USING CONCURRENT VERBALIZATION TO MEASURE MATH COMPREHENSION

by

CATHRYN COLLEY LAMBETH

A DISSERTATION

Presented to the Department of Educational Methodology,
Policy, and Leadership
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Education

June 2012

DISSERTATION APPROVAL PAGE

Student: Cathryn Colley Lambeth

Title: Using Concurrent Verbalization to Measure Math Comprehension

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Education degree in the Department of Educational Methodology, Policy, and Leadership by:

Dr. Gerald Tindal	Chair
Dr. Gina Biancarosa	Member
Dr. Scott Baker	Member
Dr. Roland Good	Outside Member

and

Kimberly Andrews Espy	Vice President for Research & Innovation/Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2012

DISSERTATION ABSTRACT

Cathryn Colley Lambeth

Doctor of Education

Department of Educational Methodology, Policy, and Leadership

June 2012

Title: Using Concurrent Verbalization to Measure Math Comprehension

The current study investigated variability in student performance on a concurrent verbalization measure based on a grade-level sample math word problem and sought to determine to what extent the variability in verbalization scores is related to scores on a reliable measure of reading (DIBELS Next) and math (easyCBM) and to student factors (e.g. sex, grade, economic status).

In light of the 2014 implementation of the Common Core State Standards and related measures of student performance, both of which contain components of language in mathematics curriculum and assessment, it was the intent of this study to identify factors associated with verbalization on sample math word problems that could be correlated with student performance on reliable, commonly used assessments of reading and math.

The sample for analysis included 105 intermediate-grade students from one elementary school in the Pacific Northwest.

Results support a relation between students' verbalizations about math word problems and benchmark assessments in reading and math. Limitations of the study, considerations for future research, and implications for practice are discussed.

CURRICULUM VITAE

NAME OF AUTHOR: Cathryn Colley Lambeth

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
George Fox University, Newberg, Oregon
Pacific University, Forest Grove, Oregon
California State Polytechnic University, Pomona

DEGREES AWARDED:

Doctor of Education, Educational Leadership, 2012, University of Oregon
Reading Endorsement, 2007, George Fox University
Master of Arts, Teaching, 2001, Pacific University
Bachelor of Arts, Political Science, 1977, California State Polytechnic University, Pomona

AREAS OF SPECIAL INTEREST:

Math Comprehension Assessments, Elementary Grades
Reading Instruction, Intermediate Grades

PROFESSIONAL EXPERIENCE:

Intervention Specialist, Learning Specialist, Canby School District 86, Canby, Oregon, 2001-present

Member, Stewardship Team for the Implementation of the Common Core State Standards, Oregon Department of Education, Salem, Oregon, 2011-2012

GRANTS, AWARDS, AND HONORS:

Canby School District Technology Grant, 2010

Canby Education Foundation Grant, 2009

Canby School District Technology Grant, 2008

ACKNOWLEDGMENTS

Oregon is a wondrous place, and the University of Oregon is its crowning jewel. Thank you to Dr. Gerald Tindal and to the faculty of the Department of Educational Methodology, Policy, and Leadership for their high expectations that led to true learning. Thanks also to the Cult of JT—Go Ducks.

For my husband, Ted, and my children, Kerry and Will, who encourage and inspire me to reach, and to my parents, Joe and Ruth Colley, who are reaching still.
And we'll not forget our Carus School.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

CHAPTER I

STATEMENT OF THE PROBLEM AND LITERATURE SURVEY

Even before Horace Mann asked, “A gentlemen has invested $140,768.25 in two estates, one of which is worth $89,329.18. What was the cost of the other?” (Mann & Chase, 1855), learning mathematics in school included both computational fluency and comprehension. Computational fluency, or the ability to recall and manipulate numbers, and comprehension, or the ability to communicate a response or solution to a problem using language, are separate mathematical skills distinguished by the addition of linguistic information (Fuchs et al., 2008). Comprehending and communicating in this language of mathematics are vital skills required of students who are preparing for college and career readiness (Clark & Shinn, 2004).

These skills of comprehension and communication are also essential parts of the curriculum in the classroom, including the intermediate elementary grades (Griffin & Jitendra, 2009; Oregon Department of Education [ODE], 2011; Rittle-Johnson, Siegler, & Alibali, 2001). Research has suggested that phonological processing underlies math skills (Fuchs et al., 2006). Making valid inferences about a student’s proficiency in math requires comprehension and communication to be measured separately; however, current educational measures in math do not consistently recognize or clearly evaluate them distinctly from other math skills (Clark & Shinn, 2004; Fuchs et al., 2008).

The importance of identifying students with deficits in math comprehension and communication is increasing as the 2014 implementation of the Common Core State

Standards (CCSS) approaches. These standards include language components in mathematics and align with research about postgraduation math application. Embedded in the CCSS are eight standards of mathematical practice that describe the level of language required for a successful performance in mathematics. For example, the Attend to Precision standard of practice describes a desired outcome of mathematically proficient students who try to communicate precisely with others, using clear definitions to state their own reasoning (Common Core State Standards Initiative [CCSS], 2011).

Proposed measures of the CCSS contain summative, performance-based assessments at each grade level. Starting in the third, fourth, and fifth grades, students will be asked to use mathematical language when taking these assessments. For example, one fourth-grade task will direct students to describe how they would configure the layout of a convenience store, including measurements of area and perimeter and a rationale of their design (SMARTER Balanced Research Consortium [SBAC], 2012; T. Alpert, personal communication, January 19, 2011). These new standards and their assessments have renewed the focus among administrators on all students' practicing mathematical language (M. Callahan, personal communication, May 1, 2012).

Students who have not acquired math skills by the end of their educational experience are denied access to important resources and opportunities in our society (National Governors Association, 2010; Schoenfeld, 2002). Given the "relative dearth of research" on math measures (Chard et al., 2005), more specific diagnostic tools are needed to identify students who may require support to reach the standards (Clarke & Shinn, 2004; Gersten, Jordan, & Flojo, 2005).

At the student level, there are external and internal factors that impact math performance. External demands of the school curriculum and its language content provide both opportunity and challenges. Internal demands impacting students' cognitive functions may support or impair math achievement (Fuchs et al., 2006). For example, working memory's processing abilities have been linked to students' ability to solve math problems from simple early arithmetic (Imbo & Vandierendonck, 2006) in the intermediate grades (Lee, Ng, Ng, & Lim, 2004) and in high school students (Kyttala, 2008). Research has explored the impact of a variety of other issues such as language or behaviors on students' processing skills as reflected in their math performance (Cirino, Fletcher, Ewing-Cobbs, Barnes, & Fuchs, 2007).

One measure of process skills used in social science research is *verbalization*, or audible verbal reporting by a participant. Verbalizations are the overt, verbal expression of normally covert mental processes (Baumann, Jones, & Seifert-Kessell, 1993). Verbalizations are an important source of information about the cognitive process that could otherwise be investigated only indirectly (Wade, 1990).

An example of a verbalization prompt is

> I'm going to think out loud while I solve this problem. That means I'm going to say everything that goes through my mind. [Model verbalization]. Now I'm going to ask you to solve a problem the same way. Just say everything that goes through your mind while you solve the problem. (Johnstone, Bottsford-Miller, & Thompson, 2006)

Verbalization has been used to elicit subjects' strategies, and to refine, clarify and validate subjects' responses (Ferrara, 2008). Verbalization has also shown promise as a measure for low-performing students (Anderson, Lai, Alonzo, & Tindal, 2011). As one student said, if she could explain the problem, she could solve it (C. Wright, personal communication, December 13, 2010).

Purpose of This Study

The purpose of the study was to investigate the relations between various components of a verbalization measure intended to collect information on students' comprehension of math word problems, and to identify the effects of student-level factors on those relations. I used a concurrent verbalization measure to document the influence of language demands and working memory. I then explored to what extent the variability recorded was correlated with scores on reliable measures of math and reading. Finally, I described the effects of student-level factors on these scores. The specific research questions are as follows:

1. Is there a relation between components of a measure used to evaluate students' ability to verbalize on the process for solving math word problems?

2. Is there a relation between students' ability to verbalize on the process for solving math word problems, and students' reading fluency or comprehension on a reliable assessment, measured by the DIBELS Next reading assessment including measures of Oral Reading Fluency, retell, comprehension, and a composite score which considers percentile rank?

3. Is there a relation between students' ability to verbalize on the process for solving math word problems, and students' ability to solve math problems on a reliable assessment, measured by easyCBM math composite score and percentile rank?

4. What is the effect of student-level factors of grade, Special Education eligibility, socioeconomic status, second language, and Talented and Gifted designation on students' ability to verbalize about a math word problem?

Literature Survey

To locate studies relating to verbalization, its application to math comprehension, and factors impacting mathematical performance in elementary students, I searched electronic databases (ERIC, Google Scholar, Academic Search Premier, PsychNet, WebScience). Table 1 lists the terms used, both alone and in combination, while searching these databases.

TABLE 1. Words and Phrases Used in Literature Search, by Topic

Verbalization	Mathematics	Internal and External Demands
Cognitive labs	Comprehension	Curriculum
Cognitive interviews	Assessment	History of math
Verbal report	Multidimensionality	Working memory
Think-aloud	K-12 elementary	Linguistic deficits
Verbalization	Problem solving	Second language
Concurrent verbalization	National Council of Teachers of Mathematics	English Language Learners
Coding		Poverty
Scoring		Socioeconomic Status
Retell		Inattentive Behavior
		Student age
		Grade
		Reading skills

As a result of these searches, over 375 articles, dissertations, and reports were located and examined. Titles and abstracts were screened according to various specific criteria. Of these articles, 113 were included because they contained one or more of the following attributes: (a) a description of the multidimensionality of mathematics; (b) a description of research relating to early identification of mathematical skills or deficits; (c) a description of the role of working memory in math problem solving; (d) a description of the impact of curriculum changes, student age, grade and gender, linguistic deficits, inattention, socioeconomic status, English language learners, or reading skills and math problem solving; (e) a description of the components of verbalization or concurrent verbalization; (f) specific information relating to verbalization and math comprehension; or (g) specific information relating to the validity of the measures used in this study.

External Demands Affecting Mathematics Performance

From infancy, humans have demonstrated an understanding of addition and subtraction. This understanding provides the basis for the development of mathematical skills (Wynn, 1992). As preschoolers, children have the capacity to learn mathematics (Cross, Woods, & Schweingruber, 2009). Entering school, children face the work of understanding all dimensions of math, filtered through their own abilities and needs. The external demands of curriculum, which students rely on for special comprehension strategies (Kintsch & Greeno, 1985; Schoenfeld, 2002), and the special features of language of math (Tindal & Anderson, 2011) each have an impact.

Curriculum

Curriculum has been defined as a sequence of learning opportunities provided to students in a field of study, playing a crucial role in student learning. Explicit instruction by a child's teachers is one of the greatest aids to fluent comprehension in major content areas (Wolf, 2007). Students recognize that special math skills, such as strategies to solve math word problems, are learned in school (Kintsch & Greeno, 1985; Kloosterman & Cougan, 1994). Research has demonstrated that curriculum has a profound effect on student achievement (Schmidt et al., 2001) and that a consistent curriculum adhered to over time and across grade levels results in student success (Datnow, Borman, & Stringfield, 2000).

In American education, the curriculum for mathematics has shifted priorities over the last 50 years, and the role of comprehension in math instruction has been continually redefined. Mathematics instruction focused primarily on algorithmic computational skills until the "New Math" of the 1960s emphasized comprehension (Klein, 2003). In 1989, the National Council of Teachers of Mathematics (NCTM, 2009) introduced its Guiding Principles and Standards for teachers of mathematics to elementary math instruction (NCTM, 2009). These standards provided a common basis for curriculum, but set off "math wars" over their emphasis on collaborative understanding at the expense of teaching traditional algorithmic tools (Klein, 2003). Over time, classroom changes mirrored the standards and curriculum fluctuations. The effect on students who started kindergarten in 2000 was exposure to up to four curriculum changes in math before graduating from high school.

These changes also affected the majority of teachers, who did not have the opportunity to establish and maintain a consistent curriculum. To be successfully adopted, a new curriculum required time and energy on the part of teachers (Datnow et al., 2000; Earl & Katz, 2000). With shifting standards and resources, teachers were found to vacillate back and forth between the old and new, using parts of both in their classroom instruction. These shifts contrast with research demonstrating that student achievement increased when content-specific curricula were consistently implemented in the elementary grades (Datnow et al., 2000).

In 2006, NCTM published curriculum focal points and process standards that identified the most critical math topics at each grade level. The focal points were intended as a step toward the implementation of the Common Core State Standards (CCSS), and included computational fluency in the intermediate grade levels (NCTM, 2011). These NCTM process standards are the partial basis for the CCSS (Common Core State Standards Initiative, 2011).

These CCSS are national standards to be implemented in adopting states by 2014. The mission statement of the CCSS initiative includes the goal of providing a “consistent, clear understanding of what students are expected to learn, so teachers . . . know what they need to do. . .” (CCSS, 2011). The standards describe the ways students will be required to use language to demonstrate math comprehension skills (ODE, 2011). Nested in the CCSS are eight standards of mathematical practice. The practice standards describe the expertise that teachers at all grade levels should seek to develop in their students, and describe expectations for teachers and students that apply across grade levels and specific curriculum topics. The practices support the expectation that students will use language to explain mathematical concepts and how they applied them, and to clearly and precisely construct viable arguments to support their reasoning and critique the reasoning of others. Examples from the practice standards for using mathematical language include the expectation that students can explain correspondences between information, construct arguments using concrete referents that make sense and are correct, and listen to the arguments of others and ask useful questions to clarify or improve the argument (CCSS, 2011).

SBAC's draft assessment of the CCSS includes measures of student performance on these eight practice standards. According to SBAC, one of the underlying reasons for the assessment is communication, especially students' "ability to *explain why* given procedures work." For example, a proposed SBAC assessment would require fourth-grade students to read and understand a scenario or context to reach and communicate a solution, as assessed by constructed responses or performance tasks (CCSS, 2011; SBAC, 2012).

Language Demands and Linguistic Features of Math Problems

For decades, researchers have been interested in the impact of linguistic features on students' ability to solve math word problems. This interest was due to the relative lack of success most students demonstrate with these problems (Nesher, 1980). In the

1980s, researchers began to identify the challenges of the language of math problems by categorizing their structures (Riley, Greeno, & Heller, 1983). Research suggested that the language of academic word problems is separate from problems students would encounter in “real life” (Nesher, 1980). More recently, language was identified as one of seven cognitive functions correlated with solving arithmetic word problems (Fuchs et al., 2006). The cognitive functions included language, nonverbal problem solving, concept formation, working memory, long-term memory, attention, and sight-work efficiency (Fuchs et al., 2006). These researchers also discussed the effect of a math problem presented with a graphic, not simply with written language.

This complexity of language has been found to impact the speed or accuracy of a solution found by students with issues relating to language (Lean, Clements, & Del Campo, 1990; Tindal et al., 2003). In English language learner (ELD) students, linguistic features have been found to have a significant negative effect on the scores of a large-scale assessment of fourth-grade students ($p < .05$: prepositions (-.146), ambiguous words (-.194), complex verbs of < 3 words (-.090), pronouns (-.148), and math vocabulary (-.184). Researchers speculated that, due to less sophisticated verbal skills, the language in math word problems may impact the performance of fourth-grade ELD students more than older students (Shaftel, Belton-Kocher, Glasnapp, & Poggio, 2006).

The complexity of language was considered in the development and validation of appropriate assessments for students with significant disabilities (Tindal et al., 2003; Tindal, Yovanoff, & Geller, 2008), in middle school students with identified learning disabilities (Helwig, Anderson, & Tindal, 2002), and for persistently low-performing students without significant disabilities (Anderson et al., 2011). Construct validity concerns were common to these studies. For example, to provide access to measure the underlying math skills, test administrators may simplify the vocabulary in a word problem to avoid entangling language skills. However, simplifying the story problem into a computational problem may not correctly capture a student's decision-making skills (Tindal & Anderson, 2011). To make participation meaningful, the administration of these assessments requires flexibility; however, this flexibility may compromise the validity of the inferences as well as the specific judgments about a student's proficiency (Tindal et al., 2008).

