

A COMPARATIVE EVALUATION OF INSTRUCTIONAL LEVELS DETERMINED
BY THE TEXT READING AND COMPREHENSION (TRC) ASSESSMENT AND AN
INFORMAL READING INVENTORY

A Dissertation
by
AMIE BROCK SNOW

Submitted to the Graduate School
at Appalachian State University
in partial fulfillment of the requirements for the degree of
DOCTOR OF EDUCATION

December 2014
Educational Leadership Doctoral Program
Reich College of Education

A COMPARATIVE EVALUATION OF INSTRUCTIONAL LEVELS DETERMINED
BY THE TEXT READING AND COMPREHENSION (TRC) ASSESSMENT AND AN
INFORMAL READING INVENTORY

A Dissertation
by
AMIE BROCK SNOW
December 2014

APPROVED BY:

Darrell Morris, Ed.D.
Chairperson, Dissertation Committee

Woodrow Trathen, Ph.D.
Member, Dissertation Committee

Devery Ward, Ed.D.
Member, Dissertation Committee

Vachel Miller, Ed.D.
Director, Educational Leadership Doctoral Program

Max C. Poole, Ph.D.
Dean, Cratis Williams Graduate School

Copyright by Amie Brock Snow 2014
All Rights Reserved

Abstract

A COMPARATIVE EVALUATION OF INSTRUCTIONAL LEVELS BY THE TEXT READING AND COMPREHENSION (TRC) ASSESSMENT AND AN INFORMAL READING INVENTORY

Amie Brock Snow
B.A., University of North Carolina-Chapel Hill
M.A., Appalachian State University
Ed.D., Appalachian State University

Dissertation Committee Chairperson: Darrell Morris

This study compared the administration of, and results yielded by, two different contextual reading instruments (an informal reading inventory [IRI] and the Text Reading and Comprehension inventory [TRC]). The purpose was to determine which instrument would provide the most valid assessment of children's reading ability. The study targeted the TRC because it recently has become the required reading assessment for primary-grade students (K – 3) in the state of North Carolina. IRI assessments have a long tradition of use for assessing children's reading levels.

A trained team of teachers administered both reading assessments to 196 first-, second-, and third-grade students. A third informal test (word recognition-timed) was administered to a randomly-selected one third of the 196 students. Word recognition-timed (WR-t) served as a neutral measure to which student performance on the IRI and TRC could be compared.

Results showed clear differences between the IRI and the TRC in regard to administration time and student reading level yielded. The TRC assessments took nearly three times as long to administer as the IRI assessments. The student reading levels yielded by the respective assessments were significantly different, with the IRI levels higher and more often in agreement with the neutral measure, WR-t. The data revealed that the TRC systematically underestimated students' reading instructional level, in most cases because testing was terminated prematurely, owing to poor student performance on written comprehension questions.

That the state-mandated TRC took an inordinate amount of time to administer (approximately 1 hour) and tended to underestimate children's reading ability are important findings and cause for concern. Several suggestions for improving the assessment are offered, including (a) eliminating written comprehension questions, (b) administering fewer reading passages, and (c) including reading rate as a factor in determining a student's instructional level.

Acknowledgments

This dissertation would not have been possible without the guidance and support of my dissertation chair and my mentor, Dr. Darrell Morris. His willingness to share his knowledge, listen to my struggles and motivate me to complete my work has helped me grow into the teacher and administrator that I am today. My ability to advocate for our students at Prince Ibrahim Elementary School is because of the time that I have spent learning from him at Appalachian State University.

Along with Dr. Morris, I am extremely thankful for the support and encouragement that Dr. Woodrow Trathen and Dr. Devery Ward have given me throughout the years I have worked with them. They helped me find the courage to share what I've learned with others and help create reading programs that support teachers and students. I would not be the professional that I have become today without them.

Collecting data for this study would have been impossible without the students and teachers at Prince Ibrahim Elementary School. Our principal, Lee Koch, completely supported my research and gave me the time and support I needed to collect the data. If it were not for Mattie Brock, Kathie Byrom, Pam Hicks, John MacMillan, Lagray Marks, Everett Martin, Jennifer Miller, Megan Pardick, and Glenda Webster, I would have never gotten each student assessed. The hours of work you have given for me can never be repaid, but I hope you know that I'm incredibly thankful for all you did for me. Our students were dedicated throughout all of our assessments, and their sweet voices as they

read to us will always be something that stays with me. The classroom teachers did so much to support this process by allowing students to work with me and the assessment team on multiple occasions, and more importantly by being my cheerleaders as I conquered this long journey. I can never thank all of you enough for the trust you have had in me as your Curriculum Coordinator and the patience you have had when I've been overwhelmed and exhausted. This dissertation would not have been possible without all of you.

I must end by thanking my family for the love and patience they have shown throughout this process. My parents, Jim Brock and Dot Crumley, have always been there to motivate me to be all that I could be, and I'm proud that they can see me achieve my biggest professional dream. Lastly, I want to thank my husband and my best friend, Scott Snow. I can honestly say that I could not have made it this far without him. It would be hard to find a husband of a teacher who knows as much about reading assessment and correction as Scott Snow. He listens to me even when he's heard it all before and continues to support me unconditionally. Thank you for all that you are and all that you do for me. I'm so glad that you and our Emma will get to celebrate this achievement with me.

Dedication

This dissertation is dedicated to my teacher, colleague and friend, Dr. Mary Proctor Hendrix, whose own dissertation was published in 2013. Mary's work, "The Relationship of Prosodic Reading to Reading Rate and Other Constructs of Reading Ability" (2013), is an incredibly powerful piece on the importance of reading rate in reading assessment and instruction. Her study strongly supports the use of reading rate in determining instructional reading levels. Her dissertation will have a lasting impact in the work that I do with teachers and students.

Mary was my first instructor when I became a student at ASU. She saw something special in me and encouraged me to work hard to learn and grow as a teacher and student. I've tried to continue to bring her energy and spirit to the work I do. I will always miss our friendship, but I know that no matter where I go, Mary is always with me. She's with me when I open a new book. She's with me when I teach children to read, and she's with me as I take on new opportunities that continue our dream of inspiring teachers to be amazing teachers of reading. She's never really left any of us. Just like one of her favorite characters, Great-Aunt Arizona:

She's always there, in a sunny room with many flowers in every window, and a hug for me every day. She never did go to the faraway places she taught us about. But my great-aunt Arizona travels with me and with those of us whose lives she touched. She goes with us in our minds. (Houston, 1997, p. 28).

Mary passed away in 2013. She was my real life Great-Aunt Arizona in so many ways. She goes with me in my mind and in my heart every single day. I may not get to talk to her like I used to, but whenever I need her it's amazing how I always sense her presence. Her energy and love for teaching and reading will continue to inspire me.

Table of Contents

Abstract	iv
Acknowledgments	vi
Dedication	viii
Chapter One: Introduction	1
Chapter Two: Literature Review	6
Chapter Three: Method	35
Chapter Four: Results	46
Chapter Five: Discussion and Implications	56
References	68
Appendix A	76
Appendix B	78
Appendix C	79
Vita	81

Chapter One: Introduction

As I look back on my educational career of almost 12 years, I believe that the 2012–2013 school year ushered in the most change. We began the year with a brand new curriculum filled with new standards, new terminology, and new expectations for our teachers and our students. That school year also brought new assessments and new pressures for teachers as they became the object of more intense scrutiny by governmental agencies at all levels—federal, state, and local. A business mentality began to predominate, with the goal of improving educational results while economizing on spending.

As curriculum coordinator in an urban, elementary school (K – 5) in central North Carolina, my job is to assist classroom teachers with assessment and instruction. The past two years I have had to attend many state- and district-sponsored meetings, where I am asked to take back (and explain) to an already overburdened teaching staff even more regulations and requirements. I have done my job. I have helped implement the new assessment and accountability procedures in my school while, at the same time, urging my teachers to stay focused on what matters most—the children’s learning. However, there are days when we become disoriented in the jumbled piles of documents, data files, and ever-changing accountability requirements.

As an educator, I am convinced that we must meet the challenges of a new curriculum and new assessment measures with energy, commonsense, and courage. We must continue to question and reflect on what is or is not working for children, and we must expect our administrators to do the same. Any new, mandated curriculum or

assessment should be able to stand up to analysis and criticism, and its proponents, be they educators, businessmen, or politicians, should welcome changes that improve student learning.

Reading Assessment in North Carolina: 2013-2014

The assessment and teaching of reading have long been debated topics. Reading is a subject (or skill) that teachers struggle to teach well when they have classrooms full of children reading at different instructional levels. There have been ongoing debates among researchers and practitioners regarding the “best” ways to assess and teach reading. Commercial programs regularly promise to raise reading achievement levels, and many school districts pay millions of dollars to implement these teacher-proof programs. Yet, many of our children (perhaps 30% on average) leave fourth grade unable to read at grade level (National Center for Education Statistics, 2013). In short, despite hand-wringing, money-making, and, recently, government mandating, teaching all children to read competently is still a major problem for our schools.

The most recent government mandate in North Carolina began during the 2013-2014 school year. The North Carolina Read to Achieve law (Excellent Public Schools Act, House Bill 950, 2012) requires all third graders to pass a standardized reading test at the end of the school year. Those children who fail to pass the reading test risk not being promoted to fourth grade. Such a test-based promotion policy obviously puts new and considerable pressure on students, their parents, and their teachers.

Recognizing the fact that passing a third-grade reading test depends, in large part, on children getting off to a good reading start in the first few years of school, the new state law requires careful monitoring of children’s reading progress throughout the

primary grades (K – 3). The law specifically requires all primary-grade teachers to administer, three times per year (fall, winter, and spring), two formative reading assessments: a skills-based assessment called Dynamic Indicators of Basic Early Literacy Skills or DIBELS (Kaminski & Good, 1996), and a passage-reading assessment called Text Reading Comprehension or TRC (Wireless Generation, 2012). Over the next few years, North Carolina schools will rely heavily on TRC data to determine student reading proficiency and growth. The TRC data will also be used to evaluate teacher performance according to Standard 6 of the state’s new Teacher Evaluation tool. In short, new mandated formative reading assessments, particularly the TRC, will significantly affect North Carolina students and teachers in the coming years.

Evaluating a Reading Test

Given the potential influence of the Text Reading Comprehension (TRC) assessment, it seems important to examine its implementation characteristics and validity. Up to now, there has been little, if any, careful study of the TRC, despite the fact that its use is being mandated state-wide by the North Carolina Department of Public Instruction. In many cases, teachers, who have been relying on other informal reading assessments, are now being prevented from using them. In the present study, I will look at how the TRC made its way into North Carolina classrooms and why so many administrative leaders and politicians support its implementation. The TRC, developed by a company called Wireless Generation (now named Amplify), promises an easily-administered test and accurate results that will directly and positively influence reading instruction in the classroom. The Wireless Generation marketing brochures (n.d.) state that TRC “will

eliminate the labor-intensive process of administering a reading assessment in the traditional paper-and-pencil manner and ensure consistency across classrooms.”

Interestingly, the computer-administered TRC bears a striking resemblance to the traditional paper-and-pencil test that it aims to replace—that is, the informal reading inventory (IRI). Both the TRC and the IRI, which has been around for 60 years, are passage reading assessments that measure oral reading accuracy, rate, and comprehension, with the goal of establishing an optimal reading instructional level for the student (e.g., third grade). The basic formats of the two tests are similar, with, however, some important differences in scoring and interpretation. It seems obvious that a direct comparison of the two assessments is needed. For example, how do the two tests agree or differ in (a) the way reading is measured, (b) the time required for administration and scoring, and (c) the results yielded that may affect student instructional placement.

The purpose of this study is to provide a careful evaluation of the TRC and IRI. By assessing a large number of primary-grade (first through third grade) students with *both* instruments, an empirical comparison of their effectiveness will be made. The study is proactive in nature. That is, it anticipates a major, state-wide change in how reading will be assessed in primary-grade classrooms over the next decade.

It also has the potential to inform North Carolina teachers, administrators and policy makers about the effectiveness of this change. For example:

- Is the TRC accurate in identifying a student’s reading instructional level? An accurate reading assessment guides correct instructional placement, which is particularly important in a period when the state is expecting *all* children to be reading on grade level by the end of third grade.

- Does the TRC take a reasonable amount of time to administer? This question is important given that the test must be administered to all children in a classroom three times per year.
- Does the present validation of the TRC—its ability to accurately measure student reading skill in a reasonable amount of time—justify its potential use as a teacher evaluation instrument?

Results from this study may support the continued use of the TRC in its present form, suggest changes to its administration and scoring that will improve the assessment, or provide reason to discontinue the use of the TRC in North Carolina schools.

Chapter Two: Literature Review

A comparative study of the Text Reading and Comprehension assessment (TRC) and an informal reading inventory (IRI) requires background information on each assessment and a description of their respective procedures for test administration and interpretation of results. I will begin with a discussion of the informal reading inventory and then show how the TRC is really an adaptation of this type of assessment instrument.

