THE UNIVERSITY OF RHODE ISLAND

University of Rhode Island
DigitalCommons@URI

School of Education Faculty Publications

School of Education

2016

Teachers' and Students' Views of Reading Fluency: Issues of Consequential Validity in Adopting One-minute Reading Fluency Assessments

Theresa A. Deeney
University of Rhode Island, tdeeney@uri.edu

Minsuk K. Shim
University of Rhode Island, mshim@uri.edu

Follow this and additional works at: https://digitalcommons.uri.edu/education_facpubs

Teachers' and students' views of reading fluency:
Issues of consequential validity in adopting one-minute reading fluency assessments

Theresa A. Deeney, Ed.D.
University of Rhode Island
School of Education
605 Chafee Hall
Kingston, RI 02881
(401) 874-2682
tdeeney@uri.edu

Minsuk K. Shim, Ph.D.
University of Rhode Island
School of Education
708 Chafee Hall
Kingston, RI 02881
(401) 874-4162
mshim@uri.edu

Abstract

In this study, we investigate assertions that use of one-minute oral reading fluency measures, such as the *Dynamic Indicators of Basic Early Literacy Skills* (DIBELS) Oral Reading Fluency (ORF) assessment and the *aimsweb Reading CBM (R-CBM)*, may generate negative educational consequences, specifically teaching to ORF tests, increased pressure and anxiety for teachers and students, and misuse of ORF measures and results. We present results of a survey designed to gain perspective on teachers' and students' views related to these potential negative consequences. Data gained from 77 teachers and 875 students in two districts generally do not support assertions of teaching to the test, but do support some caution with respect to use of ORF measurements and data, and issues of anxiety, particularly from more vulnerable students. This paper provides detailed results across teachers and students, and implications for practice.

Since the publication of the National Reading Panel report (2000), increased attention has been paid to assessing oral reading fluency (ORF). Many districts use one-minute ORF measures, such as the *Dynamic Indicators of Basic Early Literacy Skills* (DIBELS) Oral Reading Fluency assessment (Good & Kaminski, 2002) and *aimsweb Reading CBM (R-CBM)* (NCS Pearson, 2014). However, one-minute ORF measures have garnered intense scrutiny, with critics warning unintended negative effects of their use, such as teaching to the test (Tierney & Thome, 2006), increased pressure and anxiety for teachers and students (Goodman, 2006), and misuse of measures and results (Valencia, Smith, Reece, Li, Wixon, & Newman, 2010). In our search of the literature (Education Full Text, PsycINFO, PsycARTICLES, Science Direct, ProQuest), we found few empirical

studies investigating or confirming negative consequences of using one-minute measures. Our study begins to address this void by surveying teachers and students in two districts that mandate one-minute ORF measures to understand their perceptions of the measures, how measures are used in schools, and how teachers engage students in instructional practices related to ORF.

Background

Oral Reading Fluency

Definitions of oral reading fluency vary in research and practice. Some researchers (e.g. Chard, Vaughn, & Tyler, 2002) discuss rate and accuracy. Others additionally include prosody or expression (Benjamin, Schwanenflugel, Meisinger, Groff, Kuhn, & Steiner, 2013)—which consists of pitch and stress (related to intonation or expression), duration (pausing), and parsing (phrasing) (Kuhn et al., 2010). Still others argue that ORF also includes comprehension (Harris & Hodges, 1995; Samuels, 2006). The National Reading Panel (2000) suggests that fluency is "the ability to recognize words easily, read with greater speed, accuracy, and expression, and to better understand what is read" (http://www.nichd.nih.gov/research/supported/Pages/nrp.aspx/). Since ORF is a complex construct, assessing it in its entirety has proven challenging. Although strides have been made in creating valid and reliable prosody measures (Balanos, Cole, Ward, Hasbrouck, Tindal, & Schwanenflugel, 2014; Benjamin et al., 2013; Miller & Schwanenflugel, 2008; Valencia et al., 2010), in practice, most ORF assessments used in schools target only accuracy and rate. *Dynamic Indicators of Basic Early Literacy Skills* (DIBELS) ORF subtest (Good & Kaminski, 2002) and *aimsweb Reading CBM (R-CBM)* (NCS Pearson, 2014), are one-minute assessments of rate and accuracy measured in Words Correct Per Minute (WCPM). Results of these ORF measures do relate to broader reading skills, such as comprehension (Buck & Torgesen, 2003). However, assessing comprehension within the ORF task is not mandatory.

Validity and ORF Measures

ORF measures are curriculum-based measures (CBM), or, more appropriately, general outcome measures (GOM). These formative assessments are intended to measure how students are faring in reading (Deno, 1985), and to provide information for teachers to assess student progress and adjust instruction to meet student needs (Kingston & Nash, 2011). One-minute ORF measures have shown good reliability in both test-retest (.92-.97) and alternate form (.89-.94) comparisons (Baker et al., 2008). They have also shown satisfactory validity, which broadly is the degree to which a test measures what it is purported to measure (Messick, 1996). Welner (2013) advises that multiple aspects of validity be considered when evaluating assessments. Studies have investigated ORF measures' criterion validity, or the degree to which ORF measures correlate with an established criterion, typically standardized tests of reading, administered concurrently (concurrent validity) (Goffreda & DiPerna, 2010) or at a later time (predictive validity) (Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2007). ORF measures show strong correlations (58-.92 as reported in Baker et al., 2008) with tests of overall reading achievement (Daane, Campbell, Griggs, Goodman, & Oranje, 2005; Fuchs, Fuchs, Hosp, & Jenkins, 2001) and with measures of reading comprehension (Buck & Torgesen, 2003). Construct validity, the degree to which a test measures the intended construct as it is defined theoretically and through research, has also been discussed with respect to ORF measures. Critics (Hosp & Suchey, 2014; Samuels, 2007) argue that ORF measures, which assess only accuracy and rate, do not measure the full construct of oral reading fluency. Though termed "fluency" assessments, they are more accurately assessments of automaticity. In addition to measurable aspects of validity, all educational tests have intended positive and unintended negative consequences (Messick, 1996), especially consequences associated with teaching and learning (Weir, 2005). This idea of consequential validity, which is what happens as a result of test use (AERA, APA, & NCME, 1999), although not a stand-alone validity construct, should be an important part of validity evaluation (Messick, 1993; Shephard, 1997). Critics argue that ORF measures may have three particular unintended negative consequences which we will examine more thoroughly next: a narrowing definition of fluency, which is related to concerns of construct validity (Samuels, 2006); teaching to the test and associated impact on students' perceptions of reading, increased anxiety, and decreased confidence (Goodman, 2006; Kuhn et al., 2010; Tierney & Thome, 2006); and misuse of ORF measures and results (Shanahan, 2009; Valencia et al., 2010).

