

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Copyright is retained by the first or sole author, who grants right of first publication to the *Practical Assessment, Research & Evaluation*. Permission is granted to distribute this article for nonprofit, educational purposes if it is copied in its entirety and the journal is credited.

Volume 14, Number 18, October 2009

ISSN 1531-7714

Multiple Constructs and Effects of Accommodations on Accommodated Test Scores for Students with Disabilities

Stephanie W. Cawthon, Eching Ho, Puja G. Patel, Deborah C. Potvin, and Katherine M. Trundt
The University of Texas at Austin

Students with disabilities frequently use accommodations to participate in large-scale, standardized assessments. Accommodations can include changes to the administration of the test, such as extended time, changes to the test items, such as read aloud, or changes to the student's response, such as the use of a scribe. Some accommodations or modifications risk changing the difficulty of the test items or decreasing the validity of how test scores are interpreted. Questions regarding the validity of accommodated tests are heightened when scores are used in high-stakes decisions such as grade promotion, graduation, teacher merit pay, or other accountability initiatives. The purpose of this article is to review existing literature on multiple constructs that affect validity of interpretations of accommodated assessment scores. Research on assessment accommodations continues to grow but offers few conclusive findings on whether they facilitate fair and accurate measurement of student knowledge and skill. The validity of an accommodated score appears to vary depending on several factors such as student characteristics, test characteristics, and the accommodations themselves. A multiple construct approach may facilitate more accurate evaluations of the effects of accommodated test scores

The last forty years of education policy in the United States have been marked by civil rights legislation, court cases, and school reforms (e.g. Americans with Disabilities Act of 1990; Rehabilitation Act of 1973). Central to these measures is the emphasis on equal opportunities for all people, including those with disabilities. The Individuals with Disabilities Educational Act (1997, 2004) and the No Child Left Behind Act (2001) sparked the most recent focus in education: *accountability* for student outcomes. Accountability reforms depend upon standardized, high-stakes assessments to measure student knowledge and skill. In many states, test scores now influence decisions regarding student graduation and grade promotion, teacher salaries, and the allocation of school resources. In the past, students with disabilities were excluded from standardized, high-stakes assessments (Mazzeo, Carlson, Voekl, & Lutkus, 2000). This exclusion was due, in part, to the inaccessibility of some standardized test formats (Bolt & Thurlow, 2004). However, recent legislation at

the federal and state levels now requires the inclusion of all students in national and state assessments (No Child Left Behind). These regulations have been a powerful impetus to find ways that standardized tests can be a fair and valid assessment of all students' abilities (Lehr & Thurlow, 2003).

In current practice, *assessment accommodations* are frequently given to allow students with disabilities to access test material and meaningfully participate in high-stakes assessment (Bolt & Thurlow, 2004). Accommodations involve changes to the test presentation, setting, or response format. Accommodations are meant to make it easier for students with disabilities to gain access to test content without changing the difficulty of the test while at the same time not changing what is being measured by the test. Research on the effects of accommodated test scores continues to grow but offers few conclusive findings on whether they facilitate fair and accurate

measurement of student knowledge and skill (e.g. Abedi, Hofstetter, & Lord, 2004; Bolt & Thurlow). Research findings differ depending on the type of accommodation, the nature of the student's disabilities, the test content, and even the way test validity is conceptualized by the authors. Results range from demonstrating that accommodations are beneficial, that they have no effect, or that they may sometimes create an unfair advantage for students who use them (e.g. Fletcher, Francis, Boudousquie, Copeland, Young, Kalinowski, & Vaughn, 2006; Fuchs, Fuchs, Eaton, Hamlett, & Karns, 2000; Hofstetter, 2003; Schulte, Elliott, & Kratochwill, 2001; Tindal & Fuchs, 2000; Tindal, Heath, Hollenbeck, Almond, & Harniss, 1998). Given the lack of consensus in the literature, state policies often vary in the type of accommodations students are allowed to use on high-stakes assessment (Clapper, Morse, Lazarus, Thompson & Thurlow, 2005). This variability leads to mixed conclusions of the validity of interpretations of accommodated scores and how students are included in accountability reforms (Cawthon, 2007).

