



Item Discrimination (a): Drop down vs. Select Response
Linear Regression
Standard Major Axis Regression
y = 0.09 + 0.991x
R² = 0.903
Number of item pairs = 11
Select Response
Drop down
0.0
0.5
1.0
1.5
2.0


Item Discrimination (a): Form A vs. Form B
Linear Regression
Standard Major Axis Regression
$y = -0.003 + 1x$
$R^2 = 0.958$
Number of item pairs = 28
Form A
Form B
0.0
0.5
1.0
1.5
2.0

Appendix G

DIF Results

G3ELA											
			Regular				Disability				
Item	Form A Type	Form B Type	Flagged	p value	R^2 change	favored form	Flagged	p value	R^2 change	favored form
Section2item2	ordering	matching	NON	0.000311075	0.002777365		UNI	0.004103236	0.018089496	A
Section2item4	ordering	matching	NON	1.80978E-05	0.003914997					
Section2item7	drop down	selected response	UNI	0	0.015789668	A	UNI	0.004394264	0.017878173	A
Section2item9	drop down	selected response	NON	0.009657703	0.00139612		UNI	0.000149898	0.031511448	A
Section3item18	ordering	matching	UNI	0	0.016244502	A	UNI	0.00068908	0.025617612	A
Section4item3	ordering	matching	NON	0.008325438	0.001509453					

			Non-ESL				ESL			
Section2item2	ordering	matching	NON	0.000864859	0.002488979		UNI	0.002545	0.013868476	A
Section2item4	ordering	matching	NON	0.000126423	0.003290169		UNI	0.000315428	0.019714897	A
Section2item7	drop down	selected response	UNI	7.71605E-14	0.012526183	A	UNI	1.3397E-08	0.048489391	A
Section2item9	drop down	selected response	NON	0.024747576	0.001106139		UNI	2.99933E-08	0.046247034	A
Section3item18	ordering	matching	UNI	4.44089E-16	0.015015509	A	UNI	0.000235336	0.020781986	A
Section4item3	ordering	matching	NON	0.042516999	0.000936656		UNI	0.00693678	0.011240728	A

G4ELA										
			Regular				Disability			
Item	Form A Type	Form B Type	Flagged	p value	R^2 change	favored form	Flagged	p value	R^2 change	favored form
Section1item7	ordering	matching	NON	0.001722876	0.00599543					
Section1item8	categorizatior	matching	NON	0.001327388	0.015310098					
Section2item3	ordering	matching	NON	0.001493395	0.01050271		UNI	0.015502654	0.01267309	A
Section3item8	categorizatior	matching	NON	0.01465754	5.55112E-16		UNI	0.050872659	0.00000562	A
Section4item4	ordering	matching	UNI	0.006758364	6.33152E-08	A	UNI	0.01651042	0.010124022	A

			Non-ESL				ESL			
Section1item7	ordering	matching	NON	0.045289076	0.001001491					
Section1item8	categorizatior	matching	UNI	4.05E-13	0.013113572	A	NON	0.024389839	0.00633902	
Section2item3	ordering	matching	NON	0.045370002	0.000997782		UNI	0.016066751	0.007533639	A
Section3item8	categorizatior	matching	NON	3.06E-12	0.011892497		NON	0.03364982	0.005649102	
Section4item4	ordering	matching	UNI	0.000000059	0.007380337	A	UNI	0.022675986	0.006953461	A

G5ELA										
			Regular				Disability			
Item	Form A Type	Form B Type	Flagged	p value	R^2 change	favored form	Flagged	p value	R^2 change	favored form
Section2item3	ordering	matching								
Section4item1	ordering	matching	UNI	0	0.039297451	A	UNI	0.001558125	0.02761218	A
Section4item4	ordering	matching	UNI	0.001463833	0.002371198	A				

			Non-ESL				ESL			
Section2item3	ordering	matching	UNI	0.003613318	0.002173858	A				
Section4item1	ordering	matching	UNI	0	0.038126008	A	UNI	5.03676E-08	0.039652686	A
Section4item4	ordering	matching	UNI	0.000242197	0.003462309	A				

G6ELA										
			Regular				Disability			
Item	Form A Type	Form B Type	Flagged	p value	R^2 change	favored form	Flagged	p value	R^2 change	favored form
Section2item4	ordering	matching	UNI	5.82088E-10	0.007660529	A	UNI	0.030504074	0.012160763	A
Section2item8	select text	selected response2	UNI	0	0.023702227	A	UNI	0.019871729	0.014081295	A
Section4item3	ordering	matching	UNI	0.000361526	0.002538795	A				

