Florida International University FIU Digital Commons

FIU Electronic Theses and Dissertations

University Graduate School

11-9-2020

Teachers' Use of Standardized Assessments to Monitor Learning and Its Impact on Student Reading Achievement

Eilyn Sanabria Florida International University, esanabri@fiu.edu

Follow this and additional works at: https://digitalcommons.fiu.edu/etd

Part of the Curriculum and Instruction Commons, Elementary Education Commons, Elementary Education and Teaching Commons, and the Language and Literacy Education Commons

Recommended Citation

Sanabria, Eilyn, "Teachers' Use of Standardized Assessments to Monitor Learning and Its Impact on Student Reading Achievement" (2020). *FIU Electronic Theses and Dissertations*. 4553. https://digitalcommons.fiu.edu/etd/4553

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

TEACHERS' USE OF STANDARDIZED ASSESSMENTS TO MONITOR LEARNING AND ITS IMPACT ON STUDENT READING ACHIEVEMENT

A dissertation submitted in partial fulfillment of

the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CURRICULUM AND INSTRUCTION

by

Eilyn Sanabria

To: Dean Michael R. Heithaus College of Arts, Sciences and Education

This dissertation, written by Eilyn Sanabria, and entitled Teachers' Use of Standardized Assessments to Monitor Learning and Its Impact on Student Reading Achievement, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

Mido Chang

Elizabeth Cramer

Teresa Lucas

Joyce Fine, Major Professor

Date of Defense: November 9, 2020

The dissertation of Eilyn Sanabria is approved.

Dean Michael R. Heithaus College of Arts, Sciences and Education

Andrés G. Gil Vice President for Research and Economic Development and Dean of the University Graduate School

Florida International University, 2020

© Copyright 2020 by Eilyn Sanabria

All rights reserved.

DEDICATION

To Henry:

Your strength and determination were an example to follow through this endeavor.

You were right: slow and steady does win the race.

I love you.

ACKNOWLEDGMENTS

I would like to begin by thanking my committee members for helping me navigate through this journey. Dr. Fine, thank you for always being there with kind words, an open heart, and an insurmountable knowledge of the field to help guide me through this process. Your expertise allowed me to merge my two passions: reading education and assessment of student learning. I deeply appreciate your guidance in answering the millions of questions I have had throughout this journey. The road has been far from easy, but you have always been there to guide and support me. I am the reading teacher I am because of you; I could have not asked for a better mentor. Dr. Chang, thank you for your always receiving me with a smile and for your invaluable feedback to make my research stronger. Dr. Cramer, thank you for your feedback and dedication – you have no idea how much I enjoyed the courses I took with you and your honest, down-to-earth teaching style. Dr. Lucas, thank you for agreeing to come on board and dedicating time to help me. I appreciate and admire each of you.

I would also like to thank my parents, my husband, and two ladies whom I love like sisters. Mami and Papi, thank you for all you have done (and continue to do) for me and for supporting this dream of mine. I am finally done going to school! Ismael, thank you for being my cheerleader and laughing at (and usually even supporting) my craziness. You brought more laughter, adventure, and love to my life than I thought possible. I hope to continue to love, laugh, dance, and explore the world together. Lula, thank you for always standing by my side and not letting me give up in some of my darkest hours. Kathy, thank you for believing in me, talking to me endlessly, and soothing my stubbornness. I love you all immensely.

v

Last, I would like to thank my APA family. I am beyond blessed to work with such a kind and supportive team. There are too many names to mention, but you all have helped me grow, both professionally and personally, in the best of ways. Susan, you will be dearly missed, but I will never forget some of the conversations we have had – thank you for being so kind.

ABSTRACT OF THE DISSERTATION

TEACHERS' USE OF STANDARDIZED ASSESSMENTS TO MONITOR LEARNING AND ITS IMPACT ON STUDENT READING ACHIEVEMENT

by

Eilyn Sanabria

Florida International University, 2020

Miami, Florida

Professor Joyce Fine, Major Professor

Reading instruction must be "intentional, systematic, and explicit" and "implemented by a knowledgeable teacher" (Ruetzel & Cooter, 2019, p. 87). The era of accountability has brought standardized assessments to the forefront of reading instruction. However, gaps about assessment-related and instructional practices and their impact on student achievement exist in the literature. The present study aims to provide needed insights on how these practices help or hinder, specifically, historically lowperforming students.

Using student achievement and teacher survey data from the Early Childhood Longitudinal Study: Kindergarten 2011 (ECLS-K), and through the lens of data use theory (Hutchins, 1995; Spillane, 2012), hierarchical regression analyses were run to examine the relationship between fifth grade teachers' (a) use of standardized tests to monitor learning and students' reading achievement (b) reading instructional practices and use of standardized tests to monitor learning, and (c) reading instructional practices and students' reading achievement, all while controlling for students' socioeconomic status and race/ethnicity and teachers' assessment-related professional development

vii

participation. Results indicate that, though each of the three models presented is statistically significant (p < .05), historically underserved minority students continue to be at a disadvantage as it relates to standardized testing, paving the way for the reconceptualization of assessment instruction in teacher preparation programs and in professional development opportunities.

TABLE OF CONTENTS

CHAP	YTER	PAGE
I.	INTRODUCTION	1 4
П.	REVIEW OF THE LITERATURE An Era of Accountability National Landscape on Student Reading Achievement Teacher Preparation Preservice Teacher Preparation In-service Teacher Professional Development Instructional Practices and Student Characteristics Assessment Assessment of Learning and Assessment for Learning Formal Assessments and Teacher Judgments Conceptual Framework	11 13 13 13 14 15 17 19 19 19 20 21 23
III.	METHODOLOGY Research Design Research Questions Hypotheses Participants Instruments Direct Cognitive Assessment Teacher Questionnaire Statistical Treatments Limitations Summary	25 25 26 27 27 27 27 27 29 29 23 33 33
IV.	RESULTS Descriptive Statistics Student Achievement Students' Socioeconomic Status Students' Race/Ethnicity Teachers' Frequency of Use of Standardized Tests	

Teachers' Assessment-Related Professional Development	
Participation	40
Teachers' Reading Instructional Practices	43
Examining Correlations and Creating Constructs for Survey Items	49
Teachers' Frequency of Use of Standardized Tests Survey Items	50
Teachers' Assessment-Related Professional Development Participation	n
Survey Items	53
Teachers' Reading Instructional Practices Survey Items	55
Results of the Study	59
Frequency of use of standardized tests and reading achievement	60
Instructional Practices and Frequency of Use of Standardized Tests	70
Instructional Practices and Student Achievement	76
Summary	85
-	

V.	DISCUSSION	86
	Summary of Findings	86
	Implications	
	Limitations and Recommendations for Future Research	91
	Summary	
	REFERENCES	

VITA102

LIST OF TABLES

TABLE PAGE
1. Teachers' Survey Items
2. Descriptive Statistics: IRT Theta Scores
3. Descriptive Statistics: Teachers' Judgments of Student Reading Level35
4. Descriptive Statistics: Students' Socioeconomic Status
5. Descriptive Statistics: Students' Race/Ethnicity
6. Descriptive Statistics: Teachers' Frequency of Use of Standardized Tests
 Descriptive Statistics: Use of Standardized Tests to Evaluate How Well Students Respond to the Core Curriculum
8. Descriptive Statistics: Use of Standardized Tests to Monitor Progress
9. Descriptive Statistics: Use of Standardized Tests to Identify Reading Deficits39
 Descriptive Statistics: Use of Standardized Tests to Monitor Progress of Students Below Benchmarks
11. Descriptive Statistics: Use of Standardized Tests to Determine Placement40
 Descriptive Statistics: Teachers' Assessment-Related Professional Development Participation
 Descriptive Statistics: Professional Development Activities on How to Use Data to Identify Struggling Students
14. Descriptive Statistics: Professional Development Activities on How to Use Data to Guide Reading Instruction
15. Descriptive Statistics: Professional Development Activities on How to Implement the Reading Curriculum
16. Descriptive Statistics: Teachers' Reading Instructional Practices with Literature
17. Descriptive Statistics: Teachers' Reading Instructional Practices with Literary Text: Understanding Key Ideas and details44