Informal Reading Inventory (IRI)

An informal reading inventory (IRI) is an assessment instrument used by teachers to determine a student's reading instructional level. Using several pieces of data (reading accuracy, reading rate, and reading comprehension), an IRI helps teachers place students in appropriately-leveled reading materials. This study uses the IRI from the Appalachian State University Reading Clinic; however, the administration, scoring and interpretation of this IRI is consistent with many other published IRI assessments (e.g., Johns, 2012; Leslie & Caldwell, 2010; Woods & Moe, 2011).

History. In a landmark textbook, Emmett Betts (1946) introduced and popularized an assessment that has come to be known as the informal reading inventory or IRI. The IRI was essentially a series of graded passages (first grade, second grade, third grade, and so on) taken from basal readers of the time. A student was to start with an easy passage and then read successive passages until he or she was eventually frustrated by the difficulty of a given passage. The purpose of the IRI was to identify the reader's optimal instructional level, the grade or difficulty level where he or she was

challenged but not overwhelmed—the level where the student could learn and move forward.

In administering, scoring, and interpreting scores on an IRI, Betts and his colleagues focused on two measures: oral reading accuracy (percentage of passage words read correctly) and comprehension (percentage of passage questions answered correctly). Other educators of the time had emphasized these same measures (see Beldin, 1970). However, Betts’s unique contribution was to demarcate performance criteria for establishing a child’s functional reading levels: independent (too easy); instructional (just right), and frustration (too hard) (see Table 1). Obviously, the advantage to such an assessment was that the results would allow a teacher to place a student in reading material of the appropriate difficulty.

Table 1.

Performance Criteria (Percentage Correct) for Oral Reading Accuracy and Comprehension (Morris et al, 2011)

	Oral Reading Accuracy	Comprehension
Independent level	98%–100%	90%–100%
Instructional level	95%–97%	75%–89%
Frustration level	90% or below	Below 50%

Note: These criteria were based on the research of Killgallon (1942), one of Betts’s students.

Oral reading accuracy. At mid-century, the Betts criteria for establishing functional reading levels were an important contribution. The performance criteria were quickly adopted at the time and are still in wide use today (Barr, Blachowicz, Bates, Katz, & Kaufman, 2007; Lipson & Wixson, 2009). This is not to say that the oral reading accuracy criterion of 95% has not been challenged over the years. For example, Shanahan (1983) stated that the oral reading accuracy criterion of 95% is arbitrary and not based on adequate evidence. Powell (1970) reported that the 95% criterion does not hold across the grade levels. That is, first and second graders, he claimed, were able to maintain comprehension while reading with accuracy rates much lower than 95%.

Perhaps the major challenge to the quantitative standard of 95% accuracy was the “reading miscue” research that came on the scene in the 1970s. Goodman and Burke (1972), for example, argued that a child’s reading level should be determined not by a numerical count of errors, but rather by a careful analysis of the nature or quality of the errors. Unfortunately, such error (or miscue) analysis was tedious and time-consuming. Moreover, several studies showed that as a reader moved from instructional level (95% accuracy) to frustration level (below 90% accuracy), the nature of the oral reading errors changed (see Kibby, 1979; Williamson & Young, 1974). Pikulski and Shanahan (1982) concluded:

Research on the changes that occur in the pattern of oral reading errors or miscues also seems to provide some *added support for the traditional [Betts’] criteria* since in several of the studies, readers began to become inefficient and began reading mechanically, rather than for meaning, as their performance dropped below 95 percent accuracy in word recognition. (p. 106, emphasis added)

Comprehension. Betts's instructional-level criterion for comprehension (75% correct) has been less controversial over the years. This is probably due to the common-sense idea that a child, when reading at a comfortable level, should be able to answer correctly three out of four questions about a passage. Nonetheless, reading comprehension is difficult to assess, especially when the examiner is limited to just a few questions on a relatively short reading passage (150 – 250 words). While acknowledging the worth of Betts's 75% comprehension criterion, Barr et al. (2007) suggested caution in interpreting comprehension scores on an informal reading inventory. They stated:

While the print skills required across a variety of reading situations are fairly stable and determinable by a thorough initial diagnosis, comprehension skills are not. Differences in content, organization, style, and length [of passage] can make dramatically different demands on the reader, *making initial diagnosis of comprehension problems more tentative and qualified*. It is therefore less possible to make a differential diagnosis with respect to comprehension from one or two diagnostic encounters. Rather, ongoing diagnostic instruction manipulating the type and complexity of material used and degree and mode of teacher support can both pinpoint a student's abilities and disabilities and guide the appropriate sequence of instruction. (p. 179, emphasis added)

Reading rate. While oral reading accuracy and comprehension were—and still are in many circles—the key measures in an informal reading assessment, another measure began to receive attention in the 1980s, that is, reading fluency or rate (see Allington, 1983). Reading teachers had long been encouraged to record the rates at which children read assessment passages (Betts, 1946; Kress & Johnson, 1965; Wheat, 1923).

However, reading rate, as an empirical measure, was seldom used in determining a student's functional reading level (e.g., instructional, frustration). It was as if educators knew that reading speed was important but they didn't know what to do with the measure. Practically speaking, there were no widely accepted, grade-level criteria against which a student's reading rate could be compared. In the absence of comparative data, how was one to know if a third-grade child was reading a third-grade text at a fast, average, or slow pace?

The impetus for paying more attention to oral reading rate came from a movement in special education called curriculum-based measurement or CBM. Deno and his colleagues at the University of Minnesota (Deno, 1985; Deno, Marston, Shinn, & Tindal, 1983) found that timed oral reading probes (number of words read correctly in 1 minute) in curriculum-based materials could provide important information about a student's reading growth over time. More recently, CBM researchers have argued that 1-minute oral reading fluency probes are a good indicator of general reading ability, at least during the elementary grades (Deno & Marston, 2006; Fuchs, Fuchs, Hosp, & Jenkins, 2001).

From the CBM movement, there have been recent attempts to establish grade-level norms for oral reading rate. Using 1-minute reading probes, Hasbrouck and Tindal (2006) published oral reading rate norms on a large sample of students in grades 1 – 8. Although their data were obtained under less than optimal conditions (it was unclear what materials [genre and difficulty level] the students were actually reading), Hasbrouck and Tindal's grade-level norms represented an important first step in understanding elementary-school students' reading rates.

Validating craft knowledge. In a recent study, a team of researchers attempted to validate IRI performance criteria that had been handed down from one generation of reading clinicians to the next over a period of 60 years (Morris, Bloodgood, Perney, Frye, Kucan, Trathen, Ward, & Schlagal, 2011; Morris, Trathen, Frye, Kucan, Ward, Schlagal, & Hendrix, 2013). In their longitudinal study, Morris et al. (2011) administered IRIs to 250 children in second through sixth grade. Statistics (means and standard deviations) were reported each year for oral reading accuracy, reading comprehension, and reading rate. Overall, the results tended to support traditional performance criteria in reading diagnosis (see Table 2).

Notice in Table 2 that, on average, oral reading accuracy scores were 95% or 96% at each grade level, second to sixth. In addition, the spread of scores around the grade-level means was small. Thus, when a lower limit (30th percentile) for the instructional-level range was established, second graders still read orally with 93% accuracy, and third-to-sixth graders read orally with 94% accuracy. These results provide support for the traditional Betts' criterion of 95% oral reading accuracy. Regarding comprehension, scores fell within the traditionally-accepted IRI range of 75% to 90% (see Table 2), indicating that the children were reading the passages for meaning. It should be noted, however, that unlike the oral reading accuracy and rate scores, the comprehension scores in the Morris et al. (2011) study were statistically unreliable, across passages and across school years. Others have acknowledged this reliability problem and stressed the importance of being cautious when using IRI comprehension scores to determine a student's instructional level in reading (Barr et al., 2007; Paris & Carpenter, 2003).

Table 2.

*Means and Standard Deviations for Print-Processing and Comprehension Measures
(Grades 2-6)*

Grade	Oral Reading Accuracy (%)	Oral Reading Rate (wpm)	Oral Reading Comprehension (%)
Second	95 (4.8)	107 (38)	87 (17)
Third	96 (3.9)	119 (36)	85 (17)
Fourth	96 (3.6)	127 (34)	82 (18)
Fifth	96 (3.0)	128 (34)	81 (21)
Sixth	96 (2.7)	128 (35)	85 (20)

Note. Adapted from “Validating Craft Knowledge: An Empirical Examination of Elementary-Grade Students’ Performance on an Informal Reading Assessment,” by D. Morris, J. Bloodgood, J. Perney, B. Frye, L. Kucan, W. Trathen, D. Ward, and R. Schlagal, 2011, *The Elementary School Journal*, 112 (2), p. 13. Copyright, 2011 by University of Chicago Press.

Finally, notice in Table 2 that the average oral reading rates in the Morris et al. (2011) study increased steadily from second grade to fourth grade (107 wpm to 127 wpm) before tapering off between fourth and sixth grade (127 wpm to 128 wpm). These oral reading rates were higher than the widely-cited Hasbrouck and Tindal (2006) rates in grades 2 and 3, the same in grade 4, and lower in grades 5 and 6. In considering the concept of an “average grade-level reading rate,” Morris et al. raised an interesting question. At a given grade level (e.g., fourth), how slowly can a child read and still benefit from instruction and practice at that level? With this question in mind, the researchers proposed the following end-of-grade rate minimums:

- grade 2 = 80 wpm;
- grade 3 = 90 wpm;
- grade 4 = 100 wpm;
- grade 5 = 105 wpm; and
- grade 6 = 110 wpm.

At each grade level, these oral reading rate minimums are obviously lower than the average rates shown in Table 2 (they approximate the 25th to 30th percentile in the Morris et al. (2011) data set). Also note that this set of rate minimums is anchored at fourth grade by the 100 wpm criterion, a rate that Pinnell et al. (1995) have argued is necessary to support comprehension of fourth-grade material.

In summary, grade-level results in the Morris et al. (2011) study support traditional IRI criteria in the areas of oral reading accuracy and comprehension. The reading-rate results at each grade level also deserve attention because the rate data was carefully obtained through individual testing of a large number of elementary-grade students.

In this section, I have argued that the informal reading inventory evolved from clinical practice in the mid-20th century and continues to be used today. Traditionally, interpretation of performance on an IRI – that is, how to determine the student’s instructional level – has centered on contextual reading accuracy (95%) and comprehension (75%). Recently, reading rate has begun to receive the attention of researchers (e.g., Morris et al., 2011; Rasinski & Padak, 1998; Torgesen & Hudson, 2006), and many believe that rate, along with accuracy and comprehension, should become an integral part of a comprehensive reading assessment. In the next section, I will

describe the “nuts and bolts” of how to administer an IRI and how to interpret the resulting scores.

Administration of an IRI. An informal reading inventory (IRI) is a series of graded passages (each 100–250 words in length) that is used to assess reading ability. The assessment concept is simple. The examiner starts the student with an easy passage (e.g., a fourth-grade child with a second-grade passage) and then proceeds to administer progressively more challenging passages until the reader becomes frustrated, at which point the testing is stopped. The highest passage level that the student can read without becoming frustrated is designated the “instructional level.” With up to eight oral reading passages (first grade through eighth grade), the examiner needs to know at which level to begin the assessment. A previously-administered graded word recognition test can be helpful here (see Morris, 2014). The idea is that if the child can immediately recognize 80% or more of the words on a given word list (e.g., third grade), he or she should be able to read a third-grade passage with ease. Note that such a decision allows the examiner to skip the first- and second-grade oral passages, thereby saving valuable administration time.

The examiner begins the test by explaining to the child (e.g., Andrew) that his task is to read aloud a few passages and answer some questions. Next, the examiner provides a brief introduction to the first passage (“This story is about a hungry fox.”) and tells Andrew to begin reading. As Andrew, a third grader, begins to read the first passage (second-grade level), the examiner follows along on her copy. She attempts to record the child’s reading errors as he reads, but realizes that she has the tape recorder as a backup. If Andrew pauses on a given word, the examiner allows 3 seconds before providing the

word. When the child finishes reading the passage, the examiner records the number of seconds, closes the reading booklet, and proceeds to ask the comprehension questions.

At this point, the examiner must make a decision; that is, whether or not to let Andrew move forward to the next passage (third grade, in this case). If Andrew experienced difficulty reading the second-grade passage, the examiner would have him go back and try the first-grade oral passage. In effect, after each oral reading passage, the examiner must make a judgment as to whether the child was reading adequately or was frustrated, in which case the oral reading is stopped. Signs of frustration may include an increase in word-reading errors, a decrease in reading rate, or an increase in the amount of assistance the child needs to read the passage.

Scoring and interpreting results from an IRI. Three scores are derived from the child's oral reading of a passage: oral reading accuracy, comprehension, and reading rate. These scores together are used to interpret the reading level of a child.

Oral reading accuracy. In deriving an oral reading accuracy score, the examiner considers five types of errors: substitutions, omissions, insertions, self-corrections, and examiner help (see Figure 1). Repetitions are marked by underlining the repeated word or phrase (the boy was), but they are not counted as errors. Self-corrections *are* counted as errors, because they indicate an initial misreading of a word, thereby affecting reading momentum or fluency (see Barr et al., 2007; Stauffer, Abrams, & Pikulski, 1978). The oral reading score is the percentage of words read accurately in a passage (100% minus the percentage of reading error).