Definitions. Definitions of a construct should inform how we measure that construct. Samuels (2006) stresses "definitions are not "trivial, ivory tower concerns that are of no importance—how we define a construct such as fluency determines to a large extent how we will measure it" (p. 39). Although some argue that ORF measures do not adequately capture the construct of ORF as defined theoretically (Kuhn et al., 2010), since meaning is created by use, a semiotic shift may occur through test use—what the measure assesses can become the

definition of the construct (Welner, 2013; Baker, 2013). Samuels (2007) warns that use of ORF measures, which assess only accuracy and rate, has the potential to reduce the construct of fluency and redefine it as merely accuracy and rate. In our review of empirical studies to understand whether teachers view ORF as a narrow construct, we found three that purported to gain teachers' definitions of reading fluency. Two (Vesay & Gischlar, 2013; Newman-Thomas et al., 2012) did not describe teachers' definitions in the results. Meisinger, Bradley, Schwanenflugel, and Kuhn (2010) found the majority (45.2%) of third grade teachers in their study held a "basic" definition of fluency, which the researchers identified as definitions that included rate, and/or accuracy, and/or expression. However, because a "basic" definition could include prosody, their results do not allow us to support or reject critics' assertions that teachers may define fluency as simply what ORF tests measure—rate and accuracy. In addition to the potential of fluency being redefined, critics argue that use of timed ORF measures may create a shift in students' thinking about reading—that students may come to understand that the purpose of reading is speed, over comprehension (Samuels, 2007).

Teaching to the test. Tied to narrowing views of fluency, some (e.g. Goodman, 2006; Tierney & Thome, 2006; Valencia et al., 2010) raise concerns regarding Campbell's Law (Campbell, 1976), which purports that "when test scores become the goal of the teaching process, they both lose their value as indicators of educational status and distort the educational process in undesirable ways" (p. 52). In the case of ORF, they argue that the intended outcome of ORF measures will become not fluency in its broad sense, but increasing WCPM scores on tests. To achieve this, they argue that teachers may embed practice reading short texts quickly and accurately into the curriculum (Kuhn et al., 2010). Although there is some evidence that teachers teach to high stakes tests (Jones, Jones, Hardin, Chapman, Yarbrough, & Davis, 1999), we found no evidence that ORF measures are used as high stakes tests, or that teachers target instruction toward these measures. Related, Goodman (2006) warns that teaching to the test may pressure students to read quickly, which may increase students' anxiety and decrease their reading confidence, which, in turn, may affect student learning (Bandura, 1982). Reading self-efficacy has demonstrated relationship with reading achievement (Archambault, Eccles, & Vida, 2010); yet, we found no empirical research investigating relationships between ORF measures and student anxiety or reading self-perception.

Use of ORF measures. Critics also raise two arguments that ORF measures may be misused or misunderstood. Although guidelines of DIBELS and *aimsweb* recommend multiple data points to inform instruction (Dynamic Measurement Group, 2012; NCS Pearson Inc., 2012), critics express concerns that instructional decisions may be made based solely on ORF scores, which do not provide rich information. Valencia et al. (2010) caution that "making decisions on such limited data are destined to waste time and money, fail to achieve the larger goal of improved student learning, and fall short of meeting the needs of students most at risk" (p. 288). In searching the literature, we did not find data related specifically to how teachers use ORF measures to plan instruction. Second, critics worry that teachers and/or administrators may misunderstand progress-monitoring data due to the statistical properties of ORF assessments. DIBELS and *aimsweb* manuals (Dynamic Measurement Group, 2012; NCS Pearson Inc., 2012) suggest frequent progress monitoring for students below benchmark in ORF. Yet, Shanahan (2009) argues that, given the high Standard Error of Measurement (SEM) of ORF assessments (4-12 points per Poncy, Skinner, & Axtell, 2005), results of frequent progress monitoring can be difficult to interpret. As an example, given a typical gain for second graders of one WCPM per week, intervals greater than 12 weeks would be needed to ensure a gain is outside the SEM. Again, we found no empirical data regarding how teachers are using or interpreting progress-monitoring data.

In sum, critics warn about narrowing definitions of fluency and reading, teaching to the test and its associated consequences for students (increased anxiety, decreased confidence), and misuse of ORF measures and results. Because there is little empirical data confirming these unintended effects, we surveyed elementary teachers and students in two districts that mandate *aimsweb* or DIBELS to understand whether those who use/experience ORF measures confirm these potential negative consequences. Our research questions were as follows:

- What do elementary-grade (2-5) students report about the importance of various aspects of reading, classroom instructional practices, reading self-perception, and pressure/anxiety in taking ORF measures?
- What do elementary-grade (2-5) teachers report as definitions of fluency, classroom instructional practices, and use of ORF measures and results?
- What differences exist in teacher/student reports by district and reading ability?

Method

Participants

Participants were 77 teachers (62 classroom, 15 specialists who work with students who struggle) and 875

students in five elementary schools within two districts in the northeastern U.S. (Table 1). We surveyed 125 2nd-4th grade students from two small (250-374 students) K-4 schools in suburban District A. Student make up was predominantly White (85-91%), moderate income (17-19% low SES), and high achieving in reading (82-84% proficient on state tests). District A had been using *aimsweb* for three years. The 17 teachers in District A, with 5 to37 years experience, reported that they received adequate professional development (PD) and were confident in administering and scoring *aimsweb*. They also reported heavy district emphasis on *aimsweb*, and pressure to improve student scores. Although *aimsweb* scores were not used in teacher evaluations, teachers reported that improving *aimsweb* scores was a subject of many school- and grade-based meetings.

We surveyed 750 2nd-5th grade students from three large (600-700 students) K-5 schools in urban District B. Student make up was diverse (40-46% non-White), low income (80-90%), and low achieving (24-27% proficient in reading on state tests). District B had been using DIBELS for six years. The 60 District B teachers, having 1 to 36 years experience, reported adequate PD and confidence in administering/scoring the DIBELS, as well as heavy district emphasis on, and pressure to improve, DIBELS scores. Again, DIBELS results were not used for teacher evaluation, but teachers reported that district emphasis resulted in much time spent discussing lack of student progress on ORF tests.

