Title
An Examination of the Predictive Validity of the Quadrant Analysis Tool

Permalink
https://escholarship.org/uc/item/7733w0gx

Author
Atwal, Kavita Kaur

Publication Date
2014

Peer reviewed|Thesis/dissertation
An Examination of the Predictive Validity of the Quadrant Analysis Tool

A Thesis submitted in partial satisfaction of the requirements for the degree of

Master of Arts

in

Education

by

Kavita Kaur Atwal

June 2014

Thesis Committee:
Dr. Mike Vanderwood, Chairperson
Dr. Cathleen Geraghty
Dr. Cixin Wang
The Thesis of Kavita Atwal is approved:

____________________________________

____________________________________

____________________________________

Committee Chairperson

University of California, Riverside
Response-to-Intervention (RTI) is used to target reading problems in schools. Although its potential is great, various obstacles may hinder its effective implementation. One main roadblock to the implementation of RTI is the appropriate match of interventions to student skills deficits. To mitigate this problem, behavioral consultation can be used. Specifically, during the problem analysis stage, it has been recommended that problem analysis tools can be used to help in identifying interventions that are appropriately matched to student skill deficit. One such problem analysis tool that has been developed is the quadrant analysis tool. This tool uses student’s accuracy and fluency scores on Reading-CBM assessments to place students into quadrants, which in turn provide recommendations for appropriate interventions. Although this tool is used in schools today, no reliability or validity data are available for it. The purpose of this study is to examine the predictive validity of this tool. Recommendations are provided regarding the use of this tool.
# TABLE OF CONTENTS

Introduction ..................................................................................................................... 1

Method .......................................................................................................................... 20

Results ......................................................................................................................... 19

Discussion ................................................................................................................... 26

References .................................................................................................................... 30
LIST OF TABLES

Table 1, Values Used to Determine Minimally Proficient Performance Status……..19
Table 2, Descriptive Statistics for Total Sample………………………………………..20
Table 3, Frequency of Proficient and Non-Proficient Students………………………...20
Table 4, Wilks’ Lambda…………………………………………………………………22
Table 5, Classification Results…………………………………………………………..23
Literacy is a key area of focus in education. Despite this, the National Reading Panel (NRP) reports that more than 17.5 percent of children in the United States will encounter reading problems in their first three years of schooling (U.S. Department of Education, 2000). Considering that poor readers will likely always remain poor readers if they do not receive help before they enter third grade, and that a student has an 88% probability to be a poor reader in fourth grade if he or she was a poor reader in first grade (Juel, 1988), it is important to provide good early reading instruction. As a means to examine literacy instruction in the United States, Congress called for the formation of The National Reading Panel. The panel was developed through the collaboration between the National Institute of Child Health and Human Development (NICHD) and the U.S. Department of Education. The purpose of the panel was to establish a report on the effectiveness of different methods of literacy instruction (National Institute of Child Health and Human Development, 2013). After a comprehensive analysis of the existing literature base, the National Reading Panel determined the skills on which reading instruction should be focused. Later deemed the “Big 5,” these skills are: phonemic awareness, alphabetic principal or phonics, vocabulary, fluency with text, and reading comprehension (National Institute of Child Health and Human Development, 2000). These “Big 5” skills were proposed to not only be the focus of reading instruction, but also the skills evaluated in the assessment of literacy.

Unfortunately, even with steps taken by the government to target reading instruction in the United States, according to the National Assessment of Education Progress (NAEP) report only 35% of fourth graders and 36% of 8th graders are at or
above proficient in the area of reading (U.S. Department of Education, 2013). These students scoring proficient are said to have “demonstrate[d] solid academic performance and competency over challenging subject matter” (U.S. Department of Education, 2013, pp.7). The other 65% and 64%, respectively, of fourth and eighth graders in the nation scored below proficient in the area of reading. This highlights the fact that there is still much growth that needs to be made in the area of reading instruction in our nation today.

To help facilitate growth in the area of reading, assessments and interventions are used in schools. A multi-tier service delivery (MTSD) format allows for such assessments and interventions to be organized in the most resource efficient way. A MTSD “is a structure for planning and organizing the provision of increasingly intense interventions delivered in general, remedial, and special education” (Kovaleski & Black, 2010, pp. 23). Within a MTSD model, Response-to-Intervention (RTI) is used to make decisions on the provision of services to students. Although RTI can be used to target a variety of subjects, in this study, RTI will be discussed as a framework focused on reading. RTI has been described as a cohesive, preventative school-wide mechanism that provides ongoing support to all students (Glover, 2010). Moreover, “RTI service delivery models are characterized by (1) regular evaluation through universal screening and progress monitoring, (2) strategically selected interventions that increase in intensity across multiple intervention tiers and (3) data-based decision rules for changing interventions and for determining eligibility for special education services” (Daly, Hofstader, Martinez, & Anderson, 2010). An RTI framework also maintains the following five core service delivery components: multitier implementation, student
assessment and decision making, evidence-based intervention provision, maintenance of procedural integrity, and development and sustainability of systems-level capacity (Glover & DiPierna, 2007).