This paper identifies areas where the research literature supports the need to consider multiple factors when determining the validity of interpretations of accommodated test scores. The literature review focuses on how variations in test and student characteristics affect the impact of two widely used accommodations: extended time and read aloud. These findings are used to propose a multiple construct approach to accommodations validity research that systematically includes student characteristics, test characteristics, and test accommodations.

Theoretical Basis for Validity and Accommodated Assessments

Central to the discussion of fair and appropriate accommodations use is the issue of test score validity. A *valid* interpretation of an accommodated score is one where the accommodation allowed students to access an assessment without changing the construct being assessed. Validity here refers to the *interpretation* of the score because it is in how the score is used, what it is assumed to represent in terms of student proficiency, where the validity construct comes into play. However, the term validity has been used in multiple ways in the research literature, muddying the discussion of this construct. In this paper, an accommodated score will be described as to its *accuracy*, whereas an accommodation will be described with degrees of *effect* and *fairness* to keep

the distinction from validity clear. A fair accommodation must thus in some way “speak to the nature of the disability”, addressing the barriers created by the interaction between the student's disability and the test item format (Fuchs, Fuchs, & Capizzi, 2005, p. 5). A valid interpretation of the accommodated score must therefore account for both the characteristics of the test and the test taker (Abedi, Leon, & Kao, 2008a and 2008b; Middleton & Laitusis, 2007; Stone, 2009).

In order to empirically measure the effect of an accommodation on test scores, some researchers have come to rely on the idea of “differential boost” (Phillips, 1994). In this framework, a fair accommodation increases the test scores of students with disabilities more than those of students without disabilities, providing a differential boost to students with disabilities (Elliott & Marquart, 2004). This differential boost represents the interaction hypothesis: that a fair accommodation will result in an interaction between accommodation status and disability status (Sireci, Scarpati, & Li, 2005). For example, if used on an assessment of mathematical reasoning, an accommodation meant to increase access to the math skills embedded in word problems might boost scores of students with reading disabilities more than of students without disabilities. Research on differential boost has occurred both on the test level, investigating whether an accommodation boosts overall test scores, and on the item level, examining whether an accommodation provides a differential boost on specific types of test items (Bolt & Thurlow, 2006, Fletcher et al., 2006). However, research suggests that the interaction between disability and accommodation is rarely straightforward (Calhoon, Fuchs, & Hamlett, 2000; Elliott & Marquart; Helwig & Tindal, 2003; Ketterlin-Geller, Yovanoff, & Tindal, 2003).

Factors that Affect Impact of Accommodations

No single accommodation has been shown to be beneficial to *all* students with disabilities; however, certain accommodations do benefit *some* students with disabilities (Fuchs, Fuchs & Capizzi, 2005). In fact, the effects of specific accommodations varies depending on student characteristics; an accommodation that benefits one type of student may not benefit, or may even negatively impact, the performance of another. An accommodated score's accuracy may also depend on matching the accommodation to student characteristics other than disability status in order to obtain measurable

differential boost. For example, a student with a reading disability with a lower level of proficiency in the test content area, such as math, may require a different accommodation than a student with a reading disability but who is closer to grade level in the given test subject.

One example of an important test characteristic that may interact with a student's disability is the *language demand* of the test content. Language demand can refer to the length of a test passage, syntactic structure, use of metaphors, and level of vocabulary. Abedi & Herji (2004) found that the language demand of a test item can have a significant effect on a student's ability to demonstrate content knowledge, particularly for students who are not native English speakers. Reducing language demand has thus been a particular focus for test developers and researchers as a way to improve access for English Language Learners (ELLs). Language demand is relevant here because, in addition to ELLs, there are students with disabilities who may also require simplified language in order to access test content (Bolt & Thurlow, 2006; Ketterlin-Geller et al., 2007). For example, students with reading disabilities, who are deaf, hard of hearing, or with auditory processing difficulties, face processing challenges when reading test passages (or having them read aloud to them). The language demand of the assessment may therefore interact with student characteristics and test content in measures of the effects of accommodations.

