



What Do State Reading Test Specifications Specify?



Partnership for Accessible
Reading Assessment



Partnership for Accessible Reading Assessment

University of Minnesota
350 Elliott Hall
75 E. River Road
Minneapolis, MN 55455

<http://www.readingassessment.info>
readingassess@umn.edu

This work is supported, in part, by the U.S. Department of Education, Institute of Education Sciences, National Center for Special Education Research—Grant No. H324F040002. Opinions expressed do not necessarily reflect those of the U.S. Department of Education or offices within it.



**NATIONAL
CENTER ON
EDUCATIONAL
OUTCOMES**

UNIVERSITY
OF MINNESOTA
**Curriculum and
Instruction**



UC DAVIS



The University of Minnesota is committed to the policy that all persons shall have equal access to its programs, facilities, and employment without regard to race, color, creed, religion, national origin, sex, age, marital status, disability, public assistance status, veteran status, or sexual orientation.

This document is available in alternative formats upon request.

What Do State Reading Test Specifications Specify?

Christopher J. Johnstone, Martha L. Thurlow, Ross E. Moen, Danielle Matchett, Kristin E. Hausmann, and Sarah Scullin

June 2007

All rights reserved. Any or all portions of this document may be reproduced and distributed without prior permission, provided the source is cited as:

Johnstone, C. J., Thurlow, M. L., Moen, R. E., Matchett, D., Hausmann, K. E., & Scullin, S. (2007). *What do state reading test specifications specify?* Minneapolis, MN: University of Minnesota, Partnership for Accessible Reading Assessment.

Introduction

“Reading is reading.” “All reading tests measure basically the same thing anyway.” “Just pick a test.” These are the kinds of statements that professionals in the field of education sometimes make when confronted with the challenge of trying to develop a measure of student reading achievement. If this view of reading and reading tests is accurate it would make life much easier for many people. So, how well does this simple view fit the facts?

According to the Reading First program cited in the No Child Left Behind Act of 2001, reading is “a complex system of deriving meaning from print.” The Reading First definition states that deriving meaning from print involves (a) phonetic knowledge, (b) the ability to decode unfamiliar words, (c) fluent reading, (d) sufficient background and vocabulary to foster comprehension, (e) development of active strategies to construct meaning from print, and (f) the development and maintenance of motivation to read (No Child Left Behind Act of 2001). Have states found a one-size-fits-all approach to measuring student progress on this complex and varied set of knowledge, skills, abilities, and attitudes?

The No Child Left Behind Act requires all states to assess student reading, but each state is responsible for selecting what will be tested and how in their large-scale statewide assessments. As part of this process, states develop standards with which both instruction and assessments are expected to align. State standards for reading vary by definition and focus from state to state. For example, Thompson, Johnstone, Thurlow, and Clapper (2004) in their review of state standards identified 28 themes of standards that they classified into five major categories (specific skills, elements of literature, interactive activities, reading to solve problems, and reading for personal growth). Not all states had standards in all areas, and standards were frequently different from state to state and by grade level.

If reading standards differ from state to state, then it is likely that reading assessments also differ from state to state. This report examines that likelihood by analyzing the documents states use to specify what should go into their assessments. These documents, called test specifications or test “blueprints,” guide vendors in the process of designing assessments. In addition to examining how these test specifications compare from state to state, we also examined how states specifications for their assessments compared to what had been found in states’ standards. This gives a sense of the relationship between state assessments and state standards.

As important as these issues may be for work on reading assessments in general, they are especially important for work on assessing reading for students with disabilities. The interaction of disability and reading for some students with disabilities presents serious challenges for assessing the reading capabilities of these students. Sensory, intellectual, and other processing impairments along with concomitant deficits in general or specific background knowledge may cause a variety of reading issues at various points of reading

development and with various effects on the holistic process of reading. Consequently, assessments that require students to engage in such activities may be problematic for some students specifically because of their disabilities. Efforts to design accessible reading assessment for students with disabilities seek ways of circumventing disability based obstacles to seeing what these students can do without violating essential reading constructs. An important part of these efforts is understanding as clearly as possible what states consider essential to the construct of reading that they are trying to measure at each grade level.

In this paper, we report an investigation into state reading assessments that involved examining state assessment blueprints or test specifications. We looked at (1) themes related to the purposes and constructs of assessments, (2) how those themes related to state standards, (3) the number of items assigned to particular constructs, and (4) the types of items typically found in statewide assessments.

Finding Themes in State Assessments of Reading: Methods

State assessments are laid out for test developers in documents that are also often made available to the public. Test specifications are documents designed by states to inform vendors about the desired characteristics of their assessments. Test specifications typically have a short narrative portion explaining to vendors what broad information is desired on assessments. They also contain information on the constructs that should be tested, the number of items desired, and the types of items desired (e.g., forced response, multiple choice, or constructed response). States often do not have the resources to develop items in-house, so they contract with test vendors to develop items within given specifications. These specifications are available to the public and are frequently posted on State Department of Education Web sites.