18.	Descriptive Statistics: Teachers' Reading Instructional Practices with Literary Text: Understanding Craft and Structure	.45
19.	Descriptive Statistics: Teachers' Reading Instructional Practices with Literary Text: Integrating Knowledge and Ideas	.46
20.	Descriptive Statistics: Teachers' Reading Instructional Practices with Informational Texts	.46
21.	Descriptive Statistics: Teachers' Reading Instructional Practices with Informational Text: Understanding Key Ideas and details	.47
22.	Descriptive Statistics: Teachers' Reading Instructional Practices with Informational Text: Understanding Craft and Structure	.48
23.	Descriptive Statistics: Teachers' Reading Instructional Practices with Informational Text: Integrating Knowledge and Ideas	.49
24.	Correlations: Frequency of Use of Standardized Tests Survey Items	.51
25.	KMO and Bartlett's Test: Frequency of Use of Standardized Tests Survey Items	.52
26.	Communalities: Frequency of Use of Standardized Tests Survey Items	.52
27.	Total Variance Explained: Frequency of Use of Standardized Tests Survey Items	.52
28.	Reliability Statistics: Frequency of Use of Standardized Tests Survey	.52
29.	Correlations: Professional Development Survey Items	.53
30.	KMO and Bartlett's Test Professional Development Survey Items	.54
31.	Communalities: Professional Development Survey Items	.54
32.	Total Variance Explained: Professional Development Survey Items	.54
33.	Reliability Statistics: Professional Development Survey Items	.54
34.	Correlations: Instructional Practices with Literary Texts	.56
35.	KMO and Bartlett's Test: Instructional Practices with Literary Texts	.56
36.	Communalities: Instructional Practices with Literary Texts	.56

37. Total Variance Explained: Instructional Practices with Literary Texts
38. Reliability Statistics: Instructional Practices with Literary Texts
39. Correlations: Instructional Practices with Informational Texts
40. KMO and Bartlett's Test: Instructional Practices with Informational Texts
41. Communalities: Instructional Practices with Informational Texts
42. Total Variance Explained: Instructional Practices with Informational Texts59
43. Reliability Statistics: Instructional Practices with Informational Texts
44. Tests of Normality: IRT Theta Scores
45. Tests of Normality: Teachers' Judgments of Student Reading Level
46. Model Summary: Frequency of Use of Standardized Tests and Reading Achievement (IRT Theta Scores)
47. ANOVA: Frequency of Use of Standardized Tests and Reading Achievement (IRT Theta Scores)
48. Effects of Frequency of Use of Standardized Tests on Reading Achievement (IRT Theta Scores)
49. Model Summary: Frequency of Use of Standardized Tests and Reading Achievement (Teacher Judgements)
50. ANOVA: Frequency of Use of Standardized Tests and Reading Achievement (Teacher Judgements)
51. Effects of Frequency of Use of Standardized Tests on Reading Achievement (Teacher Judgements)
52. Tests of Normality: Frequency of Use of Standardized Tests
53. Model Summary: Instructional Practices and Frequency of Use of Standardized Tests
54. ANOVA: Instructional Practices and Frequency of Use of Standardized Tests73
55. Effects of Instructional Practices on Frequency of Use of Standardized Tests74

56. l (Model Summary: Instructional Practices and Student Achievement (IRT Theta Scores)	7
57. I	Effects of Instructional Practices on Student Achievement (IRT Theta Scores)7	8
58. I	Effect of Instructional Practices on Student Achievement (IRT Theta Scores)7	8
59. I (Model Summary: Instructional Practices and Student Achievement (Teacher Judgment)	1
60. <i>1</i> (ANOVA: Instructional Practices and Student Achievement (Teacher Judgment)	2
61. l (Effects of Instructional Practices on Student Achievement (Teacher Judgment)	3

CHAPTER I

INTRODUCTION

"Learning is driven by what pupils and teachers do in classrooms" (Black & William, 1998, p. 140). Classrooms, described as a "black box" by Black and William because of the obscurity of what exactly happens inside that actually helps students succeed, receive "inputs" (p. 140) from stakeholders (e.g., teachers) with the expectation that specific "outputs" (p.140) follow. These outputs are, partly, dependent on teachers' knowledge, skills, and behaviors (Black & William, 1998), as well as on students' individual characteristics.

Teacher preparation has been at the forefront of the educational debate for several decades, beginning with the publication of the Coleman Report (Coleman et al., 1966). Later, the enactment of the *No Child Left Behind Act* (NCLB) of 2001 (2002) continued to emphasize the need for improving teacher quality across the nation and required states to develop measures to ensure teacher effectiveness. Likewise, President Barack Obama's *Race to the Top* initiative focused on "recruiting, developing, rewarding, and retaining effective teachers" (U.S. Department of Education, 2009, p. 2), and the proceeding *Every Student Succeeds Act* (ESSA) signed by the President on December 10, 2015, called for further attention to teacher preparation (*Every Student Succeeds Act*, 2015). Through NCLB, *Race to the Top*, and ESSA accountability expectations officially trickled down to the classroom, one of the requirements being teachers use data to develop, and justify, instructional decisions to increase student achievement. Hence, the literature discusses the pressing need for teachers to be able to gather and interpret data as a necessary

instructional practice (Black et al., 2011; DeLuca & Johnson, 2017; Gómez-Monarrez et al., 2019; Livingston & Hutchinson, 2017; Mandinach, 2012; Reutzel & Cooter, 2019).

Teachers of reading play a major role in students' attainment of literacy skills by ensuring all students, regardless of ability, are provided with individualized instruction rooted in content and pedagogically-appropriate best practices across all six components of reading (phonemic awareness, phonics, vocabulary, fluency, comprehension, and oral language). To provide this individualized instruction, teachers must be (among other things) knowledgeable of, and implement, various assessment practices that allow for effective progress monitoring and provide insights on how reading instruction should be adjusted (Black & William, 1998). Formative reading assessments, or assessments for learning (Bloom et al., 1971; Stiggins, 2002), are amongst these data-driven monitoring practices. These instruments provide data on how students are progressing through the curriculum and insights on what adjustments, if any, need to be made. Formative assessments can take two forms: formal or informal. Formal assessments (i.e., standardized tests) have been screened for validity and reliability and have set criteria for scoring and interpreting data (Navarrete, et al., 1990). Informal assessments, on the other hand, refer to open-ended measures (e.g., portfolios, journals) (Navarrete et al., 1990). Though both formal and informal formative assessments can yield meaningful data on student learning, some (e.g., Afflerbach & Cho, 2011) criticize formal assessments by arguing they disregard classroom reading contexts, individual student characteristics, and learning experiences. However, as of 2019, 26 states continue to use student achievement data from standardized tests as a component of teacher evaluations (Ross & Walsh, 2019), making the use of the data they yield an (almost) necessary practice within

teachers' instructional decision-making repertoire. A formative data-driven process, through the use of both formal and informal assessments, is supported by the International Literacy Association's (ILA) 2017 *Standards for the Preparation of Literacy Professionals* (ILA, 2017), 2019 Florida Statutes (Assessment and Accountability, 2019), as well as existing literature (Black & William, 1998; DeLuca & Johnson, 2017; DeLuca & Klinger, 2010; Hill et al., 2017; Livingston & Hutchinson, 2017; Ross & Walsh, 2019).