Data-based decision making is a critical aspect of an RTI system, as it is used within a problem-solving framework at each tier (Burns & Ysseldyke, 2005). This data-based decision making leads to differentiated tiers, or levels of support, in the area of reading. Tier I is focused on providing a strong, evidence-based core curriculum, and should be effective for about 80% of students (Walker & Shinn, 2002). In contrast, Tiers II and III provide culturally responsive and evidence-based interventions to students who are at-risk in the area of reading. This greater level of support is in place to reduce the performance gap between at-risk readers and their low-risk peers. In order to facilitate this problem-solving approach to data-based decision making, behavioral consultation is used.

**Behavioral Consultation**

Behavioral consultation can be used within an RTI framework to aide in the development and evaluation of intervention plans and outcomes. Behavioral consultation utilizes indirect service delivery by a consultant with a consultee, and uses a problem-solving approach to treat academic and social problems. The problem-solving process in behavioral consultation consists of four stages: problem identification, problem analysis, treatment implementation, and treatment evaluation (Kratochwill & Bergan, 1990). The problem identification phase consists of specifying what the problem(s) is. The second phase, problem analysis, involves identifying what variables may help solve the problem,
as well as developing a plan to solve the identified problem. Next, during the treatment implementation phase, the plan is implemented and related data are collected to review later. In the final phase of treatment evaluation, data are reviewed to determine the effectiveness of the plan (Kratochwill & Bergan, 1990). This behavioral consultation approach to problem-solving facilitates the data decision-making process within an RTI framework when it comes to deciding what interventions should be provided, and to whom they should be provided. Specifically, these decisions are made during the problem identification and problem analysis phases of behavioral consultation.

Who Should Receive Intervention

The field of reading research has reached consensus that reading instruction and intervention should target the Big 5 skill areas of reading (Daly, Chafouleas, & Skinner, 2005). Although this is true, in order for an intervention to be effective, it is also important to select a reading intervention that matches a particular student’s instructional needs (Daly, Martens, Kilmer, & Massie, 1996). Furthermore, the most effective intervention for a particular student may be dependent on the student’s skill level (Chafouleas, Martens, Dobson, Weinstein, & Gardner, 2004). In order to match an intervention to a student’s skill level and instructional needs, it is first necessary to identify the at-risk students during the problem identification phase of problem-solving.

During the problem identification phase, data are used to identify which students are at-risk in the area of reading. This must be done prior to choosing an intervention, because without first identifying students for intervention, the “fit” of an intervention to a student’s skill deficits cannot be attained. Students receiving intervention through Tier II
and Tier III services in an RTI framework should have been previously found to have not responded to Tier 1 services, or an evidence-based core curriculum implemented with fidelity. In order to determine which students failed to respond to Tier 1 services, screening data are used.

Screening data are gathered from the systematic, universal screening of students. This universal screening consists of the school-wide assessment of all students using brief assessments of basic literacy skills. Screening usually occurs three times per year. Screening allows for the evaluation of the quality of the general education program, as well as the determination of which students the general education program is insufficient for (Hintze & Marcotte, 2010). Due to this, it is an important first step in identifying students who are at-risk for future reading difficulties (Institute of Education Sciences, 2009).

To increase sensitivity and specificity of screening data, screening should be conducted in a systematic way. A screening team should be established that will then make key decisions on the implementation of screening, and ensure that it is effectively carried out. Also of importance is what screening assessment will be used. These assessments are typically brief, and should be reliable and reasonably valid (Institute of Education Sciences, 2009). They should also demonstrate diagnostic accuracy when identifying students at risk for later, poor reading outcomes. Examples of brief, reliable, and valid screening assessments are AIMSweb and DIBELS Next measures. It is also important that individuals conducting screening are adequately trained on the administration of measures, and that interrater reliability data are collected to ensure
accurate administration. After screening occurs, resulting scores are used in conjunction with other confirming data to make decisions on which students are at-risk for failure on year-end accountability measures, and thus need intervention. The National Center on Intensive Intervention provides standards for this data-based individualization. These standards are as follow: decision rules for changing instruction should be available to help guide teachers on when to make instruction changes, decision rules for increasing goals should be established to help guide teachers on when to increase goals, screening tools should increase student performance on other achievement measures, and screening tools should help teachers in planning for and adjusting instruction to meet the needs of students (User’s Guide to Academic Progress Monitoring Tools Chart, 2013).

Curriculum-based measurements are screening tools that allow for these standards to be met.