Test specifications may be general in nature, or may be organized into one of three categories: (1) statistical specifications (provide the weight of each test item, the number of items, and the difficulty level); (2) content specifications (describe content to be covered); and (3) item specifications (refer readers to the constructs measured and the content standards for which items are to be aligned).

“Blueprints” are another term used by states to describe state requirements for assessments. This term is used interchangeably with test specifications across states without any language specifying what differentiates “specifications” and “blueprints.” Both terms are architectural in nature, and describe, for vendors, state plans for how to build an assessment.

Blueprints are frequently available for public viewing on state department of education Web sites. These documents most often are formatted as charts, which contain references to state content standards and the number of items that should assess such standards. Blueprints may also include desired test objectives, topics, or tasks; may

include subtopics within major performance standards; or may rank topics in order of importance.

For the purposes of this study, we elected to review documents from grades 4 and 8. Our selection was meant to address grades that in the early years of No Child Left Behind had assessment reporting requirements (elementary and secondary); they also reflected grades in which the National Assessment of Educational Progress was administered. Furthermore, the reading activities found in grade 4 differ from those in grade 8 because reading expectations and activities become more complex as students progress through school (Thompson et al., 2004).

Methods

The methods selected for this study were document analysis methods. In this methodology, researchers search and read all relevant documents related to a particular topic. After reviewing all documents, themes that emerge are coded with one or two-word descriptors, and used as evidence to support claims and assertions (Lincoln & Guba, 1985). In the case of a finite number of documents (e.g., documents from 50 states), tallies of themes become meaningful (e.g., a theme that arose in a majority of states may have implications versus a theme that arose in only a few states).

In this research, we investigated state assessment blueprints and specifications. The specifications and blueprints were downloaded from state Web sites or sent by request from state education departments. In total, we were able to obtain 49 state test specifications or blueprints from the World Wide Web or from contacts within the state department of education. We were unable to obtain information from one state that declined to share its test blueprints.

We organized test specification or blueprint data in a spreadsheet. As we entered data and visually inspected tables for themes, we noted that states' test specifications or blueprints addressed two major concerns: (1) the purpose of the test (the constructs or standards assessed), and (2) the breakdown of the tests (the number of items, the type of items, and the number of items dedicated to each construct or standard).

For our analysis of state test specifications and blueprints, we organized results into these two categories, "purpose" and "breakdown." Findings in the purpose category provide a snapshot of the areas of focus most often found in state blueprints and specifications. These areas of focus are then compared to earlier findings of state standard themes (see Thompson et al., 2004). Finally, the "breakdown" of items by type is reported.

Results

State Assessment Purposes

Each of the 49 state blueprints explicitly stated the constructs to be assessed in the state assessment. Each of the blueprints or specifications reviewed indicated that states assess a wide variety of competencies and skills on reading assessments. We coded these skills and competencies into three major categories: (1) foundational skills, (2) comprehension, and (3) analysis and interpretation of text. Sub-categories were also distinguished within each of these broad assessment categories. These are shown in Table 1, which portrays the number and percentage of states that directly assess each subcategory.

Foundational Skills

As noted in the Reading First definition, foundational skills are essential to the reading process. Many but by no means all states explicitly include coverage of these foundational skills in their test specifications. We distinguished four subcategories of foundational skills in our analysis—vocabulary, word identification, word analysis, and fluency.

We coded test specification information as “vocabulary” when it included terms such as “vocabulary,” “vocabulary acquisition,” “vocabulary development,” and “vocabulary strategies.” Our coding shows for both grade 4 and grade 8 that 17 out of 49 states (35%) directly test vocabulary.

Table 1. Percentage of State Blueprints or Specifications Dedicated to Assessment Purposes*

<i>Purpose (Construct)</i>	Grade 4		Grade 8	
	<i>Count</i>	<i>Percentage</i>	<i>Count</i>	<i>Percentage</i>
Foundational Skills				
Vocabulary	17	35	17	35
Word Identification	14	29	6	12
Word Analysis	8	16	8	16
Fluency	4	8	4	8
Comprehension				
Reading Comprehension	28	57	29	59
Comprehension: Literary Text	12	24	13	27
Comprehension: Expository Text	8	16	11	22
Analysis and Interpretation				
Analysis and Interpretation (general)	18	37	18	37
Analysis: Literary Text	12	24	13	27
Analysis: Expository Text	10	20	10	20

*Based on 49 states

We used the two-word code “word identification” to describe constructs in state blueprints and test specifications such as “word identification,” “word recognition,” “word study,” and “word identification skills and strategies.” Of 49 states, 14 (29%) directly assessed word identification in grade 4, and 6 (12%) directly assessed word identification in grade 8.