As previously mentioned, effective teachers have appropriate preparation and knowledge in the field to monitor student learning. Yet, the literature has mostly focused on direct relationships between student achievement and standardized tests (e.g, Dee & Jacob, 2011) without controlling for student or teacher characteristics, including teachers' assessment and instructional practices. The present study investigated (a) the impact of fifth grade teachers' use of standardized tests to monitor learning (a formative practice) on students' reading achievement while controlling for student and teacher characteristics and (b) the mediating role of fifth grade teachers' reading instructional practices on the relationship between their use of standardized tests to monitor learning and students' reading achievement (see *Figure 1*).

Figure 1

Relationship between Standardized Testing and Student Achievement (adapted from Im (2017))



Statement of the Problem

The 1970s brought about a nationwide movement for accountability in elementary and secondary education, with 33 states adopting "competency testing" (Vinovskis, 2019, p. 30) and holding students, teachers, schools, and (ultimately, by the federal government) states accountable for student achievement (Vinovskis, 2019). Since then, the landscape has continued to emphasize accountability across all levels (classroom, school, district, state, national, international; DeLuca & Johnson, 2017, Mandinach, 2012; Mandinach et al., 2006), with several mandates (e.g., NCLB, ESSA) in its support. Accountability, thus, is an integral part of the current teaching and learning environment, with assessment of student learning at its forefront.

Aside from serving as a reporting mechanism to stakeholders, assessment (when done appropriately) allows teachers to better understand student development and make instructional and diagnostic decisions (Afflerbach & Cho, 2011). In 1990, *the Standards*

for Teacher Competence in Educational Assessment of Students, published by the National Council on Education, described expectations of assessment-related knowledge teachers should have, emphasizing the ability to collect and analyze assessment data to make instructional decisions about student learning. Stiggins (1991) then coined the term "assessment literacy" (AL) to describe the assessment-related knowledge, skills, and dispositions stakeholders (e.g., teachers) should possess to impact student achievement. Research thereafter has continued to explore teachers' assessment literacy through different lenses (e.g., Xu & Brown, 2016), with the publication of a new set of standards for classroom assessment (Klinger et al., 2015) continuing to emphasize the pivotal role effective assessment plays in student learning.

Literature in the area of teacher knowledge-base and use of assessment measures to help struggling students depicts a worrisome picture and questions the preparation of pre and in-service teachers. A study conducted by Melnick and Meister (2008) with both pre- and in-service teachers, found that less than half (47%, n = 218) of experienced teachers surveyed felt prepared to use multiple assessment measures. Spear-Swerling and Cheesman (2012) found that pre- and in-service teachers demonstrated low performance on questions related to diagnosis and intervention. McCombes-Tolis and Spear-Swerling (2012) reviewed course syllabi (N = 29) of required reading courses from teacher preparation programs in one state and found that no syllabi referenced "literacy progress-monitoring measures" (p. 372). More recently, a literature review of 122 empirical articles conducted by Gómez-Monarrez et al., (2019) to identify effective teaching practices found that less than half (35%) discussed the use of evaluation to adjust

instruction and only (49%) discussed the use of evaluation to confirm learning and provide feedback.

However, though the development of standards (e.g., Klinger et al., 2015; American Federation of Teachers, 1990) and assessment of teachers' AL (e.g., Spear-Swerling & Cheeseman, 2012) are steps in the right direction to better understand the current data-driven instructional environment, transfer to instructional practices is not guaranteed because, as Xu and Brown (2016) argue, they do not "inherently advance teacher AL" (p. 153). The study of the transfer of knowledge to instructional practices, factors that influence it and, more importantly, its impact on students' reading achievement, has limited empirical evidence in the elementary grades (Hao & Johnson, 2013). A review of the literature found two recent attempts at addressing this need (Hao & Johnson, 2013; Im, 2017). Hao and Johnson (2013) examined the relationship between teachers' use of different types of assessment instruments (e.g., multiple choice, performance-based) and fourth grade students' reading achievement, gender, reading self-concept, and attitudes towards reading using data from the Progress in International Literacy Study (PIRLS) for Canada, New Zealand, England, and the United States. Results indicate teachers' reading assessment practices vary across nations and across reading constructs. Though Hao and Johnson's (2013) study advanced the knowledgebase on the impact of assessment practices on student achievement, there remained several areas to be addressed. First, they did not control for other important student characteristics aside from gender, such as students' socioeconomic status and ethnicity. As per the National Center for Education Statistics (2019), the reading achievement gap continues amongst fourth grade White, Black, and Hispanic students

and students eligible for the National School Lunch Program (NSLP), a meal program that provides free or reduced-cost meals to students determined by family's income. Second, the assessment practices employed focused on the *type* of instruments used by teachers (e.g., short answer written questions, oral questioning of students, etc.), not on the use of assessment instruments to monitor learning and make data-driven instructional decisions, a need identified in the literature (DeLuca & Johnson, 2017; DeLuca & Klinger, 2010; Hill et al., 2017; Livingston & Hutchinson, 2017; Werts et al., 2014). Lastly, it did not explore teachers' assessment-related preparation, which could impact assessment practices and, in turn, students' reading achievement (Curry, 2014). Four years later, some of these limitations were addressed by Im (2017) through the investigation of use of standardized tests to monitor learning and its impact on kindergarten student reading achievement while controlling for several instructional, student, and school-level covariates. Im's (2017) results suggest that, when effective reading instruction is implemented, the use of standardized tests to monitor learning does impact student achievement, leading the author to emphasize the implications on teachers' assessment-related professional development, variables that were not controlled for in the study even as some (e.g., traditional preparation route, credentials) have been found to impact elementary students' reading achievement (Curry, 2014). Another limitation of Im's (2017) study are the three philosophical orientations for teaching used when considering items related to kindergarten instructional practices (whole language, phonics, and balanced approach). These limit results' generalizability to other grade levels because of their grade-level specificity. Lastly, both studies (Hao & Johnson, 2013 and Im, 2017) only used a standardized measure of student reading achievement.

Incorporating teachers' judgments on students' reading proficiency, in addition to their standardized test scores, would likely yield a more robust analysis than Hao and Johnson's (2013) and Im's (2017) studies. On the basis of the current data-driven educational climate and the aforementioned limitations of the two recent studies of elementary teachers' assessment practices and student achievement, the field would benefit from examining how upper elementary teachers, while controlling for student and teacher characteristics, impact student achievement, as well as how instructional practices help mediate the use of standardized assessment practices to monitor learning and student achievement (see *Figure 1*).

Purpose of the Study

"Effective instruction in reading begins with assessment" (Cooter & Perkins, 2007, p. 6). The need to effectively scaffold learning to improve student achievement is an international practice that has been discussed in the literature (e.g., Hill et al., 2017), and assessment works hand-in-hand with instruction in the scaffolding of learning. The present study helps advance the field of elementary reading instruction by filling gaps within the literature about assessment-related and instructional practices and their impact on student achievement. Results of the present study also provide needed insights on how these practices help or hinder, specifically, historically low-performing students.

Research Questions

The present study seeks to answer the following research questions framed following the model presented in *Figure 1*:

Question 1: Does fifth grade teachers' use of standardized tests to monitor learning relate to students' reading achievement when controlling for students'

socioeconomic status and race/ethnicity and teachers' assessment-related professional development participation (see *Figure 1*)?