Curriculum-based measurements are used in the screening process, because they are reliable and valid measures of academic skills (Deno, 1985). Oral reading fluency is a specific curriculum based measurement tool used in screening, and is also the most-well established Reading-CBM tool (R-CBM) (Fuchs, Fuchs, & Hamlett, 2007). It can be used as a screening tool to measure general reading skills, (Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003) because it is a general outcome measure of overall reading skills (Hintze & Silberglitt, 2005). R-CBM measures a broad range of general skills associated with overall competence in a specific skill area (Fuchs & Deno, 1991). In a meta-analysis conducted by Reschly, Busch, Betts, Deno, and Long (2009), no significant differences were found between R-CBM and tests of comprehension, decoding, and vocabulary
(Reschly et al., 2009). Furthermore, it is predictor of general reading achievement without the need of a reading comprehension measure, because comprehension is inherent to the measure (Jenkins et al., 2003). Moderate to robust criterion-related validity of reading curriculum-based measurement for measures of broad reading achievement have been found, deeming R-CBM measures sufficient in guiding screening and for making low-stakes decisions (Christ, Zopluoglu, Monaghan, & Van Norman, 2013). They also inform instructional plans (Fuchs, et al., 2007). Additionally, scores on R-CBM have also been shown to predict future performance on high-stakes tests (Crawford, Tindal, & Steiber, 2001; Hintze & Silberglitt, 2005). Specifically, there is support for the use of R-CBM to predict students’ later performance on statewide achievement tests, as a study found moderate correlations between multiple-choice reading and math achievement tests and scores of R-CBM during the same year and R-CBM from the previous year (Crawford et al., 2001). Furthermore, errors on R-CBM assessments can be used for secondary information (Graney & Shinn, 2005). They can be used to inform how accurately a student is reading, which can then aide in identifying the specific reading skills a student may be having problems with. Using accuracy information from screening derived from how many errors a student makes can thus lead to ensuring that the intervention that is provided to a student matches his or her skill deficits.

**Issues with Screening.** Although there is an extant literature base on the use of R-CBM measures in the screening process, there are practical issues that arise during this process. These issues tend to occur during the decision-making process, or problem
identification phase, immediately following the screening period. Specifically, these issues arise with the decision-making on which students to identify as at-risk for future reading problems. Findings show that in “systems in which large numbers of students are performing below a meaningful risk criterion, identifying individual students for intervention is an error-prone endeavor” (VanDerHeyden, 2013, pp. 403). This makes it especially important to ensure that screening and screening-based decision-making are conducted accurately in schools in which a large number of students are at-risk.

Furthermore, another study of screening accuracy considered both teacher referral and an RTI model of identification in low and average-achieving classrooms. Findings showed that teacher referral and an RTI model of identification were more accurate in average-achieving classrooms compared to low-achieving classrooms; however, teacher referral was found to be less accurate and less stable than the RTI method as a referral source (VanDerHeyden & Witt, 2005). This highlights the notion that more accurate screening is necessary in both schools with large at-risk populations in the area of reading, as well as in low-achieving classrooms in general. Not only is more accurate screening-based decision-making necessary in low-achieving schools and classrooms, but it is also needed for certain students. This is true, as it has been found that students whose scores fall around the cut scores are subjects to the most error-prone screening decisions in schools (Jenkins, Hudson, & Johnson, 2007). Again this highlights the need to improve decision making based on screening data. Although it is clear that an improvement in data-based decision making is necessary, the current research is limited on how to accomplish this.

Similar limitations to those in the problem identification phase occur during the problem
analysis phase; however, these limitations arise when deciding what intervention is provided.

**What Intervention Should Be Provided**

Decision-making on what interventions are used within schools depends on what model of RTI is being implemented. There are two main models of response-to-intervention implementation that are seen, and generally accepted, within schools: the problem-solving and standard protocol approach to RTI (Fuchs, Mock, Morgan, & Young, 2003).

**Standard Treatment Protocol Response-to-Intervention.** A standard treatment protocol approach to RTI consists of delivering a preplanned, manualized intervention protocol to all students who do not meet specific criterion levels of performance, or benchmarks, on a school-wide screening assessment (Daly, Martens, Barnett, Witt & Olson, 2007; Wanzek & Vaughn, 2007). These packaged interventions contain multiple instructional components (Daly et al., 2010). Furthermore, they are evidence-based, standardized, and focus on specific skill areas (Shapiro, 2009). Students are placed into interventions based on a match between their general skills deficits and pre-existing interventions that are designed to develop skills in the same general skill area of reading (e.g.: fluency, alphabetic principle). Standard treatment protocol RTI increases ease of training, and allows for fidelity of implementation to be more easily evaluated and increased (Fuchs & Fuchs, 2006). For these and other reasons, “standard-protocol reading interventions utilized within an RTI framework may be especially promising” (Glover & DiPierna, 2007, pp. 531).
Issues with Standard Protocol RTI. Although standard protocol interventions have empirical support, they have also been shown to be unsuccessful with all students due to several factors (Daly et al., 2007). A main consideration when implementing standard protocol interventions, as with interventions in any model of RTI implementation, is that the intervention should match student need (Tilly, 2008). Although this is true, an in-depth analysis of skill deficit is not conducted within a standard protocol approach to RTI (Shapiro, 2009). The absence of an in-depth analysis of skill deficits makes it difficult to specifically match instruction to student need. One way to make standard protocol interventions more successful is to make slight modifications to the intervention plan in order to create a better fit between the instruction and the child’s level of skill proficiency (Daly et al., 2007).

Problem-Solving Response-to-Intervention. Different from a standard treatment protocol approach to RTI, a problem-solving approach to RTI focuses on targeting individual student problems. A problem has been defined as a discrepancy between what an individual currently perceives and what he or she desires (Deno, 2005). In education, problems manifest in a discrepancy between a student’s current performance level and what is expected. To target this discrepancy within problem-solving RTI, multiple stages of problem-solving are used to ameliorate reading problems. These stages are: problem identification, problem analysis, intervention implementation, and program evaluation (Kratochwill, Elliott, & Rotto, 1995). Unlike in a standard treatment protocol RTI framework, this process enables problem-solving RTI to provide individualized interventions to students as it aims at matching interventions to the skill deficits of
individual students. Overall, these individualized interventions are based on an assessment of skill deficits and instructional and environmental conditions (Tilly, Reschly, & Grimes, 1999).