A concern also arises when achievement tests are used to measure the cognitive skills affected by students' disabilities (Fuchs et al., 2005). For instance, students with learning disabilities specific to reading, poor decoding or fluency skills will have difficulty accessing assessments of reading comprehension due to the nature of their disability. The connection between basic reading skills and reading comprehension creates concern about what type of accommodations can allow for a valid evaluation of students' understanding without transforming the test from an assessment of reading comprehension into an assessment of listening comprehension (Crawford, Helwig, & Tindal, 2004; Fletcher, Francis et al., 2006). In the example provided above, the accommodation may actually alter both the test construct and how test scores are interpreted (Abedi et al., 2004; Meloy, Deville, & Frisbie, 2002). In other words, accommodations may be unfair if, in an effort to mediate the effects of a student disability, the accommodation also impedes accurate measurement of domain-specific knowledge being tested (Haladyna & Downing, 2004). The content area, or subject of the test,

may therefore be an important factor in evaluating the effect of an accommodated test score.

In the sections that follow, we review the research literature for two kinds of assessment accommodations: extended time and read aloud. In each section below, we discuss how student disability, language demand, and content area interact to produce varying conclusions about the effect of extended time and read aloud accommodations. These are two of the most documented accommodations in the research literature. These are, however, only two examples of accommodations that might be included in further research on multiple constructs that affect the validity of how accommodated test scores are used. Other accommodations, such as having test items translated in the student's native language, or providing a dictionary or glossary, also likely interact with student and test characteristics in their impact on test scores.

Extended Time

Extended time is the most frequently used and allowed accommodation (Fuchs et al., 2005) and is often given in combination with other accommodations such as read aloud, Braille, or separate testing location. Extended time can range from time and half, double time, or unlimited time on the assessment. The differential boost literature presents mixed findings on the overall effect of extended time as an accommodation for testing. Early studies suggested that students with learning disabilities benefit from extended time when compared to students without learning disabilities (Runyan & Smith, 1991) but at times benefits were deemed too large to merit the use of accommodations (Willingham, 1988). These findings indicated that students without learning disabilities do not benefit from extended time because they are already working at their maximum potential under timed conditions. However, other studies concluded that students without learning disabilities also benefit from extended time, although not as much as the students with learning disabilities (e.g. Sireci, Scarpati & Li, 2005; Stretch & Osborne, 2005; Zuriff, 2000). These authors suggest that this minimal differential boost is not sufficient to say that the accommodation removes barriers to test content for students with learning disabilities. Further studies have shown no differential effects of extended time for students with disabilities (Elliott & Marquart, 2004; Johnson, Rudner, & Sibert, 2008), while other research provides evidence for a small advantage (Chiu & Pearson, 1999).

There appears to be a greater effect of extended time on test content that involves high levels of language demand. For example, the student performance on writing assessments was shown to improve among 5th graders when more time was allotted, with increased advantages (differential boost) for students in special education (Crawford, Helwig, & Tindal, 2004). However, they also found that the performance of 8th grade students improved with the use of extended time regardless of learning disability status (non-differential boost). On timed reading tests, a task requiring lower levels of language processing than writing, overall extended time has only shown to be slightly beneficial (Runyan & Smith, 1991). The assertion that extended time is most fair for tasks involving higher-order levels of language use also gains support from research on math assessments. Extended time on computation math assessments, that typically do not involve reading a passage or a word problem, has not demonstrated differential effects for students with and without disabilities (Fuchs et al., 2000; Munger & Lloyd, 1991). However, differential effects *were* found for students with disabilities when additional time was allotted to complete complex math assessments, which included reading and writing demands (Fuchs et al.).

Math and reading proficiency have been shown to be particularly important student characteristics that influence the effects of extended time. Most of the research focuses on the impact of math and reading skills on math test performance because students use both domains in responding to word problems (i.e. in contrast with computation items that have a low language demand). For example, when using an extended time accommodation on a math test, students who had primary difficulties in reading and no documented math difficulties performed differentially better—and thus benefited more from the accommodation—than did students with difficulties in math (Fuchs et al., 2000). An additional study examined the effects of extended time on the SAT, also a standardized achievement test (Mandinach, Bridgeman, Cahalan-Laitusis, & Tripani, 2005). Results from this study demonstrated that students both with and without disabilities in the middle math ability level benefited more from the accommodation on the math section. In none of the aforementioned studies did students with low math abilities benefit from extended time, further evidencing the influence of individual abilities on the effects of accommodations.