When terms such as “word analysis,” “word structure,” “word pattern,” “structural analysis,” and “cue systems for meaning of unfamiliar words” were found in state blueprints and specifications, we coded them with the words “word analysis.” In both grades 4 and 8, eight states (16%) included word analysis skills in their blueprints or specifications.

Finally, under the broad theme category of foundational skills, the construct “fluency” (meaning the ability to read with prosody) was rarely found in state blueprints and test specifications. Only four states (8%) in grade 4 and four states (8%) in grade 8 directly mentioned the term in documents. One reason for the low number of states using the term fluency in their documents may be that it is difficult to assess fluency using typical large-scale, paper and pencil assessments.

Comprehension

Over the past five years, reports from four major educational and assessment bodies – the National Reading Panel (National Reading Panel, 2000), the RAND Reading Study Group (Snow, 2002), Progress in International Reading Literacy Study Assessment (PIRLS, 2001) and the National Assessment of Educational Progress Summary of Reading Framework (National Assessment Governing Board, 2005) have noted that comprehension of text is the primary goal of reading.

Our review of blueprints and specifications indicates that comprehension is a construct that generates much attention from states. Working within the theme category of relatively literal comprehension, we distinguished three sub-themes. The first was given the generic label of “reading comprehension.” In addition to this generic theme, states distinguish two themes by type of text: “comprehension of literary text,” and “comprehension of expository text.” We counted states as referencing a generic reading comprehension construct if their test specifications or blueprints included the terms “forming a general understanding,” “forming an initial understanding,” “determining meaning,” “comprehend,” “main idea,” “demonstrate understanding,” “basic understanding,” “comprehension strategies,” “reading comprehension,” “literal comprehension,” or “comprehension of printed materials.” In grade 4, 28 (57%) of states and in grade 8, 29 (59%) of states had some such form of generic “reading comprehension” in their blueprints or test specifications.

We counted states as referencing “comprehension of literary text” if their blueprints contained the terms “comprehend literary/recreational material,” “comprehension of literary text,” “narrative text comprehension,” “literary comprehension,” or “initial

understanding of literary text.” Twelve states (24%) in grade 4 and 13 states in grade 8 (27%) referred to comprehension of literary text in their documents.

States also included comprehension of expository (informational) text in their test blueprints and test specifications. Phrases such as “comprehend textual and functional materials,” “comprehension of informational text,” “reading for information,” “informational text comprehension,” and “initial understanding of informational text” were all counted as constructs found in the theme “comprehension of expository text.” Eight (16%) states included this construct in their grade 4 test specifications or blueprints, and 11 (22%) states included “comprehension of informational text” in their grade 8 blueprints or specifications.

Analyze and Interpret Text

In addition to relatively literal comprehension of text, state blueprints and test specifications also included constructs related to higher-level comprehension tasks of analyzing and interpreting text. The structure of sub-themes found in literal comprehension is repeated here. Specifically, some state blueprints required general text analysis, some required items related to the analysis of literary text, and some required items related to the analysis of expository text.

States that required items related to generic analysis and interpretation of text used expressions like “evaluate: setting, mood, characterization,” “demonstrating a critical stance,” “interpreting meaning,” “extending meaning,” “interpretive comprehension,” “evaluative comprehension,” “expanded comprehension,” “cognition, interpretation, critical stance, and connections,” “critical literacy,” “analysis using reading strategies and critical thinking,” “analysis of general content or structure,” and “reading and responding” to describe constructs related to analysis and interpretation of text. In both grade 4 and grade 8, 18 states (37%) requested items related to generic text analysis and interpretation.

States also specified that an important assessment construct is the analysis and interpretation of literary text. Phrases such as “literary response and analysis,” “literary analysis,” “read, analyze, and respond to literature,” “interpret and evaluate literature,” “analysis and interpretation of literary text,” “reading application – literary,” “identify literary elements,” “literary response and expression; critical analysis and evaluation,” and “recognize and use features of narrative text” were all coded as “analysis and interpretation of literary text.” Twelve (24%) states in grade 4 and thirteen (27%) states in grade 8 had references to analysis and interpretation of literary text in their test specifications or blueprints.

Finally, 10 (20%) states in grade 4 and 10 (20%) states in grade 8 described assessment requirements relating to the analysis and interpretation of expository texts in their test specifications or blueprints. Phrases such as “locate, select, synthesize information,” “informational analysis (content and practical passages),” “interpret and evaluate informational text,” “analysis and interpretation of informational text,” “informational

text – response, expression, analysis, evaluation,” “reading application – informational,” “using informational resources,” “informational- response and expression; critical analysis and evaluation,” “reading application-information,” “recognize and use features of informational text,” and “informational text- interpretation” were all coded as “analysis and interpretation of expository text.”