**Issues with Problem-Solving RTI.** Problem solving RTI has been found to be an effective model; however, its effectiveness relies on the assumption that the practitioners involved have considerable expertise in assessment and intervention (Fuchs & Fuchs, 2006). As mentioned previously, a main roadblock to an RTI system implemented with integrity is the issue of screening-based decision-making. Specifically, the issue arises when using screening data to effectively match interventions to students’ skill deficits. Problem-solving RTI has been shown to be particularly effective in improving student learning; however, this is true when it is implemented with integrity (Burns & Symington, 2002). Oftentimes, students within a problem-solving RTI model appear to be nonresponsive to intervention; however, they may just need an intervention better matched to their skills-based needs. These “nonresponsive” students contribute to the fact that problem solving response-to-intervention is likely to identify “false positives” (Fuchs & Fuchs, 2006). In order to solve this issue, more research-based guidelines need to be formed in order to aide in the accurate identification of a reading problem and the “fit” of a chosen intervention to a skill deficit problem, as they are highly related to the effectiveness of the intervention (VanDerHeyden, Witt, & Barnett, 2005).

**Limitations of Current Research**

As discussed previously, a review of the literature reveals that schools confront considerable challenges when implementing RTI, even though the potential of RTI is
great. This is true in part, because it is not only important that RTI is in place, but more importantly how it is implemented in order for it to have its greatest potential (Glover, 2010). Both standard protocol RTI and problem-solving RTI have their own methods of determining which interventions to use; however, as discussed earlier, both have issues with using screening data to make decisions on who should receive interventions and what interventions should be used. Standard protocol RTI may not individualize interventions enough, whereas problem-solving RTI may make it more difficult to implement interventions with fidelity. Although this is true, there is limited research in regards to how to mitigate the hurdle of picking interventions that match student skill deficits. There have been few studies that have looked at how teachers and interventionists should choose individual interventions for specific students, and currently R-CBM has been found to not be “useful for identifying specific skill deficits or designing individualized interventions” (Ball & Christ, 2012, pp. 234). Future research should focus both on the effects associated with systematic manipulation of individual components on student outcomes, and on the use of different approaches for informing intervention selection and student responsiveness (Glover & DiPerna, 2007). Overall, “a substantial amount of additional research is needed in this area to develop and document technically adequate problem analysis tools” (Ball & Christ, 2012, pp. 235). These problem analysis tools have the potential to greatly improve fidelity of implementation of RTI by improving screening-based decision-making through during the behavioral consultation process. Specifically, these tools can be used during the problem identification and problem analysis stages. Increase the fidelity of implementation of RTI
through the use of such tools can then lead to improved student outcomes. They can specifically be used within behavioral consultation during the problem identification phase to help identify and prioritize the target problem areas. Additionally, they can be used within the problem analysis stage to establish goals and design and implement an intervention plan. This would involve using the tool to choose an intervention that is instructionally matched to a student’s skill deficits based on the screening data.

One such proposed problem analysis tool is a quadrant analysis tool. This tool allows for problem-solving to occur within a standard protocol model of RTI – allowing for a hybrid system of problem-solving and standard protocol RTI. It has been constructed to aide in the placement of students in interventions that are instructionally matched to their individual skill deficits. The tool consists of four quadrants that are mutually exclusive and collectively exhaustive, so all students can be placed into one, and only one, quadrant. Students are placed into quadrants based on their accuracy and fluency scores on an R-CBM screening assessment. Accuracy is defined as the percentage of total words read correctly during the assessment, while fluency is defined as the number of words reads correctly during the assessment. This tool has been developed for grades K-6; however, grades 2-6 are usually the grade levels for which R-CBM is used during all screening periods as an indicator of at-risk status.

The quadrant analysis tool is indirectly supported by some research. In particular, research shows that it is important to consider accuracy and fluency when choosing interventions. Specifically, it has been shown that individuals with higher fluency and fewer errors at baseline and individuals with lower fluency and more errors at baseline
may benefit from different instructional packages. Furthermore, it has been recommended that in order to choose reading interventions for older students that are instructionally matched to the students’ needs, it is important to look at number of errors within an oral reading fluency assessment to look at accuracy (Chafoules et al., 2004). Additionally, in order to determine whether a student is or is not an accurate reader, guidelines of 93-97% accuracy during reading instruction and 70-85% accuracy during practice exercises can be followed (Gickling & Rosenfield, 1995). Fluency and accuracy scores have also been noted as predictors of reading comprehension. So, if students are fluent and accurate readers, they should be comprehending what they are reading. The quadrant analysis tool consists of four quadrants: Quadrant 1: high rate, high accuracy; Quadrant 2: low rate, high accuracy; Quadrant 3: high rate, low accuracy; Quadrant 4: low rate, low accuracy. Based on this tool, students placed in these different quadrants should have differing levels of various reading skills.

