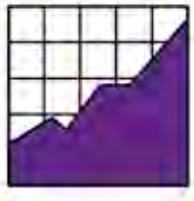
Synthesis Report 92

Results of a Cognitive Interview Study of Immediate Feedback and Revision Opportunities for Students with Disabilities in Large Scale Assessments



NATIONAL CENTER ON EDUCATIONAL OUTCOMES

In collaboration with:
Council of Chief State School Officers (CCSSO)
National Association of State Directors of Special Education (NASDSE)
Supported by:
U.S. Office of Special Education Programs

Synthesis Report 92

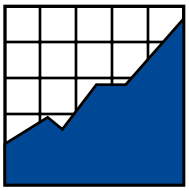
Results of a Cognitive Interview Study of Immediate Feedback and Revision Opportunities for Students with Disabilities in Large Scale Assessments

Christopher Johnstone • Chantal Figueroa
National Center on Educational Outcomes

Yigal Attali • Elizabeth Stone • Cara Laitusis
Educational Testing Service

December 2013

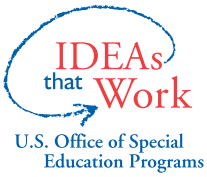
Johnstone, C., Figueroa, C., Yigal, A., Stone, E., & Laitusis, C. (2013). *Results of a cognitive interview study of immediate feedback and revision opportunities for students with disabilities in large scale assessments* (Synthesis Report 92). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.



NATIONAL CENTER ON EDUCATIONAL OUTCOMES

Funding for this report was provided by Grant No. R324A100065 from the U.S. Department of Education/National Center for Special Education Research (NCSER). The Center is supported through a Cooperative Agreement (#H326G110002) with the Research to Practice Division, Office of Special Education Programs, U.S. Department of Education. The Center is affiliated with the Institute on Community Integration at the College of Education and Human Development, University of Minnesota. The contents of this report were developed under the Cooperative Agreement from the U.S. Department of Education, but does not necessarily represent the policy or opinions of the U.S. Department of Education or Office within it. Readers should not assume endorsement by the federal government.

NCEO Project Officer: David Egnor



NCEO Core Staff

Martha L. Thurlow, Director
Deb A. Albus
Manuel T. Barrera
Laurene L. Christensen
Linda Goldstone
James Hatten
Christopher J. Johnstone
Jane L. Krentz
Sheryl S. Lazarus
Kristi K. Liu
Ross E. Moen
Michael L. Moore
Rachel F. Quenemoen
Rebekah Rieke
Christopher Rogers
Vitaliy Shyyan
Miong Vang
Yi-Chen Wu

National Center on Educational Outcomes
University of Minnesota • 207 Pattee Hall
150 Pillsbury Dr. SE • Minneapolis, MN 55455
Phone 612/626-1530 • Fax 612/624-0879
http://www.nceo.info

Executive Summary

Validly assessing students with disabilities has been a challenge for decades but is *increasingly* vital to educational policy and practice in the current era of accountability. Numerous technological and policy developments have occurred in the past several years with the emergence and decline of various forms of alternate assessments. This study was part of a larger research project originally designed to investigate an approach to Alternate Assessments with Modified Achievement Standards that provides immediate feedback and revision opportunities for students when they answer test items. After answering an item, students would receive feedback about whether the answer was correct or incorrect and would be given additional chances to correct their answer (for partial credit).

This particular study employed cognitive interviews to determine whether: (a) Feedback and Revision had a qualitative impact on student interaction with the assessment (i.e., students' interview responses suggest that the feedback and revision feature had an impact on those students' success on the test), and (b) there were qualitative differences between groups of students with and without disabilities (i.e., students' interview responses suggest that the utility and effects of feedback and revision differ for these two groups). After students had completed the test items, we asked the students post-task questions about their perceptions of the assessment and its various conditions.

Results indicated that the vast majority of students appreciated getting a second or third chance when their answer was incorrect (although four students commented that having opportunities to revise responses was "unfair"). Test and interview results suggest that feedback and revision were helpful if students had a sense of how to complete the item but were not useful for students who were demonstrating low mathematical proficiency. The latter group of students became frustrated with the test's "incorrect" feedback and simply guessed as quickly as possible when they were uncertain of how to attack an item.

In terms of feedback and revision data at the item level, most students were able to answer items correctly on two attempts. When third attempts were taken, students were primarily in a guessing mode, most likely because they did not understand the item's content. Overall, students guessed far more often on multiple choice than constructed response (open-ended) items. Overall, students also had more correct answers on multiple choice than open-ended items, despite similar content on both types of tests.

Table of Contents

Introduction

The past 20 years of assessment reform have been marked by increased participation of students with disabilities in large-scale assessments at the state level. Reforms in the Elementary and Secondary Education Act and other federal policies in the 1990s paved the way for assessment practices with almost full participation of students with disabilities on state exams. Beginning with the No Child Left Behind Act of 2001, states have been responsible for reporting on participation rates of students with disabilities for over a decade.

The purpose of large-scale assessments is to determine whether schools and school districts are reaching an adequate level of proficiency on constructs measured on exams. Because students are reported in sub-groups (including disaggregation by sex, ethnicity, English language status, and disability), data are readily available to states for making policy decisions. These data sometimes illuminate deficiencies in educational and assessment practices. For example, despite high rates of participation of students with disabilities in large-scale assessments over the past decade, students with disabilities consistently lag behind their peers without disabilities on achievement measures (Thurlow, Bremer, & Albus, 2011; Thurlow, Quenemoen, Altman, & Cuthbert, 2008).

Critics of contemporary assessment practice argue that low achievement rates on large-scale assessments for students with disabilities may indicate that students lack the requisite knowledge and skills to be successful on the assessment, which creates assessment scenarios that cannot result in valid inferences about students' abilities (Ewell, 2007; Ward Rawheiser, 2007). To this end, the U.S. Department of Education (USDE) issued a series of regulations and non-regulatory guidelines designed to create flexibility in how students with disabilities are assessed.

Historically, states have allowed testing accommodations for students with demonstrated needs. The use of accommodations can be described as an attempt to ensure that the scores received by students with disabilities are valid measures of achievement (Christensen, Braam, Scullin, & Thurlow, 2011). Accommodations are allowable in all 50 states but are carefully scrutinized to ensure that there is fidelity to constructs for both accommodated and non-accommodated test takers. Most states follow practices that align with Thurlow and Bolt's (2001) definition of accommodations as:

> changes in assessment materials or procedures that address aspects of students' disabilities that may interfere with the demonstration of their knowledge and skills on standardized tests. Accommodations attempt to eliminate barriers to meaningful testing, thereby allowing for the participation of students with disabilities in state and district assessments. (p. 3)

In 2007, new regulations were released by USDE that allowed for flexibility in how statewide assessments were administered to students with disabilities. These new regulations were released as the Alternate Assessments based on Modified Academic Achievement Standards (AA-MAS). This new flexibility allowed for modifications to assessments and their level of difficulty in order to meet the needs of students with disabilities. The AA-MAS assessments were intended for students with disabilities who were unlikely to meet grade-level proficiency.