Teachers in Districts A and B both had access to a basal reading program. Yet, walk-through visits and discussions with school principals indicated that many teachers used leveled texts in a small group guided reading format. In both districts, students struggling in reading received additional instruction with special needs teachers (for students with Individual Education Plans), and through certified reading specialists in District A and interventionists certified in elementary education in District B.

Survey Design

We designed a student and teacher survey to capture views of issues raised about ORF measures (survey items are in Appendix A). For student items, we used icons representing a four-point scale (*not very much, a little bit, a lot, a whole lot*), which we converted to numerical values (1=least, 4=most). For teacher items, we used four-point scales for frequency (*never, occasionally, weekly, daily*) and statement agreement (*strongly disagree* to *strongly agree*).

Definitions of reading and fluency. To collect information about teachers' definitions of ORF, we asked teachers to indicate their level of agreement with one statement focused on their perceptions of the importance of fluency, and two open-ended items requesting that teachers define ORF and discuss its importance. To address the potential that ORF tests may lead students to overvalue speed, we designed seven student items addressing the importance of various aspects of reading, including aspects assessed by ORF tests (speed, accuracy), comprehension, and aspects potentially diminished by a focus on speed (enjoyment, use of comprehension strategies). To add depth, we asked students to define what it means to be a good reader. To determine teacher/student agreement, we asked teachers to rate the importance their students place on comprehension and speed.

Teaching to the test. We designed similar teacher and student items to understand the frequency of instructional practices related to ORF, including those that might be designed to improve ORF scores (i.e. timed tasks), as well as practices recommended to build connected text fluency (e.g. repeated reading, buddy reading, readers' theatre), and those that may target developmental aspects of fluency (e.g. reading words in isolation/ decoding). For teachers, we specified timed/untimed where appropriate (e.g. timed versus untimed word reading) to ascertain their focus on timed tasks. For students, we included one general item asking how often students engaged in timed tasks. To understand criticisms of decreased student reading confidence/self-perception, we asked students the degree to which they feel they are good readers and like to read. To address concerns that ORF measures may generate feelings of pressure/anxiety, we asked students whether, when taking ORF tests, they pay attention to reading, understand what they read, like ORF tests, and become nervous. To determine teacher/student agreement on these issues, we asked teachers to rate their perceptions of student anxiety/pressure.

Use of ORF measures. We included five items to assess the degree to which teachers have changed their instructional practices or incorporated increased fluency instruction since using ORF measures, are more aware of fluency since using ORF measures, the measures provide important information about students' ORF, results accurately reflect students' reading abilities, teachers group students based on results, and teachers would choose to use ORF measures if given a choice. We also asked how frequently teachers administer ORF

measures, and the level of text used for progress monitoring. To add depth, we included four open items prompting teachers to discuss their understanding of fluency, how they use ORF measures, and their opinions the measures.

Data Collection

Students in grades 2-4 in District A were given informed consent materials for parent review. Students receiving parent permission were asked to complete child assents. District B chose to use the survey for curriculum reform. Therefore, we sent home passive consents in English, Spanish, and Portuguese with students in grades 2-5. Students without permission completed the survey, but results were not included in our analysis. In all schools, Author 1 administered the survey in classroom groups, read and explained directions, presented each item and choices, and monitored responses to ensure completion. While students completed the survey, the classroom teacher indicated on each student's survey his/her impression of student's reading level (above-, on-, below-grade). For teacher data, Author 1 met with teachers during faculty meetings to explain the study and request participation.

Data Analysis

We calculated frequencies and descriptive statistics for students' and teachers' responses to each item on the surveys. We used district as an independent variable for both student and teacher data to understand any differences by district. Because of the potential that students who struggle may experience more frequent assessment with ORF measures, we additionally used student reading ability as an independent variable for student data, and teacher role (classroom teacher vs. specialist) as an independent variable for teacher data, to understand any differences that exist for students at various reading abilities and for teachers who work with students who struggle.

To examine group differences (district and reading level for students, district and role for teachers), we applied multivariate analysis of variance (MANOVA) to multiple dependent variables, as there were significant correlations among items addressing the same construct on the survey (e.g. instructional practice items, anxiety items) (correlations ranged from .02-.52 for student data, and .02-.70 for teacher data). MANOVA is known to be a better statistical procedure than multiple ANOVAs as it is sensitive not only to mean differences among groups but also to the direction and size of correlations among multiple dependent variables (Grimm & Yarnold, 2000). MANOVA provides more power than separate ANOVAs and controls for Type I error inflation (Johnson & Wichern, 1998). As there were two independent variables examined in both student data (district and reading level) and teacher data (district and role), we ran a two-way MANOVA for each construct.

We analyzed open responses qualitatively (Miles & Huberman, 1994), initially using in vivo coding to capture students' and teachers' thoughts, then grouping like ideas, and further collapsing these items into broader themes. We then converted the qualitative data into percentage of respondents to enable us to more clearly discuss those findings.

Results

Definitions of Reading and Fluency

Although 12% of District A teachers and 26% of District B teachers defined fluency as rate and/or accuracy—the aspects of fluency measured by ORF assessments, the majority offered a multi-pronged definition of fluency that included rate and accuracy and/or comprehension and/or prosody, which they described as *expression, voice, tone, attending to punctuation, phrasing, smoothness, fluidity, intonation,* and *flow.* Asked why fluency is important, 83% of the teachers discussed fluency as vital to comprehension.