**Purpose of Study**

The purpose of this study is to examine the predictive validity of the quadrant analysis tool. Specifically, the predictive classification accuracy of the four quadrants that students are placed into will be examined. To do this, this study will look at whether these groups predict differential performance on later CST subtest performance, which will allow us to know whether these distinct score profiles utilizing only students’ accuracy and fluency scores predict later bands of performance on specific CST ELA subtests. If these distinct score profiles do predict later performance on specific CST ELA subtests, it could increase the utility of screening measures through improving screening-
based decision-making during the problem identification and problem analysis phases of behavioral consultation. Instead of just knowing which students are at-risk on future high-stakes tests, screening results could help place students in better instructionally matched interventions. Four research questions guided this study:

(1) To what extent do accuracy and fluency scores reliably predict future performance on the ELA subtest of “Word analysis, fluency, and systematic vocabulary development?”

(2) To what extent do accuracy and fluency scores reliably predict future performance on the ELA subtest of “Reading comprehension?”

(3) To what extent do accuracy and fluency scores reliably predict future performance on the ELA subtest of “Literary response and analysis?”

(4) To what extent does adding accuracy scores to the discriminant function increase the overall classification accuracy for each of the ELA subtests?

**Method**

**Participants**

Participants included 1,827 students from nineteen schools in an urban school district in Southern California. The district is 78.7% Hispanic or Latino, 19.7% African American, 0.1% American Indian or Alaska Native, 0.1% Asian, 0.7% Native Hawaiian or Pacific Islander, 0.1% Filipino, and 0.3% White. Additionally, 0.2% of students are of two or more races, and 0.1% of students did not report their race/ethnicity (www.ed-data.k12.ca.us). English Learners (ELs) comprise 41.4% of the students in the district. All participants were in 3rd grade.
Measures

**AIMSweb Reading Curriculum Based Measurement (R-CBM).** AIMSweb R-CBM is a brief, standardized assessment of oral reading. It is administered individually through a paper or computer-assisted administration. Three oral reading passages are given during a screening period, while only one passage is typically given for progress monitoring purposes. Students read each passage for one minute. Fluency (words read correct per minute) and accuracy (percentage of words read correctly) scores are attained for each probe administered (Pearson, 2012). Average alternate-form reliability is 0.94, average reliability of median score is 0.97, and between-season reliability for 3rd grade is .93 for fall-winter and .94 for winter-spring. Criterion validity is approximately 0.7, and classification accuracy of 3rd grade spring scores are as follow: sensitivity .77, specificity .81 (Pearson, 2012).

**California Standards Tests (CST).** The California Standards Tests are a series of criterion-referenced tests used to assess the California educational content standards. Test subjects include: English language arts (ELA), mathematics, science, and history-social science. The CSTs are a component of the STAR program which is designed to examine school and student progress in grades two and above (www.cde.ca.gov). The ELA section assesses students on the California content standards for reading and writing. This study used only the Grade 3 ELA reading portion of the CST. The ELA Reading section has subtests of *Word Analysis, Fluency, and Systematic Vocabulary Development, Reading Comprehension,* and *Literary Response and Analysis.* The *Word Analysis, Fluency, and Systematic Vocabulary Development* subtest consists of 20
questions that assess decoding and word recognition, as well as vocabulary and concept development. The *Reading Comprehension* subtest consists of 15 questions that assess structural features of informational materials, as well as comprehension and analysis of grade-level-appropriate text. The *Literary Response and Analysis* subtest consists of 8 questions that assess the structural features of literature, as well as narrative analysis of grade-level-appropriate text (California Department of Education, 2002). Students are scored based on percentage of answers correct on each of these subtests. Based on the percentages, students are determined to be proficient or not proficient in the skills measured by each of the subtests. The goal in California is to have all students performing at least at the proficient level (California Department of Education, 2011).

The California Standardized Testing and Reporting Post-Test Guide Technical Information document provides guidelines as to what percentage is used as the cut point for proficient or not proficient. The values were maintained based on the expected average percent-correct scores for each cluster from a state-representative sample of students who scored at the lowest score for proficient on the ELA section of the CST (California Department of Education, 2011).

**Procedure**

Data were collected by school employees. School employees were trained on the administration of all measures by graduate students working as Response-to-Intervention specialists in the district. The R-CBM measures were administered during the Fall, Winter, and Spring of the school year. R-CBM scores from the Spring administration were included in the analysis for this study. Each student was administered three probes
during each screening period, and the median accuracy and fluency scores were included as predictor variables in the analyses for this study. The CSTs were administered in the Spring of the same school year.

**Data Analysis.** The two independent variables, students’ fluency (words read correct per minute) and accuracy (percentage of words read correctly) were entered into a discriminant function analysis (DFA) to determine if they could reliably predict student’s performance on each of the three reading CST ELA subtests. Discriminant function analysis uses multiple independent variables to predict classification of a categorical dependent variable. In this study, DFA was used to determine how accurately student’s performance on all three reading CST ELA subtests could be reliably predicted based solely on AIMSweb fluency and accuracy scores. The relative importance of both accuracy and fluency scores was examined through looking at the standardized and structural coefficients.

Missing data were noted and addressed. 96 students had missing scores for word analysis and vocabulary, reading comprehension, and literary response and analysis scores. 72 students had missing scores for Spring fluency, and 120 students had missing scores for accuracy. Missing values were mean imputed, because less than 10% of cases were missing for each predictive and outcome variable.