As regulations were released, a series of studies was conducted to determine the validity and efficacy of AA-MAS approaches (see Lazarus, Thurlow, Christensen, and Cormier, 2007 for an overview of studies conducted in this time period). Filbin (2008), Lazarus et al. (2007), and Laitusis and Attali (2009) noted that some of the changes intended to make proficiency more attainable for students with disabilities included:

- Reducing the number of distractors on a multiple choice test
- Reducing the number of items on a test
- Selecting items with higher p-values (less difficulty)
- Developing alternate formats (e.g., portfolios)
- Providing audio supports to test takers
- Providing student choice on which items to take
- Applying strategies to increase readability of documents (e.g., reduced passage length, increased white space, shorter sentences, chunking, etc.)
- Simplifying language
- Embedding test questions within reading passages

The changes provided mixed levels of effectiveness in their ability to improve validity for students with disabilities but in time became unallowable in some states that dropped usage of the AA-MAS. By 2011, state testing practice for some students with disabilities returned to standard and accommodated administration of a single, general education assessment. Other states retained the AA-MAS through flexibility waivers but must discontinue use in 2014-2015.

Although the days of the AA-MAS were short-lived in U.S. education, several lessons were learned regarding the impact of testing changes on students with disabilities. In many cases, changes were simply an explicit and purposeful implementation of universal design principles in assessment. Universal Design of Assessment (UDA) is broadly defined as assessments that are

"designed and developed from the beginning to allow participation of the widest possible range of students, and to result in valid inferences about performance for all students who participate in the assessment" (Thompson, Johnstone, & Thurlow, 2002, p.5). Many of the interventions designed to improve readability, for example, would likely be effective for all students and could be implemented on a standard, general education assessment. Others, such as selecting only easy items for an assessment, would clearly pose problems in general assessment administration.

This research project was designed and funded by USDE during the era of AA-MAS. Although its research questions and design were clearly intended to provide evidence for a new approach to reducing test difficulty, its results provide guidance for the continued inclusion of students with disabilities on general education assessments.

Overview of Study

This research project was built on the psychological concept of feedback and its potential for influencing student engagement. Feedback (i.e., the process of receiving external information on one's thoughts or actions) has the potential to modify behavior and cognitive processes in desirable or undesirable ways (see Brunning, Schraw, & Ronning, 1999; Kulhavy & Stock, 1989; Thorndike, 1913). In the case of large-scale assessment, feedback may help students to better understand expectations, evaluate their own thoughts and responses, and possibly become aware of errors in reasoning. Further, it may facilitate student engagement and perceptions about the assessment by providing students with insights about their performance within the assessment itself.

Because students with disabilities are disproportionately at the low end of achievement scores on statewide assessments, providing feedback and revision opportunities gives students with disabilities an additional opportunity to demonstrate their knowledge on an assessment. Learners with disabilities often experience higher levels of test anxiety compared to their peers without disabilities (Moen, Liu, & Thurlow, 2008; Whitaker-Sena, Lowe, & Lee, 2007). The deleterious effect of worry on test performance is well-documented (Hembree, 1988; Seipp, 1991), but may be reduced when students understand that they can answer a second time if they are incorrect on the first attempt. Further, feedback may provide students who answer quickly with a second opportunity to re-think or re-strategize by accessing their working memory of constructs on the assessment.

This study was part of a larger research program to investigate the effectiveness of feedback and revision on assessments. The qualitative study described in this report examined the influence of feedback and revision opportunities on the test taking behavior and performance of students with and without disabilities. Under the broad conceptual goal of improving assessment validity,

we sought to understand whether students, when provided feedback and revision opportunities, (a) were personally engaged in the assessment (as evidenced by preference questions about test conditions), and (b) were academically engaged in the assessment (as evidenced by use of strategies vs. guessing). Based on the design created by the third, fourth, and fifth authors of this study, we conducted a series of cognitive interviews that compared the administration of varying assessment conditions for students with and without disabilities in order to determine students' levels of personal and academic engagement in the assessment. Results of this study informed the development of a prototype test and feedback-and-revision-capable delivery platform.

Method

Sample

Sample sizes for this study were determined by previous research on cognitive interviews with students with disabilities that found that between 5 and 10 students per subgroup yielded effective interview data (Laitusis, 2007; Nielsen, 1994; van Someren, Barnard, & Sandberg, 1994). All students participating in this study were in Grade 8 at the time they were interviewed and comprised three groups: (a) Students with Specific Learning Disabilities (Rounds 1-3); (b) Students without Disabilities (Rounds 1-3); and (c) Students with other disabilities who, according to teacher data, scored persistently low on statewide examinations (Round 3). Students in group 3 "other disabilities" included students with Autism Spectrum Disorder (ASD, n=2), Developmental Cognitive Disability (DCD, n=1), Emotional Behavior Disorder (EBD, n=2), Other Health Impairment (OHI, n=1), Speech Language Impairment (SLI, n=1), and Traumatic Brain Injury (TBI, n=1). Table 1 displays sample sizes for each round.

Table 1. Sample Size by Research Cognitive Interview Study and Subgroup

Subgroup	Round 1	Round 2	Round 3	Total
No Disability	3	11	11	25
Specific Learning Disability	4	8	9	21
Other Disability			8	8
Total	7	19	28	54

Materials

Mathematics Items

Items for the assessment were drawn from released grade-eight-appropriate mathematics items from the National Assessment of Education Progress (NAEP) and delivered using a computer-based testing platform. Educational Testing Service (ETS) researchers used NAEP Question Tool Version 3.0 (http://nces.ed.gov/nationsreportcard/itmrls/startsearch.asp) to choose items of similar difficulty level and mathematical complexity, while balancing items based on construct (geometry, etc.) and item type (multiple-choice [MC] and open-ended [OE]). Each five-item set in each condition contained items with a range of difficulty and mathematical content appropriate for Grade 8 assessment including basic algebra, geometry, estimation, units of measure, and computation. Students worked on scrap paper or used calculators to complete items, then answered directly into the computer-based platform.

Cognitive Interviews

In this study we employed cognitive interviews as a way to examine student understanding of items and to solicit feedback on the various conditions of the assessment. Over the past several years, several studies have employed cognitive interview techniques as a mechanism of understanding the interplay between large-scale assessments and cognitive processing of students with disabilities (see Johnstone, Bottsford-Miller, & Thompson, 2006; Johnstone, Liu, Altman, & Thurlow, 2007; Johnstone, Thompson, Miller, & Thurlow, 2008; King & Laitusis, 2008; Laitusis, 2007). Cognitive interviews in their contemporary form were most likely first used by psychologist Karl Duncker (1945), who originally described think aloud verbalizations as "productive thinking" and a way to understand his subjects' development of thought. Ericsson and Simon (1993) first used the term "cognitive lab" in their book *Protocol Analysis: Verbal Reports as Data*. Methods outlined in *Protocol Analysis* have guided much of the contemporary work that uses cognitive labs and interviews to better understand student assessment.