Question 2: Do fifth grade teachers' instructional practices relate to their use of standardized tests to monitor learning after controlling for the effects of students' socioeconomic status and race/ethnicity and teachers' assessment-related professional development participation

(see *Figure 1*)?

Question 3: Do fifth grade teachers' instructional practices relate to student achievement after controlling for the effects of students' socioeconomic status and race/ethnicity and teachers' assessment-related professional development participation (see *Figure 1*)?

Assumptions

As the ECLS-K: 2010-2011 manual (Tourangeau et al., 2019) does not provide a definition of formal assessments, the present study assumed they are equivalent to standardized tests, as is discussed in the literature (Navarrete et al., 1990).

Delimitations

The present study used the fifth grade Early Childhood Longitudinal Study -Kindergarten 2010-2011 (ECLS-K: 2010-2011) data set, which is delimited to fifth grade teachers' self-reporting on the teacher child-level questionnaire and fifth grade students' responses to the direct cognitive reading assessment administered as part of the ECLS-K: 2010-2011 program, both conducted in the spring of 2016 (public-use data file was released July 12, 2019). Using this large data set allowed for a decrease in the standard error of the sample, increased statistical power, and an accurate effect size, important components in educational research when drawing conclusions (Anderson & Maxwell, 2018).

Definitions and Operational Terms

Assessment Literacy (AL)

The "basic understanding of educational assessment and related skills to apply such knowledge to various measures of student achievement" (Xu & Brown, 2016).

Formal Assessment/Standardized Test

A formal assessment/standardized test is "designed to provide the best match possible to

what is perceived to be the "typical" curriculum at a specific grade level...assume a single set of expectations for all students and come with prescribed criteria for scoring and interpretation" (Navarrete et al., 1990, pp.5-6). The ECLS-K manual (Tourangeau et al., 2019) does not provide their definition for formal assessments.

Summary

Chapter I provided an overview of how the teacher preparation landscape has evolved to emphasize data-driven instructional practices and a discussion of how standardized tests are expected to be part of the data-driven decision-making process. Recent research in the area (Hao & Johnson, 2013; Im, 2017), their limitations, and existing gaps in the literature, were also discussed. Finally, the chapter provided the purpose of the study, research questions, assumptions, delimitations, and definitions.

CHAPTER II

REVIEW OF THE LITERATURE

An Era of Accountability

In 1983, A Nation at Risk, published by the National Commission on Excellence in Education (1983), deemed our educational system as "mediocre" (p. 113), which had negative long-range financial implications for us. Consequently, this report set the stage for vast school reforms across states, the development of reading standards, and determining means of assessing achievement, among them (Shanahan, 2014). In 2002, however, oversight shifted to the federal government with the passing of the No Child Left Behind Act of 2001 (NCLB). Tied to funding, this legislation focused on eliminating the achievement gap across different groups of students through the development of statelevel achievement standards, employment of "highly qualified teachers", as well as other accountability measures (NCLB, 2002; Ryan, 2004). The aim was for all students across the nation to be proficient in reading and math by the year 2014 (NCLB, 2002). NCLB also required states develop standardized measures to assess student attainment of the standards, with an emphasis on teacher quality, research-based instructional programs, and school-level accountability by demonstrating "adequate yearly progress" (Ruff, 2019, p. 3). As with any other legislation, NCLB had its critics and supporters. Ryan (2004) argued that it led to the lowering of achievement expectations across states and the promotion of school segregation (Ryan, 2004). On the other hand, Shanahan (2014) describes there being limited teacher pushback regarding the implementation of these standards and hypothesized the "substantial research supporting the effectiveness of these

approaches" (p. 10) as a possible reason. Overall, however, NCLB was instrumental in shaping the standardized test environment in K-12 classrooms across the nation.

In 2015, under President Barack Obama's urging, Congress reissued NCLB with The Every Student Succeeds Act (ESSA). Though ESSA still holds states accountable for ensuring and demonstrating 3rd through 8th grade students across their schools are learning, it provides a more flexible framework to achieve this. ESSA does not penalize schools who are struggling, but rather requires that an assessment plan using evidencebased methods be developed and implemented. Standardized testing remains a component but, as opposed to having to develop their own instruments, states can instead opt to use nationally-recognized tests to demonstrate student achievement (ESSA, 2015). This allows states a wider range of selection for metrics of proficiency that are wellaligned with the foundation of their standards. It also could, potentially (since instruments chosen would likely be a valid measure of assessing the instruction that takes place in the classroom) help remediate the "teaching to the test" phenomenon (Popham, 2001), which deviates instruction from "the body of knowledge or skills" assessed (p. 16) and focuses it on specific test items instead. Also, contrary to NCLB, ESSA acknowledges the impact of unnecessary testing in the classroom, and thus awards grants for states to conduct audits of their assessment systems across districts and schools (ESSA, 2015) -- this helps further align instruction and assessment. More importantly, ESSA calls for the improvement of instruction and learning by providing teachers with research-based professional development on how to "use data to improve student achievement" (p. 1928), indicating the legislature acknowledges the role systematically monitoring learning plays on student achievement.

National Landscape on Student Reading Achievement

Since 1992, the National Assessment of Educational Progress (NAEP) has been assessing fourth, eighth, and 12th grade students' reading proficiency periodically, with students in fourth and eighth grade assessed more often. The last assessment conducted for fourth grade students, in 2017, did not show significant average score changes from the previous administration in 2015 and shows reading scores have been relatively stagnant since 2007. On a scale of 0-500, the average fourth grade student reading score from 2007 through 2017 has ranged between 221 and 223 (McFarland et al., 2019). Essentially, as of 2017, fourth grade students' reading abilities have not improved, even with the shift in educational reform from NCLB to ESSA. In addition to average scores, NAEP also breaks down average scores by percentiles for the lowest, middle, and highest-performing students. Fourth graders at the 10th and 25th percentile, categorized as the lowest percentile groups, performed lower in 2017 than in 2015 and, most importantly, the score from students at the 10th percentile was not significantly different than the 1992 score, the first administration of this assessment (McFarland et al., 2019). This indicates we, as a nation, continue failing in improving the reading abilities of our most struggling students even with the increased accountability requirements the field has experienced during the last 20 years.

Teacher Preparation

The Coleman et al. (1966) report sparked an ongoing national discussion on teacher preparation: our teachers were not prepared, nor had the resources, to teach our diverse student population. Several initiatives by presidents (No Child Left Behind Act of 2001, 2002; Every Student Succeeds Act, 2015) and renowned organizations

(International Literacy Association, 2017) have echoed this sentiment by providing provisions and/or guidelines on how to help in-service teachers and strengthen our teacher education programs in areas related to student achievement, data-driven decisionmaking being one of them. The debate, however, is twofold: teacher quality as it relates to teacher preparation programs and their assessment content and teacher quality as it relates to in-service teachers' assessment preparation. Nonetheless, the literature (e.g., DeLuca & Klinger, 2010; Hill et al., 2017; Melnick & Meister, 2008; Spear-Swerling & Cheesman, 2012) has identified the need for strengthening both, as they both have implications on student achievement.