Students’ performance on the ELA subtests were dummy coded (0 = not proficient, 1 = at least proficient), to turn the continuous variable into a categorical variables for analysis. To discriminate between student scores that were proficient or not proficient, values for the expected average percent-correct scores for each cluster as
reported in the Post-Test Guide Technical Information document were utilized. The categories were mutually exclusive, and allowed for each case to be assigned to only category. These values were maintained from a state-representative sample of students who scores at the lowest score for proficient (California Department of Education, 2011). Values used to dummy code these variables are displayed in Table 1. Expected values were not provided for the other performance levels. Predictor variables were continuous.

Table 1

<table>
<thead>
<tr>
<th>Subtest</th>
<th>% Minimally Proficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Analysis and Vocabulary</td>
<td>80</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>73</td>
</tr>
<tr>
<td>Literary Response and Analysis</td>
<td>83</td>
</tr>
</tbody>
</table>

Results

Discriminant function analyses were performed to determine if students’ accuracy and fluency scores predicted their performance on the subtests of the CST ELA assessment. Analyses were run on data collected from 1,827 third grade students to address all research questions. Results of the analyses are presented next.

Descriptive Statistics

Prior to performing the discriminant function analyses, descriptive statistics were attained to evaluate the data. The mean, standard deviation, skewness, and kurtosis for the accuracy, fluency, and ELA subtest scores for the total sample are displayed in Table 2. As seen in the values reported, sample data for all ELA subtests appear to be platykurtic and negatively skewed, whereas sample data for accuracy and fluency scores appear to be leptokurtic and negatively skewed.
Table 2

Descriptive Statistics for Total Sample

<table>
<thead>
<tr>
<th>Measure</th>
<th>M</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Analysis and Vocabulary (N=1827)</td>
<td>68.41</td>
<td>21.41</td>
<td>-.54</td>
<td>-.49</td>
</tr>
<tr>
<td>Reading Comprehension (N=1827)</td>
<td>59.67</td>
<td>22.58</td>
<td>-.28</td>
<td>-.67</td>
</tr>
<tr>
<td>Literary Response and Analysis (N=1827)</td>
<td>59.66</td>
<td>26.40</td>
<td>-.24</td>
<td>-.87</td>
</tr>
<tr>
<td>R-CBM Fluency (N=1827)</td>
<td>111.45</td>
<td>43.04</td>
<td>-.11</td>
<td>.13</td>
</tr>
<tr>
<td>R-CBM Accuracy (N=1827)</td>
<td>95.67</td>
<td>8.48</td>
<td>-.42</td>
<td>25.06</td>
</tr>
</tbody>
</table>

After dummy coding performance on the ELA subtests (0 = not proficient, 1 = at least proficient), a frequency count was calculated for proficient and non-proficient students on each of the subtests. Frequency data are displayed in Table 3.

Table 3

Frequency of Proficient and Non-Proficient Students

<table>
<thead>
<tr>
<th>Measure</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Analysis and Vocabulary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At Least Proficient</td>
<td>708</td>
<td>38.8</td>
</tr>
<tr>
<td>Not Proficient</td>
<td>1119</td>
<td>61.2</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At Least Proficient</td>
<td>659</td>
<td>36.1</td>
</tr>
<tr>
<td>Not Proficient</td>
<td>1168</td>
<td>36.9</td>
</tr>
<tr>
<td>Literary Response and Analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>At Least Proficient</td>
<td>458</td>
<td>25.1</td>
</tr>
<tr>
<td>Not Proficient</td>
<td>1369</td>
<td>74.9</td>
</tr>
</tbody>
</table>

Next, statistical assumptions of discriminant function analysis were tested for. The assumptions of DFA are the same as those for MANOVA, and include: normality, homogeneity of variance/covariance (homoscedasticity), multicollinearity, and
independence. Violations of normality and homogeneity of variance/covariance were noted. P-P plots of variables were analyzed to assess normality. Visual analysis determined normality of all variables, with the exception of student’s accuracy scores. To assess the assumption of homogeneity of variance/covariance, Box’s M test statistic was examined for each analysis. The Box’s M test statistic for equivalence of covariance matrices was statistically significant ($p=0.000$) for all analyses, indicating that the covariance matrices did differ by group. This test statistic should be non-significant; however, this study had a large sample size which can cause small differences in covariance matrices to be significant according to Box’s M. The “Log Determinants” tables were examined next to determine if the log determinants of each group were similar, which would indicate than we can ignore the results of Box’s M. In this case, the log determinants were similar. Although these violations exist, discriminant function analysis has been reported to be robust to slight violations of assumptions (Lachenbruch, 1975).