Read Aloud

Although not as frequently used as extended time, read aloud is commonly used as an accommodation for students with a range of characteristics. The basic premise of this accommodation is that the test item is read out loud to the student in order to remove the difficulty of reading the test item from the assessment process. Possible formats of the read aloud accommodation include oral presentation of test items by the test administrator, by computer, or via a video. Despite this variability in presentation agent, there is little evidence that the various formats of read aloud accommodations result in different outcomes on test scores (Calhoun et al., 2000).

As with extended time, the subject area of assessment is important to consider when measuring the effect of read aloud accommodations on test scores. For example, Tindal et al. (1998) found that students with disabilities did better on math assessments with a read aloud accommodation compared to students without disabilities who did not receive any accommodations, thus demonstrating a differential boost. Other studies show an overall gain with oral presentation for both students with and without disabilities, also on math assessments (Elbaum, 2007; Johnson, 2000; Meloy et al., 2002). In a study of reading assessments, students in regular education outperformed students in special education with the read aloud accommodation (McKevitt & Elliott, 2003). In this study, there was no differential boost, leading to the conclusion that the read aloud accommodation was not a benefit to students with disabilities. Although read aloud for a reading assessment may change the test from a reading task to a listening task, results from McKevitt and Elliott partially alleviates the concern that the accommodation inflates student scores.

In contrast with differential boost research on overall test scores, other studies about the effects of read aloud address this empirical question with item-level analyses. Bolt and Ysseldyke (2006) analyzed the differential item functioning of difficulty (DIF) reading items when given to students with and without a learning disability. If a test item “functions differently” between two groups, this indicates that the item is potentially more difficult for one group than another. Bolt & Ysseldyke found that using a read aloud accommodation created greater DIF, or increased measurement problems with the test items. This increase in DIF was more significant for the reading/language arts section, i.e. items with higher language demand, than for the

math test items. In a further study, Bielinski, Thurlow, Ysseldyke, Freidebach, J., & Freidebach, M. (2001) found DIF concerns when a read aloud accommodation was used on a reading test for students with a reading disability. The question of whether the reading test is meant to measure decoding skills or comprehension skills (or both) remains part of the dialog about the implications of using a read aloud accommodation on a reading assessment.

As with extended time, student characteristics may also interact with test content when measuring the effect of read aloud. A student's reading *proficiency* may be a significant factor in determining the benefit of a read aloud accommodation, especially on math assessments. The focus on math assessments is relevant because there are fewer concerns for read aloud on math than on reading tests. Studies have shown that students with low reading proficiency demonstrated greater gains when using oral presentation than those who are skilled readers (Meloy et al., 2002). Further, it has been proposed that only a subgroup of students with reading disabilities may benefit from read aloud accommodation (Bielinski et al., 2001). Item difficulty, or the relative challenge of the item for low and high performing students, may also influence effects of a read aloud accommodation. For example, Bolt and Thurlow (2006) found that the read aloud accommodation had a greater benefit for student scores on items that were difficult to read. In sum, the read aloud accommodation may not have an effect for skilled readers who can already access the written form of the assessment, but may be beneficial either for poor readers or on more difficult test items.

Discussion

The purpose of this paper was to review literature on the multiple constructs that may impact the effects of accommodated standardized assessments. There are significant challenges and opportunities to integrating multiple constructs in an assessment approach. This discussion section outlines limitations to the approach taken here and issues that may be resolved through future research.

Limitations

There are some limitations regarding the literature in this area. In this discussion, the only test characteristic mentioned was language demand, and the only accommodations discussed were read aloud and extended time. To some extent, this limited scope

reflects the still-emerging status of the research on the interaction of accommodations with test and student characteristics. For example, the relationship between test item format and the effects of an accommodation is also an important consideration in assessment of students with disabilities. We do not yet know how test characteristics such as response mode (i.e. computer vs. paper) and response type (i.e. student constructed vs. multiple choice) might interact with student characteristics and accommodations to affect the validity of test scores. A full review of multiple constructs might broaden its application to students with other disabilities, students with a wider range of English language proficiency, different kinds and combinations of accommodations, and contextual factors such as opportunity to learn tested content and policies guiding accommodations use. A significant limitation to this approach is that these components can only be added to a full model for research or practice when sufficient research exists to support their inclusion.