Preservice Teacher Preparation

Preservice teacher preparation is crucial in ensuring our incoming workforce is equipped to address the varying instructional needs of our students. Several other studies (in addition to the ones previously mentioned while exploring the current problem) have explored preservice teachers' knowledge, abilities, and dispositions towards assessment, as well as reviewed the preservice teachers' reading preparation at a national level. In a study conducted in Canada, DeLuca and Klinger (2008) surveyed 288 preservice teachers on their perceived confidence in areas related to assessment practice, theory, and philosophy. Findings of this study call for teacher preparation programs to provide "direct instruction in assessment" (p. 419) and emphasize "the learning potential of assessment while also monitoring student progress" (p. 420). In 2017, Hill et al. examined the assessment learning of 27 preservice teachers enrolled in a new Master of Teaching program in New Zealand by using surveys, reviewing artifacts, and conducting focus groups. This program stemmed from a Ministry of Education call for teacher

preparation institutions to develop programs that specifically prepare teachers to address the learning needs of New Zealand's historically underachieving populations. The program's curriculum, which is guided by six principles, evolved after the review of international assessment frameworks. The fourth principle, use of evidence to scaffold learning, was the focus of Hill et al., (2017) research. As results indicate teachers use a combination of theory and practice to understand and address the learning needs of struggling students, the authors call for these two components to be embedded *across* all teacher preparation courses rather than teaching it in isolation. These findings, as Hill et al., (2017) discuss, further reiterate the view of assessment as a formative process, where "information gathered from both formal instruments...and from day to day, minute by minute observations" (p. 187) is used to make instructional decisions about student learning. Recent findings from a review of 1,047 teacher preparation programs in the United States by the National Council on Teacher Quality (NCTQ) (Drake & Walsh, 2020) support these results. After reviewing traditional, undergraduate, and graduate programs' curricula and approaches to teaching phonemic awareness, phonics, fluency, vocabulary, and comprehension, as well as the materials and opportunities provided to their students, 49%, 43%, and 67% of traditional, undergraduate, and graduate programs, respectively, did not address all areas (Drake & Walsh, 2020). This indicates gaps in the reading instruction knowledge and abilities of our existing and future teachers, which can hinder student achievement.

In-service Teacher Professional Development

A current report published by NCTQ reflects nationwide changes in teacher evaluations (e.g., supplemental measures of evaluation, frequency, etc.), per state, that

allow for better understanding of teachers' preparation and provide insights on current gaps in teacher evaluation systems (Ross & Walsh, 2019). As such, professional development plays a major role in ensuring teachers are up-to-date with instructional practices and gaps in knowledge and abilities are filled. In 2008, Melnick and Meister sought to validate Meister and Jenks' (2000) qualitative study, which examined beginning teachers' (N = 42) experience during their first year. This was done through a survey of preservice (N = 273) and in-service (N = 218) teachers that asked questions related to classroom management, workload, parent interaction, and academic preparation. Results indicate beginning teachers feel more prepared than experienced teachers to use multiple assessment methods, thus demonstrating the need to ensure our current teachers are provided with appropriate professional development in this area. Livingston and Hutchinson (2017) further explored in-service teacher preparation by examining issues related to teachers' instructional practices, capacity to use assessment to improve learning, and their understanding of assessment and its impact on student achievement. The authors situated professional development in these areas within the Career-Long Professional Learning (CLPL) context and explain how, in order for CLPL to be effective, teachers' knowledge of assessment needs to be translated to actions where practice is recognized as an "ongoing enquiry" (p. 299) dependent on the student, classroom, and school contexts and, thus, "tailored professional learning opportunities" (p. 303) are necessary. The argument by Timperley et al. (2007) supports these findings in that, in order for teachers to change their assessment practices, they need to feel part of a professional community that values not only student learning, but also their individual development as professionals. Timperley' et al. (2007) synthesized findings from

international and New Zealand studies on teachers' knowledge base to impact student achievement. One last example of how in-service teachers' professional development can help improve student achievement is Brown's (2004) study. Brown (2004) surveyed 525 elementary teachers and administrators on their conceptions of assessment and found not only that teachers agreed assessment improves teaching and learning, but that this paralleled with their rejection of assessment for accountability. Thus, implications of this study stress that professional development, aside from addressing content and pedagogical content knowledge, should also address teachers' conceptions of assessment as a means of improving teaching and learning in the elementary grades.

Instructional Practices and Student Characteristics

Comprehension refers to the understanding of printed text – it is the ultimate goal of reading instruction (Tompkins, 2017). As defined by Paris and Hamilton (2014), it is "a subset of an ill-defined larger set of knowledge that reflects the communicative interactions among the intentions of the author/speaker, the content of the text/message, the abilities and purposes of the reader/listener, and the context/situation of the interaction" (p. 32). As such, reading encompasses a variety of skills, concepts, and strategies (Reutzel & Cooter, 2019) students must employ as they progress through grade levels, encounter new material, and acquire new knowledge. In the early grades, instruction is focused on helping students *learn to read*, which is done (mostly) with narrative texts (i.e., fiction). By third grade, instruction shifts to *reading to learn* (Palacios, 2017; Toste & Ciullo, 2017), which occurs through the interaction with informational texts (i.e., non-fiction). Both of these types of texts, however, require learners understand and apply higher-order skills (e.g., analysis, inference) to create

meaning. The importance of ensuring students acquire these skills is reflected on educational standards. For example, the fifth grade Language Arts Florida Standards (LAFS) has specific standards for both narrative and informational text. Hence, this requires teachers ensure, through the implementation of effective instructional practices and monitoring of learning, students attain these skills.

The implementation and frequency of these practices, aside from being viewed by some as a quality of effective teachers (Palacios, 2017), is particularly important for lowincome students in third through fifth grade. Described as the "fourth grade slump" (Chall & Jacobs, 2003; Hirsch, 2003), this phenomenon seeks to understand why, beginning in fourth grade, reading comprehension "slumps", especially for low-income students. This is supported by recent fifth grade reading achievement data published by NCES where fifth grade students who were living below the poverty level in kindergarten scored lowest amongst all income groups assessed (Mulligan et al., 2019). An explanation for this decline is the increased difficulty in vocabulary encountered as the grade levels progress and the text shifts from narrative to informational, which puts lowincome students at a disadvantage (Chall & Jacobs, 2003) due to their limited "world knowledge" (Best et al., 2004), impairing their comprehension. Consequently, targeted interventions should be implemented, and have been found to be effective, in helping students overcome this slump (Best et al., 2004; Palacios, 2017). For example, a study (Lysynchuk, 1990) conducted with fourth and seventh grade students struggling with comprehension used reciprocal teaching of comprehension strategies as an intervention and found students who received this treatment had an increase in standardized test scores. More recently, Vernon-Feagans et al., (2018) found that students whose teachers

had participated in the Targeted Reading Intervention (TRI) program, which prepared rural elementary teachers to provide diagnostic differentiated instruction to struggling readers, performed higher than those in the control group in the areas of letter-word identification, word attack, spelling of sounds, and comprehension. Assessment, thus, plays a major role in determining how instructional practices should be adjusted.

Assessment

As Livingston and Hutchinson (2017) argue "how best to gather evidence about learning, and how to interpret and use that evidence to plan for better learning" (p. 291) is essential to both teacher and student success. Teaching and learning are interactive processes, with assessment serving as a main component of this interaction (Black & William, 1998). Assessment is the "mechanism for gauging progress" (Banta et al., 2009, p. 4) and encompasses all activities stakeholders engage in that yield information about teaching and learning (Black & William, 1998). When assessment is viewed as a means for improvement, it must "describe or diagnose the nature of student performance" and data derived "must be a valid, reliable, and an accurate description of student performance" (Brown, 2004, p. 304; Afflerbach & Cho, 2011). Assessment, therefore, plays a major role in student achievement.