The difference between a cognitive lab and a cognitive interview is slight but is important for this study and for working with school-aged children. During cognitive labs, research participants work through a cognitive challenge (i.e., a test item). Participants are asked to say everything that comes to mind during the cognitive activity. According to Ericsson and Simon, the unfiltered data in the form of research participant utterances is the most meaningful data a researcher can use to understand cognitive processes. However, Almond et al. (2009) recommended a "cognitive interview" approach that uses cognitive lab techniques combined with retrospective interview questions.

The rationale for post-hoc interview questions being included in cognitive lab procedures builds on work with young students by Branch (2000) and additional results by Fonteyn, Kuipers, and Grobe (1993). In these studies, authors found that asking subjects post-process questions helped clarify missing or incomplete information that was either uttered or omitted by research participants.

Procedure

Three rounds of cognitive interviews were administered to eighth-grade students with or without disabilities, as shown in Table 1. Before each round of interviews, a pilot administration of the round to two or three students without disabilities in sixth grade was administered in order to identify any potential problems with the test platform or the instrument itself. Each interview consisted of the administration of three five-item blocks that were built to be balanced in terms of difficulty and mathematical and graphical content.

Round 1

Round 1 took place in one school that is located in a college town in rural Wisconsin. The school was high achieving and students were highly engaged in the research process. Items were pulled from the NAEP Questions Tool, focusing on Main assessment grade 8 items and Long-Term Trend age 13 items. The selected item pool consisted of 15 four-option MC items and 10 OE items (which were not administered). These items were revised where needed (e.g., to reduce the number of options from five to four for some items) and were assembled into item blocks with the help of NAEP math test developers.

Students completed the first-round multiple-choice item blocks under three conditions: (a) Standard (ST), (b) Second-Chance (SC), and (c) Answer Until Correct (AUC). Under the ST condition, students answered items without any feedback about whether they were correct or incorrect, the same as current large-scale testing practices. Under feedback and revision conditions (SC and AUC), students received immediate feedback about the correctness of their answer. If the answer was incorrect, they were asked to revise their answer. The SC condition provided students a single additional opportunity to answer if they were incorrect. The AUC condition provided students with up to four opportunities to answer the item (i.e., students would eventually arrive at the correct option). Previously selected incorrect answers were grayed-out in both SC and AUC so that students were not able to select them again. If the student was not able to submit the correct answer within the maximum number of attempts allowed, the system notified the student and provided the correct answer.

In each administration, the three blocks were given uniquely under each of the three feedback-and-revision conditions. There were three block orders used (blocks 1, 2, 3; 2, 3, 1; 3, 1, 2) and

three condition sets for each block ordering (ST, SC, AUC; SC, AUC, ST; AUC, ST, SC), for a total of nine protocols.

Round 2

Round 2's populations were more diverse than Round 1's. In total, we collected data from five schools during this round that varied by location and overall student achievement. In this Round, achievement rates of students with disabilities and those without disabilities were closer, and both populations had lower achievement than in Round 1. Round 2 consisted of the same three conditions as in Round 1 (AUC, SC, ST). After Round 1, the decision was made to have the same items used in both MC and OE formats. Conversion of the Round 1 OE items to MC would have required the writing of options for each item, and not all MC items could be appropriately converted to OE items with responses that worked for the computer platform. Therefore, the MC items were examined to determine how many could be retained (possibly with minor revision). Additional items were then selected using the NAEP Question Tool until there were 15 unique items that could be used in both MC and OE formats. Their characteristics, provided by the Question Tool and based on their use as MC items, are displayed in Table 2. In summary, Round 2 consisted of three MC item blocks. Some of the items presented in Round 2 overlapped with those in Round 1.

Round 3

Round 3 was the largest and most diverse student sample in the study. Data were collected in six schools ranging from urban to very rural (population of less than 400) and from three distinct populations: students without disabilities, students with Specific Learning Disabilities, and students with "other" disabilities described previously. In Round 3, the OE versions of the Round 2 items were used. During this round, students took only two blocks of items (10 items), each of which required a numeric response (e.g., if asked "how many sides does a pentagon have?" students would enter the number "5" for an answer). Only two conditions, ST and SC, were considered, because allowing unlimited chances on an open-ended item could instill more frustration rather than motivation and engagement.

In Round 3, there were three item blocks, but each protocol featured only two of the feedback-and-revision conditions. There were two condition orders (ST, SC; SC, ST) and six item block orders (1, 2; 2, 3; 3, 1; 1, 3; 2, 1; 3, 2) , for a total of 12 protocols.

Table 2. Item Block Properties for Rounds 2 and 3

Block	Item	Difficulty	Complexity	Content Area
1	1	Easy	N/A	Measurement
1	2	Easy	Low	Algebra
1	3	Easy	Moderate	Number properties and operations
1	4	Medium	Low	Measurement
1	5	Medium	Moderate	Algebra
2	1	Easy	N/A	Number properties and operations
2	2	Easy	Low	Geometry
2	3	Easy	N/A	Variables and Relationships
2	4	Medium	Moderate	Data analysis and probability
2	5	Medium	Low	Geometry
3	1	Easy	N/A	Measurement
3	2	Easy	Low	Algebra
3	3	Easy	Low	Number properties and operations
3	4	Medium	Low	Algebra
3	5	Medium	N/A	Measurement

Items developed before 2005 were not categorized in terms of complexity level.

Common to All Rounds

Students were asked to complete the assessment on a computer-based format. They were reminded to verbalize everything they were thinking throughout the test-taking process. The first and second authors interviewed students using a cognitive interview protocol developed by ETS authors employing the practice of "neutral cues" and retrospective questions for students based on their assessment experiences. All conversations with students were focused on students' engagement with assessment items. During this time, researchers took notes on student test-taking strategies, comprehension of skills needed to solve problems, confidence in completing problems, and usage of a calculator. All items with significant text had the opportunity for students to listen to the item in auditory format. All cognitive interviews were audio recorded.

At the completion of each five-item block, we administered a post-test survey that asked students for their reactions to and preferences for the different item formats. The same survey instrument was used for each round.

Analysis

All response data (i.e., item responses and concurrent and retrospective think aloud responses) were entered into a spreadsheet for shared analysis. For each interview, we focused on three main areas: (1) correct/incorrect answer for item; (2) explanatory characteristics of incorrect answers (e.g., level of vocabulary, student reading difficulty, computational error, etc.); and 3) qualitative and quantitative impact of feedback and revision procedures.

As noted previously, the interview study sample was too small to conduct inferential statistical analyses, but descriptive statistics were tallied for correct and incorrect answers as well as for data on student preferences. Further qualitative data analysis was conducted to find student verbalizations that provided explanatory rationales for selections or understanding of content.

In summary, two analysis procedures were employed. First, we coded and calculated all quantitative data about correct and incorrect answers (including ratios of correct and incorrect answers by disability status, number of attempts needed to answer items correctly, etc.). Second, we used point-by-point coding (Bogdan & Biklen, 1992) of qualitative data in spreadsheets to find themes that helped explain quantitative results. This two-phased analysis reveals how results may have occurred.

Results

Achievement

Overall, the Feedback and Revision intervention appeared to lead to slightly greater performance gains across attempts for students with Specific Learning Disabilities (SLD) over students without disabilities. On first attempts, students with SLD only answered 50% of items correctly compared to 67% answered correctly by students without disabilities. Students with SLD, however, answered 35% of items correctly on their second chance compared to only 33% for students without disabilities. By the time students reached their third attempt (which was very rare), students with disabilities answered 63% of items correctly. Students without disabilities only answered 40% of items correctly.