1. Substitutions: Write the substituted or mispronounced word over the word in the text.

saw
(the boy was)

2. Omissions: Circle the omitted word.

(the big boy)

3. Insertions: Use a caret to indicate the inserted word(s).

big
(the red ball)
^

4. Self-corrections: Place a check (✓) next to the marked error to indicate that the child has self-corrected. (A self-correction is usually a substitution error that the child spontaneously corrects.)

saw ✓
(the boy was)

5. Examiner help: Place an “H” above each word that has to be provided by the examiner. The examiner should refrain from providing help unless it is clearly necessary to do so—that is, unless the child refuses to attempt the unknown word or is unsuccessful in decoding it. (Wait 3 seconds before providing help.)

H
(the boy was)

Figure 1. Coding system for oral reading errors in the IRI. Adapted from *Diagnosis and Correction of Reading Problems*, by D. Morris, Copyright 2014 by Guilford Press.

The question now is, how is an oral reading accuracy score (e.g., 95%) to be interpreted? Traditional performance criteria can be of help here (see Barr et al., 2007; Bond & Tinker, 1973; Johnson, Kress, & Pikulski, 1987).

- 98% or above indicates independent level. The child shows sufficient accuracy to read independently, that is, without teacher support.

- 95–97% indicates instructional level. The child shows sufficient accuracy to read successfully with teacher support.
- *Below 90%* indicates frustration level. That is, the child is over-challenged by text at this level of difficulty.

Note that there is a gray or borderline area for oral reading accuracy (90 - 94%). This “gray” area will become significant as I discuss the present study.

Comprehension. The comprehension score is based on the child’s answers to five or six questions about the passage he or she has just read. Answers to each question receive either full, half, or no credit. Performance criteria for comprehension are as follows:

- *90% and above* indicates independent level,
- *75–89%* indicates instructional level, and
- *below 50%* indicates frustration level.

The gray area for comprehension is between 50% and 74%.

Reading rate. Reading rate is the third factor, along with accuracy and comprehension, to be considered in setting a child’s instructional level. In this study, I use the end-of-grade-level rate minimums suggested by Morris et al. (2011, 2013):

- grade 1 = 50 wpm,
- grade 2 = 80 wpm,
- grade 3 = 90 wpm,
- grade 4 = 100 wpm, and
- grade 5 = 105 wpm.

Using IRI results to set the instructional and frustration levels. The major objective in administering an informal reading inventory is to establish a student’s reading instructional level and frustration level. Let’s say that our third-grade student (the aforementioned Andrew) attained the following scores on the second- and third-grade reading passages.

	<u>Accuracy (%)</u>	<u>Comprehension (%)</u>	<u>Rate (wpm)</u>
Second grade	95	83	88
Third grade	91	50	73

Andrew’s performance at grade level (i.e., third grade) is troubling. His reading accuracy (91%) is low in the gray area (90–94%); his comprehension (50%) borders on frustration; and his reading rate (73 wpm) is well below the third-grade minimum of 90 wpm. On the other hand, Andrew’s reading of the second-grade passage meets instructional-level criteria in each area: accuracy (95%), comprehension (83%), and rate (88 wpm). Second grade is his instructional level. In this study, I will use this type of “three-pronged” analysis (accuracy, comprehension, and rate) in setting students’ reading levels.

Text Reading and Comprehension (TRC) Assessment

History. The Text Reading and Comprehension (TRC) assessment is derived, in large part, from the traditional informal reading inventory. For example, the TRC uses graded passages to assess a student’s oral reading ability, just as the IRI has done for decades. Also similar to the IRI, as the child reads the TRC passages, oral reading errors are marked, a running time of the reading is recorded, and comprehension questions are asked. However, the TRC has a recent and important political and marketing history that bears on its design and use. In this section, I will focus on this history, along with

identifying some important differences between the TRC and IRI in scoring and interpretative procedures.

To track the creation of the Text Reading and Comprehension (TRC) assessment, we must go back to the 1990s when national concern developed over low reading achievement in our schools. One signal of concern came from the National Assessment of Educational Progress (NAEP). The NAEP, or the “Nation’s Report Card,” provides an analysis of the progress of American schoolchildren in different subject areas. The reading assessment, administered in fourth, eighth, and twelfth grade, requires students to read grade-level selections and answer comprehension questions. The NAEP data provide an overall scaled score for each student. The assessment is meant to offer a quick snapshot of national reading progress that can be used to evaluate the effectiveness of reading instruction within our public schools.

In 1992, 38% of our nation’s fourth graders scored at the Below Basic level on the NAEP. Thirty-four (34) percent scored at the Basic level, 22% at the Proficient level, and only 6% at the Advanced level. Students scoring at the Basic level “demonstrate an understanding of the overall meaning, make obvious connections, and produce simple inferences to extend understanding” (NCES, 2000, p. 14). The 1992 NAEP results showed that 72% of our students scored at the Basic level or below. Moreover, two years later in 1994, 71% of the students tested scored at these same low levels. In other words, little growth had occurred between the two testing periods; in fact, in 1994 more students scored at the Below Basic level (40%) than had done so in 1992 (38%). Our fourth-grade students weren’t progressing, and Congress determined that something had to be done. That something began with the creation of the National Reading Panel (NRP).

In 1997, in response to the low, unchanging results of the 1992 and 1994 NAEP reading assessments, Congress issued a charge to the National Institute of Child Health and Human Development (NICHD) to develop a national panel of reading experts to review research on the effectiveness of various approaches to teaching children to read (NICHD, 2012). The expectation was for the panel to present conclusions on the research and to evaluate the readiness of applying the research in the classroom. The work of the National Reading Panel (NRP) highlighted five pillars of reading instruction: phonemic awareness, phonics, fluency, comprehension, and vocabulary. The NRP report (2000) recommended that reading programs should emphasize the importance of each pillar in developing proficient, mature readers.

As is the custom, with the publication of the National Reading Panel report, commercial publishers rushed to market with new teaching products that highlighted the so-called five pillars of reading success. At the same time, reading assessments were created specifically to test a student's proficiency level in each of the five target areas. In 2000, a new company, Wireless Generation (now named Amplify), began helping school districts collect, analyze and evaluate data that could tell educators which students were successfully learning to read and which were at risk for reading failure.

Wireless Generation's first project was to offer Dynamic Indicators of Basic Early Literacy Skills (DIBELS), a popular primary-grade skills assessment, on a hand-held, computer tablet. School administrators quickly jumped at the idea of having software that could provide results quickly, along with instructional recommendations for meeting the needs of each child. The DIBELS tasks, which had been developed several years earlier by Good and Kaminski (2003), were simple one-minute probes of phoneme awareness,

decoding skill, and oral reading fluency. However, these paper-and-pencil tasks required note-taking and some analysis by busy classroom teachers. Wireless Generation made the paper and pencil recording unnecessary. The technology did all the work and the data were instantly ready for analysis, comparison, and evaluation.

Not only did Wireless Generation have technology in its corner, it also featured formative assessment, a concept that was becoming popular in the field of education. As the NRP's recommendations began to flood into local school districts, researchers began to champion regular, formative assessment in the classroom (Kaminski & Good, 1998; Reinking & Bradley, 2008). These brief assessments (e.g., the aforementioned DIBELS) were thought to provide the teacher with valuable feedback on a student's progress, feedback that could lead the teacher to continue effective instruction or modify ineffective instruction. Wireless Generation commissioned a position paper, *The Role of Formative Assessment in Pre-K through Second Grade Classrooms* (Honey, 2007), that highlighted how the combination of technology and formative assessment would allow teachers to "save significant time over traditional means, both opening up class time and encouraging the use of assessments" (p. 8).

In addition, the same position paper claimed that formative assessments could help reverse the cycle of remediation by (a) identifying struggling readers early on, (b) assessing their progress on a regular basis, and (c) adapting instruction based on the data (Honey, 2007). The NRP report (2000) had emphasized these same three objectives. Wireless Generation cited the work of Southard, Diefenbach, and Darandari (2004), which concluded that, in grades K – 2, conducting formative assessments and acting on the data would improve students' chances to be successful on summative, high-stakes

reading tests at the end of third grade. In summary, Wireless Generation developed the right kind of assessment using the right kind of tools at the exact time when school districts were struggling to find the answers.

The DIBELS probes (alphabet, phoneme segmentation, non-word decoding, and oral reading fluency), administered on Wireless Generation's hand-held devices, spread quickly across the country as more and more school administrators searched for ways to hold teachers accountable for student progress. The program continued to add more diagnostic and instructional tools that could inform teachers which literacy skills students lacked and what type of instructional support they needed. Every piece of the learning cycle seemed to be available in one program—assessment, instructional tools, regular progress monitoring, and most importantly for administrators, accountability.

In 2008, Wireless Generation took a next big step forward. With the cooperation of the Bellevue, Washington school district, the company began to test a new passage-reading assessment that could potentially replace the traditional, paper-and-pencil informal reading inventory. The new Text Reading and Comprehension assessment (or TRC) placed the traditional IRI tasks and procedures on a hand-held computer device or tablet. Data collection (e.g., coding errors, calculating oral reading accuracy and rate, evaluating comprehension responses) no longer required paper and pencil. It could now occur on the new “tablet” as the teacher followed the directions on the screen.

The founders of Wireless Generation, Larry Berger and Gregory Gunn, designed the technology for this new, innovative assessment after watching elementary school teachers administer the paper-and-pencil IRI to their students. According to Berger, the teachers were individually assessing each student, making lots of notes and checks on a

piece of paper, when he and Gunn realized that they could use technology to help teachers improve this task. Berger has argued, “Although there are no silver bullets in education, we believe that innovative technology can help teachers maximize their impact in the classroom,” (W.K. Kellogg Foundation, n.d.). Wireless Generation, in other words, could help teachers collect assessment data more efficiently, thereby leaving them more time for effective instruction.

According to Wireless Generation’s marketing brochures (n.d.), *mClass: Reading 3D* (DIBELS Next combined with the TRC assessment) offers the following advantages: (a) provides accurate, reliable assessments, (b) makes informed instructional decisions for individual students, (c) rapidly screens and progress monitors students, (d) tracks student progress from year to year and class to class, and (e) embeds professional development within the assessment process. Taken together, Wireless Generation argued, these five characteristics allow for efficient data collection that can be used to instruct, evaluate, and then remediate students’ reading performances.

The research base for the TRC is slim. Thus far, Wireless Generation cites only one validation study, which was carried out in Montgomery County, Maryland (Zhao & Von Secker, 2008). Results showed that students who were successful on the TRC passage-reading assessment at the end of second grade tended to be successful on a standardized reading test administered at the end of third grade. Wireless Generation argued that this study supports the TRC’s use as a formative assessment; that is, if students show appropriate progress on the TRC, they should do well on end-of-grade standardized assessments.

More recently Wireless Generation has reported additional findings from North Carolina that align with the data collected in Maryland. On its North Carolina webpage, Wireless Generation has shared a pie chart that claims that the TRC is “highly predictive of students’ proficiency on certain statewide ELA assessments.” According to the chart, the TRC was 79% accurate in predicting performance on the North Carolina End of Grade reading test for third grade (“Make Accurate Predictions about Reading Proficiency,” Amplify, 2014). However, no information is provided regarding the participants, methods, or data analysis used to determine this prediction. In addition, while the pie chart proclaims the TRC’s ability to predict third-grade achievement, Wireless Generation (or Amplify) does not provide formative or placement information that could guide teachers’ reading instruction, a promise that was made when the test was first developed and marketed.

In an era of high-stakes testing, Wireless Generation’s claim that the TRC is an effective predictor of standardized test performance was enticing to public school administrators and teachers. Selected schools in North Carolina had been administering the TRC for several years, and some districts were further ahead than others in the implementation process. However, in the 2013-2014 school year, each school district in the state was required to administer the mClass 3D assessment to its K – 3 students. The TRC, paired with DIBELS, had become the basic assessment used to determine North Carolina children’s reading proficiency during their first 4 years in school (K – 3).

Administration of the TRC. The TRC, like the IRI, is designed to assess oral reading accuracy and comprehension. In addition, the TRC includes written comprehension questions and error analysis for each passage. It is important to review the

TRC administration process so that a comparison can be made to the IRI. In this section, I will explain the TRC process for a second-grade reader, Mary.

To begin, Mary is given a leveled TRC reading passage (softcover book) that her teacher believes will be relatively easy for the child to read. The teacher can also use Mary's DIBELS Oral Reading Fluency score (words correct per minute) to help determine at which book level to start the TRC passage reading. Once Mary is ready to begin reading, her teacher follows the instructions on the screen—either reading aloud the directions or asking Mary to do something specific with her book before beginning the actual oral reading. For example, if Mary starts with the book, *The Statue of Liberty* (Level L or second grade), the examiner's instructions to the child are to preview the pictures and review the Table of Contents before reading the first two chapters of the book.

As Mary begins to read, the examiner must be ready to mark or code errors on her computer or handheld device. The examiner will code substitutions, omissions, insertions, and teacher helps as errors. Self-corrections and repetitions are noted but not counted as errors (see Figure 2). To record errors on the TRC, the examiner must master a series of coordinated actions on the computer screen. For example, to record a substitution, the examiner (a) taps the misread word, (b) writes the word substituted on the computer screen, (c) taps the 'sub' key to categorize the error as a substitution, and (d) taps "save." All of this coding must occur as Mary continues to read the text. There is a "pause" button at the bottom of the screen if the examiner needs to stop Mary in order to code an error before the child continues reading the text.