With respect to students' views of the importance of speed, we first ran exploratory factor analysis (see Appendix B) to see the congruence among the seven survey items designed to capture the importance students place on various aspects of reading. Six items (remembering story, understanding, figuring out all words, rereading to understand, reading w/o mistakes, enjoying reading) loaded nicely on one factor with reliability of $\alpha = .60$. Importance of speed was more closely related to items addressing timed ORF measures. Conceptually, we discuss speed importance here with other importance items. Correlations between the seven items ranged from .02-.29. As seen in Table 2, students rated understanding what one reads as most important (M=3.60), significantly more so than all other items (p<.05), and reading fast as least important (M=1.85), significantly less so than all other items (p<.01). Put another way, 94% of students reported the importance of comprehension as *a lot* or *a whole lot* versus 24% of students feeling reading fast is important. In general, students rated the importance of re-reading to understand (M=3.46), enjoying reading (M=3.41), remembering what one reads (M=3.32), and figuring out all of the words (M=3.27) as *a lot* to *a whole lot*. They felt it was *a*

lot important to read without mistakes (M=2.88). Students below grade level rated *reading fast* significantly more important than those on- or above-grade level (p<.01), although rated it as *a little bit* important (M=2.03). Students above grade level rated comprehension (p<.05) and enjoying reading (p<.01) as significantly more important than those below grade level. District B students rated remembering what is read, reread to understand, and reading without mistakes as significantly more important than students in District A (p<.01). Teachers agreed that students think comprehension (M=3.39) is more important than speed (M=2.83) (p<.01), although teachers rated their students valuing speed higher than did students themselves.

 Within student definitions of a good reader, we identified 34 aspects of readers and reading [e.g. *understands what he reads, practicing, flouece* (sic)], and collapsed these into seven themes: comprehension, accuracy, prosody/expression, practice, effort, adjusting rate, and speed. In general, students most often mentioned comprehension in their definitions (31.6%). Students in District A in all grades included comprehension in their definition more frequently than District B students. Students mentioned speed least often (8%), and less than one percent of students defined a good reader as solely one who reads fast.

Teaching to the Test

We ran separate exploratory factor analyses on student and teacher instructional practice items. Of ten student survey items, all except *timed reading tasks* formed one factor (α = .65) (see Appendix B). (Again, timed tasks were more closely related to other items addressing ORF measures). Student instructional practice item correlations ranged from .08-.33. The 12 items on the teachers survey formed two distinct factors: non-computer instruction (α = .77), and computer-based instruction (α = .68) (see Appendix C). Teacher instructional practice item correlations ranged from .02-.65.

As seen in Table 2, students reported taking a reading test as the most frequent instructional practice listed (M=2.96, *a lot*), significantly more frequently than all other practices (p<.01). In terms of percentages, 69% of students reported they take reading tests *a lot* or *a whole lot*. Students reported reading non-words as the least frequent practice (M=1.51, *not very much* to *a little bit*). They reported a moderate amount (*a little bit* to *a lot*) of reading words in isolation (M=2.77) and re-reading (M=2.46), and *a little bit* of buddy reading (M=2.22), choral reading (M=2.17), reading along with an audio recording (M=2.06), reading on a computer (M=2.06), timed reading tasks (M=2.0), and readers theatre (M=1.91). MANOVA results showed significant differences between districts, with students in District B reporting greater frequency of all instructional practices than students in District A. There were significant differences in students' responses by reading ability with students reading below grade level reporting significantly more of all practices with the exception of taking a reading test, reading isolated words, and buddy reading.

Teachers reported untimed re-reading as the most frequent of the fluency practices listed (M=3.01, *weekly*), followed by partner reading (M=2.68) (see Table 3). Re-reading and partner reading occurred significantly more frequently than all other practices (p<.01). Teachers *occasionally* engaged students in untimed word reading (M=2.07), timed re-reading (M=2.0), and readers' theater (M=1.93), but overall engaged in other fluency instructional practices (timed word reading, timed or untimed computer word reading, timed or untimed non-word reading, computerized fluency programs) *never* to *occasionally* (range = 1.18-1.61). Teachers in District A reported more computer fluency programs (p<.01) and more partner reading (p<.05) than those in District B. Specialists reported engaging significantly more frequently than classroom teachers in untimed word reading, choral reading, untimed computer word reading, and untimed non-word reading, as well as timed rereading and timed word lists.

In terms of student reader self-perception, seven items clustered together with a reliability of α = .76 (see Appendix B). Item correlations ranged from .14-.52. Students overall agreed *a lot* that they are good readers (M=3.30) and that they like to read (M=3.11) (Table 2). There were significant differences between groups, with District A students reporting a higher view of themselves as good readers (f=5.52, p<.05), and students reading above grade level expressing more positive perceptions than those below grade level regarding liking to read (f=10.42, p<.01) and self-perception (f=59.13, p<.01).

With respect to student perceptions of taking ORF measures, six items clustered together with a reliability of α = .62. Item correlations ranged from .04-.35. Overall students reported taking ORF tests *a little bit* to *a lot* (M= 2.41), with students in District B reporting significantly higher frequency (p<.01) (Table 2). Forty-six percent of students in District B (compared with 34% in District A) reported taking ORF tests *a lot* or *a whole lot*. This would be expected, given that progress monitoring is suggested more frequently for students below proficiency, and District B had a high percentage of students below proficiency. In general, students agreed that they can

pay attention (M=3.38) and comprehend (M=2.87) when taking a timed reading test, although students above grade level reported better ability to both attend (M=3.49, p<.05) and comprehend (M=3.09, p<.01) than those at or below grade level. Students reported that they like timed tests *a little bit* to *a lot* (M=2.56). District B students expressed more favorability toward timed tests (p<.05), with 52.5% reporting they like to take these tests *a lot* to *a whole lot*. Students in general reported that they get nervous *a little bit* (M=2.18). However, students in District B reported higher anxiety (p<.01); 41% of students in District B (vs. 14% in District A) reported *a lot* to *a whole lot* of anxiety. There were significant reading ability differences, with students below grade level attesting to higher anxiety (M=2.49, p<.05) than those on-grade level, who expressed higher anxiety than those above grade level (M=2.18, p<.05). Nearly half (49%) of students below grade level expressed *a lot* to *a whole lot* of anxiety. Teachers agreed that students feel pressure to read fast (M=2.91) (Table 4). There were no significant differences between districts or teacher role.

Use of ORF measures

Eight items designed to capture teacher perceptions and use of ORF measures clustered together into one factor with a reliability of $\alpha = .87$ (see Appendix C). Item correlations ranged from .23-.70. Teacher reports (see Table 4) were somewhat neutral for changes in instructional practices (M=2.35) since using ORF measures, though teachers somewhat agreed they have increased the amount of fluency instruction (M=2.71). Teachers agreed that fluency is important (M=3.38), somewhat agreed that they are more aware of fluency since using the ORF assessments (M=2.85), that the ORF measure provides important information regarding fluency (M=2.76), and that teachers use ORF results to group students for instruction (M=2.65). However, teachers disagreed that they would choose (if given a choice) to use the ORF measure (M=2.32) or that the ORF measure accurately reflects their students' reading ability (M=2.03). Although specialists who work with students who struggle reported higher agreement on all items except the importance of fluency and ORF measures and their awareness of fluency, differences were not significant, most likely due to the small sample size.