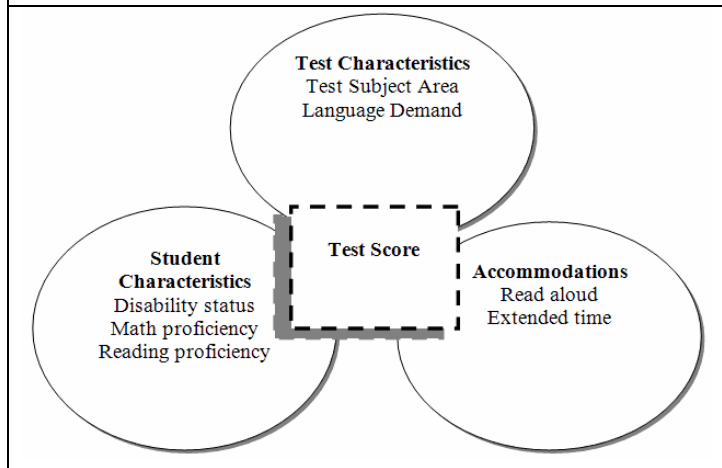
Future Research

Little research has assessed a model of accommodations' effectiveness that incorporates *multiple* item, student, and school characteristics (Schulte, Elliott, & Kratochwill, 2001). Fuchs, Fuchs, Eaton, Hamlett, & Karns (2000) have suggested that choosing an effective and fair accommodation may ultimately be dependent upon individual diagnosis. Their DATA tool begins this approach by incorporating information about the student's previous use of an accommodation when making decisions about future assessment practices. Kopriva and her colleagues developed a decision-making tool that incorporates student, school, and family language variables that affect assessments for ELL students (Kopriva, Emick, & Hipolit-Delgado, 2007). Despite the existence of these tools, there is no process yet available to systematically incorporate multiple characteristics that potentially impact accommodations' impact on test scores. Given the current research literature base on factors that affect accommodations validity, we propose the following theoretical approach for accommodations research (Figure 1).

This multiple construct framework could be used to refine research on the effects of accommodations on standardized test scores. Current models look mainly at the effects of an accommodation given one student characteristic (such as disability status or reading proficiency) and a single test characteristic (such as subject area). The model recognizes that multiple factors

may interact at the same time to affect the validity of the interpretation of an accommodated test score. Although only extended time and read aloud are listed here, this model can be applied to other accommodations (or combinations of accommodations). By including test,

Figure 1. Theoretical Model for Including Multiple Constructs in Accommodations Evaluation



Note. This figure represents a visual model of how test characteristics, student characteristics, and accommodations interact to result in the effects of an accommodated test score. Test score validity is a result of an interaction of all three of these factors. These are not all-inclusive lists, but examples given to illustrate how the multiple factor framework draws on the findings in the current research literature. Additional factors such as the level of instruction and the alignment of instruction and assessment may also play a role in the validity of the interpretation of an accommodated test score.

student, and accommodations factors, this framework can lead to a more nuanced measurement of accommodated score validity. Research results might ultimately result in more precise accommodations recommendations for students with diverse characteristics under a range of test conditions.

The complexity of assessment accommodations decisions, in conjunction with the potential scope of the multiple construct model, results in an extensive future research agenda. Areas of future research need to address not only the types of characteristics mentioned above—student characteristics, test characteristics, and accommodation characteristics—but also should address contextual issues that affect the fairness of an accommodated test score. There are practical realities that are not yet measured by empirical studies, but are still important in ensuring optimal effect of an accommodation. Some of these practical factors can be conceptualized as implementation issues, or how well the accommodation will work once the decision has

been made to use it for an assessment. For example, Individualized Education Program plan teams weigh whether or not the student has used the accommodation during instruction or other class activities when recommending the accommodation for assessment. Without prior use, the ineffectiveness (flat or decrease in score), of an accommodation may be due to implementation or logistical issues and not the accommodation itself. The interpretation of the resultant test score may thus be invalid if there are factors related to the consistency of accommodations use between instruction and assessment that affect a student’s test performance. Explicitly including the prior use of an accommodation in instruction into the multiple construct model may help to strengthen the assessment accommodation process.