			Non-ESL				ESL			
Section2item4	ordering	matching	UNI	1.76296E-09	0.007485814	A	UNI	0.013297982	0.011028439	A
Section2item8	select text	selected response2	UNI	0	0.024571581	A	UNI	0.017552049	0.010175733	A
Section4item3	ordering	matching	UNI	0.000161213	0.0029442	A				

G7ELA											
			Regular				Disability				
Item	Form A Type	Form B Type	Flagged	p value	R^2 change	favored form	Flagged	p value	R^2 change	favored form
Section1item17	categorizing	matching	UNI	0	0.032277743	A	UNI	0.001941369	0.031547495	A
Section3item8	categorizing	matrix	NON	1.63852E-08	0.006382776		UNI	0.000113206	0.048300018	B
Section3item17	background graphic	matching	UNI	2.58603E-05	0.003743598	A				
Section3item18	categorizing	matrix	NON	1.57743E-05	0.003942831		UNI	0.018828961	0.018090451	B
Section4item3	ordering	matching	UNI	9.99201E-16	0.013557288	A				

			Non-ESL				ESL			
Section1item17	categorizing	matching	UNI	0	0.030780659	A	UNI	2.11481E-07	0.041249556	A
Section3item8	categorizing	matrix	NON	1.01254E-07	0.00612315		NON	0.014938827	0.008787918	
Section3item17	background graphic	matching	UNI	0.000107192	0.003423753	A				
Section3item18	categorizing	matrix	NON	1.48625E-06	0.005280017					
Section4item3	ordering	matching	UNI	4.77396E-15	0.013912197	A	UNI	0.010976183	0.010008174	A

G8ELA										
			Regular				Disability			
Item	Form A Type	Form B Type	Flagged	p value	R^2 change	favored form	Flagged	p value	R^2 change	favored form
Section2item4	ordering	matching	NON	0.000955138	0.03220086					
Section2item5	drop down	selected response	UNI	0.001421332	0.00923266	B				
Section4item1	ordering	matching								
Section4item4	ordering	matching	UNI	0.002508991	0.000546604	A				

			Non-ESL				ESL			
Section2item4	ordering	matching	NON	0.012913464	0.001376208		UNI	0.006289669	0.012082483	A
Section2item5	drop down	selected response	NON	0.001227065	0.002337157		NON	0.001889026	0.015604314	
Section4item1	ordering	matching								
Section4item4	ordering	matching	UNI	0.005652066	0.001718013	A				

HS_ELA										
			Regular				Disability			
Item	Form A Type	Form B Type	Flagged	p value	R^2 change	favored form	Flagged	p value	R^2 change	favored form
Section1item8	categorizatior	matching	UNI	0.013525235	0.001940786	B				
Section1item16	categorizatior	matching								
Section2item3	ordering	matching								
Section3item6	ordering	matching	UNI	0.037418035	0.001384914	A				
Section3item20	ordering	matching								
Section3item21	categorizatior	matrix	UNI	0.000506711	0.003865127	B				
Section4item3	ordering	matching								

			Non-ESL				ESL			
Section1item8	categorizatior	matching	UNI	0.009685874	0.00208107	B				
Section1item16	categorizatior	matching								
Section2item3	ordering	matching								
Section3item6	ordering	matching								
Section3item20	ordering	matching								
Section3item21	categorizatior	matrix	UNI	0.000555974	0.003723477	B				
Section4item3	ordering	matching								

G3Math										
			Regular				Disability			
Item	Form A Type	Form B Type	Flagged	p value	R^2 change	favored form	Flagged	p value	R^2 change	favored form
Section1item15	labeling	matrix	UNI	0.0000166	0.002112073	A	NON	0.014949115	0.00748445	
Section1item22	ordering	matching	NON	0.000830242	0.00127611					
Section2item13	labeling	matching	UNI	0.007535958	0.000815927	A				
Section2item17	categorizatior	matrix	UNI	0.000119139	0.00169381	A				
Section2item24	ordering	matching	NON	0.005130006	0.000888455		NON	0.042898254	0.005149704	

			Non-ESL				ESL			
Section1item15	labeling	matrix	NON	0.019690766	0.000631794		UNI	0.044918207	0.004150468	A
Section1item22	ordering	matching	NON	0.000161509	0.001659875					
Section2item13	labeling	matching	NON	0.02072511	0.000623704					
Section2item17	categorizatior	matrix	UNI	0.000124327	0.00171943	A				
Section2item24	ordering	matching	NON	0.015977448	0.000671039					