In open-ended questions, teachers expressed that ORF measures provide rate/accuracy figures (33%), are helpful to compare to grade level expectations (13%), allow teachers to see growth/lack of (12%), and help teachers understand student reading levels (6%). Some expressed that they use ORF diagnostically to understand errors, strategies, etc. (21%), while others (19%) reported ORF measures are "not at all" helpful. Asked specifically how they use ORF results, 8% of District A and 27% of District B teachers wrote they "do not" use results for any purpose. Of teachers who do use results, the majority (58% District A, 41% District B) reported grouping students for instruction. Some teachers reported using results to target instructional practice for those students not meeting benchmark. Most teachers described additionally using informal and commercial means of assessing fluency, particularly stressing students reading at their instructional reading level. When asked their opinion of ORF measures, 44% of teachers provided negative, 38% mixed, and 19% positive opinions. Of positive opinions (either solely positive or mixed) most often cited were ORF as a baseline or measure of growth (31%), providing fluency data (16%), and usefulness for grouping (11%). Of those citing negative aspects of ORF measures, most indicated that ORF measures under represent reading/fluency or that other measures include more of what is important in reading (75%). Teachers also discussed perceived effects of ORF assessments on students—that assessments generate anxiety, contradict what is taught (e.g. self-correction, self-monitoring), and teach students to value speed (27%).

With respect to progress monitoring, we noted differences by district in the level of text used. District B teachers (92%) used grade level materials. District A teachers were more variable (27% grade level, 33% instructional level, and 40% both grade and instructional level passages). We also noted differences in frequency of progress monitoring. District A teachers monitored weekly with students needing "intensive intervention," and biweekly with those needing "some support." District B teachers' progress-monitoring frequency varied by school for students needing intervention. One school monitored weekly, another monthly, and another reported only using the universal screening benchmarks (fall, winter, spring).

Discussion

We undertook this study to gain perspective on teachers' and students' views related to potential negative consequences of using ORF measures. Specifically, we wanted to know whether we would find narrow definitions of fluency and reading, teaching to the test and related issues of student anxiety and reading self-perception, and misuse of ORF measures and results.

Definitions

Arguing that test construction and use may ultimately define (or redefine) the construct being measured, critics expressed concerns that teachers using ORF measures may come to define fluency as what is measured by these

assessments—accuracy and rate (Samuels, 2007). Our results do not support this for the majority of teachers, as most generated broad definitions of fluency, and described fluency as important to comprehension. The variety of ways they described fluency in general, and prosody, in particular, attests to the challenges of capturing the complexity of ORF (Kuhn et al., 2010). However, our results do lend some limited support for critics' concern as approximately one-quarter of teachers in this sample did define fluency as accuracy and rate, with several specifically mentioning WCPM scores. Concerned with the narrowness of ORF measures, critics also expressed fears that students who experience ORF measures, which value reading quickly, may come to define/ re-define reading as speed (Goodman, 2006). Although teachers in this sample agreed with this concern, our student data do not support it. Students in our study neither defined reading as speed, nor reported valuing speed as important. They described a "good reader" as one who reads, and/or one who understands what one reads.

Teaching to the Test
Critics expressed concerns that teachers, hoping to help students become more successful on ORF measures, will incorporate into instruction practice reading texts quickly and accurately (Kuhn et al., 2010). Results of this study do not support this assertion. Teachers in our sample did not report that they have changed their instruction since using ORF measures, and neither teachers nor students reported frequent practice with timed reading. Both did report relatively frequent repeated and partner reading, which are recommended as effective practices for building connected text fluency (Rasinski, 2010). Otherwise, neither teachers nor students reported frequent use of any of the fluency-related practices included in our survey. Critics' assertions of students valuing speed were in relation to their concerns that speed would be a focus of instruction. In our study, outside of the ORF test process, it seems that generally students did not engage in timed tasks, which may relate to students maintaining a focus on comprehension over speed.

Associated with teaching to the test, critics warn that students may experience anxiety and low reading self-perception. Our student data, *in general*, do not support either high anxiety over timed reading tests, or low reading self-perception. The relatively high ratings of self-perception correspond with other research suggesting that younger students have high beliefs regarding their academic ability (Archambault et al., 2010).

Use of ORF Measures
Our results regarding how teachers use ORF measures and data are equivocal. Guidelines of DIBELS and *aimsweb* (Dynamic Measurement Group, 2012; NCS Pearson Inc., 2012; Shapiro, 2008) suggest that ORF data can be used to group students for instruction. Yet, they also note that multiple sources should guide instructional decisions. Critics agree, warning against making instructional decisions based on limited data, both because ORF measures cannot suggest what, specifically, students need for instruction (Valencia et al., 2010), and because of the high examiner variability in scoring ORF measures (Cummings, Biancorosa, Scheper, & Reed, 2014). Teachers in our study both agreed that they use ORF results to group students, and reported that ORF results were useful for grouping. However, they also described using other tools and practices (e.g. informal reading inventories, running records) for assessment, and some reported using ORF measures diagnostically to better assist in planning instruction. However, since our survey items did not ask whether teachers used ORF data *in conjunction* with other data, we cannot state whether teachers group students solely based on ORF results. What we can say is that ORF results feature prominently in grouping decisions.