1. Substitutions:

Tap the word that was misread on the tablet. Write the incorrect word in the freehand space on the screen. Tap ‘Sub’. Then tap ‘Save’. The word will appear in orange with a strikethrough across the misread word. The substituted word will appear in a word bubble above the text.

2. Omissions:

Tap the omitted word. The word will appear in blue with a square outline.

3. Insertions:

Tap the word after the insertion. In the freehand space, write the word that was inserted. Tap the insertion symbol, a caret. Then tap ‘Save.’ A purple caret will now appear and the inserted word will be in a purple bubble.

4. Self-corrections:

Tap the misread word. Tap the appropriate key to categorize the initial error. Then tap ‘SC.’ Then tap ‘Save.’ The self-corrected error will now appear in green.

5. Told (Teacher Helps):

Record any attempts at the word by the student. Wait 5 seconds for a complete response. Say the word. Tap the word you told. Tap ‘Told.’ Tap ‘Save.’ The word will now be within parentheses and highlighted in red.

Figure 2. Coding system for oral reading errors on the TRC. Adapted from “Text Reading and Comprehension;” Copyright 2012 by Wireless Generation.

Once Mary finishes the oral reading, the computer prompts the examiner to provide written comprehension questions that Mary must answer on her own. The examiner encourages the child, per the screen directions, to use her book to help her answer the questions. This is an important direction because a lack of text-specific

evidence will actually lower Mary’s written comprehension score. For the story, *The Statue of Liberty*, the two written comprehension questions are:

1. Identify three text features used in the book.
2. If you could rename the Statue of Liberty, what would you call it? (Use information from the book to explain your answer.)

Mary must read the written questions without examiner help and attempt to answer them by using evidence from the text. When Mary has completed these written questions, the teacher scores her answers using a rubric (see Figure 3).

Scoring Guidelines

Score of 3: Complete Understanding	The response demonstrates a complete understanding of the text: Addresses the demands of the question Effectively uses detailed information to clarify or extend thinking
Score of 2: General Understanding	The response demonstrates a general understanding of the text: Partially addresses the demands of the question Uses general information to clarify or extend understanding
Score of 1: Minimal Understanding	The response demonstrates a minimal understanding of the text: Minimally addresses the demands of the question Uses minimal information to show understanding of the text in relation to the question
Score of 0: No Understanding	The response demonstrates no understanding of the text: The response is completely incorrect, irrelevant to the questions, or missing.

Figure 3. Scoring rubric for written comprehension questions. Adapted from “Text Reading and Comprehension,” Copyright 2012 by Wireless Generation.

Once the teacher has assessed Mary's written answers, she inputs the scores into the computer and then proceeds to ask Mary five oral comprehension questions. The examiner simply marks 'Correct' or 'Incorrect' for each oral question. Once these comprehension assessments are complete, the computer immediately determines if the examiner should go on to the next book, back up to the easier book, or stop the assessment. This is determined based on Mary's oral reading accuracy and her answers to the written and oral comprehension questions.

Scoring and interpreting TRC scores. Interpreting the TRC scores is accomplished instantly by the computer. It calculates the oral reading accuracy and then, based on the input from the examiner, interprets the student's comprehension level. For the TRC, the following benchmarks are used to determine a student's functional reading levels:

- *Frustration Level:* Accuracy score is 89% or lower; or Written Comprehension score is less than 2 points out of 3; or Oral Comprehension score is less than 4 points out of 5.
- *Instructional Level:* Accuracy score of 90-94%, a Written Comprehension score of at least 2 points out of 3, and an Oral Comprehension score of at least 4 points out of 5.
- *Independent Level:* Accuracy score of 95% or higher, a Written Comprehension score of at least two 2 points out of 3, and an Oral Comprehension score of at least 4 points out of 5 (Wireless Generation, 2012).

Depending on the time of year—beginning, middle, or end—a student’s reading level also is assigned a color to signify level of risk. For example, Mary’s reading level can be labeled red (well below benchmark), yellow (below benchmark), green (at benchmark), or blue (above benchmark). These levels of risk determine how often a student is to be progress-monitored. The risk levels are also meant to help teachers create instructional groupings that will support students’ reading growth.

Comparison of the IRI and TRC

The two assessments to be evaluated in this study have similarities and differences. Both assessments help to determine a student’s instructional level through the reading of leveled passages. However, they differ in both administration and interpretation procedures.

Administration. Both assessments count the following as oral reading errors: substitutions, omissions, insertions, and teacher helps. The marking of each error is different because of the method used: the IRI is a paper-and-pencil assessment and the TRC is a computer assessment. The one difference in error counting involves self-corrections. The IRI counts self-corrections as errors; the TRC does not.

The comprehension section of each assessment is administered differently. In the IRI, the child must answer four to six oral comprehension questions per passage. He or she cannot refer to the text, but must answer the question based on what is remembered. The purpose of the oral questions is to see if the child remembers (understands) important information in the passage.

In the TRC, comprehension is assessed in both a written and oral mode. Above the prereprimer or early-first-grade level, the student must complete two written

comprehension questions on each passage before responding to five oral questions. Unlike the IRI, the child can refer to the text in answering the written and oral TRC comprehension questions. In fact, the written questions often require the use of the book in order to provide an adequate, text-specific answer.

A final and important difference in how the two assessments are administered involves the recording of reading rate. In the IRI, the examiner records how many seconds the student takes to read a passage from beginning to end (this time is converted to a words per minute [wpm] measure). In the TRC, the computer tablet automatically records the child's reading rate in words correct per minute (wcpm). However, if the examiner pauses the reading to "catch up in recording missing errors," these pauses obviously affect the final rate.

Interpretation. The interpretation of reading scores (e.g., oral reading accuracy, rate, and comprehension) plays a critical role in determining a student's instructional level. Importantly, the interpretive rules of the IRI and the TRC differ significantly.

Regarding oral reading accuracy, the IRI and TRC use different cut-off points (or performance ranges) for instructional level. The IRI uses a 95-97% range whereas the TRC uses a 90-94% range. (Both assessments use 90% as the oral reading accuracy cut-off score in first grade). It is also important to keep in mind that the IRI, but not the TRC, counts self-corrections as oral reading errors. This makes the IRI oral reading accuracy cut-off (95%)—already higher than the TRC above first grade—a more conservative or stringent measure.

Regarding comprehension, the TRC requires that students achieve minimum scores on both the written questions (2 out of 3 points) and the oral questions (4 out of 5)

to be considered instructional at a given level. On the other hand, the IRI, which does not use written questions, requires a 75% score (3 out of 4) on the oral questions to be considered instructional. It should be noted here that it is highly unusual to use written questions in assessing the reading comprehension of primary-grade (first, second and third grade) students.

Finally, regarding reading rate in the interpretation process, the IRI requires that students read at a minimum rate (or speed) in order to be considered instructional at a given level (e.g., 80 wpm at second grade, 90 wpm at third grade). On the other hand, the TRC records the child's reading rate on each passage, but does not use the rate score in the interpretive process—that is, in setting an instructional level.

Word Recognition–timed: A Third Measure of Reading Skill

In a study that compares primary-grade students' scores on two different reading assessments (IRI and TRC), there may be performance differences. That is, a child (or a group of children) may score higher—achieve a higher instructional level—on one assessment than on the other. In this case, it would help to have a third, “neutral” reading measure to which scores on the target assessments (IRI level and TRC level) could be compared. Because the focus of this study is on print-processing skill (the accuracy and fluency with which a child reads), there is a third reading measure, *word recognition-timed*, that can be used to arbitrate differences between IRI and TRC performance.

The importance of automatic word recognition in skilled reading is widely acknowledged. For example, LaBerge and Samuels (1974) described reading as a “zero-sum” game in which the reader, at a given moment, must divide his or her attentional resources between word recognition and comprehension. The goal, according to these

researchers, is to automatize (or minimize attention to) word processing so that maximum attention can be devoted to comprehending the text. See Adams (1990), Perfetti (1985), and Stanovich (1990) for a similar theoretical stance.

Regarding assessment, the implications of automatic word recognition are clear. First, we can assess the speed with which a reader processes text through measures of reading rate (e.g., words read correctly per minute or WCPM). Second, we can obtain an isolated measure of automatic word recognition by flashing single words to a student for a fraction of a second. This is often referred to as a word recognition-timed test (Gillet, Temple, & Crawford, 2011; Stauffer, Abrams, & Pilulski, 1978).

Betts (1946) introduced the idea of a word recognition inventory (graded word lists) in his famous textbook, *Foundations of Reading Instruction*. He suggested that the teacher randomly choose 20 words from each level of a basal reader (preprimer through sixth grade). Administering a given word list (e.g., 20 second-grade words) involved flashing each word for a fraction of a second. If the child read the word correctly, the examiner moved forward. However, if, on the flash or timed presentation, the child misread the word (or failed to respond), the examiner exposed the word again and allowed more time for the child to decode. Scoring a list involved counting the number of errors and computing a percentage correct score for both the timed and untimed conditions.

Russell Stauffer, one of Betts's students, later argued that the timed score on a given list, as opposed to the untimed score, was the better predictor of contextual reading ability at that level. Stauffer et al. (1978) provided criteria for interpreting flash (or timed) performance on the word recognition inventory:

- 90% word recognition-timed on a given list is an estimate of independent reading level (appropriate for pleasure reading).
- 75% word recognition-timed is an estimate of instructional reading level (appropriate for teacher-guided reading).
- Below 50% word recognition-timed is an estimate of frustration level.

A recent study by researchers at Appalachian State University (Morris et al., 2011; 2012; 2013) provided strong validation for the word recognition-timed (WR-t) measure, showing that it was an excellent predictor of reading rate (or fluency) across the elementary grades. Correlations between WR-t and oral reading rate were as follows; second grade = .74; third grade = .70; fourth grade = .66; fifth grade = .68. Interestingly, at each grade, these correlations were significantly higher than the correlations between oral reading accuracy and reading rate.

In the present study that compares student performance on two formative reading assessments (IRI and TRC), the WR-t test will serve as an “arbiter” measure. That is, when a student’s IRI score (or level) differs from his or her TRC score, I will determine which score, if either, agrees with the arbiter score, WR-t.

Research Questions

In summary, although the IRI and the TRC are both informal passage-reading assessments, they differ in how they are administered, scored, and interpreted. The present study compared the two assessment instruments on both student performance factors and test administration factors. The following research questions were addressed:

1. Is there a difference between a reading instructional level derived from the IRI and one derived from the TRC? If so, how large is the difference?

2. If the two instruments do yield different reading levels for students, does this happen at each grade—first, second, and third?
3. If the two instruments yield different reading levels for a student (e.g., IRI = third grade, TRC = second grade), which reading level is more in agreement with the reading level provided by a third instrument, word recognition-timed?
4. How does the IRI compare to the TRC in terms of administration time (i.e., time needed to administer and score the test)?

Chapter Three: Method

The administration of, and results attained by, two different contextual reading instruments (IRI and the TRC) were compared with the purpose to determine which instrument provides the most valid assessment of children's reading ability.

Participants

The participants were 196 children (first through third grade) from an urban public school in the Winston Salem/Forsyth County school district. There were 62 first graders, 70 second graders, and 64 third graders. The school is 58% African American, 22% Hispanic, 16% Caucasian, and 4% other races. Ninety-five percent (95%) of the students qualify for free or reduced lunch.

Students at the school struggle to read on grade level but perform adequately in math. In the 2011-2012 school year for grades 3 – 5, the school was 51% proficient on the North Carolina End of Grade Reading Test and 76% proficient on the North Carolina End of Grade Math Test. These scores show the school to be more than 20% behind the state average for reading proficiency, but only 7% behind the state average for math proficiency. The Caucasian students are currently outperforming the African American and Hispanic students, while the girls are outperforming the boys in both reading and math (NC DPI, 2012).

Assessments

Teacher training. The lead researcher (author) was trained in conducting the TRC assessment at the beginning of the 2012-2013 school year. Consultants from Wireless Generation (now Amplify) and the North Carolina Department of Instruction conducted the 6-hour training with all elementary curriculum coordinators in the school district. As a curriculum coordinator, I then trained all teachers at my elementary school, including the other four examiners in this study. Each examiner was a full- or part-time literacy instructor in my school. The first two training sessions, each lasting approximately 2 hours, included an explanation of the TRC assessment and its components, a short practice session where participants were able to practice coding errors on the computer tablet, and a brief session for discussing how to score the answers to the written comprehension questions. A short review session (1 hour in length) occurred directly before beginning the data collection. The five-person assessment team began testing with the kindergarten students to ensure that each examiner had some additional practice with an actual student before collecting the data for this study (first through third grade). As issues arose, the team met and discussed how to deal with them. (*Note:* Each member of the assessment team already had extensive experience administering the IRI used in this study.)