Overall, every student struggled with open-ended items. Students with SLD and students with other disabilities correctly answered 30% and 35% of items on their first attempts, respectively, while students without disabilities answered only 47% of items correctly on their first attempt. Second attempt achievement was very low. Students with SLD answered 13% of items correctly, students with other disabilities answered 7% of items correctly, and students with no disabilities answered 10% of items correctly.

Round 1

Numbers of correct and incorrect responses for each of the 10 students in Round 1 are displayed in Appendix A. In this Round, both students with learning disabilities and those without got the majority of their first attempts correct in all conditions of the test. As was noted previously, this sample was high-achieving. Only one of three students without disabilities answered any items incorrectly. That student, who answered one of the SC items incorrectly, was not able to answer correctly on the second attempt. Although students with disabilities had scores that varied more widely, there were several cases in which students in this group were able to capitalize on the second or subsequent opportunity to revise a response. One student with disabilities was not able to engage well with the test, answering only one item correctly on the second attempt in the SC condition and not responding to any items at all in the ST and AUC conditions. Overall, the students with disabilities had more opportunities to capitalize on the feedback and revision conditions and were sometimes able to do so.

Round 2

Numbers of correct and incorrect responses for each of the 20 students in Round 2 are displayed in Appendix B. In this Round, achievement rates of students with disabilities and those without disabilities were more similar, and both groups had lower achievement than in Round 1. In this round, second and third chances were both more frequent and more effective.

Round 3

Numbers of correct and incorrect responses for each of the 28 students in Round 3 are displayed in Appendix C. Achievement on the items in Round 3 was the lowest of the study. Because the achievement levels of students were similar in Rounds 2 and 3, it is likely the discrepancy is due to the open-ended item format. Students with Specific Learning Disabilities and students with other disabilities (including Autism Spectrum Disorder, Developmental Cognitive Disabilities, Emotional Behavioral Disorders, Other Health Impairments, Speech Language Impairments, and Traumatic Brain Injury) answered approximately the same percentage of items correctly on first attempts, still well below the approximately 50% of correct first attempts made by students with no disabilities.

Achievement in the SC condition was quite low. Very low achievement in the second-chance condition may be explained by the open-ended format, which forced students to construct their own answers.

Personal Engagement

Round 1

In Round 1, students reacted very positively to feedback opportunities. Because of the relatively high achievement of Round 1 students, the majority of attempts made by students were met with the positive feedback of "correct." Even when students were incorrect, all students—with the exception of one who was unsure—perceived test feedback to be a positive experience. Two of the seven students strongly liked the testing approach. One student without a disability said, "I liked it a lot because instead of waiting for the feedback, I could just get it. I liked when I didn't know the answer, I could take an educated guess and keep trying." A student with a disability stated, "I liked to get feedback because if I got it wrong I would know right away. I like that."

Three students liked feedback and revision opportunities to a moderate extent. One student favored the concept, but believed he did not personally need second chances to respond. "[The test is] …better with feedback. I could do without, but it was nice to get." This student was one of the high achieving students without disabilities. Another student with a disability liked the idea of second chances, but perceived the "answer until correct" condition as not as effective, noting that "answering until correct might encourage people to guess." Another student without a disability liked the feedback and revision assessment very little because of a similar distaste for "answer until correct." This student stated that "four (chances) is too much. With four you can just [attempt] three times and if you keep getting them wrong it will give you the answer." One student (mentioned above) was very frustrated with the assessment and did not provide meaningful feedback. The table of reactions to Round 1 is in Appendix D.

Round 2

There were mixed results in terms of student perception of the feedback and revision experience in Round 2. Seven students (three with Specific Learning Disabilities and four without) out of nineteen very much liked the assessment approach. These students enjoyed getting positive feedback and additional opportunities to answer items correctly. Most students liked the assessment to a moderate extent but expressed some concern about frustration that was introduced when items were incorrect.

"It was frustrating when you kept getting the wrong answer over and over" noted one student, referring to multiple chances at a problem that was difficult to solve. "I liked the [correctness feedback] so that I know if I am right" mentioned one student. He continued, "the [incorrectness feedback] was a little frustrating, but the process of elimination helped. If you're going to guess and you keep getting them wrong it can be frustrating if you have to put in a new answer." Other students expressed the same mixed reactions of happiness when answers were correct

and frustration or embarrassment when incorrect. “It was embarrassing” mentioned one student, discussing her feelings when she was told she was incorrect. Another noted that “correct (feedback) made me feel smart, but incorrect was kind of frustrating when I didn’t know the answer.” The table of reactions to Round 2 is in Appendix E.

Round 3

In Round 3, students were generally positive about opportunities for feedback and revision. Out of 28 students in this sample, 13 students (three with Specific Learning Disabilities, three with other disabilities, and seven with no disabilities) liked the opportunity for feedback and revision very much. The students appreciated the opportunity to do their “best” and to see where they went right or wrong. One student with a Specific Learning Disability enthusiastically stated his approval for the assessment because “I got to think what I got to improve on!”

Four students with Specific Learning Disabilities, two students with other disabilities, and six students with no disabilities explained that they liked feedback and revision to a moderate extent. The majority of students in this group liked the idea of getting immediate feedback on test items. One student enjoyed the opportunity to engage with assessments with feedback and revision, stating that “you don’t get to think that much” on a typical statewide test. Enthusiasm was tempered in four students (two without disabilities, one with Autism Spectrum Disorder, and one with a Specific Learning Disability) who expressed anxiety or embarrassment about getting problems incorrect. One student expressed his approval for the ability to revise his answers and provided a new idea for how this assessment process should work. “It was OK,” he said about the test, “but I was expecting a little bit of a hint. When my teacher gives me feedback she gives me ideas on what to do next.” The table of reactions to Round 3 is in Appendix F.

Academic Engagement

The purpose of all accessibility and accommodations in assessment is to facilitate student opportunity to demonstrate knowledge. In this study, feedback and revision was designed to provide students who answered incorrectly with opportunities to rethink the problem and their responses a second, third, or fourth time. To an extent, this occurred. Students frequently used strategies to answer items on first attempts but, on aggregate, guessed on second and third attempts more often than they used strategies. This was especially true of students with disabilities, who guessed on 29 of 51 second attempts and 10 of 11 third attempts. In these instances, students selected answers in just a few seconds and frequently made comments about not knowing the answer. The results in Appendices G, H, and I indicate the extent of the guessing behavior, which was most likely a response to students not having a full understanding of the constructs tested.

In all three rounds, students with disabilities tended to be more likely to be observed exhibiting guessing behavior than students without disabilities. Both groups guessed more often in the second round than in the first, a phenomenon most likely attributable to the high level of proficiency of Round 1 students. Round 3 followed similar guessing patterns. Students guessed fairly often when attempting items. Guessing was more frequent on second attempts for students with other disabilities than for students without disabilities or students with a learning disability. When more attempts were allowed, there appeared to be a greater propensity for guessing. For example, only one student who participated in the study used a strategy on a third attempt—all other third attempts were guesses. Appendices G, H, and I include guessing frequencies for the three rounds.