Frequent monitoring is also recommended by test developers (Dynamic Measurement Group, 2012; NCS Pearson Inc., 2012; Shapiro, 2008), and most teachers in our sample reported that they monitor progress frequently for students below grade level. Problematic was the level of text used. The DIBELS manual directs that instructional level text be used for progress monitoring, yet, the vast majority of teachers who used DIBELS in our sample monitored student progress using grade level text. *aimsweb* recommends a process to determine appropriate text level; in general, students with a severe discrepancy (two years) between grade placement and reading level would receive monitoring at their reading level. Teaches in our sample who use *aimsweb* did vary with respect to text level used for progress monitoring, which may indicate they choose text based on students' reading ability. However, our data does prompt us to consider that confusion over progress monitoring may exist with respect to text level. Results of grade level tests, particularly for students below grade level, will not enable teachers to gain an accurate gauge of progress to assess the effectiveness of instruction (Shannahan, 2009), or help teachers choose an appropriate level of text for group instruction. Not only are the passages beyond students' reading ability, teachers may have difficulty judging student reading level based on DIBELS scores (Begeny, Krouse, Brown, & Mann, 2011), particularly when DIBELS ORF passages have been shown more difficult than other ORF measures (Duesbury, Braun-Monegan, Werblow, &

Braun, 2012). Conversely, when discussing other ways they assess fluency, teachers stressed using texts at the students' instructional reading level. Therefore, it is unclear from our data whether teachers, particularly in District B, choose to use grade level passages for progress monitoring, or perhaps were directed to do so as school/district policy. Likewise, given that that teachers used ORF measures for grouping and monitoring progress, but did not report providing many fluency-related instructional practices, we question whether grouping and progress monitoring decisions may be externally controlled, whereas, instruction may be within teacher control.

Teachers' opinions of ORF measures paint a complicated and somewhat contradictory picture. CBMs, designed for teachers to calibrate instruction to meet student needs, display good criterion validity (Roehrig et al., 2008); yet, our teachers disagreed that ORF results accurately reflect students' *reading* ability. Teachers agreed that ORF measures provide important information, and that they have become more aware of fluency since using ORF measures; yet, they disagreed that they would choose ORF measures given a choice. Our qualitative data may explain this contradiction. In open-ended responses, many teachers reported that ORF measures do not add value to their understanding of students' skills, and that they could gain more information from using longer texts and a "better" measure of comprehension. Their responses seem to indicate that ORF measures, although valuable, are too narrow. Nearly all reported that, in addition to ORF measures, they use broader measures, suggesting they recognize the need to assess fluency, but view fluency as one piece of what they seek to understand about their students. Additionally, although the majority of teachers agreed that ORF measures provide important information, when asked their opinion of the measures, the majority provided negative comments. Teacher beliefs relate to practices (Richardson & Placier, 2001). Given the low opinion many teachers in our sample expressed toward ORF measures, it is not surprising that approximately one quarter reported not using ORF results for any purpose. Although we investigated consequential validity, our teachers' perceptions of ORF measures raise questions of face validity—subjective opinions of what a measure does/does not do (Baker, 2013). Although teachers agreed they gain information from ORF measures, they did not feel the measures adequately captured students' *reading* ability. Their low opinion of ORF usefulness then may be tied more to what they believe the measures do not provide them than to validity/reliability statistics (c.f. Yell, Deno, & Marston, 1992).

Group Differences

Although our data do not, in general support that teachers teach to ORF tests or that students have high anxiety and low self-perception, there were differences by district and reading ability, which seem related. First, students in District B reported more frequent engagement in fluency-related instructional practices. Given that reading proficiency in District B is low, many of these students might need instruction to master essential fluency-related components pertinent to reading comprehension. Specialists, who work with students who struggle, also reported using more fluency-related practices, including timed practice, confirming the reports of students below-grade level. Therefore, our data do suggest differential instruction provided to students who struggle, including more timed reading practice. We cannot ascertain whether this differential instruction represents a positive consequence of ORF measures—that results are used appropriately to inform instruction—or negative—that teachers are trying to raise WCPM scores. Second, students in District B reported higher anxiety and lower self-perception. Students below grade level also reported significantly higher anxiety and significantly lower self-perception and reading enjoyment. Therefore, again, district level differences may be due to the higher number of students reading below grade level in District B. These differences prompt us to question how heightened anxiety may relate to test performance. Although we find no research showing a relationship between anxiety and ORF performance, we do find research showing relationships between anxiety and higher-level cognitive tasks, such as comprehension (Bonifacci, Candria, & Contento, 2008; Tysinger, Tysinger, Diamanduros, & Early, 2010). In our sample, students who struggle rated their ability to attend and comprehend during timed reading tasks as significantly lower than students above grade level. Therefore, it is possible that students who struggle may be at a disadvantage in adequately displaying their reading skills, particularly comprehension (if assessed) during ORF measures. An additional caution related to students who struggle is self-perception. Although students below grade level in our sample reported *a little bit* to *a lot* that they are good readers and liked to read, studies have shown that elementary students begin schooling with high motivational beliefs about the usefulness of reading and their own reading self-efficacy, but those who experience academic difficulties are likely to exhibit downward trajectories in these perceptions over the course of their school careers (Archambault et al., 2010), particularly in adolescence (Guthrie & Davis, 2003; Johnston & Winograd, 1985). Our data suggest our students who struggle already report lower self-perception,

enjoyment, and comprehension than those on or above grade level. This makes our finding of text level choice for progress monitoring more salient. We need to be mindful of the possibility that using grade level passages, out of reach for students who struggle, may hasten the downward trajectory, which, in turn may erode classroom learning (Ivey & Guthrie, 2008).

Limitations and Future Research

Our study is based on a sample of 750 students and 60 teachers in one large diverse urban district, and 125 students and 17 teachers in one small suburban district, so cannot be generalized to teachers/students in other districts. Our small sample of teachers, in particular, makes it difficult for us to determine differences between teacher perceptions and practices across roles and districts. Additionally, our data are self-reported, which may raise issues of response bias. However, the large number of students in our sample, and the degree of student/teacher agreement does offer some level of confidence in our findings. Our survey design also poses limitations to results interpretation. Although we included instructional practices related developmentally to ORF, and questions about teachers' perceptions and use of ORF measures, the items were not exhaustive. It is probable that teachers implement practices not assessed on our survey and use ORF measures in ways not apparent to us.

Although this study begins to address issues of unintended consequences of ORF measures, more research is needed to understand the nuances of how teachers use ORF measures and results. Our data show that teachers use ORF data to group students, but further research is needed to understand how ORF results, and perhaps other data, factor into grouping decisions, and how they are used to plan instruction and choose texts for those groups.