Constructs Found in State Blueprints Compared to State Standards

Three major theme categories, each with three to four sub-themes, emerged from our analysis of state test specifications and blueprints. Although organized a bit differently from the thematic structure Thompson et al. (2004) reported in their review of state standards, the three blueprint themes cover much of the same content as three of the major themes identified in states standards. A fourth theme found in state standards is only weakly represented in test blueprints and a fifth theme found in state standards is virtually missing altogether from test blueprints and specifications.

The two blueprint theme categories of foundational skills and analysis and interpretation align fairly closely with the standards theme categories of specific skills and interactive/ thinking activity. The third blueprint theme category, comprehension, does not have a direct match in the standards classification system, but it incorporates the literal comprehension sub-theme that had been classified in the standards specific skills theme category and it also has sub-themes related to comprehending literary and expository texts that the standards classification system had assigned to a theme labeled “knowledge of conventions.” The distinction between literary and expository text types is also a part of the blueprint analysis and interpretation theme category. So although the standards theme category knowledge of conventions was not identified as a theme category in blueprints analysis, its basic elements were assigned to two themes in the blueprints classification system.

In contrast, two broad themes noted by Thompson et al. (2004) as present in state standards were either completely missing from test blueprints or were present but had different applications. Thompson et al. identified “catalyst for personal growth” (p 10) as a theme related to students reflecting upon themselves as individuals through reading and becoming literate citizens. States who include these themes in standards (anywhere from 19-41 states, depending on grade level and sub-theme) expect students to develop as readers and citizens through interacting with print. Such constructs are not found in large-scale state assessments.

In the broad theme category, “reading as a problem solving tool,” Thompson et al. (2004, p. 9) found state standards requiring students to solve problems, mine information, organize information, and follow instructions in authentic research endeavors. Although such activities were found in state blueprints and were classified as part of the broad theme of analysis and interpretation, state blueprint data indicate that the tested research requirements are typically limited to information that can be gleaned from provided passages.

In sum, according to state test specifications and blueprints, most state standards are assessed in some way on state assessments. Standards related to the personal development of children are not currently measured in large-scale assessments. Likewise, standards that require authentic research activities are limited to the information provided in the test itself. Of the 27 themes Thompson et al. found in state reading standards, 22 are found in state test blueprints and assessments at least to some extent. Four of these 22 are assessed in a manner substantially different from that indicated in state standards. A visual representation of this comparison is found in Table 2.

Table 2. Constructs Found in State Standards vs. State Reading Assessment Blueprints*

Construct	Standards	Blueprints
Compare	X	X
Context	X	X
Critical Analysis	X	X
Expository Elements	X	X
Fluency	X	X
Following Instructions	X**	X
Higher Order Thinking	X	X
Inferential Comprehension	X	X
Language Conventions	X	X
Literal Comprehension	X	X
Literary Elements	X	X
Meta Literacy	X	X
Mine Information	X**	X
Organize Information	X**	X
Personal Growth Through Reading	X	
Phonemic Knowledge	X	X
Problem Solving	X**	X
Question Authors	X	X
Reflection Upon Print	X	X
Reflection Upon Culture Through Reading	X	
Reflection Upon Self Through Reading	X	
Relate to Real Life	X	
Respond to Text	X	X
Socialization Into Literary Communities	X	
Strategic, Non-Phonemic Skills	X	X
Vocabulary	X	X
Word Recognition	X	X

*"X" means that at least one state had a particular construct represented in its standards or blueprints

** Standards were intended for authentic research activities that may or may not be reflected in large-scale assessments.

Breakdown of Tests

Items by Construct

Of the 49 states with data included in this study, 28 states for grade 4 and 29 states for grade 8 had information that clarified exactly how many items were to address particular constructs in their state assessments. Using the three major areas of foundational skills, comprehension, and analysis and interpretations, we noted that there were great discrepancies in the weight (i.e., number of items) that states assigned to particular constructs.