A critical objective in this study was to determine the amount of time required to administer each assessment (TRC and IRI), and so the examiner timed the assessments (TRC and IRI) of each student, beginning with the introductory instructions. For the TRC, time started when the examiner began to read the instructions on the tablet, and time stopped when the student completed the passage reading. For the IRI, time started

when the examiner began to read the introduction to the first reading passage, and time stopped when the child reached his or her frustration level on the IRI passages. Stopping was determined by signs of frustration in the student's reading (e.g., below 90% oral reading accuracy or a very low reading rate).

Administering the IRI. Detailed descriptions of how to administer the IRI and TRC have already been provided in the literature review; therefore, a quick summary will be presented here. The IRI passages (See Appendix B) were administered as follows:

1. The examiner explained that the student should read the passage aloud at his or her normal reading rate because some questions would follow the reading.
2. The examiner read a one-sentence introduction to the passage and then signaled the child to begin reading the passages aloud. (The tape recorder was turned on at this point.)
3. As soon as the student began to read, the examiner started the timer (or stopwatch).
4. As the student read, the examiner marked or coded oral reading errors on his or her copy of the passage, following the procedures highlighted in Figure 1 (see p. 16). If necessary, the examiner could later listen to the recorded version of the child's reading, but this seldom happened.
5. Once the student read the last word in the passage, the examiner stopped the timer, turned over the reading booklet, and asked the student several comprehension questions.
6. Based on the student's performance, the examiner decided whether to move to the next, more difficult passage, drop back to a less difficult passage, or stop

the assessment because a frustration level had been reached. (The examiner used accuracy, rate, and comprehension performance to help make this decision.)

7. Once the IRI assessment was complete, the examiner used the data to determine the student's instructional and frustration levels and to note specific strengths and weaknesses in performance.

In the present IRI testing, each student, depending on grade level (first, second, or third), began the IRI at the same level and then progressed as indicated by performance. First graders began with an early-first-grade passage; second graders began with a late-first-grade passage; and third graders began with a second-grade passage. Again, the examiner made decisions on how to move a student (forward or back) based on reading performance.

Administration of the IRI yielded several measures for each student (i.e., reading instructional level, time needed to administer the test, and reason for discontinuing testing).

- *Reading instructional level (0 – 7).* Table 3 shows the seven levels with their accompanying performance criteria. For a child to be designated a Level 4 (or second-grade) reader, he or she had to achieve minimum scores in each second-grade category (i.e., 95% accuracy, 80 wpm rate, and 75% comprehension).
- *Administration time.* A second IRI measure was the amount of time it took the examiner to administer the IRI to a student and then score the performance. Six

minutes for scoring (a constant) was added to the administration time for each student.

- *Reason for discontinuing the test.* A categorical measure, the reason for discontinuing might be a low accuracy score, a low rate score, a low comprehension score, or a combination of the above.

Table 3.

Reading Instructional Levels for IRI and Accompanying Performance Criteria

Level	Grade	Performance Criteria		
		Reading Accuracy (%)	Reading Rate (wpm)	Comprehension (%)
0	Emergent	--	--	--
1	Preprimer	90	--	
2	Primer	90	40	60
3	Late-First	90	50	60
4	Second	95	80	75
5	Third	95	90	75
6	Fourth	95	100	75
7	Fifth	95	105	75

Administering the TRC. The TRC passages (see Appendix C) were administered in the following manner (see pp. 24-28).

1. The examiner began the assessment by reading aloud the instructions on the screen.
2. As the student began to read, the examiner marked or coded oral reading errors on the screen, following the instructions provided in Figure 2 (p. 26). Note that in this study, but not in normal TRC practice, the examiner recorded all oral readings for later analysis.
3. Once the student finished reading a passage orally, the examiner followed the instructions on the screen. Depending on the passage read, one of the following occurred: (a) the student answered two comprehension questions or (b) the student finished reading the selection silently and then answered two written questions. The student was encouraged to use the book in answering the questions.
4. After the student completed the written comprehension questions, the examiner, using the written rubric, quickly scored the student's answers and entered the scores into the tablet. (See written comprehension rubric on p. 27.) Then the examiner asked five oral comprehension questions concerning the passage's content. Again, the student was encouraged to use the book to answer these questions.
5. Given the input (i.e., the child's scores), the computer immediately determined whether the examiner needed to move forward to the next book (passage), back to a less difficult book, or to stop the assessment. This decision was based on the student's oral reading accuracy and comprehension scores. (See TRC scoring criteria on p. 28.)

6. Once an instructional level was identified (the highest level before the student reached frustration), the examiner conducted a Meaning, Syntax and Visual Analysis (MSV) of the instructional-level errors. A couple of steps were involved in this analysis. The examiner clicked on the MSV analysis button to go to the MSV screen. Once on this screen, the computer highlighted each of the child's errors, one at a time. For each error, the examiner quickly decided whether it preserved meaning [M], maintained syntax [S], or maintained the visual appearance of the word [V]. These entered data became a part of the student's record as well.

In the present TRC testing, each student, depending on grade level (first, second, or third), began the TRC at the same level and then progressed as indicated by performance. First graders began with an early-first-grade passage (Fountas & Pinnell Level D); second graders began with a late-first-grade passage (F & P Level I); and third graders began with a second-grade passage (F & P Level L). The device made decisions on how to move a student (forward or back) based on reading performance.

As with the IRI, administration of the TRC yielded several measures for each student.

- *Reading instructional level (0 – 7).* Table 4 shows the seven levels with their accompanying performance criteria. The table also shows how the numbered levels (1 – 7) correspond to Fountas and Pinnell levels (A through V), and to traditional grade levels (preprimer through fifth grade). For a child to be designated a Level 4 (or second-grade) reader, he or she would have to achieve the minimum scores in each second-grade category (i.e., 90%

accuracy, 2 of 3 correct responses on written comprehension, 4 of 5 correct responses on oral comprehension).

- *Administration time.* A second TRC measure was the amount of time it took the examiner to administer the TRC to a student.
- *Reason for discontinuing the test.* A categorical TRC measure, the reason for discontinuing might be a low accuracy score, a low written comprehension score, a low oral comprehension score, or a combination of the above.

Table 4.

Reading Instructional Levels for TRC and Accompanying Performance Criteria

Level	Grade	F & P	Performance Criteria		
			Reading Accuracy (%)	Comprehension Written	Comprehension Oral
0	Emergent	A, B, C	--	--	--
1	Preprimer	D, E	90	--	--
2	Primer	F, G	90	2/3	4/5
3	Late-First	H, I, J	90	2/3	4/5
4	Second	K, L, M, N	90	2/3	4/5
5	Third	O, P, Q	90	2/3	4/5
6	Fourth	R, S, T, U	90	2/3	4/5
7	Fifth	V, W	90	2/3	4/5

Note: Whereas the IRI included only one passage at each level (1, 2, 3, and so on), the TRC included multiple passages at each level (e.g., Level 1 = passages D and E; Level 2 = passages F and G; Level 3 = passages H, I and J; and so on). For comparison's sake, from second grade onward, performance on the second TRC passage at each level was used (e.g., L at Level 4; P at Level 5; S at Level 6; and so on).

Administering the WR-t. The word recognition–timed (WR-t) test was administered individually to 67 of the 196 children in this study. This subset of students was randomly selected at each of the three grade levels. Administration time was approximately 7 minutes per child.

The WR-t assessment comprised seven 20-word lists—preprimer through fifth grade (see Appendix A). Previous research (Morris et al., 2011) had demonstrated the hierarchical properties of these word lists (i.e., list 1 was easier than list 2; list 2 was easier than list 3; and so on). The test had also proven to be reliable and, at each grade second through fifth, a strong predictor of oral reading fluency (see Morris et al., 2011; 2012).

Testing began with the first word on the preprimer list. A computer flashed a single word to the child for one-half second (Barr, Blachowicz, Katz, & Kaufman, 2007; Morris, 2014; Schneider, 2013). A correct response was recorded if the child pronounced the word correctly, with no hesitation. An incorrect response was recorded if the child mispronounced the word, failed to respond, or hesitated for one second or more. If the child scored at least 50% correct on a given list, the examiner proceeded to the next test list. Testing was discontinued when the child made 11 or more errors on a given 20-word list.

The student attained a percentage correct score on each list (e.g., 15 correct responses out of 20 tries yielded a score of 75%). The critical measure yielded by the

word recognition assessment was a word recognition–timed instructional level. It was the highest level (0 – 7) at which the child met specific performance criteria (see Table 5).

Table 5.

Word Recognition–Flash Instructional Levels and Accompanying Performance Criteria

Level	Grade	% Correct
0	below Preprimer	< 60
1	Preprimer	60 ^a
2	Primer	60
3	Late-First	60
4	Second	70
5	Third	70
6	Fourth	70
7	Fifth	70

^a Instructional-level percentages recommended in Morris et al. (2011)

Research Design

The administration of each reading assessment (IRI, TRC, and WR-t) yielded an instructional level that could be put on a common 0 to 7 scale. This allowed for a comparison of performance levels on the three tasks. For example:

- How often did the IRI and TRC agree regarding a student’s instructional level? For instance, a score of 2 on the IRI vs. 2 on the TRC was a “hit” or agreement; a score of 2 vs. 3 was a “miss.”

- Did instructional-level agreement between the IRI and TRC differ across grade levels (first, second, and third)?
- When the IRI and TRC were in disagreement regarding instructional level, which of the two measures agreed more often with a third reading measure: WR-t?
- A final comparison of the IRI and TRC involved the amount of time needed to administer the assessments. Each of these questions will be addressed in the Results section that follows.

Chapter Four: Results

The major findings reported in this section involve comparisons of (a) student performance on the two reading assessments (IRI and TRC) and (b) the amount of time invested in administering the two assessments. Before turning to these results, however, it is important to establish that both tests were administered accurately by the five examiners.

Checking Scoring Fidelity for the IRI and the TRC

The IRI, a pencil-and-paper assessment, was familiar to the examiners. On the other hand, the TRC, which uses a “touch screen” to score reading errors, was relatively new to the examiners. (They had received 6 hours of training with the TRC.) One might assume that the different modes of scoring (paper/pencil vs. touch screen) might themselves produce differences (or error), particularly since each examiner only listened to a child reading a passage one time. To check for error in the examiners’ scoring, the children’s reading of all passages (IRI and TRC) was audio recorded. Later, three reading specialists listened to and scored a large subset of the recordings (386 passages across 70 children chosen randomly).

For each of the 386 passages, *a priori* criteria were used to gauge the amount of agreement between the original and rechecked scores. For oral reading accuracy, the original score had to be within 2 percentage points of the rescored passage (94% vs. 92% = acceptable; 94% vs. 91% = unacceptable). For oral reading rate, the original score had

to be within 4 wpm of the rescored passage (105 wpm vs. 109 wpm = acceptable; 105 wpm vs. 110 wpm = unacceptable).

Given these criteria, on the IRI, the percentage of oral reading accuracy agreements or “hits” was 92% (N = 213), with the average difference between the original and rechecked scores being 0.83 percent. The percentage of oral reading rate agreements was 90% (N = 195), with the average difference between the original and rechecked scores being 2.2 wpm (see Table 6).

On the TRC, the percentage of oral reading accuracy agreements was 82% (N = 173), with the average difference between the original and rechecked scores being 1.5 percent. The percentage of oral reading rate agreements was 74% (N = 173), with the average difference between the original and rechecked scores being 4.4 wpm (again, see Table 6).

In summary, the data in Table 6 indicate that the IRI results agreed more often with the rescored data than did the TRC results (92% vs. 82% for oral reading accuracy; 90% vs. 74% for oral reading rate). This finding might be attributed, in part, to the examiners’ relative inexperience with the new, computer-scored TRC assessment. Note, however, that the average differences between the IRI and rescored results and the TRC and rescored results were fairly small (accuracy = .67% [1.50 - .83]; rate = 2.2 wpm [4.43 - 2.20]).

Table 6.

Fidelity of Scores for the IRI and TRC: Original vs Rescored Passages

IRI				
	First Grade	Second Grade	Third Grade	Total
Oral Reading Accuracy (%)				
Agreement between scores	91	98	89	92
Average difference	0.75	0.61	1.13	0.83
Oral Reading Rate (wpm)				
Agreement between scores	95	88	89	90
Average difference	1.34	2.90	2.26	2.20
TRC				
Oral Reading Accuracy (%)				
Agreement between scores	84	84	78	82
Average difference	1.41	1.48	1.60	1.50
Oral Reading Rate (wpm)				
Agreement between scores	71	80	71	74
Average difference	4.00	4.10	5.20	4.43

Research Questions

Four major research questions are addressed in this section. The first three questions are concerned with a comparison of student performance on the various assessments: IRI, TRC, and word recognition-timed. The fourth question is concerned with the amount of time needed to administer the assessments.

1. Overall, is there a difference between a reading instructional level derived from the IRI and one derived from the TRC? Instructional-level performance on each of the assessments was put on a 0 – 7 scale (early first grade through fifth grade). For the 196 students in the study (grades 1 - 3), the mean IRI instructional level of 3.38 (SD = 1.67) was significantly higher than the mean TRC instructional level of 2.56 (SD = 1.32) ($t = 9.2; p < .001$). Although the difference between the means (0.82 of a reading level) is large, educationally speaking, interpretation is confounded because the analysis combines student scores across three grades. What is needed is a comparison of the instructional-level means at each grade level, first, second, and third.