Overall, the proportion of guessing increased by attempt, demonstrating that students may have had one strategy in their "tool box" and, once exhausted and incorrect, guessed on subsequent attempts. However, in several instances, students used a strategy after reading an item for the second time. This was especially true for items with open-ended formats. On OE items, nearly the entire sample of students with disabilities (15/16) and students without disabilities (6/8) used strategies at least once after guessing the first time. For MC items, using guessing after strategies was less prevalent; however, five students with Specific Learning Disabilities and three students without disabilities engaged in second-chance strategizing after guessing. Guessing pathways are documented in Appendices J, K, and L.

Summary of Personal and Academic Engagement

Overall, the feedback and revision process seemed to be personally engaging to students. The vast majority of students, both with and without disabilities, enjoyed getting feedback. For some students, engagement was increased by the "Correct!" feedback provided after answering an item correctly. For others, the opportunity to try again increased their engagement and positive psychological reaction to the assessment process.

For several students, additional opportunities were not viewed as an engaging or appealing aspect of an assessment. These students typically answered incorrectly on their first two attempts or guessed correctly on their second attempt. For students who did not have the skillset to meaningfully attempt items, a second chance did not improve personal engagement, but represented an embarrassing or frustrating experience.

In Rounds 1 and 2, students rarely reached the maximum number of tries allowed. The vast majority of students only had one or two attempts on items (this was either due to the high student achievement and the study design, in which two of the three conditions only allowed one or two attempts). In conditions where students could reach a third attempt, they often became disengaged. Six students even believed that extending opportunities to four attempts represented an

unfair testing practice and should not be allowed. All six of these students were high achieving students; three students had Specific Learning Disabilities, and three students did not.

In terms of academic engagement, guessing behavior increased by round when considering the number of guesses compared to the total items. Students with disabilities guessed more frequently than students without disabilities on first, second, and third attempts. Many times, students with disabilities attempted an item and, upon learning that they were incorrect, guessed on their second attempt. There were instances, however, in which students with Specific Learning Disabilities used a strategy for answering an item on a second attempt after guessing on a first attempt, demonstrating that second attempts may facilitate increased academic engagement in assessments.

Discussion

Like other studies of this kind, this study was intended as an exploratory attempt to understand the phenomenon of feedback and revision on assessments. It is important to note the small sample size, a common feature of cognitive interview studies. This aspect allows for more comprehensive observation of and interaction with participants, providing insights that would not be possible under other conditions, but it limits the inferences and conclusions that can be made. One limitation affecting the study was that the students in Round 1 (with the exception of one student) appeared to be very high achieving students who made few incorrect responses, providing limited evidence about feedback and revision.

Despite the inferential limitations inherent with small-scale qualitative studies, several themes appeared to cross-cut the study while others provide insights into useful further study. Across each Round of this study, it was clear that an achievement gap existed between students with and without disabilities. Students without disabilities answered more items correctly on their first attempts and were more successful on second attempts than their peers with disabilities.

In terms of feedback and revision data, the Answer Until Correct (AUC) functionality was rarely utilized because most students were able to answer items correctly on two attempts. When third attempts were taken, students were primarily in a guessing mode. Feedback on first and second attempts for this study was far more common. Some students questioned the fairness of allowing more than a second chance on an assessment. The “Correct!” feedback provided for correct answers appeared to be quite motivating for students. Students had varying opinions on second chances, ranging from great enthusiasm to concern because of increased frustration after receiving “incorrect” feedback.

The mixed achievement and engagement data in this study demonstrate a need for further research. Students with disabilities are a heterogeneous group, and the level of engagement and benefit that students receive from a feedback and revision innovation may vary. Further research will likely validate the finding that most students enjoy and benefit from further opportunities on assessments but may also demonstrate that some students with low achievement may experience frustration when they engage in multiple attempts at an item. A follow-up analysis will investigate the profiles of students (within the category of "students with disabilities") to further predict the subset of students who will most likely benefit from this innovation. Once profiles are identified and are then studied in larger and more comprehensive empirical studies, stronger conclusions can be drawn. The results of this cognitive interview study, which represents Phase I of a grant, were used to inform development of a Phase II large experimental study piloted in Spring 2012 and field tested in Spring 2013. Phase III will consist of surveys and focus groups of teachers of the participants in Phase II to determine the teachers' perceptions of the accuracy and utility of the various feedback and revision conditions.

Results of this program of research can then be compared against state accommodations policies and regulations to determine the eligibility of certain students to receive feedback and revision opportunities on statewide assessments. Further clarification of partial credit for second and third chances, the opportunity to provide students with partial scaffolding if they answer incorrectly, and the appropriateness of students without disabilities also being allowed feedback and revision opportunities are questions that further research and policy analysis will answer as this research project progresses.

References

Almond, P. J., Cameto, R., Johnstone, C. J., Laitusis, C., Lazarus, S., Nagle, K., Parker, C. E., Roach, A. T., & Sato, E. (2009). *White paper: Cognitive interview methods in reading test design and development for alternate assessments based on modified academic achievement standards (AA-MAS).* Dover, NH: Measured Progress and Menlo Park, CA: SRI International.

Bogdan, R. C., & Biklen, S. K. (1992). *Qualitative research for education. An introduction to theory and methods*. Boston: Allyn and Bacon.

Branch, J. L. (2000). Investigating the information-seeking processes of adolescents: The value of using think-alouds and think afters. *Library and Information Science Research, 22*(4), 371-392.

Bruning, R., Schraw, G., & Ronning, R. (1999). *Cognitive psychology and instruction.* Columbus, OH: Merrill Prentice Hall.

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.

Duncker, K. (1945). On problem-solving. In J. F. Dashiell (Ed.) *Psychological Monographs* (pp.1–114). Washington, DC: American Psychological Association.

Epstein, M. L., Epstein, B. B., & Brosvic, G. L. (2001). Immediate feedback during academic testing. *Psychological Reports, 88,* 889-894.

Epstein, M. L., Lazarus, A. D., Calvano, T. B., Matthews, K. A., Hendel, R. A., Epstein, B. B., & Brosvic, G. L. (2002). Immediate feedback assessment technique promotes learning and corrects inaccurate first responses. *The Psychological Record, 52,* 187–201.

Ericsson, K., & Simon, H. (1993). *Protocol analysis: Verbal reports as data.* Cambridge: MIT.

Ewell, J. (2007, May 26). Test-takers, not students. *Los Angeles Times.* Accessed online, September 10, 2008, at: www.latimes.com/news/printedition/opinion/la-oe-ewell26may26,1,3946284.story?coll=la-news-comment

Filbin, J. (2008). *Lesson from the initial peer review of alternate assessment based on modified achievement standards*. U.S. Department of Education Office of Elementary and Secondary Education.

Fonteyn, M. E., Kuipers, B., & Grobe, S. (1993). A description of think aloud method and protocol analysis. *Qualitative Health Research, 3,* 430–441.