Assessment of Learning and Assessment for Learning

Effective assessment practices revolve around data gathering, engagement from stakeholders (e.g., teachers, students), and alignment with goals (Banta et al., 2009). As previously mentioned, assessment can be either formative (assessment *for* learning) or summative (assessment *of* learning) and differentiation is based on the purpose, timing, and generalization(s) intended for the instrument (Bloom et al., 1971). Summative

assessments are used at the end of the curriculum to determine what students have achieved (Bennett, 2011). On the other hand, formative instruments are used to determine how students are progressing through the curriculum and what adjustments, if any, need to be made. Sometimes referred to as feedback, the literature discusses the different forms formative assessment is viewed as: gathering data from teacher-made instruments to make instructional decisions, providing feedback to students based on gathered evidence, or the self-regulation of students on their part (Andersson & Palm, 2017). When formative assessment is implemented, research has demonstrated improvements in student achievement (e.g., Black & William, 1998). A study conducted by Li (2016) using 2009 U.S. student assessment data from the Programme for International Student Assessment (PISA), found a positive relationship between formative assessment practices and students' reading scores, and a significantly stronger relationship in Black students' achievement. These improvements could be attributed to the instructional adjustments teachers make based on gathered evidence as they monitor learning.

Formal Assessments and Teacher Judgments

Formal assessments provide teachers with scores for ranking and categorizing and are mostly closed-ended (Afflerbach & Cho 2011). They also help teachers understand, "after the fact" (p. 323) of teaching and learning to determine if students are reaching grade-level benchmarks (Afflerbach et al., 2019). Critics of formal assessment argue it results in a scripted curriculum that negatively impacts instructional practices (Au, 2011) and, recently, the argument revolves around it creating an even wider achievement gap amongst different groups of students (e.g., Knoester & Au, 2017). For example, Pratt (2018), after conducting interviews with 12 elementary teachers about their assessment

experiences, warned that "assessment may provide a differential treatment of students rather than an equitable approach to support their individual development" (p. 516). However, some research in the literature suggests that, when explicit strategy instruction is provided, student achievement on standardized measures improves (Lysynchuk et. al., 1990), though this was later countered by Amrein and Berliner (2002), who found students do not transfer skills assessed on standardized tests to other contexts.

Aside from formal assessments, teacher judgments are also an important component of instructional decisions made regarding student learning and achievement. They set the instructional pace, support, and difficulty of instruction in the classroom and influence decisions about student placement in more or less rigorous learning environments (Meissel et al., 2017). They have also been found to account for only half of the variance in student achievement when compared to standardized tests (Brookhart, 2013). As teachers are able to take into account the entire classroom context when making judgments about individual student achievement, their judgments might not necessarily be aligned with students' standardized test scores, as these do not capture the wide range of factors that influence student performance. More than that, research suggests teacher judgments for "marginalized students" (Meissel et al., 2017, p. 48) are lower than for other students (even when controlling for performance differences in standardized achievement tests) and that teachers might be more accurate judging the performance of high-performing students instead (Hoge & Coladarci, 1989).

Conceptual Framework

Data use theory helps explain the interaction between internal (e.g., teacher preparation) and external (e.g., students' socioeconomic status) factors that influence
analysis and the decision-making process amongst stakeholders beyond the datacollection phase. These interactions play a key role in the selection, structure, and transfer of relevant information to practice (Hutchins, 1995; Spillane, 2012), as teachers must take into account, for example, individual student characteristics when analyzing gathered data and making instructional decisions. This theory is further explained by Mandinach et al.'s (2006) Conceptual Framework for Data-Driven Decision-Making (see *Figure 2*), which illustrates the data-driven decision-making process teachers, school administrators, and district personnel should engage in.

Figure 2





The foundation for this framework lies within data management and organization theories where data-driven decision-making is viewed as a cyclical, problem-solving process (Mandinach, 2012). Though the primary intent of this framework is to illustrate how technology can help facilitate the decision-making process across classrooms, schools, and districts, its application to other contexts is viable, as the processes it encompasses are not discipline/context-specific. This framework employs a three-level "data to

knowledge continuum" (Mandinach et al., 2006, p. 8) that encompasses progression through six cognitive skills (collecting, organizing, analyzing, summarizing, synthesizing, and prioritizing) to, ultimately, implement effective strategies and impact student achievement. The first level (data) focuses on merely the collection task, which then progresses to using data *for* learning (e.g., analyzing trends, external factors, etc.). Knowledge produced from analyses is then used to develop and implement instructional decisions to impact student learning. This conceptual framework and data use theory both align with the current accountability environment. Thus, using data theory as the lens to examine teachers' use of standardized tests and instructional practices, and its impact on student achievement, was appropriate because it explored how internal and external factors shape practice and impact student achievement.

Summary

The national educational landscape continues to stress accountability in the classroom, especially in the field of reading. As such, the International Literacy Association (ILA), as well as other organizations, has developed a set of standards that emphasize the data-driven decision-making practices that need to take place in the classroom to increase student achievement. ILA's Assessment and Evaluation standard, calls for teacher preparation programs to ensure candidates are able to "understand, select, and use valid, reliable, fair, and appropriate assessment tools to screen, diagnose, and measure student literacy achievement; [and] inform instruction and evaluate interventions" (International Literacy Association, 2017, p. 3)." This has implications for teacher preparation programs, schools, and school districts, as they are all responsible for ensuring preservice and in-service teachers are able to meet the learning needs of their

students. As discussed, however, both of these groups seem to not have the appropriate preparation to take on these tasks and so, even with the availability of standardized test data to monitor learning, teachers' lack of preparation could hinder their ability to adjust instructional practices to individual student needs. Teachers' ability to effectively monitor learning is particularly crucial for students of low socioeconomic status and varying ethnicities, as recent data shows they continue to be low-performing in reading (Mulligan et al., 2019). The literature, however, has not examined how teachers' use of standardized tests to monitor learning impacts student achievement, as well as teachers' judgments of their achievement, while controlling for students' socioeconomic status and race/ethnicity and teachers' professional development participation. A review of the literature also did not yield studies examining whether instructional practices mediate the relationship between teachers' use of standardized tests to monitor learning and student achievement. The next chapter will discuss the methodology the present study employed to answer the research questions.

CHAPTER III

METHODOLOGY

The present study examined the relationship between fifth grade teachers' (a) use of standardized tests to monitor learning and students' reading achievement (b) reading instructional practices and use of standardized tests to monitor learning, and (c) reading instructional practices and students' reading achievement, while controlling for students' socioeconomic status and race/ethnicity and teachers' assessment-related professional development participation. The relationship was examined using publicly-available data (ECLS-K 2011 Public-Use File K-5 Child Version 6.2.1.3) from the Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011). Results of the present study contribute to existing literature on the effects of standardized tests and teachers' instructional practices on students' reading achievement. Chapter III will discuss the research questions and hypotheses, the research design, as well as provide a description of the participants and instruments. The data statistical treatments used and limitations will also be discussed.

Research Design

The research design employed in this study was ex post facto. In this type of design, the investigation takes place after data have already been collected, thus not allowing for researcher manipulation of variables (Silva, 2010).

Research Questions

This study sought to answer the following research questions:

Question 1: Does fifth grade teachers' use of standardized tests to monitor learning relate to students' reading achievement when controlling for students' socioeconomic status and race/ethnicity and teachers' assessment-related professional development participation?

Question 2: Do fifth grade teachers' instructional practices relate to their use of standardized tests to monitor learning after controlling for the effects of students' socioeconomic status and race/ethnicity and teachers' assessment-related professional development participation?

Question 3: Do fifth grade teachers' instructional practices relate to student achievement after controlling for the effects of students' socioeconomic status and race/ethnicity and teachers' assessment-related professional development participation?

Hypotheses

This study was guided by the following hypotheses:

Hypothesis 1: Fifth grade teachers' use of standardized tests to monitor learning will be corelated to students' reading achievement, when controlling for students' socioeconomic status and ethnicity and teachers' assessment-related professional development participation.