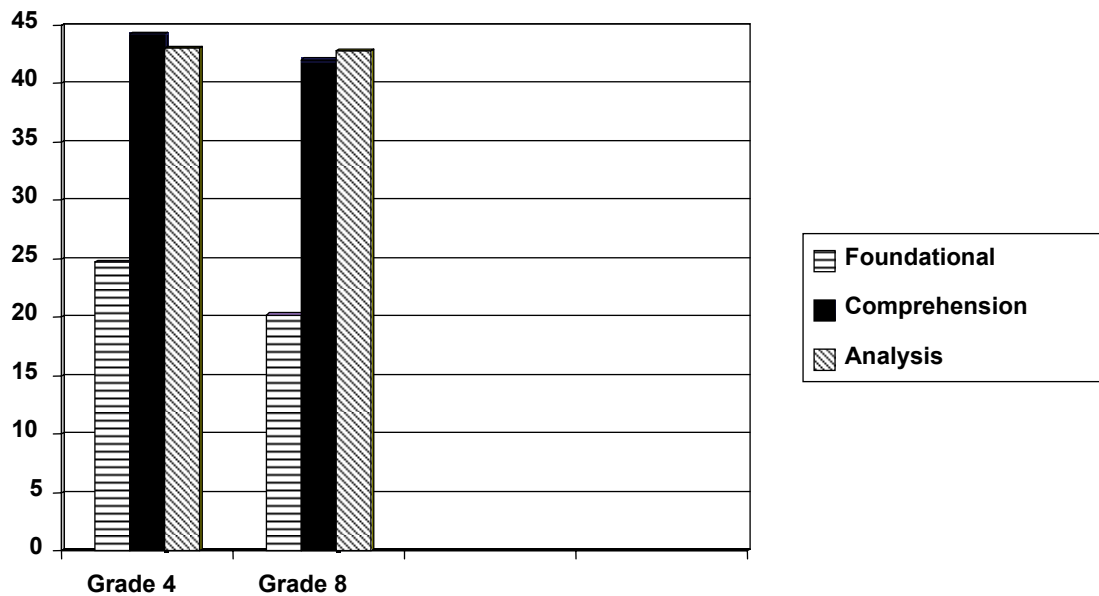
For example, among the 28 states in grade 4 that indicated that foundational skills would be assessed, states ranged from designating 0% of their items to foundational skills to assigning 60% of their items to foundational skills. The mean percentage of items designated for foundational skills in grade 4 for these states was 24%; the median was 22%. In grade 8, twenty-nine states that required measurement of foundational skills designated from 0% to 47% of items for measuring foundational skills. The mean percentage of items designated for foundational skills in grade 8 for these states was 20%; the median percentage was 24%.

There was also considerable variability among states in the emphasis placed on comprehension constructs. Among the 27 states that were classified as requiring measurement of reading comprehension, in both grade 4 and grade 8 the percentage of items designated as measuring comprehension ranged from 0% to 100% of items. Note that for both grades, states that placed zero emphasis on comprehension placed a proportionally high emphasis on analysis and interpretation constructs. That is, no state had a test made up of all foundational skills. The mean percentage of items designated for comprehension was 41% in grade 4 and 41% in grade 8. There was relatively little difference in weightings between grades 4 and 8, demonstrating that, among states that quantify importance of constructs, comprehension appears equally important in grades 4 and 8.

Similar to comprehension constructs, analysis and interpretation constructs were frequently found in blueprints and test specifications. In both grades 4 and 8, specific item requests for analysis and interpretation constructs ranged from zero items (0%) to all items (100%). The median percentage for grade 4 was 42% of items designated for analysis and interpretation, while the median percentage for grade 8 was 50%. In both grade 4 and grade 8, the mean number of items was 43%.

The preceding paragraphs document considerable variability among states in what they emphasize. There are large ranges in the percentages of items different states designate for measuring various constructs. A converse image of two key points of convergence is shown in the graph in Figure 1. First, foundational skills, comprehension, and analysis and interpretation constructs are all found in both grade 4 and 8 assessments. Second, in both grade 4 and grade 8, comprehension and analysis and interpretation items are found to a greater extent in state assessments than items testing foundational items.

Figure 1. Percentage of Items Dedicated to Particular Constructs



Item Types

Of the 49 states from which we obtained data, 23 had specific information pertaining to item types in the test specifications or blueprints. Among these 23 states that provided specific information on item types, multiple choice items appeared to be the item type preferred. The mean proportion of multiple choice items was 83% in grade 4 and 89% in grade 8. State blueprints and test specifications do not stipulate the proportion of time that would be spent completing multiple choice items, but multiple choice items clearly are an important feature in many state assessments.

Although multiple choice was the most common item type found in blueprints by states that provided item-specific information, there was a small but notable presence of requests for short answer, constructed response, and open-ended items in state blueprints and test specifications. For the purposes of this report, “short answer,” “constructed response,” and “open-ended” item requests were collapsed, because state information was often not sufficient to differentiate these types of items.

In grade 4, 18 states of the 23 providing item-specific information requested constructed response items. The percentage of items ranged from 5-100% of the total assessment. The mean percentage of constructed response items in grade 4 assessments was 17%.

In grade 8, nine states of the 23 providing item-specific information requested *no* constructed response items indicating that these states’ tests are completely multiple choice. Fourteen states, however, requested *some* constructed response items. The percentage of constructed response items ranged from 3% to 41% of the total assessment. The mean percentage of assessments made up of constructed response items was 12% in grade 8; the median was 10%.

In summary, 23 states requested a specific number (or proportion) of items in their test specifications or blueprints to be dedicated to either multiple choice or constructed response items. In both grade 4 and grade 8, multiple choice items were the dominant mode of assessment; the average percentage of items designated as multiple choice was more than 80% in each grade. Constructed response items made up on average 17% in grade 4 and 12% in grade 8 of items on assessments. While these numbers may seem counterintuitive, the actual score points assigned to items result in more constructed response points in grade 8 than in grade 4.