Hallahan, D. P., Kauffman, J. M., & Pullen, P. C. (2012). *Exceptional learners: An introduction to special education* (12th ed.). Upper Saddle River, NJ: Pearson Education.

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research, 58,* 7-77.

Johnstone, J., Bottsford-Miller, N., & Thompson, S. (2006). *Using the think aloud method (cognitive labs) to evaluate test design for students with disabilities and English language learners* (Technical Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Johnstone, C., Liu, K., Altman, J., & Thurlow, M. (2007). *Student think aloud reflections on comprehensible and readable assessment items: Perspectives on what does and does not make an item readable* (Technical Report 48). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes

Johnstone, C. J., Thompson, S. J., Miller, N. A., & Thurlow, M. L. (2008). Universal design and multi-method approaches to item review. *Educational Measurement: Issues and Practice, 27*(1), 25-36.

Kang, S. H. K., Carpenter, S. K., Rohre, D., & Pashler, H. (June, 2009). *Does guessing when one doesn't know the answer hurt subsequent learning.* Poster presentation at IES Research Conference, Washington, DC.

King, T., & Laitusis, C. C. (2008*). Sample protocol for cognitive interviews on tests of reading comprehension*. Downloaded on September 15, 2008, from www.ets.org/dara

Koretz, D. M., & Hamilton, L. S. (2006).Testing for accountability in K12. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed). Westport, CT: Praeger.

Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review, 1*(4), 279-308.

Kulhavy, R. W., Yekovich, F. R., & Dyer, J. W. (1976). Feedback and response confidence. *Journal of Educational Psychology, 68*(5), 522-528.

Laitusis, C. C. (2007). A variety of approaches for studying accommodations on assessments. In C. C. Laitusis & L. L. Cook (Eds.), *Accommodating students with disabilities on state assessments: What works?* (pp.73-81) Arlington, VA: Council for Exceptional Children.

Laitusis, C., & Attali, Y. (2009). *Proposal to Institute on Education Sciences for study of feedback and revision of assessment items (unpublished document).* Princeton, NJ: ETS.

Lazarus, S. S., Thurlow, M. L., Christensen, L. L., & Cormier, D. (2007). *States' alternate assessments based on modified achievement standards (AA-MAS) in 2007* (Synthesis Report 67). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Moen, R., Liu, K. K., & Thurlow, M. L. (2008, June). *Identifying less accurately measured students.* Council of Chief State School Officers National Conference on Student Assessment Pre-session Workshop (Designing Accessible Reading Assessments for Students with Disabilities), Orlando, FL.

Nielson, J. (1994). Estimating the number of subjects needed for a thinking aloud test. *International Journal of Human-Computer Studies, 41*(3), 385-397.

No Child Left Behind Act of 2001, 20, U.S. C. 6301 et seq. (2001).

Pressey, S. L. (1926). A simple apparatus which gives tests and scores - and teaches. *School and Society, 23*(586), 373-376.

Pressey, S. L. (1950). Development and appraisal of devices providing immediate automatic scoring of objective tests and concomitant self-instruction. *The Journal of Psychology, 29*, 417–447.

Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. *Anxiety Research, 4*, 27-41.

Skinner, B. F. (1968). *The technology of teaching*. New York: Appleton-Century-Crofts.

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.

Thompson, S. J., Johnstone, C. J., Thurlow, M. L., & Altman, J. R. (2005). 2005 *State special education outcomes: Steps forward in a decade of change*. Minneapolis, MN: University of Minnesota.

Thorndike, E. L. (1913). *Educational psychology. Volume 1: The original nature of man*. New York: Columbia University, Teachers College.

Thurlow, M., & Bolt, S. (2001). *Empirical support for accommodations most often allowed in state policy* (Synthesis Report 41). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Thurlow, M. L., Bremer, C., & Albus, D. (2011). *2008-09 publicly reported assessment results for students with disabilities and ELLs with disabilities* (Technical Report 59). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Thurlow, M. L., Quenemoen, R., Altman, J. R., & Cuthbert, M. (2008). *Trends in the public reporting of state assessment data (2001-02 through 2004-05)* (Technical Report 50). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

U. S. Department of Education (2007, July 20). *Modified academic achievement standards: Non regulatory guidance*. Washington, DC: Office of Elementary and Secondary Education, U.S. Department of Education. Available at www.ed.gov/admins/lead/account/saa.html#regulations

van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think-aloud method: A practical guide to modeling cognitive processes*. San Diego, CA: Academic Press Ltd.

Ward Rawheiser, P. (2007, June 5). Testing is too much. *USA Today.* Accessed at http://blogs.usatoday.com/oped/2007/06/testing_is_too_.html

Whitaker-Sena, J. D., Lowe, P. A., & Lee, S. W. (2007). Significant predictors of test anxiety among students with and without learning disabilities. *Journal of Learning Disabilities, 40*(4), 360-376.

Appendix A

Round 1: Number of Correct Answers Out of All Attempts

	ST	SC		AUC		
ID	**1st attempt**	**1st attempt**	**2nd attempt**	**1st attempt**	**2nd attempt**	**3rd attempt**
Students with No Disability (3 students)						
101	3/5	4/5	0/1	5/5	NA	NA
102	5/5	5/5	NA	5/5	NA	NA
103	5/5	5/5	NA	5/5	NA	NA
Students with Learning Disabilities (4 students)						
104	4/5	4/5	0/1	5/5	NA	NA
105	2/5	4/5	1/1	3/5	1/2	0/1
106	0/0	0/5	1/5	0/0	NA	NA
107	4/5	4/5	0/1	4/5	1/1	NA

Appendix B

Round 2: Number of Correct Answers Out of All Attempts

	ST	SC		AUC		
ID	**1st attempt**	**1st attempt**	**2nd attempt**	**1st attempt**	**2nd attempt**	**3rd attempt**
Students with No Disability (11 students)						
109	1/5	2/5	1/3	3/5	0/2	1/2
110	5/5	5/5	NA	4/5	0/1	1/1
113	5/5	4/5	1/1	5/5	NA	NA
201	4/5	3/5	1/2	4/5	1/1	NA
202	1/5	3/5	1/2	3/5	2/2	NA
203	3/5	5/5	NA	5/5	NA	NA
405	5/5	5/5	NA	5/5	NA	NA
406	3/5	3/5	2/2	2/5	3/3	NA
500	2/5	3/5	1/2	5/5	NA	NA
501	4/5	4/5	0/1	4/5	0/1	0/1
502	2/5	5/5	NA	3/5	2/2	NA
Students with Learning Disabilities (8 students)						
108	2/5	4/5	0/1	2/5	1/3	1/2
111	2/5	1/3	0/1	0/0	NA	NA
112	2/5	4/5	0/1	4/5	0/1	1/1
301	3/5	4/5	0/1	4/5	0/1	1/1
303	3/5	2/5	1/3	4/5	1/1	NA
304	2/5	2/5	2/3	2/5	2/3	0/1
305	3/5	5/5	NA	5/5	NA	NA
407	2/5	3/5	0/2	2/5	2/3	1/1