The SBAC (2012) draft assessment of the CCSS in math described the intentionality of the assessment design to minimize the impact of unnecessary linguistic complexity of math items as a source of irrelevant interference for ELD students and students with a disability. In addition to the assessment design considerations, SBAC suggested allowing students to use the language in which they are the most proficient, due to the high level of verbal and written communication skills required to demonstrate competency in math.

Internal Student-Level Factors Affecting Mathematics Performance

Student-level factors impact students' mathematical performance. These factors, including student age and grade, working memory (Berg, 2008; Swanson, 2006; Swanson & Beebe-Frankenberger, 2004), inattentive behaviors (Swanson, 2006), linguistic deficits (Fox, Ericsson, & Best, 2011), socioeconomic status (Jordan, Kaplan, Olah, & Locuniak, 2006), and English as a second language, have all been related to mathematical fluency (Fuchs et al., 2008; Swanson & Beebe-Frankenberger, 2004). A student's reading skills have also been demonstrated to be a predictor of both computational and problem-solving skill (Fuchs et al., 2008).

Student Age

Given that our current education model strongly links age and grade, older students have increased access to and practice with the special skills of math problem solving, leading to higher ability in math (Kintsch & Greeno, 1985). More recent research relating to working memory found that beginning in third grade, students increasingly rely on the visuo-spatial aspect of working memory for math reasoning, specifically on word problems (Meyer, Salimpoor, Wu, Geary, & Menon, 2010). As students age, growth in their executive processing capacity has been a more important predictor of their problem-solving skills than reading or calculation skills (Swanson, 2006).

Working Memory

Working memory has been defined as the short-term storage of information used to perform a task (Wolf, 2007). While our understanding of the cognitive mechanisms in working memory that specifically support math are limited, the functions of a student's working memory have been confirmed as an important predictor of problem-solving skills both generally (Swanson & Beebe-Frankenberger, 2004) and specifically for the intermediate grades (Lee et al., 2004; Meyer et al., 2010; Passolunghi & Siegel, 2001). A student's working memory has been described as a pivotal source of variance in elementary math performance, especially in problems featuring language tasks involving manipulation and transformation of information (Siegel & Ryan, 1989).

In reading, processing speed was linked to the quality and quantity of semantic knowledge activated. In other words, the more a reader knew about a word, the faster it was recognized and understood (Wolf, 2007). The processing speed of intermediate-grade students' working memory during reading decoding strongly predicted their math computational fluency (Hecht, Torgesen, Wagner, & Rashotte, 2000). Working memory has also been described as a major determinant of the differences between an average student and a student with math difficulties, underlying most math disabilities (Swanson & Jerman, 2006).

Recent research has focused on the interaction of the components of working memory and specific dimensions of math. The Baddeley model first described working memory as made up of two temporary storage areas, the auditory-verbal and the visuo-spatial, accessed by a central executive that prioritizes information and guides decision-making. Later, Baddeley (2010) described an "episodic buffer," linking working memory with awareness. The central executive was described as the "core" of working memory (Salmon et al., 1996). This central executive was demonstrated to have as much influence on math performance as literacy or an overall cognitive score, and was predictive of intermediate age students' performance on algebraic word problems (Lee et al., 2004).

The visuo-spatial "sketchpad," primarily responsible for the generation and manipulation of mental images (Meyer et al., 2010; Swanson & Beebe-Frankenberger, 2004), has been connected with algebraic, geometric, and other word problem-solving abilities in older students (Kyttala, 2008; Reuhkala, 2001). As children enter the intermediate grades, visuo-spatial representations have been found to play an increasingly important role in math reasoning. Where second-grade students rely on the central executive and phonological components for math reasoning, third-grade students increasingly use visuo-spatial functions for these same skills (Meyer et al., 2010). Brain research on adult subjects confirmed the role of the visuo-spatial function during working memory tasks (Salmon et al., 1996), and research on student subjects continued to seek to untangle the role of working memory and other cognitive functions in solving math word problems (Fuchs et al., 2006).

Inattentive Behavior

Attentive behavior was called a "critical cognitive determinate" in all facets of math performance, including the computational skills required to solve math word problems (Fuchs et al., 2006). Behavioral inattention has impacted student performance on most measures of arithmetic skills (Cirino et al., 2007). Research has identified retrieval speed and prioritizing as sources for the lower mathematical tasks performance of students with attention deficits. Students with attention disorders are primarily slower in math facts retrieval, which influences the speed at which a student is able to solve math word problems (Zentall, 1990).

Researchers have speculated that the required prioritizing of information in the solution of a math word problem created difficulty for an inattentive student. When students are unable to attend, the function of working memory is interrupted. As a result, students have a decreased ability to filter and prioritize the information required, resulting in lowered ability to solve a math word problem (Passolunghi & Siegel, 2001; Raghubar et al., 2009). While research established inattentive behavior's impact on computational fluency (Fuchs et al., 2008; Kercood, Zentall, & Lee, 2004; Swanson, 2006), the underlying nature of the impact of inattention on math performance is not yet clear (Fuchs et al, 2006).

Language Deficits and Disabilities

Studies of the impact of students' language deficits and disabilities on math performance generally follow two themes: intervention and assessment. Studies of interventions describe effective instructional practices for students with disabilities (e.g., Mercer & Miller, 1992; Montague & Bos, 1986). Studies relating to assessment focus on the ability to appropriately measure a student's knowledge and skills in math, including comprehension, differentiated from a student's disability (Tindal & Anderson, 2011). These studies verify the correlation between an identified disability and math achievement. The SBAC proposed assessment of the CCSS acknowledged the need to design assessments and provide accommodations to "get around" barriers created by disabilities (SBAC, 2012).

Research described types of disabilities that interfere with performance on math assessments: learning disabilities, emotional disorders, and speech-language disorders. Learning disabilities are clearly connected with math performance by research. Research has attempted to identify patterns of math performance within this disability. Lower visuo-spatial and mathematical abilities across grade levels were found in students with an identified learning disability. However, males with a learning disability have demonstrated strengths in mathematical reasoning even while their computational fluency remains below the mean (Vogel, 1990). Under the Individuals With Disabilities Education Improvement Act (2004), eligibility for a learning disability was defined by the discrepancy between their performance in math or other academic areas and their potential in these areas.

The impact of a speech-language disability on math achievement has also been examined by researchers. One longitudinal study compared the academic growth through the intermediate grades of students with speech-language impairments and that of students with a developmental delay. Though both groups received early preschool interventions and continued to receive services during their primary and intermediate school years, by age 10 (fifth grade) the students with speech or language impairment continued to have significantly higher mean scores on a normed assessment in math than students with a developmental delay (Carlson, Jenkins, Bitterman, Keller, & Lauer, 2011).

Students with identified emotional and behavioral disorders generally experience large academic achievement deficits in all content areas. Over time, from the end of the intermediate grades through high school, these students demonstrate a growing deficit in math scores, measured by performance on a normed test (Nelson, Benner, Lane, & Smith, 2004).

Socioeconomic Status

A study on the impact of poverty on math achievement found that kindergarteners from low-income families were over 17 times more likely to score below grade level on math assessments. As the study followed the students through their elementary years, these students generally made parallel gains in math compared to their higher-SES peers, but they did not make enough gains to demonstrate the same levels of achievement. Unlike growth in other math skills such as computational fluency, these students demonstrated almost no growth in their ability to comprehend math word problems (Jordan et al., 2006).

Young children have the capacity to learn mathematics, but that potential is often not realized due to a lack of math exposure or opportunity in their environments (Cross et al., 2009). Because of this gap between potential and achievement, some researchers describe this lack of exposure to mathematical thinking as a specific type of deficit. Low math achievement has been described as "pronounced" in students from low-income households to the extent of being called a civil rights issue (Schoenfeld, 2002).

Frequent moves, with attendant changes in curriculum among other more specific variables, have had an almost uniformly negative impact on the academic achievement of intermediate-grade students (Ingersoll, Scamman, & Eckerling, 1989). For example, in a study of Arizona student mobility in Grades K-12, students eligible for free or reduced lunch moved almost twice as many times during their elementary years (Fong, Bae, & Huang, 2010).

Sociocultural modifications have also been suggested for students who live in poverty (D. Beagle, personal communication, September 1, 2011). While linked to research in poverty and language, accommodations related to socioeconomic status are not specifically addressed in discussions of proposed assessments of the CCSS (SBAC, 2012).

English as a Second Language

Researchers have continued to explore why students learning English as a second language generally score lower than the mean on math word problems. A proposed CCSS assessment suggested providing multiple opportunities for students learning English as a second language (ELD) to communicate their understanding through performance tasks or other multiple domain expressions (SBAC, 2012), acknowledging a concern about second language and math word problems. Research has examined the relationship between complex linguistic factors and students' English language skills (Lean et al., 1990). However, language simplifications of math word problems have not consistently boosted ELD scores (Shaftel et al., 2006).

One study of fourth- and eighth-grade English language learners suggested that the level of their parent's education was more highly related than language to student performance in reading. In this study, the mean reading scores of students whose parents had not completed 12th grade was over 15 points lower (M = 25.2, SD = 14.1) than ELD students whose parents had completed postgraduate work (M =40.4, SD = 19.6; Abedi, 2004). The variance in scores on large-scale tests has been attributed primarily to other factors such as student mobility and economic status (Abedi, Courtney, & Leon, 2003; Lord, Abedi, & Poosuthasee, 2000).

Researchers have described how ELD students averaged more mobility than other subcategories of students, including those in poverty (Fong et al., 2010), compounding the negative impact of frequent moves on the achievement of these students (Ingersoll et al., 1989). Other researchers suggested that the sociocultural context of math word problems, not simply language translation, is needed to appropriately measure ELD student skills (Solano-Flores & Trumbull, 2003).

Reading Skills

Reading difficulties are an important correlate of math difficulties (Swanson & Jerman, 2006) and of outcomes in math on large-scale assessments (Castillo, Torgesen, Powell-Smith, & Al-Otaiba, 2009). Students with difficulty in both subjects experienced deficits in both computation and problem solving (Fuchs & Fuchs, 2002; Jordan & Hanich, 2000, 2003).

Recent research has sought to tease out the differences between student disabilities in math and reading (Fuchs et al., 2008). Students with difficulties only in math may perform better than students with math and reading disabilities because they are simply more able to follow the math problem’s narrative, literally helping them understand the problem better (Powell, Fuchs, Fuchs, Cirino, & Fletcher, 2009).

Notably, research has confirmed that students who are on target in reading are not immune from deficits in math. These measurable, specific skill deficits do not significantly improve as students age (Swanson & Jerman, 2006). While students who struggle in both math and reading lag farthest behind in math scores, a separate subset consistently underperforms in math despite normal reading scores (Hanich, Jordan, Kaplan, & Dick, 2001; Jordan & Hanich, 2000; Landerl, Bevan, & Butterworth, 2004; Swanson & Beebe-Frankenberger, 2004). Research confirms the need to identify the "core deficit" of mathematics disability (Chiappe, 2005). Further, math disabilities are an "underestimated" research topic (Gregoire & Desoete, 2009).

The relationship between math comprehension and communication, and the variables impacting student math achievement, including curriculum changes, language demands of assessments, student age and grade, cognitive functions of working memory, inattentive behaviors, linguistic deficits, socioeconomic status, second language skills, and reading skill, are not well understood. The complexity of this relationship includes the multidimensional nature of both the math problems and the students who are solving them.

Current Measures of Math Comprehension

Proficiency in mathematics is a combination of knowledge and skills (Cross et al., 2009). A meta-analysis conducted by Fuchs et al. (2008) questioned whether the separate skills in mathematics could be measured through a single type of assessment, concluding that assessing computational and problem-solving skills separately is of "critical importance" in order to identify specific difficulties. Mathematics professionals have long expressed concerns about whether current tests of mathematics adequately emphasize comprehension (Romberg, 1992). The Fuchs study noted the need for an exploration of alternative forms of assessment specifically related to comprehension (Fuchs et al., 2008).

Currently, both curriculum-based measures and large-scale, standards-based assessments are commonly used to measure student math performance. Multiple studies support the use of Curriculum-Based Measures (CBMs) to measure student progress and to inform and guide instruction (Marzano & Haystead, 2008). CBMs are used to screen and monitor the progress of students who may be at risk (Anderson, Alonzo, & Tindal, 2009; Anderson et al., 2011). CBMs may also be used to establish standards for all students (Deno, Fuchs, Marston, & Shin, 2001). In primary grades, CBMs using number sense have been confirmed as a reliable early indicator of a student's elementary math achievement (Chard et al., 2005; Clarke & Shinn, 2004; Jordan et al., 2006).

Standards-based measures such as the Oregon Assessment of Knowledge and Skills (OAKS) and the Texas Assessment of Knowledge and Skills (TAKS) have been used to compare individual student math performance against a set of expectations at grade and benchmark levels, and have included both computation and comprehension components using multiple-choice, computer-based tests (Carr-George, Vannest, Willson, & Davis, 2009; ODE, 2010a; Texas Education Agency [TEA], 2011). Reliability and validity of the TAKS was described in the Texas National Comparative Data Study, conducted to compare the achievement of Texas students to a national benchmark. This study reported the reliability of the TAKS in math by grade, including the intermediate grades, and found strong evidence of reliability ($r \leq .920$) in the intermediate grades (TEA, 2008).

Concurrent Verbalization as a Measure of Comprehension

Ericsson and Simon (1984, 1993) first classified types of verbalization measures. After examining verbalization studies, they identified different levels of verbalizations and described the process wherein subjects are asked to describe or explain what they are thinking. These studies demonstrated that participants were able to verbalize their thoughts about a task during the task, including mathematical tasks, without changing the accuracy of their performance. From their studies, these researchers developed recommendations for using verbalization as a measure, including using similar tasks as a warm-up exercise to give subjects practice in expressing their thoughts directly without explanation or indirect information.

A meta-analysis of 94 studies based on Ericsson and Simon's (1984, 1993) work affirmed that when subjects were asked to verbalize their thoughts during a task, their performance on the task was not altered. An effect size "indistinguishable from zero" (r = -.03) was found, even across factors such as type of task. This meta-analysis included verbalization on a wide range of math problems of varying difficulty (Fox et al., 2011).

A connection was also established between the speed of a subject's verbalizations and the difficulty of the task. When subjects were working under a heavy cognitive load, they tended to either stop verbalizing or provide less complete verbalizations (Fox et al., 2011). The difficulty of a test item impacted the verbalization because highly skilled students work automatically and less-skilled students have trouble explaining why they do not understand. Overall, verbalizations appeared to be an effective measure of comprehension with a variety of students on items of low to moderate difficulty (Johnstone et al., 2006).

Retell verbalization, or verbalization following reading, is a simple and efficient tool that, when used with a fluency measure, provided a more complete picture of reading comprehension (Baumann et al., 1993) and added significantly to validity of related inferences (Marcotte & Hintze, 2009). Students responding to a retell prompt must be able to recall information read in text and express it orally. This recall differs from simply recognizing information read, as assessed by multiple-choice questions referring back to text. Retell verbalization is an equivalent item across passages and often resembles classroom instruction (Reed, 2011; Roberts, Good, & Corcoran, 2005). DIBELS Next is an example of an assessment that uses retell verbalization to measure reading comprehension (McKenna & Good, 2003).

The Fox meta-analysis proposed that the most accurate method of understanding a subject’s cognitive processes is *concurrent verbalization,* or the verbalization of task-relevant thoughts generated between the start and completion of a task. The purpose of concurrent verbalization is to access the information contained in a subject’s short-term memory. Unlike other verbalization study methods such as probes or reflections, concurrent verbalization does not measurably interfere with a subject’s cognitive process, as it does not involve a reflection or other verbal or nonverbal encoding that may disrupt the task or influence the accuracy of performance (Fox et al., 2011). While validated using math problems, the focus of research using concurrent verbalizations has primarily been on reading and writing, where it has been used to examine the processes and strategies underlying the act of reading in students of both average and below-average abilities (Medina, 2008).

CHAPTER II

METHODS

Setting and Participants

The methods section describes the district and school where the study was set. It includes an explanation of the selection process, followed by a description of the participants and administrators. Finally, it describes the data-collection procedures in which students were administered a concurrent verbalization (CV) measure and a benchmark measure in both reading and math. These data were then analyzed using relational matrices, and significant relations were identified.

Setting

This study took place at a rural elementary school located in an approximately 5,000-student school district in the Pacific Northwest. The district operates six elementary schools, one middle school, one high school, and a K-12 alternative education academy. The district mandated a minimum of 60 minutes a day for core reading and core math instruction in the intermediate elementary grades. Adherence to the mandate was confirmed by administrative classroom observations during the 2011-2012 school year.