Results Summary

Our review of state assessment blueprints and test specifications leads us to the following conclusions. First, there is enough commonality among state blueprints and test specifications to permit examination of three kinds of information: (1) the nature of the content or constructs that states stipulate should be included in their assessments, (2) the amount of emphasis that states want placed on different constructs, and (3) the extent to which different types of test items will be used in state's assessments. Second, content that can be classified as relating to foundational skills, comprehension, or analysis and interpretation constructs can be found in statewide assessments as it was in state standards. In contrast, constructs related to the personal development of readers through literary activities that was found in state standards are not found in state assessment documents and the state standards constructs related to using reading as a problem solving tool have weak representation in assessment specifications. Third, blueprint and test specification data indicate that comprehension and analysis/interpretation items are found approximately two times as frequently as foundational skill items. Fourth, although some states use constructed response items extensively, multiple choice items are the dominant mode of assessment in most states. Finally, despite the commonalities that can be identified, there is wide variability among states in the amount of emphasis they put on different constructs and in the extent to which they use different types of assessment item formats.

Implications for Students with Disabilities

The interaction of disability and reading presents a challenge for researchers wishing to make general conclusions for assessment purposes. Sensory, intellectual, and processing impairments may cause a variety of reading issues at various points of reading development and with various effects on the holistic process of reading.

For example, many state reading assessments clearly place some emphasis on foundational skills. Students with speech and language impairments, however, often show poor phonological skills and are frequently unable to connect symbols and sounds (Gillon, 2002). Likewise, the most common challenge for students with learning disabilities relates to decoding print (Vaughn, Bos, & Schumm, 2004). Decoding print is

also a challenge for students with behavioral disorders (Barton-Arwood, Wehby, & Falk, 2005). Finally, students who have mental retardation face significant challenges when attempting to connect sounds and symbols in print (Browder & Xin, 1998).

For some students with visual impairments, the main struggle is with seeing print. The difficulties these students have with gaining messages from printed material are often addressed through accommodations such as braille or large print tests. Braille and large print tests are standard offerings across states for assessments (Clapper et al., 2005).

Other students with sensory impairments (visual impairments/blindness or deaf/hard of hearing) often have comorbid disabilities (Cline, Johnstone, & King, 2006), which may influence their foundational reading skills. For example, students who are deaf have significant challenges in terms of foundational reading skills. Students who cannot hear sounds naturally struggle with the ability to connect sounds to symbols. As such, a recent policy decision in Maryland allowed deaf students to not be scored for sections of a large scale assessment that required students to demonstrate proficiency in rhyming activities (activities that required students to connect print to sounds) (Thurlow, Johnstone, Thompson, & Case, in press).

Gersten, Fuchs, Williams, and Baker (2001) found that students with learning disabilities also struggle with reading comprehension activities. Gersten et al.'s meta-analysis of research attributed many of the comprehension problems faced by students with learning disabilities to a lack of foundational skills (i.e., foundational skills act as a gateway to comprehension). Edyburn's (2000) assistive technology research found that students with learning disabilities can comprehend printed material at a higher rate when it is presented in audio formats.

Students with disabilities also face challenges when attempting to analyze and interpret text. Algozzine and Wood (1994) found that the cognitive delays associated with mental retardation create analysis difficulties for students in this disability category. Specifically, students have difficulty understanding complex text or text that is not personally engaging (Katims, 2000). Students with autism spectrum disorders have a wide range of intellectual functioning (approximately 20% of students with autism spectrum disorders are within the normal range of fluent and spoken language – Tager-Flusberg, Paul, & Lord, 2005). A common feature of students with autism, however, is an inability to produce communicative and empathic responses to text. Specifically, students with autism may have a degree of social isolation coupled with few, deep interests. Reading activities that require students to connect and analyze the actions of characters in literature, therefore, may be exceptionally challenging to this population (Kluth & Darmody-Latham, 2003).

The challenges faced by students with disabilities in terms of reading are wide-ranging and specific to disability categories. Organic learning difficulties may be one reason why students with disabilities often struggle on large-scale assessment (National Center on Educational Statistics, 2003). It is evident that students with disabilities have difficulties

in the three broad areas of (1) foundational skills, (2) comprehending the messages and themes found in text, and (3) analyzing and interpreting text. To this end, it would seem that assessments that require students to engage in these activities may be problematic for students with disabilities. At the same time, minimizing requirements in any of these areas is in conflict with the No Child Left Behind Act and has implications for the validity of assessments. A perennial challenge to educators, policymakers, and assessment designers lies in how to design accessible assessments without violating essential reading constructs.