An additional area for future research returns to the concept of differential boost that grounds much of the accommodations research literature. Within this framework, focus lies on the relative differences between accommodated and un-accommodated test scores of students with and without disabilities. The degree of boost and its implications for validity is currently an unexplored area. For example, how large a gap in scores is big enough to support the benefit of an accommodation for the students with disabilities? It may be that if the accommodation either has no differential impact or only very little, a case is made not to administer the accommodation because it does not appear to provide access to test content. On the other hand, a very large gap in scores may lead one to determine that the accommodation is giving an unfair advantage to students with disabilities, making the test content easier than for students who do not use the accommodation (Willingham, 1988). What are the lower and upper bounds to this “effective” differential boost? Studies that look at how different degrees of differential boost result in valid interpretations of test scores, particularly in the context of norms and cutoffs in the overall range of scores, would be particularly valuable in the current high stakes assessment context.

Conclusion

Accommodations research over the course of the last fifteen years leads us to an awareness of the need for specificity when looking at whether an accommodation fulfills its purpose: to increase access to test content for students with disabilities. The field recognizes that some accommodations are likely to be useful and fair, but that there are many factors that affect how cautious to be

about an accommodated test score. While early studies focused on “students with disabilities” as a group and, at times “accommodations” as a group, more recent work illustrates how a student’s individual characteristics or the language demand of the test item can have a strong influence on how much access an accommodation gives to test content. This approach reflects an overall perspective that could be generalized beyond students with disabilities. For students who are English Language Learners, factors at the student and school level, such as language of instruction and language of assessment, may interact with different test formats and supplemental dictionaries or glossaries. Focusing on the interaction of factors across different levels of the education context, from test item and format to student characteristics to classroom instruction, is an empirical challenge but reflects the complexity behind measurement of student knowledge and skill (Willingham, Pollack, & Lewis, 2002). If the field is to use standardized test scores to draw conclusions about the gain in academic proficiency for a diverse, inclusive student body, such an approach could lead us towards a stronger assessment and accountability system.