Hypothesis 2: Fifth grade teachers' instructional practices will be correlated their use of standardized tests to monitor learning after controlling for the effects of students' socioeconomic status and race/ethnicity and teachers' assessment-related professional development participation.

Hypothesis 3: Fifth grade teachers' instructional practices will be correlated to student achievement after controlling for the effects of students' socioeconomic status and race/ethnicity and teachers' assessment-related professional development participation.

26

Participants

The Early Childhood Longitudinal Study – Kindergarten: 2011 (ECLS-K: 2011) is the third study in a series funded by the National Center for Education Statistics (NCES). By gathering student, teacher, parent, and school administrator data, this study sought to understand elementary students' development, school readiness, and experiences (Tourangeau et al., 2019). The study followed a nationally-representative sample of students from both public and private schools from their kindergarten "cohort" year (2010-11) through fifth grade (2015-16). The cohort year included students who were in kindergarten for the first time as well as those who were repeating the grade level. Later iterations of data collection (i.e., first grade, second grade, etc.) assessed these same students regardless of their actual grade level. The last data collection round took place in spring of 2016 with the fifth-grade students, who had all participated in the 2010-11 (base year) data collection rounds (Tourangeau et al., 2019). The present study used fifth-grade students' reading achievement assessment data and their reading teachers' questionnaire responses to answer the research questions. The public-use data file with these data was released on July 12, 2019, and no student, teacher, parent, or school identifiable information were used.

Instruments

Students' Direct Cognitive Assessment reading scores and teacher responses to the child-level questionnaire were used to answer the research questions.

Direct Cognitive Assessment

In spring 2016, fifth grade students were administered a direct cognitive reading assessment, which included many of the same items from previous assessments as well as

27

new items with increased difficulty. Curriculum experts, national and state standards, commercial instruments, the 2011 National Assessment of Educational Progress (NAEP) reading frameworks, and reading standards from Texas, California, New Jersey, Florida, and Virginia were used to validate the direct cognitive reading assessment. Pools of items were developed, further examined by discipline experts, and a field test conducted prior to producing the final instrument. Students were individually administered the direct cognitive reading assessment by a trained assessor in about a 60-minute period per child; responses were collected via a computer-assisted interviewing (CAI) program. Overall, questions assessed students' basic reading skills, vocabulary knowledge, and comprehension. Since all fifth-grade students assessed were part of the kindergarten cohort, a language screener for students whose home language was not English was not deemed necessary. The assessment employed a two-phase approach. An initial, broadrange 12-item routing session was implemented to determine the difficulty of the second phase (low, middle, or high). The assessment asked students questions related to images containing pictures, words, or short sentences, and to passages read from a booklet. However, though the items completed by students varied as a result of the two-step approach, a subset of the same items was completed by all students – all items in the routing phase and a set of common items in the second phase (Tourangeau et al., 2019).

Item Response Theory (IRT) procedures were used for data analysis, as it allowed for the calculation and comparison of students' overall scores regardless of the test items administered. As Tourangeau et al., (2019) describes "IRT uses the pattern of right and wrong responses to the items actually administered in an assessment and the difficulty, discriminating ability, and "guess-ability" of each item to estimate each child's ability on

28

the same continuous scale" (p. 3-3). Using IRT procedure, three different scores to describe students' reading abilities were formulated, with a spring 2016 reliability score of 0.86 (Tourangeau et al., 2019):

- IRT theta scores: An estimate of a student's ability using performance on actual completed items. Scores are normally distributed and not dependent on item difficulty (Tourangeau et al., 2019).
- IRT standard errors of measurement (SEM) scores: In IRT, the standard error of measurement can vary depending on the number of items students respond to (i.e., SEM precision increases as the number of answered items increases). The IRT SEM scores are "a measure of uncertainty of the theta score estimate [with] an approximate 95 percent confidence interval or range of values that is likely to include the true theta score" (Tourangeau et al., 2019, p. 3-4).
- 3. IRT-based overall scale scores: Under the assumption students were administered all (205) distinctive router and second phase questions for all grade levels (K-5) reading assessment, IRT-based overall scale scores are an estimate of the number of questions students would have answered correctly. In essence, IRT-based overall scale scores are the sum of students' theta scores (Tourangeau et al., 2019).

As IRT theta scores represent students' actual performance, the theta score was the one selected for the statistical analyses.

Teacher Questionnaire

During the spring of 2016, students' reading teachers completed a child-level questionnaire that was directly linked to a sampled child. The questionnaire included

questions regarding the specific child's behavior and reading abilities and the teachers' instructional practices and preparation, among others. The present study will use several survey items (see Table 1), along with students' reading direct cognitive assessment scores, to answer the research questions.

Table 1

Category	Main Question	Sub-Questions	Answer Choices
Frequency of Use of Standardized Tests	How often do you use a formal assessment in reading for the following purposes?	To evaluate how well each student is responding to the core curriculum provided in the general education classroom (RespCoreCur) To monitor each student's progress on specific skills over the school year (MntrPrgrss) To identify the deficits in specific skills of struggling students (IdentDfcts) To monitor the progress of students who fall below benchmark levels (MntrPrgBelBnchmrk) To determine whether students need placement in a more or less intensive level of instruction (DetPlacement) How to use	(per sub-question) Never Once a year 2 times a year 3 to 4 times a year 5 to 8 times a year 1 to 2 times a month 1 to 2 times a week
professional	participate in professional	assessment data to identify students who	Once 2 times

development activities	development activities covering the following topics in the last 12 months?	are struggling or at risk of failure in reading (IdentStrgStdnts) How to use and apply assessment data to guide reading instruction (GuideRdngInst) How to implement the reading curriculum (ImpmtRdngCurr)	3 to 4 times More than 4 times
Instructional Practices	From the first day of school until today, please indicate how many days you have covered each of the following reading skills and concepts in this child's class:	Literature: Understanding key ideas and details (LIT_IDEA) Understanding craft and structure (LIT_CRFT) Integrating knowledge and ideas (LIT_KNOW) Informational text: Understanding key ideas and details (IT_IDEA) Understanding craft and structure (IT_CRFT) Integrating knowledge and ideas (IT_KNOW)	Not yet taught or not taught in this grade On 1-10 days On 11-20 days On 21-40 days On 41-80 days On more than 80 days
Teacher Judgment	Overall, how would you rate this child's academic skills in each of the following areas, based on curriculum	Reading	Below Grade Level About on Grade Level Above Grade Level

standards for his/her current grade level?

Statistical Treatments

Aside from providing descriptive statistics for each variable, several statistical treatments were used to test the hypotheses. First, Pearson's Product-Moment Correlation was used to determine the direction and strength of the relationship between variables related to teachers' use of standardized tests, assessment-related professional development participation, and instructional practices, all continuous scale variables. Though it does not describe causation, Pearson's Product-Moment Correlation coefficient (r) "quantifies the direction and degree of linear relationship between two variables" (Walk & Rupp, 2010, p. 3). Exploratory Factor Analysis (EFA) was then performed to determine the number of constructs (i.e., factors) in survey questions related to teachers' use of standardized tests, their assessment-related professional development participation, and their instructional practices, thus evaluating their construct validity (Reio and Shuck, 2015; Watkins, 2018). Conducting an EFA was appropriate because the covariance structure had not been hypothesized (Hayasbi & Yuan, 2010), variables used adequately represented the domains, and there were at least three measured variables for identifying a factor, two best practices discussed by Watkins (2018). Cronbach's Alpha tests were then conducted to assess internal consistency (i.e., reliability), as it is a suitable method for scale items (Ercan et al., 2007). Finally, emerging factors from the EFA were used to conduct hierarchical regression. Hierarchical regression is an appropriate technique because it is used when examining connections between two or more independent

variables and one dependent variable (Segrin, 2010) while controlling for the effects of other variables.