Given the inevitable tension between issues related to reading for students with disabilities and the essential constructs of reading that need to be assessed for accountability, it is important to consider what constructs and assessment approaches are actually found in states. The results of this study have implications for both accommodations policy and future research. For example, the purposes of assessments may directly interact with accommodations policies for particular states. More than one third of the states explicitly assess some form of foundational skills. Students with a wide variety of disabilities may have difficulty with the foundational aspects of reading and may require print reading accommodations. Oral presentation accommodations, however, may change the construct of items that directly assess foundational skills. Furthermore, extended time accommodations may also interfere with the intended construct of rapid sight word recognition. Finally, sign language interpretation of foundational skill items could either inadvertently provide students with correct answers, or mislead them because of the disconnect between manual and printed language. Items related to foundational skills appear to pose great challenges in terms of providing students with disabilities with accommodations that do not change the construct.

Comprehension and analysis/interpretation items may also pose problems for students with disabilities. In the case of comprehension and analysis/interpretation, allowable accommodations often relate to the state's definition of "reading." When reading comprehension or analysis/interpretation items are considered an extension of foundational skills, read aloud and sign language accommodations will invalidate the test (although extended time, separate setting, and familiar examiner accommodations would not). On the other hand, if states define the constructs of comprehension or analysis/interpretation as the ability to understand text, no matter how it is presented, then read aloud accommodations are valid. There is no way to determine, from test specifications, whether states deem foundational skills as "gateway" skills to comprehension, or if foundational and comprehension skills are separate constructs, ruled by separate allowances for accommodations. A recent study from the National Accessible Reading Assessment Projects by Cahalan-Laitusis (2006) investigated the validity and effectiveness of read aloud accommodations for students with learning disabilities, and sought to add scientific data to the thorny issues related to the effectiveness and validity of read aloud accommodations (Sireci, Scarpati, & Li, 2003).

Finally, reading activities related to the development of the reader through engaging and authentic activities may be the most motivating of state standards. These standards,

however, are not found in state assessments or are assessed in different ways than intended in standards. A closer examination of how to bring personally engaging and authentic aspects of reading to large-scale assessments may make assessments more accessible to a variety of students.

Implications for the assessment of students with disabilities also lie in the types of items selected by states and vendors to assess reading. Among states that specified item types in their blueprints and test specifications, multiple choice items were most commonly chosen. To this end, there may be a variety of issues present for students with disabilities. Accommodations such as familiar examiners, sign language interpretation of answer choices, visual cues, and additional examples may all unnecessarily aid or hinder a student's ability to answer multiple choice questions. Furthermore, preliminary data from a National Accessible Reading Assessment Projects distractor study conducted by Abedi, Leon, & Kao (in press) indicate that students with disabilities may approach multiple choice items differently than their non-disabled peers, causing a differential distractor function (DDF) to be present in assessment items.

Finally, a portion of statewide assessments in some states will be “constructed response.” Such items have obvious implications for students with dysgraphia or other expressive language difficulties. In cases such as these, it is unknown whether providing accommodations such as scribes or technological aids will disrupt the constructs of reading. Most likely, response related accommodations will not invalidate a reading test, but careful examination of the types of tasks students are asked to complete is important to consider in the process of test development.

Overall, this study has provided data from which to design further research studies. After a careful review of state blueprints and test specifications, we know that any given state may place different degrees of emphasis on foundational, comprehension, and analysis/interpretation constructs in their state assessment. Because of this, it is important to consider what factors (if any) can increase the accessibility of these types of items for all students including students with disabilities.

Accessible options for foundational skills may include alternative scoring options for students who are unable to perform foundational activities of reading, but otherwise are able to comprehend and analyze/interpret text. Another option for increasing accessibility of foundational items may be to include a wide variety of grade-level appropriate foundational skills to allow students to access at least some of the items on an assessment. A third option may be to follow the example of National Assessment of Educational Progress 2009 framework (National Assessment Governing Board, 2005) and imbed foundational skills in comprehension and analysis/interpretation items. Further research is needed to determine the most valid and effective approaches.

Likewise, comprehension and analysis/inferential items may have a variety of approaches to increase accessibility. Approaches such as improving the motivational level of passages, reducing or extending the length of passages, and removing or including

graphics in passages may help to increase the accessibility of comprehension and analysis/interpretation items without changing the constructs of items. In addition, allowances for assistance from technology or human readers for comprehension or analysis-specific items may help to increase accessibility but require careful definitions of intended assessment constructs.

In terms of item types, allowances such as letting students answer multiple-choice questions on the test booklet, decreasing the number of foils that are present on items, and removing items that are found to have differential distractor functioning may be ways of increasing accessibility of multiple choice items without threatening validity. Accessibility for open-ended items may be increased by specifying the tasks that students must complete clearly, allowing for flexibility in response, ensuring that reading-related constructs are being assessed, or allowing student choice in reading materials and response items.