Implications for Practice

Our results suggest implications for practice that relate specifically to the two districts investigated, and may relate to other districts that use ORF CBM measures. First is the need for professional development for progress monitoring, grouping, and using results. Teachers, and perhaps administrators, need assistance in keeping with ORF test developers' recommendations for progress monitoring. The purpose of progress monitoring is to show growth and assess the effectiveness of intervention. To do this, particularly with students who are significantly below grade level, teachers need to use texts at the students' instructional reading level, rather than texts at students' grade level. Additionally, in schools where student reading proficiency is low, many students will be identified as at risk on ORF measures, complicating grouping decisions. It is important that districts/schools ensure that teachers have the expertise to use multiple data points (Ardoin, Christ, Morena, Cormier, & Klingbeil, 2013) to not only group students, but to determine the instruction provided within each group (Valencia et al., 2010). Although our survey was not designed to determine how instructional decisions are made for students, we can state that teachers engaged students *occasionally* in fluency related practices. ORF CBMs are intended to provide information for teachers to assess student progress and adjust instruction to meet student needs (Deno, 1985). Therefore, we might expect some instruction targeted to results of these measures. Students overwhelmingly reported frequent testing—more frequent than any practice listed. Yet, neither students nor teachers report frequent instructional practices to address fluency at either the single-word level (automaticity) or connected text. It seems that fluency may be assessed, but not addressed. Students who struggle with reading proficiency most likely need instruction to master essential components pertinent to reading comprehension—some of which will include ORF. Those who provide intervention, particularly, may benefit from professional development in identifying/targeting areas of need. With respect to instruction and text choice, despite schools in these two districts having an adopted basal reading program, most teachers relied on small group guided reading instruction using leveled texts. As part of this approach, teachers may be accustomed to assessments that have a more direct relationship between results and recommended text level, and may have difficulty converting ORF results to text choices (Begeny et al., 2011). Therefore, teachers who use ORF measures may need assistance in understanding how to translate results to appropriate level of texts for students.

Second, our data raise concerns for vulnerable students. Although we cannot suggest that the students who struggle in our sample are less invested or engaged in school, we hearken to the plethora of data documenting progressive disengagement for students who struggle (c.f. Johnston & Winograd, 1985). Educators should be mindful of students' self-perception and anxiety, and the role they may play in learning. We should also be mindful of research-based practices that could build both fluency and self-efficacy (Guthrie, McRae, & Klauda, 2007).

Lastly, we are concerned that the time devoted to administering, scoring, and interpreting ORF data may be

wasted. The majority of teachers in this study would not choose to use ORF measures and, qualitatively, many described wanting assessments that provide richer information. Secretary Duncan has recently denounced duplicating effort, acknowledging, "where tests are redundant, or not sufficiently helpful for instruction, they cost precious time that teachers and kids can't afford" (Duncan, 2014). Although ORF measures are used widely, districts and schools may need to examine the use of resulting data to determine, and possibly increase, the efficacy of these measures' use for administrative and instructional needs.

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Archambault, I., Eccles, E., & Vida, M. (2010). Ability self-concepts and subjective value in literacy: Joint trajectories from grades 1-12. *Journal of Educational Psychology, 102*(4), 804-816.

Ardoin, S. P., Christ, T. J., Morena, L. S., Cormier, D. C., & Klingbeil, D. A. (2013). A systematic review and summarization of the recommendations and research surrounding Curriculum-Based Measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology, 51*, 1–18.

Baker, E. L. (2013). The chimera of validity. *Teachers College Record, 115*(9).

Balanos, D., Cole, R., Ward, W., Hasbrouck, J,, Tindal, G., & Schwanenflugel, P. (2013). Human and automated assessment of oral reading fluency. *Journal of Educational Psychology, 105*(4), 1142-1151.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*(2), 122-147.

Begeny, J., Krouse, H., Brown, K., & Mann, C. (2011). Teacher judgments of students' reading abilities across a continuum of rating methods and achievement measures. *School Psychology Review, 40*(1), 23-38.

Benjamin, R., Schwanenflugel, P., Meissinger, E., Groff, C., Kuhn, M., & Steiner, L. (2013). A spectrographically grounded scale for evaluating reading expressiveness. Reading Research Quarterly, 48(2), 105–133.

Bonifacci, P., Candria, L., & Contento, S. (2008). Reading and writing: What is the relationship with anxiety and depression? *Reading and Writing: An Interdisciplinary Journal, 21*(6), 609-625.

Buck, J. & Torgesen, J. (2003). *The relationship between performance on a measure of oral reading fluency and performance on the Florida Comprehensive Assessment Test*. Tallassee, FL: Florida Center for Reading Research.

Campbell, D. (1976). Assessing the impact of planned social change. Paper #8, The Public Affairs Center, Dartmouth College.

Chard, D. J., Vaughn, S., & Tyler, B. J. (2002). A synthesis of research on effective interventions for building reading fluency with elementary students with learning disabilities. *Journal of Learning Disabilities*, *35*, 386-407.

Cummings, K., Biancarosa, G., Shaper, A., & Reed, D. (2014). Examiner error in curriculum-based measurement of oral reading. *Journal of School Psychology, 52*, 361-375.

Daane, M.C., Campbell, J.R., Grigg, W.S., Goodman, M.J., & Oranje, A. (2005). *Fourth-grade students reading aloud: NAEP 2002 special study of oral reading* (NCES 2006-469). U.S. DOE, IES, NCES. Washington, DC: Government Printing Office.

Deno, S. (1985). Curriculum based measurement: The emerging alternative. *Exceptional Children, 52*, 219-232.

Duesbury, L., Braun-Monegan, J., Werblow, J., & Braun, D. (2012). Practical issues in field based testing of oral reading fluency at upper elementary grades. *Studies in Educational Evaluation, 38*, 121-126.

Duncan, A. (2014). A back-to-school conversation with teachers and school leaders. Retrieved 3/12/15 from
        http://www.ed.gov/blog/2014/08/a-back-to-school-conversation-with-teachers-and-school-leaders/

Dynamic Measurement Group (2012). *Progress monitoring with DIBELS Next*. University of Oregon Center for Learning and Teaching.

Edformation (2002). *AIMSweb Standard Reading Assessment Passages*. Eden Prairie, MN: Edformation.

Fuchs, L.S., Fuchs, D., Hosp, M.D., & Jenkins, J.R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*(3), 239–256.

Goffreda, C., & DiPerna, J. (2010). An empirical review of psychometric evidence for the   Dynamic Indicators of Basic Early Literacy Skills. *School Psychology Review, 39*(3),   463-483.

Good, R., Kaminski, R. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.