The setting is a K-6 elementary school with a population of 311 students. Its attendance area includes established farms, developments of larger homes, and two mobile home parks. Students are generally assigned to general education classrooms by their age. In addition to the 12 general education classrooms, the school hosts a selfcontained class for students with severe disabilities and an early-intervention preschool program, both operated by the educational service district, and a private parent cooperative preschool.

Table 2 lists the school's student population and subgroups for the entire school, and for the intermediate grades from which the participants were selected.

TABLE 2. School Enrollment by Subgroup

Subgroup	Number of Students				% of Total Population			
	K-6	3	4	5	K-6	3	4	5
Total students enrolled	311	41	44	45	100	13	14	14
Male	167	21	25	23	57	13	15	14
Female	144	20	19	22	46	14	13	15
Special Education (SpEd)	36	4	3	6	12	1	3	2
Free or reduced lunch (F/RL)	113	13	18	17	36	12	16	14
English Language Learners (ELD)	16	3	4	0	5	2	3	0
Gifted (TAG)	8	2	2	4	3		3	5

Note. Based on January 2012 school enrollment data.

The school was designated Title I-A for the 2011-2012 school year and targeted reading support to identified students in kindergarten, first and second grades. Students in the intermediate grades did not directly receive Title I-A support; reading support in intermediate grades was provided through the school's Response to Intervention program. Since 2010, the school continued its instructional focus of increased student engagement. In 2011-2012, it provided intermediate-grade teachers with professional development focused on improved math instructional practices, and implemented a CCSS-based curriculum.

Participants

Intermediate-grade students at the school were assigned by grade to one of two classrooms at Grades 3, 4, and 5. The intermediate grades are the earliest to take the state's large-scale assessment in math. Teachers in these six classrooms included four females and two males, all White, non-Hispanic. A fourth-grade teacher was in his second year of teaching and his first year at the school. The other five teachers had taught at the school from six to 27 years. All the teachers were licensed and designated highly qualified for their positions by the state.

Of the six teachers, five provided math instruction. In third grade, students were assigned to a leveled group for math, and instruction was then provided by the grade-level teachers. In fourth grade, math was taught in each homeroom. In fifth grade, math instruction was provided by one teacher to two leveled groups on a rotating basis. The fifth-grade teacher who provided math instruction participated in extensive, continuing district-led professional development about the CCSS implementation during the 2010-2011 school year. All intermediate-grade teachers participated in ongoing professional development targeting student engagement in math. All teachers participated in weekly data team meetings, where short-term learning targets for math were identified and student progress was measured. For the school year this study was conducted, progress was measured by growth demonstrated by students eligible for free or reduced lunch assistance.

All intermediate-grade students in these six classrooms were given the opportunity to participate in the study. Of the 128 students in the intermediate grades, 105 (or 82%) participated in the study. The recruitment procedures were approved in writing by the University of Oregon's Office for Protection of Human Subjects, the superintendent of the district, and the building administrator. Active consent was obtained from each participant prior to data collection. "Opt-out" consent letters were also mailed to parents and guardians. These letters explained the research study, outlined activities that would be involved, and detailed the anticipated benefits and possible risks of participation. The parent consent letters explicitly stated that participation was voluntary and that they had the right to revoke consent at any time during the study. The study was further described in the weekly school newsletter and in the administrator's bi-monthly email to parents.

Prior to the start of data collection, the study was explained verbally to students by the principal investigator in their classrooms. The explanation included a list of tasks they would be asked to perform, the assent process, and assurances that they could decline to participate at any time. They were also told their parents had been informed of the study. Participants were given a small incentive (e.g., a 25¢ gift certificate at the student store) for participation.

Table 3 describes the participants by student-level subgroups, in total and disaggregated by grade. The table reports the participants in both total number and percentage of the participant population.

TABLE 3. Student Participants by Subgroup

Subgroup	Total by Subgroup	Grade 3		Grade 4		Grade 5	
		N	%	*n*	%	*n*	%
Grade	105	40	38	28	27	37	35
Male	43	21	56	14	50	18	49
Female	52	19	48	14	50	19	51
SpEd (Language)	5	3	>.1	1	>1	1	>1
SpEd (Attention)	5	1	>.1	1	>1	3	8
SpEd (other)	3	0	0	0	0	3	8
F/RL	38	12	30	13	46	13	35
ELD	7	2	>.1	3	1	2	>1
TAG	6	2	5	0	0	4	1

Note. Based on January 2012 school enrollment data.

Measures

This section describes the CV measure that is the subject of this study. Benchmark measures of reading and math used to compare outcomes are also described.

Concurrent Verbalization Measure for Math

The CV measure used a set of specific protocols based on DIBELS Next retell, materials from concurrent verbalization studies (Johnstone et al., 2006; Johnstone, Liu, Altman, & Thurlow, 2007; Zucker, Sassman, & Case, 2004) and recommendations of the Fox et al. (2011) meta-analysis for concurrent verbalization protocols. The CV measure was also modified as a result of a pilot study. Each modification was targeted and addressed specific recommendations, considering that differences in wording of retell prompts have produced significant effects in student responses (Reed, 2011).

The scoring protocols of the CV measure were based on the DIBELS Next protocols, and contained a fluency word count and four scored components: (a) whether the participant read the math question correctly, (b) whether an appropriate strategy or math concept was provided, (c) whether a correct answer was provided, and (d) an assigned quality score.

The CV fluency word count was included to record the number of words used by participants to express task-relevant thoughts. Recording these thoughts has been defined as the intent of concurrent verbalization (Fox et al., 2011). Based on the pilot study, the scoring protocols counted individual words to 45, and a category for more than 45 words.

The scored component recording whether a participant successfully read the word problem aloud was included to determine whether participants' verbalizations were based on a correctly decoded problem. If a participant read the problem aloud with one or fewer errors, the administrator scored the problem as correctly read. Self-corrections and repeated or added words were not considered errors. The administrators did not begin counting words for the fluency score until after the student read the problem aloud; otherwise, words read aloud were not counted in the fluency word count.

The scored components of providing a strategy or math concept and of providing a correct solution were included based on the CCSS standards of mathematical practice. Measures under the CCSS assess students' ability to make connections between information, and to construct arguments that refer concretely to information, make sense, and are correct.

If participants provided a strategy that in the administrator's judgment could result in a correct solution, the administrator credited the participants with providing a strategy or concept. Suggested strategies and math concepts for each problem were included on the scoring sheet, but administrators were directed to also credit participants who provided a plausible strategy or concept that was not listed. For example, one verbalization problem related to estimating the total number of toy cars. Potential solutions provided included three number sentences a student could express to reach the solution. However, students who stated a different number sentence aligned with an appropriate strategy were given credit for providing a strategy.

When participants provided the correct answer, listed on the scoring protocols, they were credited. This component was included because the CCSS include the desired outcome that students will use mathematical thinking to reach a correct conclusion about a math problem.

The quality score was assigned by the administrator based on the number of details, not words, provided by participants, and the organization of the presentation. A quality score of 4 was assigned if participants provided three or more details and a correct answer in a manner that described the math strategy or concept in an organized way. A quality score of 3 was assigned if participants described three or more details that described the math strategy or concept in an organized way. A quality score of 2 was assigned if participants described three or more details about the problem. A quality score of 1 was assigned if participants described two or fewer details about the problem and their verbalization lacked a solving process or they appeared to guess. The quality score was included in the scoring protocol to compare participants' fluency to the content of their verbalizations.

The protocols from DIBELS Next were initially modified by changing the quality of retell score descriptions to reference math comprehension instead of reading. The DIBELS Next scoring sheet is found in Appendix B. After the pilot study, the protocols and scoring sheet were further modified based on interviews of students and administrators in order to support content validity (Reed, 2011). The protocol script used for the study included a description of the purpose of the verbalization, modeling by the administrator, an opportunity for the subject to practice, and a reminder of the purpose (Johnstone et al., 2006). The script also used a practice question at each grade level as its subject, selected from the TAKS practice tests (TEA, 2011). The script, practice questions and subject questions are seen in Appendices C, H, and I.

The TAKS sample problems were selected from sample tests provided by the TEA online (TEA, 2011). The problems were first sorted by grade level. Problems with graphics such as charts were excluded. This exclusion was based on concerns regarding the variable of nonverbal problem solving, or the ability to complete patterns presented visually, which has been identified as a unique predictor of word problem-solving skills in primary grades (Fuchs et al., 2006).

As a method of minimizing concerns regarding language load and vocabulary, problems were excluded if they included more than one specialized math vocabulary word, defined in this study as words that, used in context, have a specialized mathematical meaning (Helwig, Rozek-Tedesco, & Tindal, 1999). For example, a problem was considered if it stated, “Kim is building a table. The *perimeter* is 48 ft. What is the length of one of the sides?” but excluded if it stated, “Kim is building a table in the shape of an *octagon*. The *perimeter* is 48 ft. What is the length of one of the sides?” Sample problems are seen in Appendix I.

After these exclusions, a set of six or seven problems at each grade level remained from the TEA sample TAKS test set. These problems were then checked for alignment with the grade-level standards in Oregon. Problems were screened to conform in content and difficulty. Each set of grade-level problems had an identical number of steps to solution. For example, if a problem at third grade required the participant to compute the perimeter based on the provided measurement of two of the sides (one step), a problem at fourth grade was excluded if it required computing the perimeter, then using it to solve an equation (two steps).

After screening, three problems were selected at each grade level. Answers to the word problems were not provided to the subjects, following the pilot study findings. Appendix I contains the nine problems used for participants’ verbalizations.

Benchmark Measures

Two measures, DIBELS Next reading assessment and easyCBM math tests, were used to provide benchmark scores in reading and math. The criteria for selection for the benchmark measures included being research-based and widely available in elementary settings.

Reading Benchmark

DIBELS Next is a curriculum-based assessment of grade-level reading skills (Good & Kaminski, 2002; Kaminski & Cummings, 2007). At each grade level, DIBELS Next measures and reports reading fluency, reading comprehension, and a composite score reflecting both fluency and comprehension scores and percentile rank. DIBELS Next benchmark goals were developed based on data from a large-scale study, and reflect the probabilities of students meeting later important reading outcomes (DMG, 2011). Fluency is measured by providing subjects with three grade-level passages to read aloud while an administrator records the words read correctly in one minute. To measure comprehension, the DIBELS Next has two subtests: retell and DAZE (Dynamic Measurement Group [DMG], 2011). Retell is a type of verbalization (Fox et al., 2011). DIBELS Next retell protocols require the retell measure to be administered immediately following oral reading of each of the fluency passages. The retell scores included retell fluency, which measures the number of on-topic words students use to describe the passage they just read. Retell quality is scored based on the number of details provided, the order in which they were provided, and whether they contained the main idea of the passage. The scoring process is identical in each of the intermediate grades (DMG, 2011). An example of the scoring sheet from 5th-grade DIBELS Next can be seen in Appendix B.

Math Benchmark

EasyCBM is a curriculum-based assessment program for math and reading. EasyCBM focused on conceptual understanding of a student's knowledge and skills, targeted in the Focal Point Standards of NCTM (Nese et al., 2010). EasyCBM was found to be a reliable measure of math skills in the intermediate grades (Anderson et al., 2009). Its scores are strongly correlated to student scores on large-scale assessments in math (Anderson, Alonzo, & Tindal, 2010a; Nese et al., 2010), both historically and under current Oregon math standards (Anderson, Alonzo, & Tindal, 2010b; Anderson et al., 2011; Park, Irvin, Anderson, Alonzo, & Tindal, 2011). At each intermediate grade level, easyCBM provides three math tests based on grade-level standards, curriculum, and large-scale assessments. Research-based grade-level targets are included in the program (Anderson et al., 2009; Anderson et al., 2010b; easyCBM, 2009). The sum of the three subtest scores may be used, with the percentile rank, to compare student scores (G. Tindal, personal communication, February 17, 2012).

EasyCBM is an effective way to measure persistently low-performing students under the four guidelines of validity, reliability, accessibility, and alignment (Anderson et al., 2011). While there are anticipated changes to the math standards found in the CCSS, easyCBM retains the potential to identify students who are at risk (Anderson et al., 2010b).

Pilot Study

In order to increase the reliability of the CV measure, a pilot study was conducted. The pilot study included three practice administrations: Fall 2010, Winter 2010, and Fall 2011. As with the study, the pilot was administered in conjunction with the DIBELS Next screener in reading.

Pilot Administrators and Training

For the fall 2010 and winter 2010 pilot phases, administrators were three certified teachers, four instructional assistants and two work-study students from an education program at a local community college. The teachers had between 5 and 10 years of experience in education. In addition to their licenses in elementary multiple subjects, one teacher had a license endorsement in reading, one in special education and one in both. The instructional assistants were all designated highly qualified by the state of Oregon based on their college-level course work and experience in education. Of the two work-study students, one was completing her second year of a planned 4-year degree in education and the other was in the last two semesters of a program to certify educational instructional assistants.

Prior to the first pilot phase, administrators were trained on the DIBELS Next retell protocols. The instructional assistants attended a one-hour training session. The training included a review of the DIBELS Next administration guide, a period for questions and answers, followed by 15 minutes of guided practice with modeling of retell administration and scoring. The teachers participated in a 15-minute training period specifically relating to DIBELS Next retell verbalization protocols. In addition to their experience in administering assessments, all teacher-administrators had experience and training in administering the DIBELS Next reading fluency measures.

Pilot Concurrent Verbalization Measure Development

For the first pilot phase, a problem in the strand of Numbers and Operations was selected at each grade level from the OAKS practice tests (ODE, 2010b). Each grade level's problems were printed from the ODE website with the answers. The problem and the prompts were cut and pasted into a format similar to that of the DIBELS Next reading measure, including the DIBELS Next scoring protocols (see Appendices A and B). Specifically, the retell prompt and procedures were copied from the DIBELS Next reading retell with the substitution of the word "problem" for the word "story." Administration protocols from the DIBELS Next reading retell were followed. Students were allowed one minute to read the problem, and one minute to retell. The problem was removed from the student's view after the first one-minute period. Administrators recorded both retell fluency and quality of retell scores, based on the reading retell protocols.

Following the first pilot phase, students and administrators were interviewed. Both groups reported confusion over specific aspects of the pilot measure and protocols.

Students reported that they did not understand that the intent of the instructions was to describe the components of the problem. Administrators reported that when students were shown the math word problem with the answers, students would frequently respond by providing the answer they selected—e.g., “12” or “d.” Time for a student’s response would expire as he or she struggled to understand the task, and additional prompting or explanation was often required to obtain an on-topic student response.

Administrators also reported that about one in three students finished reviewing the single math word problem before the one-minute period was over, creating a delay between their review and their prompt to retell. During this delay, students engaged in off-task behaviors, including conversing with administrators, asking questions unrelated to the testing, and looking around.

Administrators further reported a lack of clarity in assigning a retell quality score to a response based on a math word problem instead of a reading passage. They reported difficulty in applying the reading retell criteria to a math retell. Administrators noted that students would provide three or more details that seemed equally meaningful and the sequence of the details, which are important in comprehending a story or passage, but this provision did not relate to the student’s understanding of the math problem.

Modifications for the Second Phase of Pilot Administration

In the second pilot administration, the problems were again selected from the grade-level sample tests provided by the sample OAKS tests (ODE, 2010b). The number of problems was increased from one to four, in order to address student off-task behavior observed in the first pilot phase when only one question was presented. The answers to the problems were initially included and visible to the students to evaluate the impact on student verbalizations. After one round of testing, the answers were removed and not made visible to the students. Students were given one minute to review the problems and one minute to retell.

When interviewed, students reported that the prompts did not clearly describe the task. Administrators reported that student responses were more likely to be on topic and their behaviors on task during the response period with the multiple problems presented. They reported that when the problems remained visible to students, students would reference the problems to confirm facts. Administrators further reported that when the answers were not visible, student responses were more likely to include verbalization about the problem, rather than a statement of their selection of the answer.