References

- Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometrics issues. *Educational Assessment*, 8(3), 231-257.
- Abedi, J. & Hejri, F. (2004). Accommodations for students with limited English proficiency in the National Assessment of Educational Progress. *Applied Measurement in Education*, 17(4), 371-392.
- Abedi, J. Hofstetter, C., & Lord, C. (2004). Assessment accommodations for English language learners: Implications for policy-based empirical research. *Review of Education Research*, 74(1), 1-28.
- Abedi, J., Leon, S., & Kao, J. C. (2008a). *Examining differential item functioning in reading assessments for students with disabilities*. (CRESST Tech. Rep. No. 744). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Retrieved May 19, 2009 from <http://www.cse.ucla.edu/products/rsearch.asp>.
- Abedi, J., Leon, S., & Kao, J. C. (2008b). *Examining differential distractor functioning in reading assessments for students with disabilities*. (CRESST Tech. Rep. No. 743). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Retrieved May 19, 2009 from <http://www.cse.ucla.edu/products/rsearch.asp>.
- Americans with Disabilities Act of 1990 (Public Law 101-336).
- Bielinski, J., Ysseldyke, J. E., Bolt, S., Friedebach, J. & Friedebach, M. (2001). Prevalence of accommodations for students with disabilities participating in a statewide testing program. *Assessment for Effective Intervention*, 26(2), 21-28.
- Bolt, S. E. & Thurlow, M. L. (2004). Five of the most frequently allowed testing accommodations in state policy. *Remedial and Special Education*, 25(3), 141-152.
- Bolt, S. E. & Thurlow, M. L. (2006). Item-Level Effects of the Read-Aloud Accommodation for Students With Reading Disabilities. *Assessment for Effective Intervention*, 33, 15-28.
- Bolt, S. E. & Ysseldyke, J. E. (2006). Comparing DIF across Math and Reading/Language Arts Tests for Students Receiving a Read-Aloud Accommodation. *Applied Measurement in Education*, 19 (4), 329-355.
- Calhoon, M. B., Fuchs, L. S. & Hamlett, C. L. (2000). Effects on computer based test accommodations on mathematics performance assessments for secondary students with learning disabilities. *Learning Disability Quarterly*, 23, 271-282.
- Cawthon, S. (2007). Hidden benefits and unintended consequences of No Child Left Behind policies for students who are Deaf or hard of hearing. *American Educational Research Journal*, 44(3), 460-492.
- Chiu, C. W., & Pearson, P D. (1999). Synthesizing the effects of test accommodations for special education and limited English proficient students. Paper presented at the *National Conference on Large Scale Assessment*, Snowbird, UT.
- Clapper, A. T., Morse, A. B., Lazarus, S. S., Thompson, S. J., & Thurlow, M. L. (2005). *2003 state policies on assessment participation and accommodations for students with disabilities* (Synthesis Report 56). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Available at <http://education.umn.edu/NCEO/OnlinePubs/Synthesis56.htm>
- Crawford, L., Helwig, R., & Tindal, G. (2004). Writing performance assessments: How important is extended time. *Journal of Learning Disabilities*, 37, 132-142.
- Elbaum, B. (2007). Effects of an oral testing accommodation on the mathematics performance of secondary students with and without learning disabilities. *Journal of Special Education*, 40(4), 218-229.
- Elliott, S. N., & Marquart, A. M. (2004). Extended time as a testing accommodation: Its effects and perceived consequences. *Exceptional Children*, 70, 349-367.
- Fletcher, J., Francis, D. J., Boudousquie, A., Copeland, K., Young, V., Kalinowski, S., & Vaughn, S. (2006). Effects of accommodations on high-stakes testing for students with reading disabilities. *Council for Exceptional Children*, 72(2), 136-150.