Appendix C

Round 3: Number of Correct Answers Out of All Attempts

	ST	SC	
ID	**1st attempt**	**1st attempt**	**2nd attempt**
Students with No Disability (11 students)			
705	1/5	3/5	0/2
706	2/5	2/5	0/3
707	5/5	2/5	1/3
801	2/5	3/5	0/2
903	3/5	2/5	0/3
1201	3/5	3/5	0/2
1202	2/5	2/5	0/3
1203	1/5	3/5	1/2
1204	3/5	2/5	1/3
1205	2/5	2/5	0/3
1206	1/5	2/5	0/3
Students with Learning Disabilities (9 students)			
602	1/5	0/5	0/5
604	2/5	2/5	1/3
605	2/5	1/5	0/4
606	1/5	1/5	0/4
607	3/5	2/5	1/3
701	2/5	1/5	0/4
802	2/5	1/5	1/4
803	1/5	3/5	0/2
902	2/5	3/5	1/2
Students with Other Disabilities[1] (8 students)			
601 (ASD)	3/5	0/5	1/5
1002 (ASD)	2/5	1/5	1/4
704 (DCD)	0/5	1/5	0/4
702 (EBD)	0/5	1/5	0/4
703 (EBD)	1/5	3/5	0/2
901 (OHI)	5/5	2/5	0/3
603 (SLI)	1/5	3/5	0/2
1001 (TBI)	1/5	1/5	0/4

[1] ASD=Autism Spectrum Disorder; DCD=Developmental Coordination Disorder; EBD=Emotional Behavioral Disorder; OHI=Other Health Impairment; SLI=Speech Language Impairment; TBI=Traumatic Brain Injury.

Appendix D

Round 1: Student Perceptions of Feedback and Revision

ID	Reactions to Feedback and Revision Conditions Summary
Students with No Disability (3 students)	
101	Liked very much – enjoyed knowing correct answer immediately
102	A little – believed answer until correct encouraged guessing
103	Moderately liked – enjoyed seeing answers to items after, but "could do without"
Students with a Learning Disability (4 students)	
104	Moderately liked – only needed second chance once but liked knowing item answers. Did not like the idea of four chances because it encouraged guessing.
105	Liked very much – enjoyed getting more chances to try again if incorrect.
106	"Don't know" – student did not engage in interview because of frustration with assessment
107	Moderately – inspired confidence to go on.

Appendix E

Round 2: Student Perceptions of Feedback and Revision

ID	Reactions to Feedback and Revision Conditions Summary
Students with No Disability (11 students)	
109	Moderately liked – enjoyed correct feedback but was "embarrassed" when incorrect
110	Moderately liked – enjoyed knowing when correct but frustrated by multiple guesses when did not know answer
113	Moderately liked – enjoyed hearing correct feedback, but did not want to re-do incorrect problems
201	Liked it moderately when given a second chance to try again. Liked "correct" feedback a lot
202	Liked it a lot because it reinforced when correct and gave a second chance when incorrect
203	Liked it a lot because it built confidence
405	Liked moderately first round, then a lot second round. Liked knowing results immediately
406	Liked very much – liked positive reinforcement and reduced anxiety
500	Liked positive reinforcement very much but did not like finding out he was wrong
501	Liked second chance a lot because got another chance. Did not like until correct at all because "it gave away the answer." Became frustrated with wrong answers.
502	Moderately – liked hearing when he was correct but did not like being told he was wrong. Liked the second chance idea but thought multiple chances was too much
Students with a Learning Disability (8 students)	
108	Liked second chance a little – but felt nervous with only one more chance. Liked Answer Until Correct a lot – felt relaxed
111	Student asked to be excused before feedback interview questions
112	Moderately liked – enjoyed "correct" feedback but felt frustrated when incorrect. Thought AUC condition was "too many chances"
301	Liked it moderately – "it helped me to get right answers"
303	Like it – likes to know results
304	Liked moderately – liked second chance but did not like answer until correct
305	Liked very much for second chances, felt wasn't necessary for items that were easy
407	Liked it a lot – liked to keep working toward correct answer

Appendix F

Round 3: Student Perceptions of Feedback and Revision

ID	Reactions to Feedback and Revision Conditions Summary
Students with No Disability (11 students)	
705	Very much – liked knowing if correct
706	Liked very much knowing if right or wrong
707	Moderately liked, but got frustrated when heard he was wrong
801	Liked very much knowing if right or wrong
903	Very much – if you get it wrong you get a second chance
1201	Very much – gives you a chance to concentrate more on the question to get it right
1202	Moderately – better than state test which "don't let you think that much"
1203	Very much – it gives you a chance to improve
1204	Moderately – would have preferred a hint for second chance
1205	Liked very much to have second chance to get items correct
1206	Moderately – liked getting a second chance but was tense when learning he was incorrect
Students with a Learning Disability (9 students)	
602	Moderately liked in order to know if right or wrong
604	Liked very much – enjoyed getting answers back to check results
605	Not at all – items were hard. It was only good when getting one right.
606	Very much – "I got to think what I got to improve on"
607	Moderately – good because a second chance might help
701	Liked very much because it made her feel like she tried her best
802	Moderately – became nervous of getting wrong answers
803	Moderately – liked knowing the answer
902	A Little – it is OK for people to get second chances
Students with Other Disabilities[2] (8 students)	
601 (ASD)	Moderately liked – enjoyed correct feedback but was "embarrassed" when incorrect
1002 (ASD)	Very much – on the state test you put the answer and hope for the best
704 (DCD)	Moderately – it was pretty fun
702 (EBD)	Very much – liked to know when he got it wrong so could fix it
703 (EBD)	Moderately – liked to know when he got it right
901 (OHI)	Liked very much being able to see the answer immediately
603 (SLI)	Did not like at all – could not figure out problems and extra chances did not help
1001 (TBI)	Did not know – but was nervous about getting wrong answers

[2] ASD=Autism Spectrum Disorder; DCD=Developmental Coordination Disorder; EBD=Emotional Behavioral Disorder; OHI=Other Health Impairment; SLI=Speech Language Impairment; TBI=Traumatic Brain Injury.

Appendix G

Round 1: Student Guessing

ID	Fraction of Items on Which Guessing Occurred on the		
	First Attempt	**Second Attempt**	**Third Attempt**
Students with No Disability (3 students)			
101	1/15	NA	NA
102	0/15	NA	NA
103	0/15	NA	NA
Students with a Learning Disability (4 students)			
104	0/15	NA	NA
105	6/15	1/2	1/1
106	10/10	NA	NA
107	0/15	2/2	NA

Appendix H

Round 2: Student Guessing

	Fraction of Items on Which Guessing Occurred on the		
ID	**First Attempt**	**Second Attempt**	**Third Attempt**
Students with No Disability (11 students)			
109	6/15	3/5	2/2
110	1/15	1/1	1/1
113	0/15	NA	NA
201	1/15	0/3	NA
202	3/15	3/4	NA
203	0/15	NA	NA
405	0/15	NA	NA
406	3/15	0/5	NA
500	0/15	NA	NA
501	2/15	2/2	1/1
502	1/15	0/3	NA
Students with a Learning Disability (8 students)			
108	13/15	3/4	NA
111	0/15	1/1	NA
112	8/15	0/2	0/1
301	1/15	2/2	1/1
303	1/15	2/4	2/2
304	4/15	2/6	1/1
305	1/15	NA	NA
407	3/15	4/5	1/1