Modifications for the Third Phase of Pilot Administration

In the third administration, the math word problems were drawn from the TAKS sample tests (TEA, 2011) to address concerns about student exposure to the OAKS sample test questions. Students were provided verbal modeling and a sample problem on which to practice the verbalization task (Fox et al., 2011). The prompts were modified to describe a concurrent verbalization rather than a retell verbalization task. The prompts were also modified to clarify the task for intermediate students. Three problems were presented to each student with an increasing number of steps; i.e., the first problem shown had one step, the second problem had two steps, and the third problem had three steps. Students were prompted to read the problem aloud prior to verbalizing, in order to record their access to the problem.

The Quality of Verbalization score was emphasized and further defined. The administrator's training and scoring included (a) whether the subject stated the mathematical function or principle correctly (e.g., "I need to add" or "I need to find the pattern"); (b) the number of pertinent, separate facts or strategies the subject provides; and (c) whether the subject provided the correct answer.

Following the third phase of the pilot, administrator interviews confirmed improvement in students' ability to comprehend the task using the revised prompts and protocols. Student responses were also compared to previous pilot data to confirm the increased understanding of the task.

Study Protocols

Based on revisions from the pilot study, the protocols included a script and coding sheet based on DIBELS Next procedures and modified for the CV administration. This script and coding sheet included the administrator prompts, and the protocols for timing and discontinuance. The scoring sheet shows the problems the participants used to verbalize, and a set of numbers to record the number of on-topic words the participants used to describe their strategy to solve the problem. The sheet also included a scoring area to record the quality score and mark whether the participant read the problem correctly, provided a strategy, or provided the correct answer.

A “yes” score of correctly read was assigned if the participant read the problem aloud to the administrator with one or fewer errors. Administrators provided words a participant was unable to read after 3 seconds, following the DIBELS Next protocols (DMG, 2011). If an administrator provided more than one word, or if the participant was unable to read more than one word correctly, the participant would receive a score of “no.” If the participant did not read the problem after 20 seconds, the administrator would prompt the participant according to the protocols. After 10 additional seconds, if the participant did not respond, the administrator would discontinue scoring.

Participants who provided a mathematical strategy or described a mathematical concept related to the problem received a “yes” score in the strategy/concept category. For example, if a participant said, “I need to add these together,” or “11+11+11,” or “If you put these together over and over, then you’ll find the answer,” the participant would have been scored as having provided a strategy. While the scoring sheet listed suggested strategies and concepts, a participant score was not based on the assessor’s judgment of whether the strategy or concept was correct, or aligned with the examples provided.

Participants who provided the answer listed on the scoring sheet within the 2-minute time period were marked as providing the correct answer. Participants who provided the correct answer after the 2-minute time period expired were scored “no,” as they did not provide the correct answer in the scoring period.

Administrators assigned a quality score of 1-4 for each participant’s verbalization according to the guidelines provided. These guidelines were based on the number of on-topic details provided by the participant, the order in which they were provided, the connection stated by the participant to a math process, and whether the correct answer was provided. In order to receive a quality score of 4, a participant was required to provide the correct solution to the problem. In order to receive a score of 3, a participant provided three or more organized details about the problem with a strategy or concept to solve, but did not provide the correct solution. In order to receive a score of 2, a participant provided three or more details lacking organization or a strategy or mathematical concept. In order to score a 1, a participant provided two or fewer details in no particular order or in an attempt to guess. Appendix C contains copies of the script and coding sheets.

Study Administrators

Administrators for this study included staff members from the subject school and a retired teacher from a nearby district. The administrators were all designated by the State of Oregon as highly qualified for their positions in elementary education, reading, or special education, and had experience administering DIBELS Next reading assessment.

Administrator Training

Prior to the study, all administrators attended a 30-minute training session. Training materials included the DIBELS Next reading retell scoring sheet, the CV script and scoring sheet, a video of individual students responding to the CV math retell prompt, an integrity checklist and written general guidelines for the study administration.

During the training session, the lead investigator described the CV measure and protocols. Time for questions and answers were provided. Administrators then individually score three videotaped examples of student retell. After each example, the scores were discussed in order to promote interrater reliability. During the third video sample, the lead investigator scored the administrators using an Assessment Accuracy Checklist (see Appendix G) adopted from the DIBELS Next checklist (see Appendices E and F). The administrator who did not demonstrate consistency in rating student retell received additional opportunities to practice after the training period until consistency with the scoring protocols was demonstrated. Prior to the sample, all administrators demonstrated consistency of 95% or greater with the Assessment Accuracy Checklist.

Administering the Measure

The study was conducted in January 2012, during a one-week period within the winter testing window described in the DIBELS Next manual (DMG, 2011). DIBELS Next and the CV measures were administered in the school library, an approximately 20 x 40 meter carpeted room with four 5-meter-wide floor-to-ceiling dividers down the center width of the room. Six tables approximately 2 meters long and .5 meters wide were positioned at least 5 meters apart. No participant had a direct sight line to any other participant. The room was closed to other uses, limiting nonparticipant access. No fire drills, announcements, or other interruptions occurred during the study administration.

Before participants arrived in the library, each administrator was provided a packet containing a set of scripts and scoring sheets for DIBELS Next (see Appendices A and B) and the concurrent verbalization measure (see Appendix C). The packet included reminders of protocols, the language of the specific instruction, prompts and the frequency of the allowable prompts. The packet also included a set of TAKS sample questions for each grade level. A supply of sharpened pencils and timers were readily available.

Participants were led by a staff member to the library in groups of four to six. On reaching the library, if an administrator was not immediately available, participants were directed to wait in an area separated from the administrators by approximately 6 meters and facing away from the active participants. As administrators become available, participants were randomly directed to their tables for testing.

DIBELS Next Administration

The DIBELS Next reading fluency and retell subtests were administered first. Administrators followed the DIBELS Next protocols, reminding participants that a retell verbalization would be prompted after each fluency sample. Participants were directed to read each of three passages aloud for one minute. Immediately following each timed reading, participants were directed to “tell me [the administrator] everything you can about the story you just read.” Administrators recorded the participant’s responses on a coding sheet (DMG, 2011). This coding sheet included protocols and reminders about what and how often prompts were allowed (see Appendices A and B).

CV Administration

Immediately following the three DIBELS Next reading fluency and retell samples, the concurrent verbalization measure was administered. Participants were first provided a sample problem to review while the administrator modeled concurrent verbalization (see Appendix C). Participants were then asked to practice the verbalization on a different sample problem (see Appendix H).

After modeling and practice, participants were provided the three sample TAKS math problems (see Appendix I) and directed to begin verbalizing. As they spoke, the administrator recorded the responses on the CV coding sheet (see Appendix C). After 2 minutes, participants were directed to stop and the data gathering ceased.

After the DIBELS Next and CV measures were administered to all participants, the scoring sheets were collected and reviewed for omissions. Both DIBELS next and CV scores were assigned following the DIBELS Next protocols, which state that after scoring three samples, the mean score from each sample is identified and recorded as the participant's score on that sample. These protocols also state that each scoring category is considered independently (DMG, 2011). For example, if a participant verbalized on the three sample problems and received quality scores of 2, 3 and 3, a quality score of 3 would be assigned for that participant. If he or she provided a correct answer, but on the problem that received a quality score of 2, a "yes" for providing a correct answer would be recorded. If participants were unable to complete verbalizing on the three math problems before the time period elapsed, a score of 0 was assigned on the unfinished verbalization scores and the mean score was selected.

EasyCBM Administration

The easyCBM measure was administered online in the school's computer lab during the same month as the reading and math verbalization sample. Participants completed the math tests designated at their grade level for the winter term. After participants completed the on-line test, their scores were posted automatically on the easyCBM website and accessed via password by the principal investigator. The administration followed the protocols described in the easyCBM handbook, including not allowing participants to use a calculator. EasyCBM math tests are not timed (easyCBM, 2009). Participants who did not choose to complete the test in one session were provided multiple opportunities to complete the test at other times.

Data Analyses

Correlational analyses have been used in educational research when seeking to investigate the association between variables (Babbie, 2007). The current study was designed to test the concurrent verbalization measure by identifying and describing correlations between CV scores and scores on existing reliable measures. It is for this reason that an examination of descriptive statistics of the measure's outcomes was conducted, and a correlational matrix was created and analyzed.

Prior to the analysis, all data were entered into an SPSS data file. There were no missing data. The scores on the variables relating to the CV measure were identified using the same method as DIBELS Next scores. Three scores were collected, and the median score in each category was used (DMG, 2011).

Two categories of variables were used in the study. The first category describes the features of individual participants, or the student-level variables. The student-level variables and definitions of the codes are provided in Table 4.

TABLE 4. Student-Level Variable Names, Descriptions, and Coding Definitions

Variable Name	Variable Description	Coding
Number	Randomly assigned participant number	
Teacher	Classroom teacher of the of participants	1, 2 = 3rd grade 3, 4 = 4th grade 5, 6 = 5th grade
Grade	Male or Female	1 = Female 2 = Male
SpEd	Special Education Eligibility	0 = not identified 1 = identified
F/RL	Free or Reduced Lunch Eligibility	0 = not eligible 1 = eligible
ELD	English Language Learners Students currently eligible for services in the district’s English language development program, by scoring lower than a proficient level on the English Language Performance Assessment.	0 = not eligible 1 = eligible
TAG	Talented and Gifted Eligibility Students scoring in the 99th percentile or above on a standardized measure.	0 = not eligible 1 = eligible

The second category of variable describes the scores obtained on components of the measure, or the *measures variables.* The measures variables and definitions of the codes are provided in Table 5.

TABLE 5. Measures' Variable Names, Descriptions, and Coding Definitions

Variable Name	Variable Description	Coding
DIBELS ORF	Oral Reading Fluency on DIBELS Next Number of words read correctly in one minute.	
DIBELS re-fl	Retell Fluency on DIBELS Next Number of on-topic words provided in one minute.	
DAZE adj	DAZE adjusted score Composite score reflecting comprehension on a three minute timed passage.	
DIBELS comp	DIBELS Next composite score Composite score of fluency, comprehension, and percentile rank.	
CV ORF	Fluency on Concurrent Verbalization sample Number of on-topic words provided in two minutes.	
CV Rd y/n	Correctly read aloud the Concurrent Verbalization problems	0 = no 1 = yes
CV Strat y/n	Provided a strategy to solve the CV problems	0 = no 1 = yes
CV Solv y/n	Provided a correct solution for the CV problems	0 = no 1 = yes
CV quality	Quality score on the CV problems Composite score of fluency, reading, strategy, and solution	1-4
Math Comp	easyCBM math composite score	
Math PR	easyCBM math percentile rank score	

CHAPTER III

RESULTS

Analyses were conducted by grade level on data collected from all 105 participants. Descriptive statistics are reported for the benchmark measures, the verbalization measure, and the student-level variables. Scatterplots are described for specific variables.

Descriptive Statistics

To report and examine the data relevant to the research questions, I created tables for the variables at each grade level. I first compared the participant scores on each component of the CV measure. I then reported participants' scores on the benchmark measures and compared them to the benchmark targets to confirm that the study data are aligned with these targets. I also reported the effect of student-level variables on participant scores on the CV measure.

Descriptive Statistics for the Concurrent Verbalization Measure

Statistics for each component of the CV measure are reported at each grade level. In addition, I examine cross-tabulations and report specific findings of interest.

Third Grade

The data indicate a difference of .72 in the mean of the dichotomous variables relating to participants who provided a strategy (.92) and participants who provided a solution (.20). All third-grade participants read the measure correctly. Descriptive statistics for the third grade are reported in Table 6.

TABLE 6. Descriptive Statistics for the Concurrent Verbalization Measure, Internal Relations—Third Grade (n = 40)

Variable	*M*	*SD*	Minimum	Maximum	Range
CV Fluency	26.98	13.11	7	45	38
CV Read	1	0	1	1	0
CV Strategy	.92	.27	0	1	1
CV Solved	.20	.41	0	1	1
CV Quality	2.75	.93	1	4	3

Fourth Grade

The data indicate a difference of .53 in the mean of the dichotomous variables relating to participants who provided a strategy (.82) and participants who provided a solution (.29). All fourth-grade participants read the measure correctly. Table 7 reports the descriptive statistics for fourth grade.

Fifth Grade

The data indicate a difference of .72 in the mean of the dichotomous variables relating to participants who provided a strategy (.62) and participants who provided a

TABLE 7. Descriptive Statistics for the Concurrent Verbalization Measure, Internal Relations—Fourth Grade (*n* = 28)

Variable	*M*	*SD*	Minimum	Maximum	Range
CV Fluency	25.36	12.43	6	45	39
CV Read	1	0	1	1	0
CV Strategy	.82	.39	0	1	1
CV Solved	.29	.46	0	1	1
CV Quality	2.64	1.13	1	4	3

solution (.03). A review of a cross-tabulation of these two variables revealed that of the 37 participants, only one provided a correct solution. All fifth-grade participants read the measure correctly. Descriptive statistics for the fifth grade are reported in Table 8.

TABLE 8. Descriptive Statistics for the Concurrent Verbalization Measure, Internal Relations—Fifth Grade (*n* = 37)

Variable	*M*	*SD*	Minimum	Maximum	Range
CV Fluency	27.51	15.25	3	3	42
CV Read	1	0	1	1	0
CV Strategy	.62	.49	1	0	1
CV Solved	.03	.16	1	0	1
CV Quality	2.08	1.14	4	1	3

Descriptive Statistics for the Benchmark Measures

Descriptive statistics for the DIBLES Next, CV, and easyCBM variables are presented in Tables 9, 10, and 11. The data indicate the dispersion of DIBELS Next and easyCBM scores are consistent with previous research of their reliability. All scores are within the standard deviations of the benchmark targets set by each measure (DMG, 2011; easyCBM, 2009).

Third Grade

At this grade, the mean of the participant scores on DIBELS Next oral reading fluency was 17.2 words above the benchmark measure target. The scores were also above targets for DIBELS Next retell (14.1 words above) and DAZE adjusted score (3.6 points). Table 9 shows the results of the benchmark measures for third grade.

Fourth Grade

At this grade, all DIBELS Next scores were above the targets set by the benchmark measure except the DAZE adjusted score, which was 1.5 point below the target. The data indicate the dispersion of DIBELS Next and easyCBM scores are consistent with previous research of their reliability (DMG, 2011; easyCBM, 2009). The results for the fourth grade are shown in Table 10.

TABLE 9. Descriptive Statistics for the Benchmark Measures—Third Grade (*n* = 40)

Variable	*M*	Target	*SD*	Range	District *M*	District *SD*
DIBELS ORF	103.2	86	30.7	133	99.1	40
DIBELS retell fluency	40.1	26	16.6	74	39.9	16.5
DIBELS retell quality	2.95	2	.78	2	2.9	.8
DAZE adjusted score	14.6	11	7.2	31	14.4	7.2
DIBELS composite	340.7	285	86.9	377	342.9	89.9
easyCBM composite	35.9	35	5.3	25	0.0[a]	0.0[a]
easyCBM percentile rank	55.5	50	26.4	96	0.0[a]	0.0[a]

Note. Adapted from DMG (2011), easyCBM (2009), and easyCBM (n.d.).

[a]easyCBM is not administered district-wide.

TABLE 10. Descriptive Statistics for the Benchmark Measures—Fourth Grade (*n* = 28)

Variable	*M*	Target	*SD*	Range	District *M*	District *SD*
DIBELS ORF	102.9	103	30.8	129	109.2	39.2
DIBELS retell fluency	41.4	30	19.2	72	40.1	17.3
DIBELS retell quality	2.5	2	.7	3	2.6	.7
DAZE adjusted score	15.5	17	5.9	22	16.7	6
DIBELS composite	352.3	330	86.7	313	361.4	82.5
easyCBM composite	34.8	35	6.4	29	0.0[a]	0.0[a]
easyCBM percentile rank	49.4	50	30.7	98	0.0[a]	0.0[a]

Note. Adapted from DMG (2011), easyCBM (2009), and easyCBM (n.d.).

[a]easyCBM is not administered district-wide.

Fifth Grade

The data for fifth grade contain scores above and below targets for the benchmark measures. In reading, DIBELS Next oral reading fluency, DIBELS retell fluency, and DIBELS composite scores are above the targets. For math, both the easyCBM composite and percentile rank are below target, the latter by 17 points. The data indicate the dispersion of DIBELS Next and easyCBM scores are consistent with previous research of their reliability (DMG, 2011; easyCBM, 2009). Table 11 reports the results for the fifth grade.