2. Are there differences between IRI and TRC instructional levels at each grade level—first, second, and third? Table 7 shows comparisons of instructional-level means (IRI and TRC) at each grade level. Note in the table that there is a significant difference between the IRI instructional-level mean and TRC instructional-level mean at each grade level, first - third. These mean differences decrease across the grades—0.97 at first to 0.80 at second to 0.72 at third. Still, even at third grade, the children are performing almost three-fourths of a level higher on the IRI assessment.

3. How do the instructional reading levels derived from the IRI and TRC compare to instructional levels derived from a third reading assessment—Word Recognition-timed (WR-t)? When two assessments, administered to the same children at the same time, yield different results, an obvious question arises: Which results better represent the students' true ability? In the present case, should we pay more attention to the IRI instructional level or the TRC instructional level? To address this question, I administered a third assessment, word recognition-timed (WR-t). The idea was to see

Table 7.

Comparison of Instructional-Level Means (IRI and TRC) at Grade Levels 1 – 3

	N	Mean	Std. Dev..	<i>t</i>	<i>p</i>
First Grade					
IRI	62	2.08	1.26	6.92	.000
TRC	62	1.11	.81		
Second Grade					
IRI	70	3.56	1.26	5.42	.000
TRC	70	2.76	.71		
Third Grade					
IRI	64	4.45	1.59	4.05	.001
TRC	64	3.73	.84		

which measure, IRI instructional level or TRC instructional level, was a better match with a third reading measure (WR-t instructional level).

I began by conducting a repeated measures analysis of three variables (IRI instructional level, TRC instructional level, and WR-t instructional level) for the subset of 67 students who received all three assessments. Table 8 shows means and standard deviations for the three variables. The IRI mean of 3.31 in the table (N = 67) is very close to the IRI mean of 3.38 in the total population (N = 196). Similarly, the TRC mean of 2.54 in Table 8 is almost identical to the TRC mean of 2.56 in the total population. The similarity of both sets of scores for both assessments indicates that the 67-student subset is representative of the total population of students in this study.

Table 8.

Means and Standard Deviations for IRI, TRC, and WR-t Variables

Variable	N	Mean	Standard Deviation
IRI	67	3.31	1.48
TRC	67	2.54	1.27
WR-t	67	3.39	1.75

When considering the three variables in a repeated measures analysis, there was an overall significant difference ($F = 19.71; p < .001$). Table 9 shows paired contrasts, indicating where significant differences occurred between paired variables. In the subset of 67 students, there was again a significant difference between the two passage-reading measures, IRI and TRC. But what about the arbiter variable, WR-t, and its relationship to each of the passage-reading variables? Results showed that there was a significance difference between WR-t and TRC, but not between WR-t and IRI. In other words, WR-t performance more closely matched performance on the IRI than it did performance on the TRC.

To further examine the relationship or match between WR-t level and IRI and TRC levels, I used the SignTest, a non-parametric test that is concerned with how many values match (e.g., 3 vs. 3) and how many values are below or above a corresponding value (e.g., how many 2's and how many 4's).

Table 9.

Tests of Paired Differences Between IRI, TRC, and WR-t Variables

Paired Contrast	Mean Square	F	Significance
IRI vs. WR-t	.37	.45	.504
TRC vs. WR-t	48.49	22.46	.000
IRI vs. TRC	40.36	26.20	.000

First, I compared the IRI levels with the WR-t levels. There were 32 matches, 20 positive differences (WR-t > IRI), and 15 negative differences (WR-t < IRI). There was no significant difference between the number of positive and negative differences ($z = 1.02$) (see Table 10). So, in nearly half the cases (32 of 67), the WR-t and IRI instructional levels matched, and there was no significant difference between the positive and negative instances.

Next, I compared the TRC levels with the WR-t levels. There were only 10 matches, with 43 positive differences (WR-t > TRC) and 14 negative differences (WR-t < TRC) (see Table 10). This time, there was a significant difference between the number of positive and negative instances ($z = 3.71; p < .001$). So, in only 15% of the cases did the TRC and WR-t levels match. Moreover, there were significantly more occurrences where the TRC level was lower than the WR-t than where it was higher. In other words, the TRC, when compared to WR-t, tended to provide a lower estimate of the children's reading level.

Table 10.

A Comparison of Matches Between WR-t level and IRI and TRC Levels

	<u>IRI</u>		<u>TRC</u>	
	N	(%)	N	(%)
Matched with WR-t	32	48	10	15
Lower than WR-t	20	30	43	64
Higher than WR-t	15	22	14	21

A third approach to examine the relationship between the reading levels produced by WR-t, IRI and TRC is to perform a chi-square analysis on the frequencies (Ns) found in Table 10. Is there a disproportionate representation of frequencies within the rows for the two assessment approaches? The chi-square value produced by the frequencies in Table 10 is 19.45 (df = 2), $p < .001$. So, there is a disproportionate representation of the frequencies in the table. Since there is no difference in the frequencies in the row “Higher than WR-t,” the difference must be between the “Matched with WR-t” and “Lower than WR-t” rows. Again, we are left with the conclusion that more matches resulted for the levels produced by the IRI and WR-t than for the levels produced by TRC and WR-t. Also, the TRC, when compared to WR-t, provided a lower estimate of the children’s reading level.

4. How does the IRI compare to the TRC in terms of time needed to administer the assessment? Table 11 shows the average number of minutes it took to administer the IRI and TRC to a single child. Even a cursory look at the table shows that the differences in administration time were large, and statistically significant, at each grade, first—third. To be fair, the computer-administered TRC automatically scores the child’s reading performance as the test proceeds, whereas the paper-and-pencil administered IRI has to be manually scored by the examiner *after* the assessment is completed. Nonetheless, if we add a generous 8 minutes of IRI scoring time to the IRI administration times shown in Table 11, we find that the TRC, compared to the IRI, took twice as long to administer and score in first grade (40 to 21 minutes), three times as long in second grade (69 to 23 minutes), and nearly four times as long in third grade (86 to 23 minutes). The large difference in IRI and TRC administration times was a surprising and consequential finding. Reasons for this difference will be considered in the Discussion section.

With the TRC, each student had an individual testing time (see averages and standard deviations in Table 11). Interestingly, in an effort to save testing time, the TRC manual encouraged the teacher to assess multiple students at the same time when possible. To do this, the teacher began an oral reading passage with one student, sent him or her off to complete the written comprehension questions, and then began an oral reading passage with a second student. Because the teacher was testing two, sometimes three students simultaneously with the TRC, I decided to calculate an average per student testing time for the teacher.

Table 11.

Average Time (Minutes) Needed to Administer the IRI and TRC Assessments Across Grades 1 - 3

Level	IRI		TRC		t	Sig.
	M	s. d.	M	s. d.		
First Grade (N = 60).	13.2	5.0	40.1	34.4	- 6.5	.000
Second Grade (N = 70)	15.2	5.7	69.0	37.4	-11.9	.000
Third grade (N = 62)	14.9	5.4	86.3	45.1	-12.5	.000
Total (N = 192)	14.5	5.4	65.6	43.2	-16.7	.000

Five teachers (or examiners) took ten full days to complete the TRC assessment of 196 students (grades 1 – 3). Allowing for breaks, each examiner tested for approximately 4.5 hours each day or 45 hours across the ten days. For the five examiners, this amounted to a total of 225 hours of TRC testing (5 x 45). When one divides 225 TRC hours by 196 students, the result is 1.1 hours of teacher testing time for each student. It is important to remember that the TRC is administered not once, but three times during the school year (beginning, middle, and end). Given a class of 20 students, this would amount to 60+ hours (9 full days or 18 half days) of TRC testing during the year. Conversely, the IRI assessment averaged far less teacher time per student (15 minutes), which would total 15 hours a year (a little more than 2 full days or 4 half days), saving teachers important instructional time.

Chapter Five: Discussion and Implications

This study compared a new passage-reading assessment (TRC) with a traditional one (IRI) in terms of ease of administration and validity of results. One hundred and ninety-six students (grades 1 – 3) were administered both the TRC and the IRI. A randomly chosen subset of these students ($N = 67$) was administered a third assessment—word recognition-timed or (WR-t). In this section, I discuss the major findings in the study and implications for using these assessments in the future.

Major Findings

Students achieved a higher reading instructional level on the IRI than they did on the TRC. Across 196 students in grades 1 – 3, there was a significant difference in the instructional level yielded by the TRC and that yielded by the IRI. Overall, the TRC was .82 of a reading level lower. More important, the average TRC instructional level was significantly lower than the average IRI level at each grade level—almost a full level lower at first grade (.97), four fifths of a level lower at second grade (.80), and nearly three fourths of a level lower at third grade (.72). Educationally speaking, these are consequential differences. For example, if the assessment results in this study were closely adhered to, the TRC might designate a second-grade child as a late-first-grade reader, whereas the IRI would designate the same child as a mid-second-grade reader. Underestimating a child’s reading level by four fifths of a level (approximately 8 months) is not acceptable. Primary-grade children need to be taught at the “cutting edge” of their

reading ability (i.e., the correct instructional level) if they are to make adequate progress in a given school year (Allington, 2006; Morris, 2014).

IRI performance, compared to TRC performance, more closely matched performance on a third reading assessment, WR-t. Students achieved a higher reading instructional level on the IRI than they did on the TRC. To determine which was the more valid instructional level, IRI or TRC, I administered a third arbiter assessment (WR-t) to approximately one third of the students. Results clearly showed that the WR-t instructional level more closely matched the IRI level than it did the TRC level. The WR-t and IRI levels agreed in almost half of the cases, while the WR-t and the TRC levels agreed in only 15% of the cases. Moreover, while WR-t mismatches with the IRI were evenly distributed above and below the target, WR-t mismatches with the TRC were consistently above the target. In other words, the TRC levels were consistently below the student's measured word recognition level.

There are at least two reasons why the WR-t and IRI levels matched more closely. First, both of these assessments measure automaticity or processing speed. The WR-t does this by flashing single words for a half second and requiring an immediate response. The IRI measures automaticity by timing how long (wpm) it takes a child to read a short passage. On the other hand, the TRC does not consider rate, or automaticity, in setting a child's instructional level.

A second reason why the WR-t level more closely matched the IRI level than it did the TRC level has to do with how the TRC was administered. Examiner adherence to strict and elaborate TRC comprehension criteria often led to a student's test being

terminated before the child had reached his instructional level in print-processing. I will discuss this issue further in the next section.

Why did the TRC consistently underestimate students' reading instructional level? Even while administering the assessments, the examiners sensed that the TRC was frequently stopping students before they reached their true instructional level in reading. Knowing this to be the case, I searched through our data to determine why students seemed to be stopped too soon on the TRC assessment. My search revealed that students were stopped early on the TRC for several reasons, including low oral reading accuracy, poor oral comprehension, and poor written comprehension.

Students were sometimes stopped on the TRC for low oral reading accuracy (90–94%) even if they maintained appropriate oral and written comprehension. For example, even if the examiner wanted to continue testing a first-grade child who read a first-grade passage with 93% accuracy, the computer would not allow the test to continue. The TRC's assumption was that if the reader performed at only 93% accuracy, he or she would not be able to read the next, more difficult passage with instructional-level accuracy. In contrast, on the IRI, the examiner did not stop the test for low oral reading accuracy unless the child's score fell below 90%. (This is traditional practice in reading diagnosis.) The thinking, here, is that the child may have made some careless, inconsequential errors, leading to a 93% oral reading accuracy score. Therefore, if his or her accuracy was above 90%, and comprehension and fluency scores were acceptable, then the next IRI passage could be administered. This type of examiner judgment was not allowed on the TRC (i.e., the computer simply stopped the test).

Some students were stopped on the TRC because of poor oral comprehension. Even if they were reading accurately (95% or higher) and fluently, they were stopped at a given level (e.g., on a second-grade passage) if they failed to answer correctly at least four out of five oral comprehension questions. Furthermore a child may have read with 98% accuracy and answered three of five comprehension questions correctly, and still, the testing would be terminated. As with oral reading accuracy, the examiner could not exercise judgment in this situation. In contrast, on the IRI, a child who read with 98% accuracy and 60% comprehension would proceed to the next passage where he might conceivably read with 96% and 80% comprehension, an instructional-level performance.

The third and most prevalent reason for stopping the TRC assessment was poor written comprehension. Testing was terminated when the student did not score at least two points out of three on both written comprehension questions that accompanied a given passage. These written questions were problematic on several accounts. First, the children were required to read and interpret each written question without examiner support. Then they had to craft a written answer to the question. They were allowed to look back in the text for the answer. It seems obvious that a low reader might have trouble reading and interpreting the question without help. And, even an able primary-grade reader might have trouble expressing his comprehension of a passage in writing.

For example, a second-grade child might read a late-first grade passage with good accuracy and fluency, answer four of five oral comprehension questions correctly, and still be stopped (designated frustrated) on that passage (or level). This is because the child had failed to answer adequately one of the two written comprehension questions that accompanied the passage. Terminating the TRC testing because the student did not meet

criterion on written comprehension clearly led to many children not being able to attempt passages that otherwise they may have been able to read and comprehend. In fact, in 56% of the cases where the TRC underestimated a student’s reading instructional level, written comprehension was the determining factor (see Table 12).