**Research Question 1 Results**

As displayed in Table 4, Wilks’ lambda was significant for both fluency and accuracy by the F test ($p = .000$). Wilks’ lambda was smaller for fluency (0.711) than accuracy (0.910) scores, indicating that fluency scores contribute more to the discriminant function than do accuracy scores for this ELA subtest. Furthermore, this means that fluency scores contribute more than accuracy scores to do the classification of students’ performance on this CST ELA subtest. The overall model as a whole is significant, because both independent variables are significant. The canonical correlation
(.504) is medium, showing that there is a medium correlation between the discriminant function and the groups. Not all of the variance in the discriminant scores can be attributed to group differences explained by the function. Wilks’ Lambda for the model as a whole was significant (0.708, \(p=.001\)), indicating that the overall model is significant, and thus is discriminating. Standardized canonical discriminant function coefficients were examined to see the degree to which each variable contributed to the function. Fluency scores contribute more to the model than do accuracy scores. Structural coefficients indicated that fluency scores were also more correlated with the discriminant function than are accuracy scores.

Table 4

*Wilks’ Lambda*

<table>
<thead>
<tr>
<th></th>
<th>Wilks’ Lambda</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Word Analysis and Vocabulary</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluency</td>
<td>.71</td>
<td>.00</td>
</tr>
<tr>
<td>Accuracy</td>
<td>.91</td>
<td>.00</td>
</tr>
<tr>
<td>Test of Function</td>
<td>.71</td>
<td>.00</td>
</tr>
<tr>
<td><strong>Reading Comprehension</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluency</td>
<td>.75</td>
<td>.00</td>
</tr>
<tr>
<td>Accuracy</td>
<td>.93</td>
<td>.00</td>
</tr>
<tr>
<td>Test of Function</td>
<td>.75</td>
<td>.00</td>
</tr>
<tr>
<td><strong>Literary Response and Analysis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluency</td>
<td>.79</td>
<td>.00</td>
</tr>
<tr>
<td>Accuracy</td>
<td>.95</td>
<td>.00</td>
</tr>
<tr>
<td>Test of Function</td>
<td>.78</td>
<td>.00</td>
</tr>
</tbody>
</table>

The results of the discriminant function analysis predicting performance on the ELA subtest of “word analysis, fluency, and systematic vocabulary development” produced high classification rates. Group membership was correctly predicted for 74.4%
of proficient students and 75.2% of non-proficient students. 74.9% of students were correctly classified overall. This classification accuracy was compared to a “by chance alone” criterion. This criterion was calculated by multiplying prior probabilities by group sizes, adding these together for all groups, and dividing the resulting sum by N. This calculation resulted in 50%. The classification accuracy of this discriminant function (74.9%) is higher than chance alone. Classification accuracy results for all research questions are displayed in Table 5.

Table 5

<table>
<thead>
<tr>
<th>Classification Results</th>
<th>Percentage of Cases Correctly Classified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word Analysis and Vocabulary</td>
<td>74.9</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>72.5</td>
</tr>
<tr>
<td>Literary Response and Analysis</td>
<td>73.9</td>
</tr>
</tbody>
</table>

Research Question 2 Results

As displayed in Table 4, Wilks’ lambda was found to be significant for both fluency and accuracy by the F test ($p = .000$). Wilks’ lambda was smaller for fluency (0.749) than accuracy (0.926) scores, indicating that, similar to the previous ELA subtest, fluency scores contribute more to the discriminant function than do accuracy scores for this ELA subtest. Both independent variables are significant, indicating that the overall model is significant. The canonical correlation (.504) is medium, showing that there is a medium correlation between the discriminant function and the groups. Wilks’ Lambda for the model as a whole was significant (0.746, $p = .001$), indicating that the overall model is significant, and thus is discriminating.
Based on the standardized canonical discriminant function coefficients, fluency scores contribute more to the model than do accuracy scores. Structural coefficients indicated that fluency scores were also more correlated with the discriminant function than were accuracy scores.

Group membership for the reading comprehension subtest was correctly predicted for 72.7% of proficient students and 72.2% of non-proficient students. 72.5% of students were correctly classified overall. The classification accuracy of this discriminant function (72.5%) is higher than chance alone (50%).

**Research Question 3 Results**

As displayed in Table 4, Wilks’ lambda was found to be significant for both fluency and accuracy by the F test \((p = .000)\). Wilks’ lambda was smaller for fluency (0.790) than accuracy (0.950) scores, indicating that, similar to the findings on the other ELA subtests, fluency scores contribute more to the discriminant function than do accuracy scores for this ELA subtest. The canonical correlation (.465) is medium, showing that there is a medium correlation between the discriminant function and the groups. Wilks’ Lambda for the model as a whole was significant (0.784, \(p=.00\)), indicating that the overall model is significant, and thus is discriminating. Based on the standardized canonical discriminant function coefficients, fluency scores contribute more to the model than do accuracy scores. Structural coefficients indicated that fluency scores were also more correlated with the discriminant function than were accuracy scores.

Group membership for the reading comprehension subtest was correctly predicted for 73.8% of proficient students and 74.2% of non-proficient students. 73.9% of students
were correctly classified overall. The classification accuracy of this discriminant function (73.9%) is higher than chance alone (50%).