G4Math										
			Regular				Disability			
Item	Form A Type	Form B Type	Flagged	p value	R^2 change	favored form	Flagged	p value	R^2 change	favored form
Section1item11	ordering	matching	NON	0.000529416	0.042587412					
Section1item13	labeling	matrix	UNI	0.002492558	0.0000108	A				
Section1item26	labeling	drop down	UNI	0.005242863	1.78E-10	A				
Section1item29	background g	matching	UNI	0.002243406	0.0000306	B	UNI	0.009551049	0.010115707	B
Section2item17	background g	matching								
Section2item20	categorizatior	matrix	UNI	0.002781206	0.00000406	B				
Section2item22	labeling	drop down								

			Non-ESL				ESL			
Section1item11	ordering	matching	UNI	0.000495411	0.001663514	A				
Section1item13	labeling	matrix	UNI	0.00000116	0.003239735	A				
Section1item26	labeling	drop down	UNI	9.74E-09	0.004509807	A	UNI	0.044277	0.003498	A
Section1item29	background g	matching	UNI	0.00000334	0.002966008	B				
Section2item17	background g	matching								
Section2item20	categorizatior	matrix	UNI	0.0000476	0.002301288	B				
Section2item22	labeling	drop down								

G5Math										
			Regular				**Disability**			
Item	Form A Type	Form B Type	Flagged	p value	R^2 change	favored form	Flagged	p value	R^2 change	favored form
Section1item14	labeling	matching	UNI	0	0.02387358	A	UNI	7.63465E-09	0.043201748	A
Section1item17	ordering	matching	UNI	0.011049451	0.00066797	B				
Section1item22	ordering	matching	NON	0	0.007183535		UNI	0.004885029	0.010416513	A
Section2item4	labeling	matrix	UNI	1.27445E-11	0.004762516	A				
Section2item15	labeling	matching					UNI	0.029465266	0.006230583	B
Section2item29	labeling	matching								

			Non-ESL				**ESL**			
Section1item14	labeling	matching	UNI	0	0.026237302	A	UNI	1.52457E-05	0.015419369	A
Section1item17	ordering	matching	UNI	0.017875967	0.000609215	B				
Section1item22	ordering	matching	NON	1.33227E-15	0.00681924		NON	0.038425954	0.003526132	
Section2item4	labeling	matrix	UNI	5.87523E-11	0.004669274	A	UNI	0.044761487	0.003372568	A
Section2item15	labeling	matching								
Section2item29	labeling	matching								

G6Math										
			Regular				**Disability**			
Item	Form A Type	Form B Type	Flagged	p value	R^2 change	favored form	Flagged	p value	R^2 change	favored form
Section1item4	drop down	selected response								
Section1item6	labeling	drop down	NON	0.00010897	0.001697732					
Section3item21	ordering	matching	NON	1.02E-12	0.005634705		NON	0.001639316	0.014445116	

			Non-ESL				**ESL**			
Section1item4	drop down	selected response								
Section1item6	labeling	drop down	NON	0.000114039	0.001762813		UNI	0.00000906	0.019004478	B
Section3item21	ordering	matching	NON	1.03E-14	0.006907663		UNI	0.00000096	0.023189108	A

G7Math										
			Regular				Disability			
Item	Form A Type	Form B Type	Flagged	p value	R^2 change	favored form	Flagged	p value	R^2 change	favored form
Section3item12	matching	click categorization								
Section3item15	background g	matching								
Section3item18	ordering	matching	NON	0.000223689	0.001691376					

			Non-ESL				ESL			
Section3item12	matching	click categorization								
Section3item15	background g	matching	UNI	0.024401512	0.000647907	A				
Section3item18	ordering	matching	NON	0.000077	0.002010154					

G8Math										
			Regular				Disability			
Item	Form A Type	Form B Type	Flagged	p value	R^2 change	favored form	Flagged	p value	R^2 change	favored form
Section2item2	background g	matching	NON	0.00115649	0.001964491					
Section3item2	labeling	matching	UNI	0.00680484	5.76E-14	A				
Section3item23	ordering	matching	NON	0.001228386	0.001451504					

			Non-ESL				ESL			
Section2item2	background g	matching	NON	0.000486865	0.001524838					
Section3item2	labeling	matching	UNI	1.62E-14	0.007376832	A				
Section3item23	ordering	matching	NON	0.003704865	0.001058823					

HS Math										
			Regular				Disability			
Item	Form A Type	Form B Type	Flagged	p value	R^2 change	favored form	Flagged	p value	R^2 change	favored form
Section1item14	labeling	drop down	UNI	0.00000984	0.003704412	A				

			Non-ESL				ESL			
Section1item14	labeling	drop down	UNI	0.0000107	0.003649595	A				