Table 12.

Reasons for Terminating the TRC Assessment

	First	Second	Third	Total
Oral Reading Accuracy (%)	44	0	0	14
Oral Reading Comprehension (%)	0	6	11	6
Written Comprehension (%)	39	72	56	56
Oral Reading Comprehension and Written Comprehension (%)	11	16	22	16
Oral Reading Accuracy and Oral Reading Comprehension (%)	6	0	11	6
Oral Reading Accuracy and Written Comprehension (%)	0	6	0	2

While the written comprehension questions often stopped students from reaching their true instructional level on the TRC, the written questions occasionally pushed a few students on to passages that were too hard. For example, one third-grade boy was judged

by the TRC to be proficient at the second-grade reading level because he could read at that level with adequate accuracy, and answer both the oral and written comprehension questions by determinedly looking back and perusing the text for answers. On the other hand, the child read the second-grade passage at a rate of 43 wpm, a borderline rate for a mid-first-grade reader. Given time to look back at the text—lots of time—he could compensate for his print-processing difficulties and answer comprehension questions. However, no impartial observer would designate, as the TRC did, this child as a second-grade reader. He would struggle mightily at this level because of his lack of reading fluency. This case raises an important question: Why does the TRC exclude reading rate as a factor in determining instructional level, especially when the computerized assessment could easily record rate and use it in interpreting the child's performance (see Morris et al., 2013)?

In summary, there were several reasons why the TRC tended to underestimate a child's reading instructional level. However, the factor that stands out in the data, and the one that is hardest to understand, is the TRC's use of written questions to assess the reading comprehension of primary-grade students. Measuring comprehension of short reading passages is a difficult task (see Barr et al., 2007; Paris & Carpenter, 2003). Historically, reading educators have used oral questions to obtain an informal estimate of a child's comprehension of short passages. The TRC's use of written questions introduces a new set of variables, such as (a) can the child read and understand the question and (b) can he or she effectively express his or her understanding in written form? In effect, one must ask if the written comprehension questions are measuring reading comprehension or writing competence? This is a question of crucial importance

since the purpose of the TRC is to determine a student's reading level, not his or her writing level.

How do the TRC and IRI compare regarding administration time?

Throughout the course of data collection, I timed all of the assessments. Results showed that the TRC, when compared to the IRI, took twice as long to administer and score in first grade, three times as long in second grade, and nearly four times as long in third grade. Several characteristics of the TRC contributed to this huge difference in administration time, including (a) long book introductions and picture walks, (b) long passages, (c) multiple passages at each reading level, and (d) written comprehension questions.

In the early TRC levels (kindergarten, early-first grade), the examiner often gave an extensive introduction and the child took a "picture walk" through the book *before* the actual reading began. The picture walk could go quickly or quite slowly, depending on the student. Often, children took their time gazing at each picture before returning to the beginning of the book to start reading orally. Students at the lowest TRC levels (A – E) were required first to orally read the text as the examiner recorded errors on the computer pad; then they were asked to reread the text before giving an oral retelling of the story. These rereads could take considerable time, especially if the child had to move through several early reading passages. (Students at more advanced TRC levels [F and above] did not do picture walks.)

Length of reading passages was another factor that contributed to TRC administration time. At the lower levels (A – D), TRC passage length was reasonable, averaging 85 words. However, average passage length increased to 134 words in late-first

grade, to 190 words at late-second grade, to 195 words at third-grade level and higher. The corresponding passage lengths for the IRI were 100 words at late-first grade, 118 words at second grade, and 144 words at third grade.

Passage length aside, the TRC featured multiple passages at each reading level, which also contributed to the length of the assessment. At first grade, the TRC had seven passages (D – J); the IRI had only three passages (preprimer, primer, and late-first). At second grade, the TRC had four passages, while the IRI had one. And at third grade, the TRC had three passages to the one for the IRI. Consider the following scenario. The examiner starts a first-grade child at TRC level D and moves him up to Level J. This child ends up reading seven passages (five oral questions for each passage and two written questions for each passage above Level E). The same child, on the IRI, reads only three passages (four oral questions for each passage, no written questions). Given the clear difference in number of passages read and questions posed and answered, it is not hard to see why the TRC took much longer to administer.

Finally, the written comprehension component contributed significantly to the length of administration time for the TRC. For each TRC passage (mid-first grade and higher), a student had to answer, in writing, two written questions. There was no time limit, so many children took their time in looking back through the text for information and then composing a written answer. It is true that when a child was working independently on his or her written questions, the examiner could be working with another student on the oral reading part of the TRC. Still, when the first student had finished his or her written questions, the teacher had to stop and score the answers using a somewhat complicated rubric. Juggling the simultaneous testing of two or more children

was challenging for the examiner, and it took time, much more time than the straightforward, one-student-at-a-time administration of the IRI.

Practical Suggestions for Improving the TRC

This study has pointed out some significant problems with the TRC reading assessment as currently designed and used. According to the results, the TRC frequently underestimates primary-grade students' reading instructional levels, and it takes an inordinate amount of time to administer, particularly when compared to the traditional IRI. In the near future, the state of North Carolina appears to be committed to using the TRC for the assessment of student reading performance and the evaluation of teacher effectiveness. If the state plans to continue its partnership with Amplify (formerly Wireless Generation), then the TRC needs to be modified in order to provide a more efficient and valid assessment of schoolchildren's reading. In this section, I mention three possible modifications.

Eliminate written comprehension questions. Answering questions in writing to express one's comprehension of a reading passage is an important skill for students to learn. In fact, when students enter the upper-elementary and middle-school grades, they should be able to formulate a written response to a short answer or essay question. However, primary-grade students, particularly first and second graders, often cannot fully express in writing what they understand in their minds. They are better able to reveal their understanding or comprehension in the oral mode. A reading assessment should assess reading, not writing. Moreover, if a writing assessment strategy is used, this can lead, as these findings show, to an underestimation of a child's true reading ability.

Include rate as a factor when determining a student's reading instructional level. The TRC should include reading rate in its criteria for determining instructional level. Reading rate or fluency is a crucial component of the reading process. Moreover, rate can be assessed easily and reliably with a simple timer. Primary-grade readers need to recognize printed words automatically so that they can devote their attention to the message in the text (LaBerge & Samuels, 1974; Perfetti, 1985). If the child is unable to process text efficiently (at a minimum speed), both comprehension and motivation to read can be adversely affected. The TRC computer already calculates a reading rate for each passage the child reads. It should be a simple matter to begin using this rate information in assigning to the child a reading instructional level.

Use fewer passages at each reading level. Currently, the TRC uses multiple passages, especially at the beginning levels, to determine a student's instructional level. For example, in first grade, the TRC uses seven passage levels, while the IRI does the job with three (beginning, middle, and end of first grade). Similarly, at second grade, the TRC uses four passage levels while the IRI uses only one. Reducing the number of passages at each grade should not affect the diagnostic (or level-setting) effectiveness of the TRC. At the same time, using fewer passages should significantly reduce the time needed to administer the assessment.

Limitations

This study was conducted in an urban elementary school where 95% of the students receive free and reduced lunch. On entering kindergarten, the school's students function at least a year behind State Department of Education expectations. It is possible that the same study, conducted in a school where most of the children read at or above

grade level, would have produced different results—both in the determination of instructional levels (IRI vs. TRC) and in the time needed to administer the tests. That said, there are many schools in the state of North Carolina that are similar demographically to my school, and results in the present study should generalize to these student populations.

A second possible limitation was that the examiners in this study were administering the TRC for the first time (although they did receive considerable training prior to administering the test). It may be that the lengthy administration time of the TRC can be reduced somewhat as examiners gain more experience with the assessment.

A third limitation, at least for some readers, might be that the “arbiter” variable (WR-t) was an isolated measure of automatic word recognition rather than a contextual measure of reading comprehension or fluency. My rejoinder, regarding the use of WR-t as the arbiter variable, is that the loss in complexity (or complication) was a gain in simplicity (or clarity). Moreover, previous work has shown that WR-t is a direct predictor of contextual reading fluency (e.g., Carver, 2000; Morris et al., 2011; 2012), and an indirect predictor of reading comprehension (Perfetti, 1985; 2007).

Conclusions

The purpose of this study was to provide a careful evaluation of a new reading assessment (TRC) that, by state mandate, is currently being used in every primary-grade classroom in North Carolina. My research strategy was to administer the new TRC, along with two other reading assessments (IRI and WR-t), to a large number of children in grades 1 – 3, and then to compare reading performance on the three instruments.

There were two major findings. First, the TRC, compared to the traditional IRI, tended to underestimate students' reading instructional level. There were several reasons for this, notably the TRC's requirement that young children respond in writing to comprehension questions. The purpose of a formative test like the TRC is to help teachers place students appropriately in a leveled reading curriculum. Therefore, the finding that the TRC consistently underestimated the children's reading instructional level is a cause for concern.

The second major finding was that the TRC took a long time to administer, on average, almost three times as long as the IRI. The long administration time was influenced by several factors, including lengthy passage introductions, multiple passages at each reading level, and written comprehension questions. The bottom line, however, is that it should not take up to an hour to assess a child's reading level. The problem is exacerbated when one takes into account that teachers must administer the TRC three times during the school year to each child in the classroom. Such extensive and time-consuming assessment can only take valuable time away from the direct reading instruction that low-reading primary-grade students badly need.

This dissertation is only one evaluation of the TRC—its characteristics and its effectiveness—and, of course, other studies are needed. It is curious, though, that thus far my dissertation appears to be the only comprehensive evaluation of TRC, a lengthy test that is administered three times per year to all primary-grade students in the state. This situation can and should be addressed immediately by more independent studies of the TRC. The overused dissertation phrase, “more research is called for,” truly warrants attention in this case.

References

- Adams, M. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Allington, R. L. (1983). Fluency: The neglected goal. *The Reading Teacher*, 36, 556–561.
- Allington, R. L. (2006). Fluency: Still waiting after all these years. In S. J. Samuels & A. Farstrup (Eds.) *What research has to say about fluency instruction* (pp. 94-105). Newark, DE: International Reading Association.
- Amplify (2014). *Make accurate predictions about reading proficiency*. Retrieved from <http://www.Wireless Generation.com/assessment/mclass-reading-3d>
- Barr, R., Blachowicz, C., Bates, A., Katz, C., & Kaufman, B. (2007). *Reading diagnosis for teachers: An instructional approach* (5th ed.). Boston: Allyn and Bacon.
- Beldin, H. (1970). Informal reading testing: Historical review and review of the research. In W. Durr (Ed.), *Reading difficulties: Diagnosis, correction, and remediation* (pp. 67–84). Newark, DE: International Reading Association.
- Betts, E. (1946). *Foundations of reading instruction*. New York: American Book Company.
- Bond, G., & Tinker, M. (1973). *Reading difficulties: Their diagnosis and correction*. Englewood Cliffs, NJ: Prentice Hall.
- Carver, R. (2000). *The causes of high and low reading achievement*. Mahwah, NJ: Erlbaum.

- Deno, S. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52, 219–232.
- Deno, S., & Marston, D. (2006). Curriculum-based measurement of oral reading: An indicator of growth in fluency. In S. J. Samuels & A. Farstrup (Eds.), *What research has to say about fluency instruction* (pp. 179–203). Newark, DE: International Reading Association.
- Deno, S., Marston, D., Shinn, M., & Tindal, G. (1983). Oral reading fluency: A simple datum for scaling reading disability. *Topics in Learning and Learning Disabilities*, 2, 53–59.
- Excellent Public Schools Act of 2012*. General Assembly of North Carolina Session 2011 Session Law 2012-142, H. B 950, Section 7A (2012) (enacted).
- Fuchs, L., Fuchs, D., Hosp, M., & Jenkins, J. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, 5, 239–256.
- Gillet, J., Temple, C. A., Temple, C., & Crawford, A. (2011). *Understanding reading problems* (8th ed.). New York: Pearson.
- Good, R. H., & Kaminski, R. A. (2003). *DIBELS: Dynamic indicators of basic early literacy skills*. Longmont, Colorado: Sopris West.
- Goodman, Y. M., & Burke, C. (1972). *Reading miscue inventory: Manual for diagnosis and evaluation*. New York: Macmillan.
- Hasbrouck, J., & Tindal, G. (2006). Oral fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher*, 59, 636–644.

- Hendrix, M. P. (2013). *The relationship of prosodic reading to reading rate and other constructs of reading ability* (Unpublished doctoral dissertation). Educational Leadership Doctoral Program, Appalachian State University, Boone, NC.
Retrieved from
http://libres.uncg.edu/ir/asu/f/Hendrix,%20Mary_2013_Dissertation.pdf
- Honey, M. (2007). *The role of formative assessment in pre-k through second grade classrooms*. New York: Wireless Generation.
- Houston, G. (1997). *My great-aunt Arizona*. New York: HarperCollins.
- Johns, J. (2012). *Basic Reading Inventory: Pre-primer through grade twelve and early literacy assessments*. Dubuque, IA: Kendall Hunt Publishing.
- Johnson, M., Kress, R., & Pikulski, J. (1987). *Informal reading inventories*. Newark, DE: International Reading Association.
- Kaminski, R. A., & Good, R. H. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review, 25*, 215-227.
- Kaminski, R. A., & Good, R. H. (1998). Assessing early literacy skills in a problem solving model: Dynamic indicators of basic early literacy skills. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement*, (113-142). New York: Guildford Press.
- Kibby, M. W. (1979). Passage readability affects the oral reading strategies of disabled readers. *The Reading Teacher, 32*, 390-396.
- Killgallon, P. A. (1942). *A study of relationships among certain pupil adjustments in reading situations* (Unpublished doctoral dissertation). Pennsylvania State University, State College, PA.