**Research Question 4 Results**

Initial analyses showed that the combination of using accuracy and fluency scores reliably predicted students’ performance on each of the CST ELS subtests. Furthermore, fluency scores were found to contribute more than accuracy scores to the discriminant function for all three subtests. To examine whether or not the inclusion of accuracy scores increased the overall classification accuracy, analyses were run first without accuracy scores and then with them. For the Word analysis, fluency, and systematic vocabulary development subtest, 74.5% of all cases were correctly classified when using fluency scores alone. Including both fluency and accuracy scores together, 74.9% of cases were correctly classified. The inclusion of accuracy scores did not greatly increase the overall classification accuracy. For the reading comprehension subtest, 72.1% of cases were correctly classified when using fluency scores alone. The addition of accuracy scores into the equation increased this percentage to 72.5%. Again, the classification accuracy remained almost the same. For the literary response and analysis subtest, 72.9% of cases were correctly classified when using fluency scores alone. The addition of accuracy scores for prediction of performance on this subtest increased the overall classification accuracy by 1% (to 73.9%). Overall, the results of these analyses show that the addition of accuracy scores to the equation did not increase the overall classification accuracy for any of the ELA subtests.
Discussion

Fluency and accuracy scores predicted performance levels on the “Word analysis, fluency, and systematic vocabulary development,” “reading comprehension,” and “literary response and analysis” CST ELA subtests at 72.9%, 72.5% and 73.9%, respectively. All of these rates are higher than what would be predicted by chance alone. Additionally, fluency scores contributed more to the discriminant functions than accuracy scores did for all three subtests. Although this is true, accuracy scores also significantly contributed to the discriminant functions. When examining whether the overall classification accuracy would increase when also considering accuracy scores into the equation, results showed that fluency scores alone predicted performance on all three ELA subtests with almost the same classification accuracy than when also adding in accuracy scores.

As hypothesized, both fluency and accuracy scores contributed significantly to the discriminant functions for each subtest; however, fluency scores contributed more than the accuracy scores did for all subtests. This may be due to the fact that R-CBM is a general outcome measure of reading, even when not considering accuracy scores. Although the accuracy scores may provide some extra information, fluency scores are predictive enough to get the information we need from these measurement tools. When examining the overall classification accuracy after adding in accuracy scores to the equation, again the accuracy scores did not provide extra information than what was attained with the fluency scores alone. The ELA subtests are aimed at individually measuring: decoding, word recognition, vocabulary, and concept development, structural
features of informational materials and comprehension, and the structural features of literature and narrative analysis of text. The results found in this study suggest that using fluency scores alone are predictive of each of these sub areas assessed within the ELA section of the CST.

The quadrant analysis tool aims to use both fluency and accuracy scores within the problem identification and problem analysis stages of problem-solving in an RTI framework. This is done with the hopes of improving instructional match of interventions to student skill deficits, which in turn will lead to improved student outcomes. This study shows that using the results of Reading-CBM can improve instructional fit of interventions, as classification rates were high when predicting proficient performance on each of the CST ELA subtests; however, fluency scores alone are sufficient to accomplish this. Furthermore, although fluency scores tend to be used to predict overall reading future reading performance, this study shows that they are also predictive of each of the individual subtest assessment areas.

Implications for Practice

Results of this study indicate that fluency and accuracy scores reliably predict student performance on all three subtests of the CST ELA subtest. Therefore, the quadrant analysis tool that has been developed to classify students for interventions based on their fluency and accuracy scores can be effective in discriminating what interventions students should be placed in. Classification rates were high for both students who were proficient on all subtests, as well as those who were not-proficient on all subtests. The implication of this is that students who are at higher or lower risks of reading skills can
be classified using their accuracy and fluency scores with a high level of accuracy. Although this is true, when examining how much the addition of accuracy scores into the equation increased the overall classification accuracy, substantial differences were not noted. This shows that the consideration of accuracy scores by the quadrant analysis tool may be unnecessary. Considering accuracy scores when making decisions on what interventions students need is a unique consideration of the quadrant analysis tool. Results of this study show that this may be unnecessary. Although the results of this study show this, it is not to say that considering accuracy in developing interventions is not appropriate. Accuracy scores may inform how well a child is comprehending text. Considering this, other methods of identifying students for interventions should be explored. For example, other pieces of data representing students’ accuracy, or comprehension ability, can be used in conjunction with students’ fluency scores to help better inform intervention placements. Also important to consider is that the classification accuracy was not 100% for any subtest. Considering this, it is important that school staff utilize other pieces of data, such as teacher judgment, before placing students in interventions if they do decide to use this quadrant analysis tool.

**Limitations**

Several limitations are necessary to discuss for this study. Firstly, the use of secondary data is a limitation of this study. Due to the nature of using secondary data, researchers could not monitor and control for standardization of AIMSweb passage delivery. Specifically, interrater data were not collected during the administration of the AIMSweb measures. It is important to confirm accurate data. This lack of interrater
reliability means that inaccurate testing, and thus inaccurate data, may have been included in this study. Results of this study may have been affected by the lack of interrater data.

Another limitation is the limited sample to include only third grade students. In terms of reading skills, there are differences among students of different grade levels. Furthermore, Reading-CBM has varying reliabilities for different grade levels. Future research should be conducted to consider the classification accuracy of using students’ fluency and accuracy scores to predict their performance on CST ELA subtests at other grade levels.

A third limitation of this study is that the ELA subtests that were utilized in this study are comprised of small numbers of questions. This small number of items may result in less-reliable scores than overall test scores (California Department of Education, 2011).
References


*User’s guide to academic progress monitoring tools chart.* (2013). In *National center on intensive intervention.* Retrieved from