TABLE 11. Descriptive Statistics for the Benchmark Measures—Fifth Grade (n = 37)

					District	
Variable	*M*	Target	*SD*	Range	*M*	*SD*
DIBELS ORF	128.8	120	36.6	205	128	32.6
DIBELS retell fluency	47.8	36	19.6	77	46.3	20.8
DIBELS retell quality	2.76	3	.86	3	2.8	.8
DAZE adjusted score	18.1	20	7.5	38	18	8.1
DIBELS composite	403.4	372	98.8	500	398.6	101.4
easyCBM composite	30.7	35	5.9	23	0.0[a]	0.0[a]
easyCBM percentile rank	35.0	50	26.7	2	0.0[a]	0.0[a]

Note. Adapted from DMG (2011), easyCBM (2009), and easyCBM (n.d.).

[a]easyCBM is not administered district-wide.

Descriptive Statistics for the Student-Level Variables

The number of participants in each student-level subgroup—gender (female or male), special education identified, eligible for Free or Reduced Lunch, English language learners (ELD), and eligible for Talented and Gifted (TAG) programs—were compiled. The statistics for these student-level variables, by total number and percentage of the participants at each grade, are reported in Table 12.

TABLE 12. Descriptive Statistics—Student-Level Variables

	Grade 3		Grade 4		Grade 5	
Variable	*n*	%	*n*	%	*n*	%
Total participants	40		28		37	
Female	19	48	14	50	19	50
Male	21	52	14	50	18	50
SpEd	4	10	2	7	7	19
F/RL	12	30	13	46	13	35
ELD	2	5	3	10	2	5
TAG	2	5	0	0	4	11

Correlations

I conducted an analysis using correlation matrices to identify relations with the CV scores, the scores for each benchmark measure, and the student-level variables. This analysis was conducted separately for the third, fourth, and fifth grades.

Relations Within the Concurrent Verbalization Variables

I examined the results from the variables within the CV measure, in order to identify relations. These results are reported to describe relations within the concurrent verbalization measure.

Third Grade

In third grade, relations were identified with CV quality and all other components of the measures. All third-grade participants successfully read the math sample problems. Table 13 reports the relations in the third grade.

TABLE 13. Relations for the Concurrent Verbalization Measure—Third Grade (*n* = 40)

Variable Name	CV Fluency	CV Strategy	CV Solved	CV Quality
CV Fluency	1	-.140	.233	.346*
CV Strategy	-.140	1	.142	.441**
CV Solved	.233	.142	1	.546**
CV Quality	.346*	.441**	.546**	1

*Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).

Fourth Grade

Relations were identified with CV quality and all other components of the measures in fourth grade. All participants in this grade successfully read the math sample problems. Table 14 reports the relations for the fourth grade.

TABLE 14. Relations for the Concurrent Verbalization Measure—Fourth Grade (n = 28)

Variable Name	CV Fluency	CV Strategy	CV Solved	CV Quality
CV Fluency	1	.227	.299	.381
CV Strategy	.227	1	.295	.523**
CV Solved	.299	.295	1	.703*
CV Quality	.381*	.523**	.703*	1

*Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).

Fifth Grade

In fifth grade, relations were not identified with CV quality and CV solved, however relations were identified with CV quality and all other components of the measures. Unlike third and fourth grades, in fifth grade, relations were found with CV strategy and fluency (see Table 15). All participants in this grade successfully read the math sample problems.

TABLE 15. Relations for the Concurrent Verbalization Measure—Fifth Grade (n = 37)

Variable Name	CV Fluency	CV Strategy	CV Solved	CV Quality
CV Fluency	1	.416*	.194	.555**
CV Strategy	.416*	1	.130	.602**
CV Solved	.194	.130	1	.285
CV Quality	.555**	.602**	.285	1

*Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).

Relations of the Concurrent Verbalization Variables With Benchmark Measures

I examined the CV measure and the benchmark assessments scores. Tables 16, 17 and 18 contain the relations between the scores of these measures.

Third Grade

In third grade, comparisons with the reading benchmark identified two significant relations ($p < .01$) between the CV providing a solution and DIBELS Next oral reading fluency ($r = .450$), and composite score (r = 452). Relations identified included those between CV Fluency and DIBELS Next retell quality. I then created a histogram to examine the dispersion of the DAZE adjusted scores. This examination revealed a positively skewed distribution, with a Kurtosis value of -.54 and a Skew value of 1.16.

TABLE 16. Benchmark Measures and CV Relations—Third Grade (*n* = 28)

Variable Name	CV Fluency	CV Strategy	CV Solved	CV Quality
DIBELS ORF	.279	-.136	.450**	.186
DIBELS retell fluency	.394*	-.028	.309	.325*
DIBELS retell quality	.362*	-.018	.275	.336*
DAZE adjusted score	.108	-.334*	.332*	.048
DIBELS composite	.291	-.069	.452**	.291
easyCBM math composite	.110	.142	.321*	.329*
easyCBM math percentile rank	.146	.173	.345*	.348*

*Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).

TABLE 17. Benchmark Measures and CV Relations—Fourth Grade (*n* = 28)

Variable Name	CV Fluency	CV Strategy	CV Solved	CV Quality
DIBELS ORF	.112	.423*	.478**	.511*
DIBELS retell fluency	.340	.052	.645**	.487**
DIBELS retell quality	.527**	.230	.547**	.348
DAZE adjusted score	-.054	.421*	.628*	.495**
DIBELS composite	.187	.513**	.671**	.587**
easyCBM math composite	.173	.402*	.564**	.594**
easyCBM math percentile rank	.198	.437*	.608**	.663**

*Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).

TABLE 18. Benchmark Measures and CV Relations—Fifth Grade (*n* = 37)

Variable Name	CV Fluency	CV Strategy	CV Solved	CV Quality
DIBELS ORF	.314	.367*	.378*	.595**
DIBELS retell fluency	.523**	.136	.350*	.453**
DIBELS retell quality	.350*	.366*	.243	.529**
DAZE adjusted score	.366*	.302	.446**	.444**
DIBELS composite	.450**	.284	.429**	.542**
easyCBM math composite	.369*	.329*	.357*	.409*
easyCBM math percentile rank	.359*	.312	.431**	.399*

*Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).

Fourth Grade

In reading, significant relations ($p < .01$) were identified between nine components of the CV and the DIBELS Next measures. Relations identified included a moderately strong relation ($r = .527$) between CV Fluency and DIBELS Next retell quality. In math, significant relations ($p < .01$) were identified between four of these components, and relations to $p <.05$ were identified between two others.

Table 17 shows the benchmark reading and math measures and the level of significance with each CV variable for the fourth grade.

Fifth Grade

An examination of the reading benchmark data identified relations ($p < .05$) between 15 of the 20 components of the CV and DIBELS Next measures, including moderately strong relations with CV quality and DIBELS Next oral reading fluency (r = .595) and retell quality (.529). Moderately strong relations were also identified with CV Fluency and DIBEL Next retell fluency (r = .523).

With the math benchmark, relations to at least p <.05 were observed between five of the six components of the measures. Table 18 shows the benchmark reading and math measures and the level of significance with each CV variable for the fifth grade.

Student-Level and Concurrent Verbalization Variables

In order to identify the potential effect of student-level variables on the CV measure outcomes, I examined the data from the measure for participants in each subgroup. I first examined the effect of grade, reflective of student age, on the CV outcomes. I then examined the effect of gender, special education identification and socioeconomic status, identified by free or reduced lunch eligibility on CV scores. Due to an insufficient number of participants, SpEd, ELD, and TAG data are not reported.

Grade and Concurrent Verbalization Outcomes

To screen for the effect of the student-level variable of grade on the CV scores, I examined the data at each grade level. Table 19 shows the relation at each grade level and each component of the CV measure.

TABLE 19. Means of the Concurrent Verbalization Measure by Grade

Variable	Grade 3	Grade 4	Grade 5
CV Fluency	26.9	25.3	27.5
CV Read	1	1	1
CV Strategy	.9	.8	.6
CV Solved	.2	.3	.03
CV Quality	2.7	2.6	2.1

Third Grade

In third grade, no significant positive relations ($p < .01$) were found. Significant negative relations ($p < .01$) were identified between the CV fluency of verbalization and student socioeconomic status, defined by Free/Reduced Lunch eligibility ($r = -.332$). Table 20 lists the CV and student-level variables of gender, Special Education identification and Free/Reduced Lunch eligibility, and the level of significance of each relation.

TABLE 20. Concurrent Verbalization and Student-Level Variable Relations—Third Grade ($n = 40$)

Variable Name	CV Fluency	CV Strategy	CV Solved	CV Quality
Gender	.095	.109	.225	.232
SpEd	.084	.095	.250	.273
Free / Reduced Lunch	-.332*	-.021	-.055	-.238

Fourth Grade

In fourth grade, no relations of any significance were found ($p < .01$ or $p < .05$) in the student-level data. Table 21 lists the CV and student-level variables and the level of significance of each correlation for the intermediate grades.

TABLE 21. Concurrent Verbalization and Student-Level Variable Relations—Fourth Grade ($n = 28$)

Variable Name	CV Fluency	CV Strategy	CV Solved	CV Quality
Gender	.175	-.280	-.158	.064
SpEd	-.088	-.233	-.175	-.036
Free / Reduced Lunch	-.191	-.127	-.272	-.088

Fifth Grade

In fifth grade, a negative relation to $p < .05$ was identified with CV quality and students identified in special education ($r = -.403$). Table 22 presents student-level variables and the level of significance of the CV variables for the fifth grade.

TABLE 22. Concurrent Verbalization and Student-Level Variable Relations—Fifth Grade ($n = 37$)

Variable Name	CV Fluency	CV Strategy	CV Solved	CV Quality
Gender	-.288	-.021	.171	-.118
SpEd	-.209	-.050	-.081	-.403*
Free / Reduced Lunch	-.055	-.009	-.123	-.103

Summary of Results

The findings of the current study include descriptive statistics that show participant scores at or above the targets set by the benchmark measures, with two exceptions. The scores on the DAZE comprehension component were slightly below target in the fourth and fifth grades. The scores on easyCBM composite were slightly below target in fourth grade; however, the fifth-grade scores show a larger distance from the targets, both composite (*M* = 30.7, target = 35) and percentile rank (*M* = 33rd, target = 50th).

The findings identified relations between concurrent verbalization scores and the benchmark assessment scores on DIBELS Next reading and easyCBM math, although the relations varied between grades. The data failed to indicate a relation between concurrent verbalization fluency and math scores, except in fifth grade. However, relations were found between the concurrent verbalization quality of response scores and benchmark math scores at each grade level.

Of the student-level variables, the effect of participants' grade on CV strategy, solved, and quality scores was identified. The effect of other student-level variables was noted in isolation, and was not consistently identified across grade levels.

CHAPTER IV

DISCUSSION

Researchers continue to seek to define the multiple skills necessary for students to solve and communicate about solutions of math word problems. While the math skills of comprehension and communication have been called essential to classroom math instruction (Griffin & Jitendra, 2009; ODE, 2011; Rittle-Johnson et al., 2001), practitioners and measures do not always evaluate them directly, distinct from other math skills (Clark & Shinn, 2004; Fuchs et al., 2008). Multiple studies have implicated working memory as a predictor of math comprehension skills (Siegel & Ryan, 1989; Swanson & Beebe-Frankenberger, 2004), yet a recent meta-analysis stated that "no studies" have teased out the relations of the various functions of students' working memory on performance to math comprehension tasks (Fuchs et al., 2008). Meanwhile, focus on these skills is an integral part of the Common Core State Standards, adopted by the majority of states (ODE, 2011), and these current changes in national educational standards and assessments have moved identifying these skills to the forefront of practitioners' conversations.

After a decade of shifting math standards and curriculum (Klein, 2003), the 2014 implementation of the CCSS appears to be aligning the external demands of a shifting curriculum with learning the complex language of math. The CCSS includes standards for mathematical practices on which to base a language-rich math curriculum, and includes requirements to precisely communicate about math concepts, processes, and solutions (CCSS, 2011). Assessments of the CCSS are intentionally designed to minimize the impact of unnecessary linguistic complexity, but may provide an opportunity for students to demonstrate mathematical proficiency using the specialized math language gained by consistent curriculum and instruction (SBAC, 2012). As the CCSS push the use of complex, math-specific language into intermediate grade math instruction, students will be consistently asked to use the language of math to verbally justify why a conclusion is true, or explain a mathematical rule and how they applied it to their solutions (ODE, 2011).

While the CCSS provide “clear signposts” toward a goal or college and career readiness for every student, they do not define methods or materials to support students who are not yet meeting grade-level expectations, or specific formative assessments on which to base measures of student progress relative to these signposts (SBAC, 2012; ODE, 2011). And internal demands like working memory function (Fuchs et al., 2006) or socioeconomic concerns (Schoenfeld, 2002) continue to impact students’ math achievement.

A consortium creating proposed assessments of these new standards is planning to provide formative assessment tools, but current proposals are limited to summative assessment criteria (SBAC, 2012). Because of the approaching CCSS implementation without a clear indication of an accompanying nonsummative assessment, I set out to examine whether a concurrent verbalization measure could be used to score students as they “think aloud” about a math problem. This type of measure is reliable and often resembles classroom instruction (Fox et al., 2011; Reed, 2011). I intended to compare scores on this measure with scores on reliable reading and math measures, with the goal of identifying relations between the scores. I also intended to examine the effects of student-level variables on the scores of the verbalization measure. In this chapter, I explain the findings of the study in relation to the research questions and the context of current educational research. I also describe the limitations of the study and considerations for future research.

Findings Regarding the Research Questions

The research questions centered on examining data collected for internal and external relations between the components of a verbalization measure of students' ability to solve math word problems, and students' reading and math skills as measured by benchmark assessments. I also sought to examine the effect of student-level factors on performance of the verbalization task.

Relations Within Concurrent Verbalization Measure

The first research question investigated the relation between the internal CV components. These five components were based on previous research and best practices, enumerated in a 2011 meta-analysis of verbalization measures (Fox, 2011). These protocols were also modeled on the DIBELS Next reading measure (DMG, 2011), the reading measure that sparked this study. The protocols were created and applied for the purpose of this study, so an examination of the data resulting from the measure is appropriate.

I found low variability in the means, standard deviations and ranges of CV scores at each grade. For example, the means of the CV fluency component, which recorded the number of words a participant used to verbalize about a problem, were clustered in a range of 2.15 words, with standard deviations over 12. The scores may confirm that the sample problems selected provided an accessible base for students' verbalizations and were aligned with the curriculum at each grade. They also suggest that participants in each grade demonstrated similar ability to respond to the verbalization task, and the means of the CV fluency scores indicate that most student participants were able to talk on-topic about the sample math problems. In addition, the moderate relation between CV fluency and quality scores in fifth grade, not evident in third or fourth grade, suggests that older students may be more able to more meaningfully explain how they would solve a math problem.

All participants successfully read the selected math problems aloud. Examples of math vocabulary included in the problems were "estimate" (third grade), "how much" (fourth grade) and "find the difference" (fifth grade). Problems containing words requiring a cultural context, like "museum" and "Frisbee," were included due to the limited number of problems available after the screening process was completed. While I did not discover direct evidence of an impact of linguistic complexity of the word problems on this study, concerns about this impact continue to be addressed by researchers and assessment developers (Abedi et al., 2003; SBAC, 2012).

One potential explanation for the success in participants' reading is the reading intervention program implemented at the subject school in 2008, the school year that the third-grade participants entered first grade. The increase in DIBELS Next reading fluency rates from third to fifth grade may also reflect the success of this implementation and instruction.

The increase in reading fluency from third to fifth grade suggests an expectation of a similar increase in verbalization fluency scores across grades, evidencing the connection between reading and math skills (Powell et al., 2009) and between reading and problem-solving skills (Fuchs et al., 2008). However, while the math verbalization fluency scores are stable from third through fifth grade, they do not demonstrate the growth that is seen in reading fluency. This may indicate that reading decoding speed and thinking aloud about a math problem are separate skills that do not improve at the same rate. It may include the impact of a subset of students who are on-target in reading, but underperform in math (Hanich et al., 2001; Jordan & Hanich, 2000; Landerl et al., 2004). It may reflect the percentage of students eligible for Free/Reduced Lunch, which would align with previous findings of the limited growth over time in word problem comprehension with economically disadvantaged students (Jordan et al., 2006).