- Kress, R., & Johnson, M. (1965). *Informal Reading Inventories*. Newark, DE: International Reading Association.
- LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293-323.
- Leslie, L., & Caldwell, J. S. (2010). *Qualitative reading inventory*. New York: Pearson.
- Lipson, M. & Wixson, K. (2009). *Assessment and instruction of reading and writing difficulties: An interactive approach* (4th ed.). New York: Pearson.
- Morris, D. (2014). *Diagnosis and correction of reading problems*. New York: Guilford Press.
- Morris, D., Bloodgood, J., Perney, J., Frye, B., Kucan, L., Trathen, W., Ward, D., & Schlagal, R. (2011). Validating craft knowledge: An empirical examination of elementary-grade students' performance on an informal reading assessment. *The Elementary School Journal*, 112 (2), 205-233.
- Morris, D., Trathen, W., Frye, E., Kucan, L., Ward, D., Schlagal, R., & Hendrix, M. (2013). The role of reading rate in the informal assessment of reading ability. *Literacy Research and Instruction*, 52, 1, 52-64
- Morris, D., Trathen, W., Lomax, R., Perney, J., Kucan, L., Frye, E., Bloodgood, J., Ward, D., & Schlagal, R. (2012). Modeling aspects of print-processing skill: Implications for reading assessment. *Reading & Writing*, 25 (1), 189-215.
- National Center for Education Statistics. (2000). *The nation's report card: Fourth grade reading 2000*. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education. Retrieved from <http://nces.ed.gov/nationsreportcard/pdf/main2000/20001499.pdf>

- National Center for Education Statistics. (2013). *A first look: 2013 mathematics and reading*. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education. Retrieved from <http://nces.ed.gov/nationsreportcard/subject/publications/main2013/pdf/2014451.pdf>
- National Institute of Child Health and Human Development. (2012). *National Reading Panel Report*. Washington, DC: Author. Retrieved from <http://www.nichd.nih.gov/research/supported/Pages/nrp.aspx>
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. Bethesda, MD: Author.
- NC DPI, school report card. Retrieved from North Carolina Department of Public Instruction: <http://www.ncreportcards.org>
- Paris, S., & Carpenter, R. (2003). FAQs about IRIs. *The Reading Teacher*, 56, 578–580.
- Perfetti, C. (1985). *Reading ability*. New York: Oxford University Press.
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11 (4), 357-383.
- Pikulski, J., & Shanahan, T. (1982). Informal reading inventories: A critical analysis. In J. Pikulski & T. Shanahan (Eds.), *Approaches to the informal evaluation of reading* (pp. 94–116). Newark, DE: International Reading Association.
- Pinnell, G., Pikulski, J., Wixon, K., Campbell, J., Gough, P., & Beatty, A. (1995). *Listening to children read aloud*. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

- Powell, W. (1970). Reappraising the criteria for interpreting informal reading inventories. In D. DeBoer (Ed.), *Reading diagnosis and evaluation* (pp. 100–109). Newark, DE: International Reading Association.
- Rasinski, T. V., & Padak, N. D. (1998). How elementary students referred for compensatory reading instruction perform on school-based measures of word recognition, fluency, and comprehension. *Reading Psychology: An International Quarterly, 19*, 185-216.
- Reinking, D., & Bradley, B. A. (2008). *On formative and design experiments: Approaches to language and literacy research*. New York: Teachers College Press.
- Schneider, N. (2013). *Reading fluency assessment: The role of word-level automaticity* (Unpublished doctoral dissertation). Educational Leadership Doctoral Program, Appalachian State University, Boone, NC. Retrieved from <http://libres.uncg.edu/ir/asu/listing.aspx?styp=ti&id=15543>
- Shanahan, T. (1983). A critique of P. A. Killgallon's study: A study of relationships among certain pupil adjustments in reading. In L. Gentile, M. Kamil, & J. Blanchard (Eds.), *Reading research revisited* (pp. 577–582). Columbus, OH: Merrill.
- Southard, M., Diefenbach, B., & Darandari, E. (April 12, 2004). *Does the reading coaches model improve elementary students' reading proficiency?* Paper presented at the meeting of American Educational Research Association, San Diego, California.

- Stanovich, K. (1990). Concepts in developmental theories of reading skill: Cognitive resources, automaticity, and modularity. *Developmental Review, 10*, 72-100.
- Stauffer, R., Abrams, J., & Pikulski, J. (1978). *Diagnosis, correction and prevention of reading disabilities*. New York: Harper & Row.
- Torgesen, J. K. & Hudson, R. (2006). Reading fluency: critical issues for struggling readers. In S.J. Samuels and A. Farstrup (Eds.). *Reading fluency: The forgotten dimension of reading success*. (pp. 130-158). Newark, DE: International Reading Association.
- Wheat, H. G. (1923). *The teaching of reading*. New York: Ginn.
- Williamson, L. K. & Young, F. (1974). The IRI and RMI diagnostic concepts should be synthesized. *Journal of Reading Behavior, 6* (2), 183-194.
- Wireless Generation. (n.d.). mCLASS reading 3D. New York: Author. Retrieved from <http://www.doe.in.gov/sites/default/files/assessment/mclassreading3dbrochure.pdf>
- Wireless Generation. (2008). *Bellevue School District, Bellevue, WA: Strengthening a balanced approach to literacy with mCLASS solutions*. New York: Author. Retrieved from http://www.amplify.com/media/bellevue_school_district.pdf
- Wireless Generation. (2012). *Text reading and comprehension*. New York: Author.
- W. K. Kellogg Foundation. (n.d.). Wireless generation ushers in new wave of educational technology. Battle Creek, MI: Author. Retrieved from <http://www.wkkf.org/what-we-do/featured-work/wireless-generation>
- Woods, M., & Moe, A. (2011). *Analytical reading inventory: Comprehensive standards-based assessment for all students including gifted and remedial*. Boston: Allyn & Bacon.

Zhao, H., & Von Secker, C. (2008). *Evaluation of the criterion-related validity of Montgomery County Public Schools assessment program in primary reading*. Rockville, MD: Montgomery County Public Schools Office of Shared Accountability.

Appendix A
Word Recognition Assessment

Word Recognition Lists (Preprimer through Fifth Grade)

<u>Preprimer</u>	<u>Primer</u>	<u>First</u>	<u>Second</u>
and	back	leg	able
cat	eat	black	break
me	sun	smile	pull
is	bird	hurt	week
go	pat	dark	gate
play	saw	white	felt
where	feet	couldn't	north
like	lake	seen	rush
thing	hid	until	wrote
old	cut	because	perfect
your	about	men	change
up	one	winter	basket
said	rain	shout	shoot
big	water	glass	hospital
for	two	paint	spill
by	how	children	dug
dog	window	table	crayon
not	need	stand	third
who	that's	head	taken
here	mother	drove	prize

Morris, Bloodgood, Perney, Frye, Kucan, Trathen, Ward, & Schlagal (2011).

Word Recognition Lists (continued)

<u>Third</u>	<u>Fourth</u>	<u>Fifth</u>
accept	average	labor
favor	hamster	cripple
seal	select	hasten
buffalo	tobacco	frontier
slipper	brilliant	riverbed
receive	liberty	settlement
legend	prance	absent
haircut	solemn	dissolve
dresser	disease	plea
icy	impress	surrender
customer	miracle	organization
thread	wrestle	evidence
plop	coward	width
bandage	explode	rampaging
further	opinion	horseshoe
moat	suffer	grammar
closet	vast	assorted
unroll	relationship	soybean
storyteller	furnace	troublesome
yarn	clan	circumstance

Morris, Bloodgood, Perney, Frye, Kucan, Trathen, Ward, & Schlagal (2011).

Appendix B
Passage Reading Inventory – IRI

Grade Level	Author, Passage Title, and Publisher
Preprimer	Randell, B. (2003). <i>Baby bear goes fishing</i> . Austin, TX: Rigby.
Primer	Lobel, A. (1978). <i>Mouse tales</i> . (pp. 18-23). New York: HarperCollins.
Late-First	Lobel, A. (1984). <i>Frog and toad all year</i> . (pp. 30-33). New York: HarperCollins.
Second	Woods & Moe (2003). <i>Analytical reading inventory</i> (“Busy Road”). Columbus, OH: Merrill.
Third	Woods & Moe (2003). <i>Analytical reading inventory</i> (“The Cave”).
Fourth	Woods & Moe (2003), <i>Analytical reading inventory</i> (“Crossing the River”).
Fifth	Woods & Moe (2003), <i>Analytical reading inventory</i> (“The Bicycle Race”).

Morris, Bloodgood, Perney, Frye, Kucan, Trathen, Ward, & Schlagal (2011).

Appendix C
Passage Reading Inventory – TRC Levels B – T

Guided Reading Level	Author, Passage Title and, Publisher
B	Smith, A., & Giles, J. (2013). <i>At the zoo</i> . Austin, TX: Rigby.
C	Randell, B. (2006). <i>Wake up father bear</i> . Austin, TX: Rigby. Smith, A. (2013). <i>The big plane</i> . Austin, TX: Rigby.
D	Giles, J. (2000). <i>The merry-go-round</i> . Austin, TX: Rigby. Smith, A. (2002). <i>Nick’s snowman</i> . Austin, TX: Rigby.
E	Randell, B. (2013). <i>The cat and the mice</i> . Austin, TX: Rigby. Giles, J. (2013). <i>Little rabbit’s party</i> . Austin, TX: Rigby.
F	Giles, J. (2013). <i>The helpful bulldozer</i> . Austin, TX: Rigby. Smith, A. (2013). <i>Stuck in the ditch</i> . Austin, TX: Rigby.
G	Giles, J. (2013). <i>Buying a new house</i> . Austin, TX: Rigby. Giles, J. (2013). <i>Tom’s train ride</i> . Austin, TX: Rigby.
H	Smith, A. (2013). <i>The water slide</i> . Austin, TX: Rigby. Randell, B. (2013). <i>The fox and chicken-to-go</i> . Austin, TX: Rigby.
I	Smith, A. (2013). <i>Skip goes to the rescue</i> . Austin, TX: Rigby. Randell, B. (2013). <i>Great lion and tiny mouse</i> . Austin, TX: Rigby.
J	Smith, A. (2013). <i>The greedy dog and the bone</i> . Austin, TX: Rigby. Smith, A. (2013). <i>The vacation surprise</i> . Austin, TX: Rigby.
K	Smith, A. (2013). <i>The old cabin in the forest</i> . Austin, TX: Rigby. Smith, A., & Giles, J. (2013). <i>The skating twins</i> . Austin, TX: Rigby.
L	Giles, J. (n.d.). <i>Kwan the artist</i> . Austin, TX: Rigby. Giles, J. (n.d.). <i>The Wind and the Sun</i> . Austin, TX: Rigby.

Passage Reading Inventory – TRC Levels (continued)

Guided Reading Level	Author, Passage Title and, Publisher
M	Smith, A. (n.d.). <i>Tricks with a kite</i> . Austin, TX: Rigby.
N	Smith, A. (n.d.). <i>The miller, his son, and their donkey</i> . Austin, TX: Rigby.
O	Perkins, L.F. (n.d.). <i>The cave beside the waterfall</i> . Austin, TX: Rigby.
P	Lang, A. (n.d.). <i>Jack and the beanstalk</i> . Austin, TX: Rigby.
Q	Smith, A. (n.d.). <i>A great sense of smell</i> . Austin, TX: Rigby.
R	Kipling, R. (n.d.). <i>Rikki-tikki-tavi</i> . Austin, TX: Rigby.
S	Perkins, L.F. (n.d.). <i>Tracks by the stream</i> . Austin, TX: Rigby.
T	Perkins, L.F. (n.d.). <i>Fording the river</i> . Austin, TX: Rigby.

Vita

Amie Brock Snow is currently a resident of King, North Carolina. She received her Bachelor of Arts in Elementary Education from The University of North Carolina at Chapel Hill in 2002 and her Master of Arts in Reading Education from Appalachian State University in 2006. Dr. Snow serves as the Curriculum Coordinator at Prince Ibrahim Elementary School in Forsyth County, North Carolina, as well as teaches part-time at Appalachian State University in the Department of Reading Education and Special Education. In addition she has recently become the Project Director for the 21st Century Learning Centers Grant for Winston Salem/Forsyth County Schools.

Dr. Snow has received several awards for her work as both a teacher and a student, including the Graduate Student of the Year Award at Appalachian State University and the Teacher of the Year Award at Prince Ibrahim Elementary. She currently serves as the chair of the Intervention Support Team for Ibrahim, leads Professional Learning Team meetings for all classroom teachers and specialists, plans and presents staff development at her current school as well as other schools in Forsyth and Davie Counties.