Appendix I

Round 3: Student Guessing

	Fraction of Items on Which Guessing Occurred on the	
ID	**First Attempt**	**Second Attempt**
Students with No Disability (11 students)		
705	0/10	2/3
706	2/10	NA
707	1/10	2/3
801	0/10	NA
903	3/10	2/3
1201	0/10	1/2
1202	1/10	1/3
1203	2/10	2/2
1204	0/10	2/3
1205	2/10	1/3
1206	3/10	2/3
Students with a Learning Disability (9 students)		
602	2/10	2/4
604	0/10	1/3
605	2/10	2/4
606	1/10	2/4
607	5/10	0/3
701	1/10	1/4
802	1/10	2/4
803	2/10	2/2
902	3/10	1/2
Students with Other Disabilities[3] (8 students)		
601 (ASD)	0/10	NA
1002 (ASD)	0/10	1/4
704 (DCD)	4/10	4/4
702 (EBD)	3/10	3/4
703 (EBD)	1/10	1/2
901 (OHI)	0/10	1/3
603 (SLI)	5/10	2/2
1001 (TBI)	3/10	4/4

[3] ASD=Autism Spectrum Disorder; DCD=Developmental Coordination Disorder; EBD=Emotional Behavioral Disorder; OHI=Other Health Impairment; SLI=Speech Language Impairment; TBI=Traumatic Brain Injury.

Appendix J

Round 1: Strategy and Guessing Pathways (SC, AUC Conditions)

			S[4]	S	G	G	S	S	S	S	G	G	G	G
	S	G	S	G	S	G	S	G	S	G	S	S	G	G
ID							S	S	G	G	S	G	S	G
Students with no Disability (3 students)														
101	14	0	1	0	0	0	NA	NA	NA	NA	NA	NA	NA	NA
102	15	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
103	15	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Students with a Learning Disability (4 students)														
104	14	0	0	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA
105	8	4	0	1	1	0	0	0	0	1	0	0	0	0
106	0	0	0	0	0	5	NA	NA	NA	NA	NA	NA	NA	NA
107	13	0	1	1	0	0	NA	NA	NA	NA	NA	NA	NA	NA

S= Strategy (i.e., student used a strategy – correct or incorrect – to solve problem..

G = Guess (i.e., student did not use any strategy but simply guessed at answer)

NA = Not Applicable (student completed item before reaching third attempt or attempted item with only one or two allowable attempts.

[4] Multiple letters represent strategy and guess use for multiple attempts at a single item. For example, “S – S” means that a student used a strategy on her/his first attempt, was incorrect, and tried a strategy on her/his second attempt.

Appendix K

Round 2: Strategy and Guessing Pathways (SC, AUC Conditions)

			S[5]	S	G	G	S	S	S	S	G	G	G	G
	S	G	S	G	S	G	S	G	S	G	S	S	G	G
ID							S	S	G	G	S	G	S	G
Students with No Disability (13 students)														
109	7	3	0	1	0	2	0	0	0	0	0	0	0	2
110	14	0	0	0	0	0	0	0	0	0	0	0	0	1
113	14	0	1	0	0	0	NA	NA	NA	NA	NA	NA	NA	NA
201	12	0	1	2	0	0	NA	NA	NA	NA	NA	NA	NA	NA
202	11	0	0	1	1	2	0	0	0	0	0	0	0	0
203	15													
401	15	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
402	13	0	0	0	1	1	NA	NA	NA	NA	NA	NA	NA	NA
405	15	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
406	10	0	2	0	3	0	NA	NA	NA	NA	NA	NA	NA	NA
501	12	1	0	0	0	0	0	0	0	1	0	0	0	1
502	13	1	1	0	0	0	NA	NA	NA	NA	NA	NA	NA	NA
500	14	1	6	0	0	0	NA	NA	NA	NA	NA	NA	NA	NA
Students with a Learning Disability (10 students)														
108	11	0	1	0	0	1	0	0	0	1	0	0	0	1
111	3	2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
112	9	3	0	0	1	0	1	0	0	0	1	0	0	0
301	11	3	0	1	0	0	0	0	0	0	0	0	0	1
302	10	2	0	0	1	2	NA	NA	NA	NA	NA	NA	NA	NA
303	7	2	2	2	1	1	0	0	0	0	0	0	0	0
304	7	3	3	1	0	0	0	0	0	0	0	0	0	1
305	14	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
403	10	0	0	0	3	2	NA	NA	NA	NA	NA	NA	NA	NA
407	6	4	2	1	0	1	0	0	0	0	0	0	0	1

S= Strategy (i.e., student used a strategy – correct or incorrect – to solve problem.

G = Guess (i.e., student did not use any strategy but simply guessed at answer).

NA = Not Applicable (student completed item before reaching third attempt or attempted item with only one or two allowable attempts.

[5] Multiple letters represent strategy and guess use for multiple attempts at a single item. For example, “S – S” means that a student used a strategy on her/his first attempt, was incorrect, and tried a strategy on her/his second attempt.

Appendix L

Round 3: Strategy and Guessing Pathways (SC, AUC Conditions)

ID	S	G	S-S[6]	S-G	G-S	G-G
Students with No Disability (11 students)						
705	7	1	2	0	0	0
706	7	0	1	2	0	0
801	8	0	2	0	0	0
802	6	1	1	1	0	1
903	6	1	1	0	0	2
1201	8	0	1	1	0	0
1202	7	0	2	0	0	1
1203	7	1	0	1	0	1
1204	7	0	1	2	0	0
1205	8	2	NA	NA	NA	NA
1206	5	2	1	1	0	1
Students with a Learning Disability (9 students)						
602	4	1	3	2	0	0
604	7	0	3	0	0	0
605	4	2	2	2	0	0
606	6	0	2	1	0	1
607	4	3	1	1	2	0
701	6	0	3	0	0	1
707	7	0	2	1	0	0
803	7	2	0	1	0	1
902	7	1	0	1	0	1
Students with Other Disabilities (8 students)						
601 (ASD)	5	0	5	0	0	0
1002 (ASD)	6	0	3	1	0	0
704 (DCD)	6	0	0	1	0	3
702 (EBD)	5	1	1	1	0	2
703 (EBD)	7	1	1	1	0	0
603 (SLI)	6	2	0	2	0	0
901 (OHI)	7	0	2	1	0	0
1001 (TBI)	5	1	0	1	0	3

S= Strategy (i.e., student used a strategy – correct or incorrect – to solve problem.

G = Guess (i.e., student did not use any strategy but simply guessed at answer).

NA = Not Applicable (student completed item before reaching third attempt or attempted item with only one or two allowable attempts.

[6] Multiple letters represent strategy and guess use for multiple attempts at a single item. For example, “S – S” means that a student used a strategy on her/his first attempt, was incorrect, and tried a strategy on her/his second attempt.

COLLEGE OF EDUCATION
+ HUMAN DEVELOPMENT

UNIVERSITY OF MINNESOTA

NCEO is an affiliated center of the Institute on Community Integration