The lack of gains in math verbalization may also be attributed to curriculum. Research has indicated that a consistent curriculum adhered to over time, such as the reading curriculum at the school, results in student growth (Datnow et al., 2000). The exposure of these student participants to multiple math curricula over their elementary years did not provide that consistency. Further, the curricula used did not consistently contain the requirements now enumerated in the CCSS for communicating mathematical understanding (SBAC, 2012). Students likely have not been consistently asked to verbalize about math during teaching and learning, as previous standards did not require that skill. This lack of practice in verbalizing about math, as reflected by the equivalent scores across grades, aligns with research on the similarity between classroom instruction and verbalization (Reed, 2011). Since the quantity and quality of students' semantic knowledge impact their understanding of words (Wolf, 2007), as the language-rich curriculum and practices based on the CCSS are implemented, it is likely that students will demonstrate improved verbalization skills in math.

Of the third-, fourth- and fifth-grade participants, 79% provided a strategy, but only 16% provided a correct solution. This is reflected in a gap in the means of these two scores at each grade level. One explanation for the gap may lie in specific parts of the verbalization measure's script that administrators used to model and practice the verbalization with the participants. This script includes statements such as "We are not as interested if your answers are right, we just want to know how you are thinking about the problem." The script was modified to include these statements from reliable verbalization measures (Johnstone et al., 2006) after the pilot study administrators found that student verbalizations were frequently limited to providing an answer without disclosing their thinking.

The scoring protocols relating to the two variables of strategy and solution are also different. Protocols for scoring the solution provided a specific answer on which to base the score. Administrators could view the answer and credit the participants for providing the answer at any time during the verbalization. Scoring protocols for the strategy provided two open-ended examples of a likely strategy or concept.

Administrators credited participants with providing a strategy based on the entire content of the 2-minute verbalization. Administrators were trained to credit a participant when a strategy was provided that aligned with the problem, without consideration of whether the strategy resulted in a correct solution. During interviews following the sample, administrators affirmed that participants were credited for providing strategies potentially aligned with solutions even though the strategies were not clearly included in the scoring protocol examples.

Finally, the time limits of the CV measure may have contributed to the gap in scores. Each participant was provided 2 minutes to verbalize about three separate math problems. One administrator noted that the requirement that participants read the problem aloud "ate up the clock," leaving a reduced time for the participant to think aloud. The limited duration of verbalization may have been compounded by the modeling and practice protocols, which encouraged the participants to read the problem aloud, verbalize about a strategy to solve, and then finally to provide an answer. After the sample was conducted, an examination of the scoring protocols found that time had expired before most participants had completed verbalizing on all three sample problems. Fifty percent of the third-grade participants, 70% of the fourth-grade participants, and 80% of the fifth-grade participants did not complete all three problems. These participants' scores were based on the problem or problems they completed. While studies found that adequate time limits in verbalization measures did not affect the outcomes (Fox et al., 2011), the data and administrator observations infer that this measure's 2-minute limit may not have been adequate.

The inadequate time to verbalize may have adversely affected students with lower processing speeds, including students with increased cognitive loads due to second language demands as well as students with a learning disability. The working memory's processing speed has been demonstrated to affect performance on math tasks (Berg, 2008; Swanson & Beebe-Frankenberger, 2004) and was found to be predictive of third-, fourth-, and fifth-grade students' performance on algebraic word problems (Lee et al., 2004). Due to the limited number of participants, an examination of the impact of the time expiring on the scores of students identified for ELD or special education would not be appropriate. However, since administrators estimated a significant number of students ran out of time, this impact apparently was not limited to these subgroups. Modifications to the protocols, including the administrator reading the problem aloud, may address concerns both about the time limit and students' ability to read the problems.

At fifth grade, only one of the participants provided a correct solution, resulting in a mean of .03 for that component. The time limit may have had an increased effect at this grade because of the specific math problems presented. Research has noted that, as students age, growth in executive processing capacity of working memory increasingly relates to math problem-solving skills (Swanson, 2006). The problems selected from a fifth-grade assessment may have been relatively more difficult to solve and their processes and solutions more difficult to describe.

Relations With Concurrent Verbalization and Benchmark Measures

The second and third research questions investigated the relations between the concurrent verbalization components and student scores on benchmark measures of DIBELS Next for reading and easyCBM for math. An examination for predictive validity was not considered in this study.

On the reading benchmark measure, participants' scores on the means reported either exceeded or were slightly below but well within the standard deviations of the DIBELS Next targets. These scores indicate strong reading fluency skills demonstrated by the participants. However, participants' scores on reading fluency were not significantly related to their scores on the math verbalization fluency. This may indicate that reading decoding and thinking aloud about a problem are not related skills. It may reflect participants' familiarity with the DIBELS Next assessment, and their unfamiliarity with the verbalization tasks; however, research would counter that concern (Fox et al., 2011). This may suggest inconsistencies in scoring. While the protocols for both fluency measures direct administrators to record "on-topic" verbalizations, administrators were all experienced in scoring only the reading retell. As discussed above, this may also reflect differences in the curricula and teaching practices commonly used for reading and math.

The DIBELS Next retell fluency and CV fluency scores record the number of words in a participant's retell or concurrent verbalization. The DIBELS Next retell quality and CV quality both assign a quality score to the content of the verbalization based on a number of on-topic details provided. Low/moderate relations in all grades are observed between scores on these two most similar reading and math verbalization tasks.

An example of relations observed is the low/moderate relations with a student's verbalization fluency and the quality of a verbalization in reading and math in third grade. The relations may suggest that similar skills were measured. Specifically, both measure the number of on-topic words and the number of on-topic details provided by participants.

In contrast, the DIBELS Next composite score reflects a student's overall reading skills, and includes comprehension and retell verbalization components. The CV quality score is not a composite score, defined as a measure based on more than one data item (Babbie, 2007), as it contained more than one data item only for the score of 4. A meaningful comparison of these scores may require amending the concurrent verbalization protocols to include a composite score.

Turning to the math benchmark measure, third- and fourth-grade scores on easyCBM grade-level tests exceed or nearly hit the targets set by the measure. In fifth grade, however, participant scores are barely within the standard deviation for the composite score, and the percentile rank of the mean is 15 points below the target.

Moderate relations are also seen in the relations between the math benchmark scores and the CV components. This may reflect the reliability of concurrent verbalization measures generally (Fox et al., 2011). In fifth grade, relations are identified in all areas but one. Because easyCBM has been established as a reliable measure of math performance (Anderson et al., 2009, 2010a, 2010b), the percentile rank of the easyCBM scores and related CV components may also indicate a cause for concern about the skills of participants in fifth grade. This concern was echoed in the participants' 2012 OAKS results, as 57% of fifth graders met the performance standard, 6% to 8% fewer than in third or fourth grade.

Regarding the relations with math verbalization fluency and the easyCBM scores, moderate relations are seen in fourth and fifth grade, but not in third grade. This may demonstrate that students who are capable of understanding and solving math problems are able to state strategies and solutions, but their statements may be succinct. These mixed relations may be connected to the different skills being measured. The CV fluency records the fluency of students' verbalizing about their thinking. The easyCBM scores reflect a student's ability to read and solve grade-level math problems, problems that were intentionally screened to minimize language loads and are presented without time limitations (Anderson et al., 2009; easyCBM, 2009). Since research has described the impact of the complexity of language on assessment of students with a range of disabilities (Anderson et al., 2011; Tindal et al., 2003) and in second language learners (Lean et al., 1990; Solano-Flores & Trumbull, 2003), further research is needed to discern whether math verbalization fluency alone is a reliable indicator of students' math performance on curriculum or large-scale measures.

Relations—Concurrent Verbalization and Student-Level Variables

The fourth research question investigated the effects of student-level factors on math verbalization scores.

After an examination of the descriptive statistics, data about the variables of Grade, Gender and Free/Reduced Lunch were examined, based on the number of participants at each grade level. Little evidence of the effect of a participant's gender was observed.

For participants eligible for Free or Reduced Lunch assistance, generally negative but not significant effects were noted, except the moderate relation identified between CV Fluency and Free/Reduced Lunch eligibility in third grade. However, by fourth grade the effect is reduced and a relation is not identified, and in fifth grade, the effect is near zero. This reduction may reflect the impact over time of the close monitoring and adjustment in math instructional practices implemented at the school, targeting students with F/RL eligibility. A larger participant population would be required in order to examine this effect appropriately.

The effect of grade level on students' CV scores was identified by comparing the means of each CV component at each grade. The stability of the CV fluency component across grades has been described. Three other components show a decline in the mean over grades. As students age from grade to grade, they may appear less able to provide a strategy, solution or adequate details about a math problem. Potential explanations include the increase in difficulty of the individual problems (Fox et al., 2011) and the time limits imposed by the CV measure's protocols (Berg, 2008; Swanson & Beebe-Frankenberger, 2004). Finally, the concern remains that, while students are presumably gaining mathematical skills as they progress through the grades, this growth is not evident with an increase in the scores across grade levels. A potential explanation for the lack of growth remains the lack of a consistent curriculum and related teaching practices for the 4 years prior to this study (Datnow et al., 2000).

Limitations and Future Research

The results of this study must be interpreted considering its limitations. Limitations to its reliability include the number of participants, and the setting of the study. When the data is disaggregated into grade levels, the sample size is small. Due to the number of participants, the grade-level data did not allow a complete analysis of the impact of student-level factors on the data, especially related to students eligible for special education and English language learners. Further studies related these subgroups would help define the measure's reliability. The established reading intervention program may reduce the generalizability to schools where this support is not in place.

The math problems on which the verbalizations were based were selected from an assessment bank provided from a 2009 large-scale test, aligned with the NCTM standards but not the CCSS (TEA, 2008). Future studies of this measure should be centered on the CCSS assessment and focus on the high-priority assessment clusters identified by the assessment developers (SBAC, 2012).

This study did not attempt to identify the curriculum or instructional practices used in the building relating to math word problems. Further studies may wish to include a discussion of these instructional practices, especially after curriculum and instructional practices based in the CCSS are in place. As assessments of the CCSS become available, future studies should also consider the relation of these performance-based assessments with these findings. The CCSS testing items (SBAC, 2012) could be used in future research as the basis for concurrent verbalization studies. Researchers should consider the curriculum and instructional strategies implemented to convey this content during the transition to teaching under the CCSS in order to align the measure with classroom practices (Reed, 2011). Finally, as CCSS assessment results are published, research may wish examine these results for relations with this and future CV assessment data.

Specific changes in the concurrent verbalization protocols used in this measure have been described above, but include an alignment within the quality score, the addition of a composite score which considers all components of the measure, and an extension of the amount of time provided to each participant for verbalization. This extension may require changes in the scoring sheets, specifically to the number of words recorded for each verbalization. Researchers may also consider modifying the check for participants' ability to read the problem as a way to address concerns about the time limit.

Conclusion

The CCSS and related assessments will soon be implemented, with their increased language emphasis in mathematics. During this transition, as educators continue to strive to make sound instructional decisions and provide appropriate support in math, research should continue to investigate the assessments on which these decisions are based.

The current findings support a relation between verbalization relating to a math problem and benchmark assessments in reading and math. As assessments relating to language, math and the CCSS continue to be developed, it may be helpful to consider verbalization as a component of these assessments.

APPENDIX A

DIBELS NEXT ORAL READING FLUENCY SCRIPT

Say these specific directions to the student:

▶ ***I would like you to read a story to me. Please do your best reading. If you do not know a word, I will read the word for you. Keep reading until I say "stop." Be ready to tell me all about the story when you finish.*** (Place the passage in front of the student.)

▶ Begin testing. ***Put your finger under the first word*** (point to the first word of the passage). ***Ready, begin.***

Timing	1 minute. Start your stopwatch after the student says the first word of the passage. Place a bracket ( ]) and say ***Stop*** after 1 minute.
Wait	If no response in 3 seconds, say the word and mark it as incorrect.
Discontinue	If no words are read correctly in the first line, say ***Stop***, record a score of 0, and do not administer Retell. If fewer than 10 words are read correctly on passage #1, do not administer Retell or passages #2 and #3. If fewer than 40 words are read correctly on any passage, use professional judgment whether to administer Retell for that passage.
Reminders	If the student stops (not a hesitation on a specific item), say ***Keep going***. (Repeat as often as needed.) If the student loses his/her place, point. (Repeat as often as needed.)

APPENDIX B

DIBELS NEXT ORAL READING FLUENCY CODING SHEET

▶ ***Now tell me as much as you can about the story you just read. Ready, begin.***

Timing	1-minute maximum. Start your stopwatch after telling the student to begin. Say ***Stop*** after 1 minute.
Wait/ Reminder	If the student stops or hesitates for 3 seconds, select *one* of the following (allowed 1 time): —If the student has not said anything at all, provides a very limited response, or provides an off-track response, say ***Tell me as much as you can about the story.*** —Otherwise, say ***Can you tell me anything more about the story?***
Discontinue	After the first reminder, if the student does not say anything or gets off track for 5 seconds, say ***Thank you*** and discontinue the task.

0 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23 24 25
26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48
49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71
72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94

Retell Total: ________

Quality of Response: (Note: If the student provides only a main idea, it is considered one detail.)

1 Provides 2 or fewer details

2 Provides 3 or more details

3 Provides 3 or more details in a meaningful sequence

4 Provides 3 or more details in a meaningful sequence that captures a main idea

APPENDIX C

SAMPLE CONCURRENT VERBALIZATION SCRIPT AND SCORING SHEETS

Script - Introduce the task

▶***"We are curious about how students solve problems on tests, so we want to ask you and other students to talk about some math problems. We want to listen to how you think out loud about the problems.***

We are not as interested if your answers are right, we just want to know how you are thinking about the problems."

Students should not feel the slightest sense of being judged or of having to obtain a particular result.

Model and Practice

Show the student the teacher model question to familiarize him or her with thinking aloud while working through a task. First you solve a problem and then ask the student to solve theirs.

▶"***First, I am going to read this problem out loud.***

Jake counted 8 oranges, 7 pears, and 4 apples in a fruit bowl. What was the total number of oranges and apples in the fruit bowl?

Then I am going to think out loud while I work on it. That means I'm going to say out loud everything that goes through my mind.

I see that Jake counted oranges, pears, and apples in the bowl. I think this is an addition problem. But I see that I should not count the pears. So he counted 8 oranges plus 4 apples. 8 plus 4 is 12. So there were 12 oranges and apples in the bowl."

▶***"Now I'm going to ask you to practice working on your problems the same way. Read the problem out loud, then think out loud about how you would solve it. Go ahead."***

(Student practices on Sample Problem B. If no words are read correctly, say ***Thank you*** and discontinue the measure.

Joey has 8 books. Roberto has twice as many books as Joey has. How many books does Roberto have?

Scoring practice:
Strategies ex. mental picture, grouping, repeated adding, multiplication, subtraction

Concept: multiplication with carrying, subtraction
Answer: 16 x 2 – 8 = 24

THIRD GRADE - Administer the measure
▶"Now tell what you are thinking about each of these problems. Just say everything that goes through your mind while you work on the problems. Ready, begin." Place the three test problems in front of the subject.

Timing	Start your stopwatch after saying *Ready, begin.* Say *Stop* after 2 minutes.
Wait/ Reminder	If the student stops or hesitates, up to **20** seconds, select one (allowed 1 time): -If the student has not said anything at all, provides a very limited, or off-track response, say ***Tell me as much as you can about the problems.*** -Otherwise, say ***Can you tell me anything about the problems****?*
Discontinue	After the first reminder, if the student does not say anything or gets off track for **10** seconds, say ***Thank you*** and discontinue the task (30 secs. elapsed).

As the student talks, use the scoring sheet to record their verbalizations and count their words.

		Read correctly	Strategy /concept	Correct answer	Quality Score
1	Mr. Johnson's company built 4 office buildings. Each building had 43 windows. What was the total number of windows of these 4 office buildings? *Repeated adding; multiplication, multiplication w/ regroup* *Answer: 172* 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 43 43 44 45+	o Yes (>1 error) o No (<1 error) o Discontinue	o Yes o No	o Yes o No	1 2 3 4
2	Jose had 17 toy cars. His father bought him 27 more toy cars. Then his friend gave him another 13 toy cars. What is a good *estimate* of the number of cars he has? *Adding, estimating, rounding* *Solutions:15+25+15=55, 20+30+10+=60, 15+30+15=60* 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 43 43 44 45+	o Yes (>1 error) o No (<1 error) o Discontinue	o Yes o No	o Yes o No	1 2 3 4
3	Look at the pattern below: 48, 45, 42, 39, 36, ____, ____ If the pattern continues, what two numbers will come next? *Reverse adding, subtracting, difference* *Solution: 33, 30* 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 43 43 44 45+	o Yes (>1 error) o No (<1 error) o Discontinue	o Yes o No	o Yes o No	1 2 3 4

Quality Score:

1. Two or fewer details. Solving process not apparent or student appeared to guess.
2. Three or more details. Details presented randomly.
3. Three or more details. Details presented in organized manner with strategy for solving.
4. Three or more details. Presented a correct solving process with underlying math concepts, patterns or generalizations.