Goodman, K. (2006). A critical review of DIBELS. In K. S. Goodman (Ed.), *The truth about   DIBELS: What it is, what it does*. Portsmouth, NH: Heinemann.

Grimm, L. G., & Yarnold, P. R. (2000). Introduction to multivariate Ssatistics. In L. Grimm & P. Yarnold (Eds.) *Reading and understanding more multivariate statistics* (pp. 3-21).   Washington DC: American Psychological Association.

Guthrie, J., & Davis, M. (2003). Motivating struggling readers in middle school through an   engagement model of classroom practice. *Reading and Writing Quarterly, 19*, 59-85.

Guthrie, J.T., McRae, A. & Klauda, S.L. (2007). Contributions of Concept Oriented Reading   Instruction to knowledge about interventions for motivations in reading. *Educational   Psychologist, 42*, 237-250.

Harris, T. L., & Hodges, R. E. (1995) *The literacy dictionary: The vocabulary of   reading and writing*. Newark, DE: International Reading Association.

Hosp, J. & Suchey, N. (2014). Reading assessment: Reading fluency, reading fluently, and comprehension. School Psychology Review, 43(1), 59-68.

Ivey, S., & Guthrie, J. (2008). Struggling readers: Boosting motivation in low achievers. In J.   Guthrie (Ed.), *Engaging adolescents in reading* (pp. 115–129). Thousand Oaks, CA: Corwin Press.

Johnson, R. A. & Wichern, D. W. (1998). *Applied multivariate statistical analysis* (2nd ed.).   Englewood Cliffs, NJ: Prentice Hall.

Johnston, P. & Winograd, P. (1985). Passive failure in reading. *Journal of Reading Behavior,   17*(4), 279-301.

Jones, G., Jones, B., Hardin, B., Chapman, L., Yarbrough, T., & Davis, M. (1999). The impact   of high-stakes testing on teachers and students in North Carolina. *Phi Delta Kappan*, 81, 199-203.

Kame'enui, E.J. & Simmons, D.C. (2001). Introduction to this special issue: The DNA of   reading fluency. *Scientific Studies of Reading, 5,* 203-210.

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice, 30*(4), 28-37.

Kuhn, M., Schwanenflugel, P. J., & Meisinger, E. B. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and the definitions of fluency. *Reading Research   Quarterly, 45*, 230-251.

Messick, S. (1996). Validity and washback in language testing. *Language Testing, 12*(3), 241-56.

Meisinger, E. B., Bradley, B. A., Schwanenflugel, P., & Kuhn, M. (2010). Teachers' perceptions of word callers and related literacy skills. *School Psychology Review, 39*(1), 54-68.

Miles, M.B. and Huberman, A.M. (1994). *Qualitative data analysis: An expanded sourcebook*   (2nd ed.). Newbury Park, CA: Sage Publications.

Miller, J., & Schwanenflugel, P. (2008). A longitudinal study of the development of reading   prosody as a dimension of oral reading fluency in early elementary school children.   *Reading Research Quarterly, 43*, 336-354.

NCS Pearson, Inc. (2014). *aimsweb Reading*. New York: Pearson.

NCS Pearson, Inc. (2012). *aimsweb Progress monitoring guide*. Bloomington, MN: Pearson.

Newman-Thomas, C. N., Smith, C. A., Zhao, X., Kethley, C. I., Rieth, H. J., Swanson, E. A., &   Heo, Y. (2012). The efficacy of technology to teach oral reading fluency assessment to   preservice teachers. *Journal of Special Education Technology, 27*(1), 15-32.

Poncy, B. C., Skinner, C. H., & Axtell, P. K. (2005). An investigation of the reliability and   standard error of measurement of words read correctly per minute using curriculum-   based measurement. *Journal of Psychoeducational Assessment, 23*, 326–338

Rasinski, T. (2010). *The fluent reader* (2nd ed.). New York: Scholastic.

Richardson, V., & Placier, P. (2001). Teacher change. In V. Richardson (Ed.), *Handbook of   research on teaching* (4th ed., pp. 905-947). Washington, DC: AERA.

Roehrig, A., Petscher, Y., Nettles, S., Hudson, R., & Torgesen, J. (2008). Not just speed reading: Accuracy of the DIBELS oral reading fluency measure for predicting high-stakes third grade reading comprehension outcomes. *Journal of School Psychology, 46*, 343–366.

Samuels, S. J. (2006). Toward a model of reading fluency. In S. J. Samuels & A. E. Farstrup (Eds.), *What research has to say about fluency instruction* (24-46). Newark, DE: IRA.

Samuels, S. J. (2007). The DIBELS tests: Is speed of barking at print what we really mean by reading fluency? *Reading Research Quarterly, 42*, 563-566.

Shanahan, T. (2009, Aug. 20). Yes, Virginia, you can DIBEL too much! [Web log post]. Retrieved from http://www.shanahanonliteracy.com/search/label/DIBELS

Shapiro, E. S. (2008). Best practices in setting progress monitoring goals for academic skill improvement. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology*, 5th Ed. (pp. 141-158). Bethesda, MD: National Association of School Psychologists.

Shephard, L. A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice, 16*(2), 5-8.

Tierney, R. J., & Thome, C. (2006). Is DIBELS leading us down the wrong path? In K. S. Goodman (Ed.), *The truth about DIBELS: What it is, what it does* (pp. 50-59). Portsmouth, NH: Heinemann.

Tysinger, J., Tysinger, P. D., Diamanduros, T., & Early, A. (2010). The effect of anxiety on the measure of reading fluency and comprehension. *Georgia Educational Researcher, Online Edition, 8*(1).

Valencia, S., Smith, A., Reece, A., Li, M., Wixson, K., & Newman, H. (2010). Oral reading fluency assessment: Issues of construct, criterion, and consequential validity. *Reading Research Quarterly, 45*(3), 270-291.

Vesay, J. P., & Gischlar, K. L. (2013). The big 5: Teacher knowledge and skill acquisition in early literacy. *Reading Horizons, 52*, 281-303

Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Basingstoke: Palgrave Macmillan.

Welner, K. G. (2013). Consequential validity and the transformation of tests from measurement tools to policy tools. *Teachers College Record, 115*(9).

Yell, M., Deno, S. L., & Marston, D. (1992). Barriers to implementing curriculum-based measurement. *Diagnostique*, 18(1), 99.