All of the methods described in this paper are merely theoretical approaches to increasing accessibility, and will be examined in future studies by the National Accessible Reading Assessment Projects. The aim of this study was to lay the groundwork for understanding broad themes found in reading assessment blueprints and test specifications. This study provides a framework for understanding the current constructs assessed and item types found in state assessments. Information gleaned from this study is best used as a springboard for future studies related to foundational, comprehension, and analysis items that are presented in both multiple choice and constructed response formats.

References

- Abedi, J., Leon, S., & Kao, J. (in press). *Examining differential distractor functioning across student disability status*. Minneapolis, MN: Partnership on Accessible Reading Assessment.
- Algozzine, B., & Wood, K.D. (1994). Reading and special education in the twenty-first century: Time to unify perspectives. In K. D. Wood & B. Algozzine (Eds.), *Teaching reading to high-risk learners: A unified approach* (pp. 1-8). Boston: Allyn and Bacon.
- Barton-Arwood, S.M., Wehby, J., & Falk, K. (2005). Reading instruction for elementary age students with emotional and behavioral disorders: Academic and behavioral outcomes. *Exceptional Children*, 72(1), 7-27.
- Browder, D. M., & Xin, Y. P. (1998). The meta-analysis and review of sight word research and its implications for teaching functional reading to individuals with moderate and severe disabilities. *The Journal of Special Education*, 32(3), 130-53.
- Cahalan, C. (2006). *An examination of the validity of a read aloud accommodation for a standardized reading assessment using differential boost and predictive validity as criteria*. Paper presented at National Council on Measurement in Education Meeting, April 7-11, 2006.
- Cline, F., Johnstone, D., & King, T. (2006). *Gathering public reactions to a definition of reading proficiency*. Paper presented at American Educational Research Association Meeting, April 7-11, 2006.
- Clapper, A., Morse, A., Lazarus, S., Thompson, S., & Thurlow, M. (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.
- Edyburn, D.L. (2000). Assistive technology and students with mild disabilities. *Focus on Exceptional Children*, 32(9), 1-24.
- Gersten, R., Fuchs, L., Williams, J., & Baker, S. (2001). Teaching reading comprehension to students with learning disabilities: A review of research. *Review of Educational Research*, 71, 279-320.
- Gillon, G. (2000). The efficacy of phonological awareness intervention for children with spoken language impairment. *Language, Speech and Hearing Services in Schools*, 31, 126-141.
- Katims, D.S. (2000). Literacy instruction for people with mental retardation: Historical highlights and contemporary analysis. *Education and Training in Mental Retardation and Developmental Disabilities*, 35, 3-15.

- Kluth, P., & Darmody-Latham, J. (2003). *Beyond sight words: Literacy opportunities for students with autism*. *The Reading Teacher*, 56(6), 532-535.
- Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- National Assessment Governing Board. (Spring, 2005). *Specifications for the 2009 NAEP Reading Assessment*. Washington, DC: Author. Retrieved February 18, 2006 from http://www.nagb.org/pubs/reading_fw_08_05_prepub_edition.doc
- National Center on Educational Statistics (2003). *NAEP data: National reading composite, grade 4* (2003, 2002, 2000, and 1998). Washington, DC: U.S. Department of Education. Retrieved from the World Wide Web on August 16, 2004 at <http://nces.ed.gov/nationsreportcard/naepdata/getdata.asp>.
- National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: National Institute of Child Health and Human Development.
- No Child Left Behind Act of 2001, 20 U.S.C. 6301 e seq (2001) PL 107-110).
- Pressley, M. (2006). *Reading instruction that works: The case for balanced teaching*. New York: Guilford Press.
- Progress in International Reading Literacy Study Assessment (PIRLS) (2001). *PIRLS International report*. Retrieved from the World Wide Web March 16, 2006. http://timss.bc.edu/pirls2001i/pdf/P1_IR_Introduction.pdf
- Sireci, S.G., Scarpati, S., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research*, 75(4), 457-490.
- Snow, C. (2002). *Reading for understanding: Toward an R & D program in reading comprehension*. Arlington, VA: RAND.
- Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and communication in autism. In F. Volkmar, R. Paul, A. Klin, and D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders-3rd Ed.* (pp. 335-364). New York: Wiley.
- Thompson, S. J., Johnstone, C. J., Thurlow, M. L., & Clapper, A. T. (2004). *State literacy standards, practice, and testing: Exploring accessibility* (Technical Report 38). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.
- Thurlow, M.L., Johnstone, C.J., Thompson, S.J., & Case, B. (in press). Using universal design research and perspectives to increase the validity of scores on large-scale assessments. In Karchmer, M. (Ed.). *Assessing deaf students' academic achievement in an age of accountability*. Washington, DC: Gallaudet University Press.

Vaughn, S., Bos, C., & Schumm, V. (2004). *Teaching exceptional, diverse, and at risk students in the general education classroom*. Upper Saddle River, NJ: AB Longman.