(Note: if the student provides only a solution, it is considered one detail.)

FOURTH GRADE - Administer the measure

▶***"Now tell what you are thinking about each of these problems. Just say everything that goes through your mind while you work on the problems. Ready, begin." Place the three test problems in front of the subject.***

Timing	Start your stopwatch after saying *Ready, begin.* Say *Stop* after 2 minutes.
Wait/ Reminder	If the student stops or hesitates, up to **20** seconds, select one (allowed 1 time): -If the student has not said anything at all, provides a very limited, or off-track response, say ***Tell me as much as you can about the problems.*** -Otherwise, say ***Can you tell me anything about the problems****?*
Discontinue	After the first reminder, if the student does not say anything or gets off track for **10** seconds, say ***Thank you*** and discontinue the task (30 secs. elapsed).

As the student talks, use the scoring sheet to record their verbalizations and count their words.

		Read correctly	Strategy /concept	Correct answer	Quality Score
1	Mrs. Cahill chose books for 7 students in her reading group. She chose 11 books for each student. How many books in all did Mrs. Cahill choose? *Repeated adding; multiplication* *Answer: 77* 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 43 43 44 45+	o Yes (>1 error) o No (<1 error) o Discontinue	o Yes o No	o Yes o No	1 2 3 4
2	Belinda bought a pack of pencils for $1, a box of paints for $3, and 6 folders. What information is needed to find the total amount of money Belinda spent? *Find the difference, missing information* *Solutions: the cost of the 6 folders, the price of a folder* 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 43 43 44 45+	o Yes (>1 error) o No (<1 error) o Discontinue	o Yes o No	o Yes o No	1 2 3 4
3	On Monday, 36 people visited an art museum. On Tuesday, 15 fewer people visited the museum than on Monday. Each person paid $5 for admission. How much money was paid for admission to the museum on Tuesday? *Multiplication, subtraction* *Solution: 36-15=21x$5 = $105* 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 43 43 44 45+	o Yes (>1 error) o No (<1 error) o Discontinue	o Yes o No	o Yes o No	1 2 3 4

Quality Score:

1. Two or fewer details. Solving process not apparent or student appeared to guess.
2. Three or more details. Details presented randomly.
3. Three or more details. Details presented in organized manner with strategy for solving.
4. Three or more details. Presented a correct solving process with underlying math concepts, patterns or generalizations.

(Note: if the student provides only a solution, it is considered one detail.)

FIFTH GRADE - Administer the measure

▶***"Now tell what you are thinking about each of these problems. Just say everything that goes through your mind while you work on the problems. Ready, begin." Place the three test problems in front of the subject.***

Timing	Start your stopwatch after saying *Ready, begin.* Say *Stop* after 2 minutes.
Wait/ Reminder	If the student stops or hesitates, up to **20** seconds, select one (allowed 1 time): -If the student has not said anything at all, provides a very limited, or off-track response, say ***Tell me as much as you can about the problems.*** -Otherwise, say ***Can you tell me anything about the problems****?*
Discontinue	After the first reminder, if the student does not say anything or gets off track for **10** seconds, say ***Thank you*** and discontinue the task (30 secs. elapsed).

As the student talks, use the scoring sheet to record their verbalizations and count their words.

		Read correctly	Strategy /concept	Correct answer	Quality Score
1	Sarah and Jen participated in the Frisbee toss on field day. Sarah threw the Frisbee 30.95 meters. Jen threw the Frisbee 39.31 meters. How much farther did Jen throw the Frisbee than Jen? *Subtraction, find the difference, counting up* *Solution: the difference is 8.36* 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 43 43 44 45+	o Yes (>1 error) o No (<1 error) o Discontinue	o Yes o No	o Yes o No	1 2 3 4
2	Olga bought 25 T-shirts for $8 each. She sold them all for $12 each. What is the difference between the amount of money Olga made and the amount of money she spent on these 25 T-shirts? *Multiplication, find the difference* *Solution:25 x $8 =$200; 25 x 12=$300.* *The difference between $200 and $300 is $100.* 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 43 43 44 45+	o Yes (>1 error) o No (<1 error) o Discontinue	o Yes o No	o Yes o No	1 2 3 4
3	Mrs. Kline was putting wallpaper on her kitchen walls. She used 5 rolls of wallpaper and 2 feet of another roll to cover half the kitchen. What information is needed to fin the total number of feet of wallpaper Mrs. Kline deed to cover her whole kitchen? *Subtraction, difference* *Answer: the number of feet in a roll of wallpaper* 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 43 43 44 45+	o Yes (>1 error) o No (<1 error) o Discontinue	o Yes o No	o Yes o No	1 2 3 4

Quality Score:

1. Two or fewer details. Solving process not apparent or student appeared to guess.
2. Three or more details. Details presented randomly.
3. Three or more details. Details presented in organized manner with strategy for solving.
4. Three or more details. Presented a correct solving process with underlying math concepts, patterns or generalizations.

(Note: if the student provides only a solution, it is considered one detail.)

APPENDIX D

CONCURRENT VERBALIZATION TRAINING MATERIALS AND ADMINISTRATION PROTOCOLS

Training materials:

a. The DIBELS Next reading retell instructions and scoring sheet (Figures 1, 2)
b. The concurrent verbalization script and coding sheet (Figure 3)
c. The concurrent verbalization training protocols. (Figure 4)
d. The DIBELS Next Assessment Accuracy Checklist – ORF (Figure 5)
e. The DIBELS Next Assessment Accuracy Checklist – retell (Figure 6)
f. The concurrent verbalization Assessment Accuracy Checklist (Figure 7)
g. The TAKS practice and sample word problems (Figures 8, 9)
h. Video clips modeling the concurrent verbalization administration.
i. Guidelines for general administration (DMG, 2010)

Agenda:

1. Introduction, purpose and general guidelines.
2. Practice scoring and discussion. Administrators individually score three video examples. After each example, the scores will be shared and discussed with the stated goal of scoring consistency and reliability between administrators. During the third video sample, observers will score the administrators using the checklist. Administrators who have not demonstrated consistency in rating student retell will receive additional opportunities to practice after the training period until consistency with the scoring protocols is demonstrated.

APPENDIX E

DIBELS ORAL READING FLUENCY ASSESSMENT ACCURACY CHECKLIST

DORF Assessment Accuracy Checklist

Consistently	Needs practice	Does the assessor:
☐	☐	1. Position materials so that student cannot see what is being recorded?
☐	☐	2. State standardized directions exactly as written? ***I would like you to read a story to me. Please do your best reading. If you do not know a word, I will read the word for you. Keep reading until I say "stop." Be ready to tell me all about the story when you finish.*** (Place the passage in front of the student.) Begin testing. ***Put your finger under the first word*** (point to the first word of the passage). ***Ready, begin.*** Begin testing (2nd and 3rd passages). ***Now read this story to me. Please do your best reading. Ready, Begin.***
☐	☐	3. Start the timer when the student reads the first word of the passage?
☐	☐	4. Score student responses correctly according to the scoring rules?
☐	☐	5. Use reminder procedures correctly and appropriately?
☐	☐	6. Say the word and put a slash over it if the student fails to say it correctly within 3 seconds?
☐	☐	7. Write "sc" above a previously slashed word if the student self-corrects within 3 seconds?
☐	☐	8. Discontinue if the student does not read any words correctly in the first row of the passage?
☐	☐	9. Place a bracket (]) after the last word the student read before the minute ran out and tell the student to stop?
☐	☐	10. Correctly calculate the total number of words read (correct and errors) and record it on the scoring sheet?
☐	☐	11. Correctly add the number of errors and record it on the scoring sheet?
☐	☐	12. Correctly subtract the errors from the total words and record the words correct on the scoring sheet?
☐	☐	13. Record both scores on the front of the scoring booklet?

APPENDIX F

DIBELS RETELL FLUENCY ASSESSMENT ACCURACY CHECKLIST

DORF Assessment Accuracy Checklist: Retell

Consistently	Needs practice	Does the assessor:
☐	☐	14. Administer Retell if the student read 40 or more words correct?
☐	☐	15. Remove the passage and then state the standardized Retell directions exactly as written? ***Now tell me as much as you can about the story you just read. Ready, begin.***
☐	☐	16. Start the stopwatch after saying ***Begin***?
☐	☐	17. Use reminder procedures correctly and appropriately?
☐	☐	18. Mark the number or words in the student's response and circle the total number of words?
☐	☐	19. Tell the student to stop if he/she is still retelling at the end of one minute?
☐	☐	20. Record the number of correct words at the bottom of the scoring booklet?
☐	☐	21. Record the score on the front cover of the scoring booklet?

APPENDIX G

CONCURRENT VERBALIZATION ASSESSMENT ACCURACY CHECKLIST

Consistently	Needs Practice	Does the assessor:
☐	☐	Position materials so that student cannot see what is being recorded?
☐	☐	State standardized directions exactly as written?
☐	☐	Starts timer after instructing student to "*Begin*".
☐	☐	Score student verbalizations according to the scoring rules?
☐	☐	Uses reminder procedures correctly and appropriately?
☐	☐	Discontinue if the student does not verbalize after 30 seconds?
☐	☐	Tell the student to stop if he/she is still verbalizing at the end of one minute?
☐	☐	Correctly calculate the total number of words verbalized and record them on the scoring sheet?
☐	☐	Correctly record the quality of verbalization score on the scoring sheet?
☐	☐	Shadow score verbalization fluency with the examiner. Is he/she within 2 points on the final fluency score and within 1 point of the final quality score?

APPENDIX H

TAKS SAMPLE QUESTION FOR CONCURRENT VERBALIZATION MODELING AND PRACTICE

Side One – Teacher Model

Jake counted 8 oranges, 7 pears, and 4 apples in a fruit bowl. What was the total number of oranges and apples in the fruit bowl?

Side Two – Student Model

Remember: Read the problem then think about how you would solve it.

Joey has 8 books. Robert has twice as many books as Joey has.

How many books does Roberto have?

APPENDIX I

TAKS PROBLEMS FOR CONCURRENT VERBALIZATION

Grade 3 problems:

Mr. Johnson's company built 4 office buildings.

Each building had 43 windows.

What was the total number of windows of these 4 office buildings?

Jose had 17 toy cars.

His father bought him 27 more toy cars.

Then his friend gave him another 13 toy cars.

What is a good *estimate* of the number of cars he has

Look at the pattern below:

48, 45, 42, 39, 36, ____, ____

If the pattern continues, what two numbers will come next?

Grade 4 problems:

Mrs. Cahill chose books for 7 students in her reading group.

She chose 11 books for each student.

How many books in all did Mrs. Cahill choose?

Belinda bought a pack of pencils for $1, a box of paints for $3, and 6 folders. What information is needed to find the total amount of money Belinda spent?

On Monday, 36 people visited an art museum.

On Tuesday, 15 fewer people visited the museum than on Monday.

Each person paid $5 for admission.

How much money was paid for admission to the museum on Tuesday?

Grade 5 problems:

Sarah and Jen participated in the Frisbee toss on field day.

Sarah threw the Frisbee 30.95 meters.

Jen threw the Frisbee 39.31 meters.

How much farther did Jen throw the Frisbee than Jen?

Olga bought 25 T-shirts for $8 each. She sold them all for $12 each.

What is the difference between the amount of money Olga made

and the amount of money she spent on these 25 T-shirts?

Mrs. Kline was putting wallpaper on her kitchen walls.

She used 5 rolls of wallpaper and 2 feet of another roll to cover half the kitchen.

What information is needed to find the total number of feet of wallpaper Mrs. Kline needs

to cover her whole kitchen?

REFERENCES CITED

Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher*, *33*(1), 4–14.

Abedi, J., Courtney, M., & Leon, S. (2003). *Research-supported accommodation for English language learners in NAEP* (CSE Technical Report No. 586). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Anderson, D., Alonzo, J., & Tindal, G. (2009). *Diagnostic efficiency of easyCBM math: Oregon* (Technical Report No. 0915). Retrieved from http://www.brtprojects.org/publications/technical-reports

Anderson, D., Alonzo, J., & Tindal, G. (2010a). *easyCBM mathematics criterion related validity evidence: Oregon state test* (Technical Report No. 1011). Retrieved from http://www.brtprojects.org/publications/technical-reports

Anderson, D., Alonzo, J., & Tindal, G. (2010b). *Diagnostic efficiency of easyCBM math: Oregon* (Technical Report No. 1009). Retrieved from http://www.brtprojects.org/publications/technical-reports

Anderson, D., Lai, D.-F., Alonzo, J., & Tindal, G. (2011). Examining a grade-level math CBM designed for persistently low-performing students. *Educational Assessment*, *16*(1), 15–34.

Babbie, E. (2007). *The practice of social research.* Belmont, CA: Thomson Higher Education.

Baddeley, A. (2010). Working memory. *Current Biology*, *20*(4), R136–R140.

Baumann, J. F., Jones, L. A., & Seifert-Kessell, N. (1993). Using think alouds to enhance children's comprehension monitoring abilities. *The Reading Teacher*, *47*(3), 184–193.

Berg, D. H. (2008). Working memory and arithmetic calculation in children: The contributory roles of processing speed, short-term memory and reading. *Journal of Experimental Child Psychology*, *99*(4), 288–308.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 30–32.

Canby School District. (2009). *Canby mathematics K-12*. Retrieved from http://wiki.canby.k12.or.us/groups/canbymathematicsk12/

Carr-George, C., Vannest, K. J., Willson, V., & Davis, J. L. (2009). The participation and performance of students with emotional and behavioral disorders in a state accountability assessment in reading. *Behavioral Disorders, 35*(1), 66–78.

Castillo, J. M., Torgesen, J. K., Powell-Smith, K. E., & Al-Otaiba, S. (2009). Examining the decision reliability and validity of three reading fluency measures for predicting outcomes on statewide reading accountability tests. In R. K. Wagner, C. Schatschneider, & C. Phythian-Sence (Eds.), *Beyond decoding: The behavioral and biological foundations of reading comprehension* (pp. 195–226). New York, NY: Guilford Press.

Carlson, E., Jenkins, F., Bitterman, A., Keller, B., & Lauer, K. (2011). *A longitudinal view of receptive vocabulary and math achievement of young children with disabilities* (NCSER 2011-3006). Washington, DC: U.S. Department of Education. Retrieved from ies.ed.gov/ncser/pubs/20113006/pdf/20113006.pdf

Chard, D. J., Clarke, B., Baker, S., Otterstedt, J., Braun, D., & Katz, R. (2005). Using measures of number sense to screen for difficulties in mathematics: Preliminary findings. *Assessment for Effective Intervention, 30*(2), 3–14.

Chiappe, P. (2005). How reading research can inform mathematics difficulties: The search for the core deficit. *Journal of Learning Disabilities, 38*(4), 313–318.

Cirino, P. T., Fletcher, J. M., Ewing-Cobbs, L., Barnes, M. A., & Fuchs, L. S. (2007). Cognitive arithmetic differences in learning difficulty groups and the role of behavioral inattention. *Learning Disabilities Research and Practice, 22*(1), 25–35.

Clarke, B., & Shinn, M. R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review, 33*(2), 234–248.

Common Core State Standards Initiative. (2011). *Standards for mathematical practice*. Retrieved from http://www.corestandards.org/the-standards/mathematics/introduction/standards-for-mathematical-practice/

Cross, C. T., Woods, T. A., & Schweingruber, H. (Eds.). (2009). *Mathematics learning in early childhood: Paths toward excellence and equity*. Washington, DC: National Academies Press. Retrieved from http://www.nap.edu/catalog.php?record_id=12519#toc

Datnow, A., Borman, G., & Stringfield, S. (2000). School reform through a highly specified curriculum: Implementation and effects of the Core Knowledge Sequence. *Elementary School Journal, 101*(2), 167–191. Retrieved from http://www.jstor.org/stable/1002341

Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish standards for students with learning disabilities. *School Psychology Review, 30*(4), 507–524.