- Fuchs, L., Fuchs, D., & Capizzi, A. M. (2005). Identifying appropriate test accommodations for students with learning disabilities. *Focus on Exceptional Children*, 37, 1-8.
- Fuchs, L., Fuchs, D., Eaton, S., Hamlett, C. L., & Karns, K. (2000). Supplementing teachers' judgments of mathematics test accommodations with objective data sources. *School Psychology Review*, 29, 65-85.
- Haladyna, T. M. & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice* 23(1), 17-27.
- Helwig, R. & Tindal, G. (2003). An experimental analysis of accommodation decisions on large-scale mathematics tests. *Exceptional Children*, 69(2), 211-225.
- Hofstetter, C. (2003). Contextual and mathematics accommodation test effect for English language learners. *Applied Measurement in Education*, 16(2), 159-188.
- Individuals With Disabilities Education Act Amendments of 1997, 20 U.S.C. § 1400 *et seq.*
- Individuals with Disabilities Education Improvement Act, as amended, 20 U.S.C. § 1400 *et seq.* (IDEIA)
- Johnson, E. S. (2000). The effects of accommodations on performance assessments. *Remedial & Special Education*, 21, 261-268.
- Johnson, K., Rudner, L. & Sibert, R. (2008). *Do accommodated GMAT test takers have an unfair advantage*. GMAC Research Report Series RR-88-02, July, 11, 2008.
- Ketterlin-Geller, L., Yovanoff, P., & Tindal, G. (2007). Developing a new paradigm for conducting research on accommodations in mathematics testing. *Exceptional Children*, 73(3), 331-347.
- Kopriva, R. J., Emick, J. E., & Hipolito-Delgado, C. P. (2007). Do proper accommodation assignments make a difference? Examining the impact of improved decision making on scores for English language learners. *Educational Measurement: Issues and Practice*, 26(3), 11-20.
- Lehr, C., & Thurlow, M. (2003). Putting it all together: Including students with disabilities in assessment and accountability systems (Policy Directions No. 16). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Mandinach, E., Bridgeman, B., Cahalan-Laitusis, C., & Tripani, C. (2005). *The impact of extended time on SAT test performance*. New York: The College Board.
- Mazzeo, J., Carlson, J. E., Voekl, K. E., & Lutkus, A. D. (2000). *Increasing the participation of special needs students in NAEP* (NCES 2000-473). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.
- McDonnell, L., McLaughlin, M. & Morison, P. (1997) *Educating one & all: Students with disabilities and standards-based reform*. Washington, DC: National Academy Press.
- McKevitt, B. C. & Elliott, S. N. (2003). Effects and perceived consequences of using read aloud and teacher recommended testing accommodations on a reading achievement test. *School Psychology Review*, 32, 583-600.
- Meloy, L. L., Deville, C. & Frisbie, D. A. (2002). The effects of a read aloud accommodation on test scores of students with and without a learning disability in reading. *Remedial and Special Education*, 23(4), 248-255.
- Middleton, K., & Laitusis, C. C. (2007). *Examining test items for differential distractor functioning among students with learning disabilities*. (Research Report No. 07-43). Princeton, NJ: Educational Testing Service.
- Munger, G. F., & Lloyd, B. H. (1991). Effect of speededness on test performance of handicapped and nonhandicapped examinees. *Journal of Educational Research*, 85, 53-58.
- No Child Left Behind Act of 2001 (Public Law 107-110).
- Phillips, S. E. (1994). High-stakes testing accommodations: Validity versus disabled rights. *Applied Measurement in Education*, 7(2), 93-120.
- Rehabilitation Act of 1973 (Public Law 93-112)
- Runyan, M. K., & Smith, J. (1991). Identifying and accommodation learning disabled law school students. *Journal of Legal Education*, 41, 317-349.
- Schulte, A. G., Elliot, S. N., & Kratochwill, T. R. (2001). Effects of testing accommodations on standardized mathematics test scores: An experimental analysis of the performances of students with and without disabilities. *School Psychology Review*, 30(4), 527-547.
- 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*, 75(4), 457-490.
- Stone, E. (2009). *Examining the fairness of an English-language arts assessment for students who are deaf or hard of hearing using differential distractor functioning analysis*. Paper presented at the annual meeting of the National Council on Measurement in Education; San Diego, CA.
- Stretch, L.S. & Osborne, J.W. (2005). Extended time test accommodation: Directions for future research and practice. *Practical Assessment, Research & Evaluation*, 7, 1-6.
- Tindal, G. & Fuchs, L. (2000). A summary of research on test changes: An empirical basis for defining accommodations. Report for Mid-South Regional Resource Center, University of Kentucky. Available at <http://www.rfcnetwork.org/content/view/526/463/>
- Tindal, G., Heath, B., Hollenbeck, K., Almond, P. & Harniss, M. (1998). Accommodating students with disabilities and large scale tests: An empirical study of student response and test administration demands. *Exceptional Children*, 64, 439-450.

Willingham, W. (1988). Testing handicapped people: The validity issue. In Wainer, H. & Braun, H. (Eds.) *Test Validity*. NJ: Earlbaum.

Willingham, W., Pollack, J., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement*, 39(1), 1-37.

Zuriff, G. E. (2000). Extra examination time for students with learning disabilities: An examination of the Maximum Potential Thesis. *Applied Measurement in Education*, 13, 99-117.

Note

The preparation of this manuscript was supported by a University of Texas at Austin Research Grant awarded to the first author. The authors would like to thank the following individuals for their helpful feedback and contributions to this manuscript: Deborah Tharinger, Tim Keith, Judy Wan, Saul Natada, Sarah Pettit, and Emma Cole.

Citation

Cawthon, Stephanie W., Ho, Eching, Patel, Puja G., Potvin, Deborah C., and Trundt, Katherine M. (2009). Multiple Constructs and the Effects of Accommodations on Standardized Test Scores for Students with Disabilities. *Practical Assessment, Research & Evaluation*, 14(18). Available online: <http://pareonline.net/getvn.asp?v=14&n=18>.

Corresponding Author

Authorship is listed in alphabetical order

For correspondence, contact:

Stephanie W. Cawthon
Department of Educational Psychology
The University of Texas at Austin
1 University Station, Mail Code D5800
Austin, Texas, 78712
Stephanie.cawthon [at] mail.utexas.edu.