Dynamic Measurement Group. (2011). DIBELS Next technical manual. Retrieved from https://www.dibels.org/next/index/php

Earl, L., & Katz, S. (2000). Changing classroom assessment: Teachers struggle. In N. Bascia & A. Hargreaves (Eds.), *The sharp edge of educational change: Teaching, leading and the realities of reform* (pp. 97–111). New York, NY: RoutledgeFalmer.

easyCBM. (2009). *easyCBM teachers' manual*. Retrieved from http://easycbm.com/teachers/auth/index.php

easyCBM. (n.d.). *Detailed percentile lookup table for easyCBM*. Eugene, OR: Behavioral Research and Teaching,

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis*. Cambridge, MA: MIT Press.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis* (rev. ed.). Retrieved from http://cognet.mit.edu.libproxy.uoregon.edu/library/books/mitpress/0262550237/cache/chpt2.pdf

Ferrara, S. (2008, January). *Thinking about the think-aloud method to guide development of assessments of modified achievement standards*. Presentation at the meeting of OSEEP GSEG project directors, Washington, DC. Retrieved from www.cehd.umn. edu/NCEO/Presentations/.../7Ferrara2percentKickoff.ppt

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin, 137*(2), 316–344.

Fong, A. B., Bae, S., & Huang, M. (2010). *Patterns of student mobility among English language learner students in Arizona public schools*. Washington, DC: Institute of Education Services, National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=REL2010093

Fuchs, L. S., & Fuchs, D. (2002). Mathematical problem-solving profiles of students with mathematics disabilities with and without comorbid reading disabilities. *Journal of Learning Disabilities, 35*, 563–573.

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., . . . Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98*(1), 29–43.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Lambert, W., Stuebing, K., & Fletcher, J. M. (2008). Problem solving and computational skill: Are they shared or distinct aspects of mathematical cognition? *Journal of Educational Psychology, 100*(1), 30–47.

Gregoire, J., & Desoete, A. (2009). Mathematical disabilities—An underestimated topic? *Journal of Psychoeducational Assessment, 27*(3), 171–174.

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities, 38*(4), 293–304.

Good, R. H., & Kaminski, R. A. (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Retrieved from http://dibels.uoregon.edu/

Good, R. H., & Kaminski, R. A. (2009). *What are DIBELS?* Retrieved from http://www.dibels.org/dibels.html

Griffin, C. G., & Jitendra, A. K. (2009). Word problem-solving instruction in inclusive third-grade mathematics classrooms. *Journal of Educational Research, 102*(3), 187–202.

Hanich, L. B., Jordan, N. C., Kaplan, D., & Dick, J. (2001). Performance across different areas of mathematical cognition in children with learning disabilities. *Journal of Educational Psychology*, *93*(3), 615–626.

Hecht, S. A., Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2000). The relations between phonological processing abilities and emerging individual differences in mathematical computation skills: A longitudinal study from second to fifth grades. *Journal of Experimental Child Psychology*, *79(*2), 192–227.

Helwig, R., Anderson, L., & Tindal, G. (2002). Using a concept-grounded, curriculum-based measure in mathematics to predict statewide test scores for middle school students with LD. *Journal of Special Education*, *36*(2), 102–112.

Helwig, R., Rozek-Tedesco, M. A., & Tindal, B. H. (1999). Reading as an access to mathematics problem solving on multiple-choice tests for sixth-grade students. *Journal of Educational Research*. *93*(2), 113–126.

Imbo, I., & Vandierendonck, A. (2006). The development of strategy use in elementary school children: Working memory and individual differences. *Journal of Experimental Child Psychology, 96,* 284–309.

Ingersoll, G. M., Scamman, J. P., & Eckerling, W. D. (1989). Geographic mobility and student achievement in an urban setting. *Educational evaluation and policy analysis*, *11*(2), 143–149.

Individuals With Disabilities Education Improvement Act of 2004, Pub. L. No. 108-446. (2004). Retrieved from http://idea.ed.gov/explore/view/p/%2Croot%2C statute%2C

Johnstone, C. J., Bottsford-Miller, N. A., & Thompson, S. J. (2006). *Using the think aloud method (cognitive labs) to evaluate test design for students with disabilities and English language learners* (Technical Report No. 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from http://www.cehd.umn.edu/nceo/OnlinePubs/Tech44/#background

Johnstone, C. J., 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 No. 48). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from http://eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=ED499410&ERICExtSearch_SearchType_0=no&accno=ED499410

Jordan, N. C., & Hanich, L. B. (2000). Mathematical thinking in second-grade children with different forms of LD. *Journal of Learning Disabilities, 33*(6), 567–578.

Jordan, N. C., & Hanich, L. B. (2003). Characteristics of children with moderate mathematics deficiencies: A longitudinal perspective. *Learning Disabilities Research and Practice, 18*(4), 213–221.

Jordan, N. C., Kaplan, D., Olah, L. N., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development, 77*(1), 153–175.

Kaminski, R., & Cummings, K. D. (2007). *DIBELS: Myth and facts.* Retrieved from http://www.dibels.org/papers/Myths_0208.pdf

Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review, 92*(1), 109–129.

Kercood, S., Zentall, S. S., & Lee, D. L. (2004). Focusing attention to deep structure in math problems: Effects on elementary education students with and without attentional deficits. *Learning and Individual Differences, 14*(2), 91–105.

Klein, D. (2003). A brief history of American K-12 mathematics education in the 20th century. Retrieved from http://www.csun.edu/~vcmth00m/AHistory.html

Kloosterman, P., & Cougan, M. C. (1994). Students' beliefs about learning school mathematics. *Elementary School Journal, 94*(4), 375–388.

Kyttala, M. (2008). Visuospatial working memory in adolescents with poor performance in mathematics: Variation depending on reading skills. *Educational Psychology, 28*(3), 273–289.

Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: A study of 8-9-year-old students. *Cognition, (93)*2, 99–125.

Lean, G. A., Clements, M. A., & Del Campo, G. (1990). Linguistic and pedagogical factors affecting children's understanding of arithmetic word problems: A comparative study. *Educational Studies in Mathematics, 21*,165–191.

Lee, K., Ng, S.-F., Ng, E.-L., & Lim, Z.-Y. (2004). Working memory and literacy as predictors of performance on algebraic word problems. *Journal of Experimental Child Psychology, 89*(2), 140–158.

Lord, C., Abedi, J., & Poosuthasee, N. (2000). *Language difficulty and assessment accommodations for ELLs*. Dover, DE: Delaware Department of Education.

Marcotte, A. M., & Hintze, J. M. (2009). Incremental and predictive utility of formative assessment methods of reading comprehension. *Journal of School Psychology, 47*(5), 315–335.

Mann, H. M., & Chase, P. E. (1855). The grammar school arithmetic: Containing much valuable commercial information. Philadelphia, PA: E. H. Butler. Retrieved from http://books.google.com/books?id=TorvgymEfAAC&pg=PR5&lpg=PR5&dq=horace+mann+the+grammar+school+arithmetic&source=bl&ots=VamfV__87p&sig=Z1lkJ4EZIPfN1NsJPuWPK7AggOI&hl=en&ei=yQP9TeegBMrjiAKG0ZjrBA&sa=X&oi=book_result&ct=result&resnum=1&ved=0CB0Q6AEwAA#v=onepage&q&f=false

Marzano, R. J., & Haystead, M. W. (2008). *Making standards useful in the classroom.* Association for Supervision and Curriculum Development. Retrieved from http://books.google.com/books?id=QKR8dvbupH4C&pg=PA7&lpg=PA7&dq=marzano+haystead+2008+standards+useful&source=bl&ots=bKJUfK6nky&sig=cfvHxPtVdz--10Wm5VYyZKBNqJg&hl=en&ei=rgT9TdSJGaPViAKA5JSqBA&sa=X&oi=book_result&ct=result&resnum=1&ved=0CBYQ6AEwAA#v=onepage&q&f=false

McKenna, M. K., & Good, R. (2003). *Assessing reading comprehension: The relation between DIBELS oral reading fluency, DIBELS retell fluency and Oregon state assessment scores.* Eugene, OR: University of Oregon. Retrieved from https://dibels.org/papers/DIB_refs_081308.pdf

Medina, A. D. (2008). *Concurrent verbalization, task complexity, and working memory: Effects on L2 learning in a computerized task* (Doctoral dissertation). Retrieved from http://gradworks.umi.com/33/15/3315458.html

Mercer, C. D., & Miller, S. P. (1992). Teaching students with learning problems to acquire, understand, and apply basic math facts. *Remedial Special Education, 13*(3), 19–35.

Meyer, M. L., Salimpoor, V. N., Wu, S. S., Geary, D. C., & Menon, V. (2010). Differential contribution of specific working memory components to mathematics achievement in 2nd and 3rd graders. *Learning and Individual Differences, 20*(2), 101–109.

Montague, M., & Bos, C. S. (1986). The effect of cognitive strategy training on verbal math problem solving performance of learning disabled adolescents. *Journal of Learning Disabilities,19*(1), 26–33.

National Council of Teachers of Mathematics. (2009). *Guiding principles for mathematics curriculum and assessment.* Retrieved from http://www.nctm.org/standards/content.aspx?id=23273

National Council for Teachers of Mathematics. (2011). Curriculum focal points for prekindergarten through Grade 8 mathematics. Retrieved from http://www.nctm.org/standards/content.aspx?id=270

National Governors Association. (2010). *Reaching higher: The common core state standards validation committee.* Retrieved from http://www.corestandards.org/ about-the-standards

Nelson, J. R., Benner, G. J., Lane, K., & Smith, B. W. (2004). Academic achievement of K-12 students with emotional and behavioral disorders. *Exceptional Children, 71*(1), 59–73.

Nese, J. F. T., Lai, D. F., Anderson, D., Jamgochian, E. M., Kamata, A., Saex, L., & Tindal, G. (2010). *Technical adequacy of the easyCBM mathematics measures: Grades 3–8, 2009-2010 version* (Technical Report No. 1007). Retrieved from http://www.brtprojects.org/publications/technical-reports

Nesher, P. (1980). The stereotyped nature of school word problems. *For the Learning of Mathematics 1*(1), 41–48.

Oregon Department of Education. (2010a). *Mathematics test specifications and blueprints 2010-2011*. Retrieved from http://www.ode.state.or.us/search/ page/?id=496#MA

Oregon Department of Education. (2010b). OAKS online practice test site. Retrieved from https://oakspt.tds.airast.org/student/

Oregon Department of Education. (2011). Common core state standards for mathematics. Retrieved from http://www.ode.state.or.us/search/page/?id=1527

Pajares, F. (1996). Self-efficacy beliefs and mathematical problem-solving of gifted students. *Contemporary Educational Psychology, 21*, 325–344.

Park, B. J., Irvin, P. S., Anderson, D., Alonzo, J., & Tindal, G. (2011). *A cross-validation of easyCBM mathematics cut scores in Oregon: 2009-2010* (Technical Report No. 1104). Retrieved from http://www.brtprojects.org/publications/technical-reports

Passolunghi, M. C., & Siegel, L. S. (2001). Short-term memory, working memory, and inhibitory control in children with difficulties in arithmetic problem solving. *Journal of Experimental Child Psychology, 80*, (44–57).

Powell, S. R., Fuchs, L. S., Fuchs, D., Cirino, P. T., & Fletcher, J. M. (2009). Do word-problem features differentially affect problem difficulty as a function of students' mathematics difficulty with and without reading difficulty? *Journal of Learning Disabilities, 42*(2), 99–110.

Raghubar, K., Cirino, P., Barnes, M., Ewing-Cobbs, L., Fletcher, J., & Fuchs, L. (2009). Errors in multi-digit arithmetic and behavioral inattention in children with math difficulties. *Journal of Learning Disabilities, 42*(4), 356–371.

Reed, D. K. (2011). A review of the psychometric properties of retell instruments. *Educational Assessment*, *(16)* 123–144. doi: 10.1080/10627197.2011.604238

Reuhkala, M. (2001). Mathematical skills in ninth-graders: Relationship with visuo-spatial abilities and working memory. *Educational Psychology, 21*(4), 387–399.

Riley, M. S., Greeno, J. G., & Heller, J. I. (1983). Development of children's problem-solving ability in arithmetic. In H. Ginsberg (Ed.), *The development of mathematical thinking* (pp. 153–196). New York, NY: Academic Press.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93*(2), 346–362.

Roberts, G., Good, R., & Corcoran, S. (2005). Story retell: A fluency-based indicator of reading comprehension. *School Psychology Quarterly*, *20*(3), 304–317.

Romberg, T. A. (1992). *Mathematics assessment and evaluation: Imperatives for mathematics educators*. Albany, NY: State University of New York Press.

Salmon, E., Van der Linden, M., Collette, F., Delfiore, G., Maquet, P., Degueldre, C., . . . & Franck, G. (1996). Regional brain activity during working memory tasks. *Brain, 119*(5), 1617–1625.

Schmidt, W. H., McKnight, C. C., Houang, R. T., Wang, H., Wiley, D. E., Cogan, L. S., & Wolfe, R. G. (2001). *Why schools matter: A cross-national comparison of curriculum and learning*. San Francisco, CA: Jossey-Bass.

Schoenfeld, A. H. (2002). Making math work for all children: Issues of standards, testing, and equity. *Educational Researcher, 31*(1), 13–25.

Shaftel, J., Belton-Kocher, E., Glasnapp, D., & Poggio, J. (2006). The impact of language characteristics in mathematics test items on the performance of English language learners and students with disabilities. *Educational Assessment, 11*(2), 105–126.

Siegel, L. S., & Ryan, E. B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child Development, 60*(4), 973–980.

SMARTER Balanced Assessment Consortium. (2012). Mathematics draft content specifications. Retrieved from http://www.smarterbalanced.org/wordpress/wp-content/uploads/2011/12/MathContentSpecifications.pdf

Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher, 32*(2), 3–13.

Swanson, H. L. (2006). Cross-sectional and incremental changes in working memory and mathematical problem solving. *Journal of Educational Psychology, 98*(2), 265–281.

Swanson, H. L., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for math disabilities. *Journal of Education Psychology, 96*(3), 471–494.

Swanson, H. L., & Jerman, O. (2006). Math disabilities: A selective meta-analysis. *Review of Educational Research, 76*(2), 249–274.

Texas Education Agency. (2008). *Texas national comparative data study*. Rolling Meadows, IL: Riverside.

Texas Education Agency. (2011). TAKS released tests archive, last modified 10/4/2011. Retrieved from http://www.tea.state.tx.us/student.assessment/taks/released-tests/archive/

Tindal, G., & Anderson, D. (2011). Validity evidence for making decisions about accommodated and modified large-scale tests. In S. N. Elliot, R. J. Kettler, P. A. Beddow, & A. Kurz (Eds.), *Handbook of accessible achievement tests for all students* (part 3, pp. 183–200). New York, NY: Springer.

Tindal, G., McDonald, M., Tedesco, M., Glasgow, A., Almond, P., Crawford, L., & Hollenbeck, K. (2003). Alternate assessments in reading and math: Development and validation for students with significant disabilities. *Exceptional Children, 69*(4), 481–494.

Tindal, G., Yovanoff, P., & Geller, J. P. (2008). Generalizability theory applied to reading assessments for students with significant cognitive disabilities. *Journal of Special Education, 44*(1), 3–17.

Vogel, S. (1990). Gender differences in intelligence, language, visual-motor abilities, and academic achievement in students with learning disabilities. *Journal of Learning Disabilities, 12*(1), 44–52.

Wade, S. E. (1990). Using think alouds to assess comprehension. *The Reading Teacher, 43*(7), 442–451.

Wolf, M. (2007). *Proust and the squid*. New York, NY: HarperCollins.

Wynn, K. (1992). Addition and subtraction in human infants. *Nature, 358,* 749–750.

Zentall, S. S. (1990). Fact-retrieval automatization and math problem solving by learning disabled, attention-disordered, and normal adolescents. *Journal of Educational Psychology, 82*(4), 856–865.

Zucker, S., Sassman, C., & Case, B. J. (2004). *Cognitive labs* (Technical Report). Retrieved from http://www.education.pearsonassessments.com/pai/ai/research/resources/tmrsresources?squery=cognitive+labs