Limitations

This study conducted secondary analyses of existing data. Thus, the researcher was not able to manipulate the variables, as data had already been collected. In addition, teachers' judgment of students' reading levels, used as a measure of student achievement, is limited to teachers' self-report.

Summary

Using data from a nationally-representative sample of students from both public and private schools, this study used ex post facto design to explore, quantitatively, the relationship between fifth grade teachers' (a) use of standardized tests to monitor learning and students' reading achievement (b) reading instructional practices and use of standardized tests to monitor learning, and (c) reading instructional practices and students' reading achievement, while controlling for students' socioeconomic status and race/ethnicity and teachers' assessment-related professional development participation. To test the hypotheses and best answer the research questions, hierarchical regression was used.

CHAPTER IV

RESULTS

The purpose of the present study was to explore, quantitatively, the relationship between fifth grade teachers' (a) use of standardized tests to monitor learning and students' reading achievement, (b) reading instructional practices and use of standardized tests to monitor learning, and (c) reading instructional practices and students' reading achievement, all while controlling for students' socioeconomic status and race/ethnicity and teachers' assessment-related professional development participation. To do so, and grounded in data use theory (Hutchins, 1995; Spillane, 2012), student achievement and demographics were analyzed, as well as teacher survey response data from the Early Childhood Longitudinal Study – Kindergarten: 2011 (ECLS-K:2011) using hierarchical regression. Chapter IV includes descriptive statistics of variables and the results of the statistical analyses conducted to answer the research questions and test the hypotheses.

Descriptive Statistics

Student Achievement

Student achievement was measured using (a) student performance on a direct cognitive reading assessment (via IRT theta scores, reported on a range of -4 to 4) and (b) teacher judgement of students' reading skills (reported as below, at, or above grade level). A total of 11,427 valid theta scores were available with a mean of 1.452 (see Table 2), and a total of 10,381 valid responses were available to teachers' judgment of students' reading level. The largest percentage of students (42.6%) were judged by their teacher to be reading on grade level, with 31.7% judged to be reading above grade level and 25.7% below (see Table 3).

Descriptive Statistics: IRT Theta Scores

IRT_THETA						
Ν	Valid	11427				
	Missing	6747				
Mean	L	1.452275				
Std. Deviation		.3536123				
Rang	e	2.1812				
Minir	num	.0246				
Maxi	mum	2.2058				

Table 3

Descriptive Statistics: Teachers' Judgments of Student Reading Level

					Cumulative
		Frequency	Percent	Valid Percent	Percent
Valid	Below grade level	2664	14.7	25.7	25.7
	About on grade level	4423	24.3	42.6	68.3
	Above grade level	3294	18.1	31.7	100.0
	Total	10381	57.1	100.0	
Missing	Not ascertained	67	.4		
	System	7726	42.5		
	Total	7793	42.9		
Total		18174	100.0		

Students' Socioeconomic Status

Students' socioeconomic status was provided via a composite score with a mean of 0 and a standard deviation of 1 based on the average of five components: father/male guardian's education, mother/female guardian's education, father/male guardian's occupational prestige, mother/female guardian's occupational prestige, and household income. The mean socioeconomic status for students in the sample was -.059577 with a standard deviation of .8164454 (see Table 4).

Table 4	
Descriptive Statistics: Students'	Socioeconomic Status

					Std.
	Ν	Minimum	Maximum	Mean	Deviation
SES	10220	-2.7359	2.2729	059577	.8164454
Valid N (listwise)	10220				

Students' Race/Ethnicity

The composite variable for race/ethnicity shows students in the sample (N =

18,174) were from diverse backgrounds, with the largest three groups being White, non-

Hispanic (N = 8,488), Hispanic (N = 4,207), and Black, non-Hispanic (N = 2,396) (see

Table 5). As this is a categorical variable, it was recoded into "dummy" variables to use

for the hierarchical regression analyses.

Table 5

Descriptive Statistics: Students' Race/Ethnicity

	Ν	Sum
WhiteNonHispanic	18174	8488
BlackNonHispanic	18174	2396
Hispanic	18174	4207
AsianNonHispanic	18174	385
NativeHawaiianOtherPacificIslanderNonHisp	18174	1543
anic		
$\label{eq:laska} American Indian Or Alaska Native Non Hispanic$	18174	117
TwoOrMoreRacesNonHispanic	18174	168
Unknown	18174	827
Valid N (listwise)	18174	

Teachers' Frequency of Use of Standardized Tests

Teachers were asked five questions, with seven possible responses each, related to their frequency of use of standardized tests. Table 6 shows the sample size (N) per

question, along with the means. Using standardized tests to monitor progress

(MntrPrgrss) had the highest mean (M = 5.45), followed by to evaluate how well students are responding to the core curriculum (RespCoreCur, M = 5.41), to identify deficits in struggling students (IdentDfcts, M = 5.36), to monitor progress of students falling below benchmark levels (MntrPrgBelBnchmrk, M = 5.29), and to determine placement in more or less intensive levels of instruction (DetPlacement, M = 4.61).

Table 6

Descriptive Statistics: Teachers' Frequency of Use of Standardized Tests

	Minimu					Std.
	Ν	Range	m	Maximum	Mean	Deviation
RespCoreCur	10304	6	1	7	5.41	1.392
MntrPrgrss	10316	6	1	7	5.45	1.294
IdentDfcts	10299	6	1	7	5.36	1.412
MntrPrgBelBnchmr	10318	6	1	7	5.29	1.407
k						
DetPlacement	10309	6	1	7	4.61	1.667
Valid N (listwise)	10264					

Table 7

Descriptive Statistics: Use of Standardized Tests to Evaluate How Well Students Respond to the Core Curriculum

					Cumulative
		Frequency	Percent	Valid Percent	Percent
Valid	Never	145	.8	1.4	1.4
	Once a year	129	.7	1.3	2.7
	2 times a year	402	2.2	3.9	6.6
	3-4 times a year	2526	13.9	24.5	31.1
	5-8 times a year	1322	7.3	12.8	43.9
	1-2 times a month	3007	16.5	29.2	73.1
	1-2 times a week	2773	15.3	26.9	100.0
	Total	10304	56.7	100.0	
Missing	Not ascertained	123	.7		
	System	7747	42.6		
	Total	7870	43.3		

A total of 10,304 teachers responded to how often they used standardized tests in reading to evaluate how well each student is responding to the core curriculum. Responses (29.2%) indicate they are used most 1-2 times a month, with 1-2 times a week following with 26.9% of responses (see Table 7).

Table 8

					Cumulative
		Frequency	Percent	Valid Percent	Percent
Valid	Never	73	.4	.7	.7
	Once a year	76	.4	.7	1.4
	2 times a year	408	2.2	4.0	5.4
	3-4 times a year	2377	13.1	23.0	28.4
	5-8 times a year	1638	9.0	15.9	44.3
	1-2 times a month	3174	17.5	30.8	75.1
	1-2 times a week	2570	14.1	24.9	100.0
	Total	10316	56.8	100.0	
Missing	Not ascertained	111	.6		
	System	7747	42.6		
	Total	7858	43.2		
Total		18174	100.0		

Descriptive Statistics: Use of Standardized Tests to Monitor Progress

A total of 10,316 teachers responded to how often they used standardized tests in reading to monitor student progress on specific reading skills. Responses (30.8%) indicate they are used most 1-2 times a month, with 1-2 times a week following with 24.9% of responses (see Table 8).

					Cumulative
		Frequency	Percent	Valid Percent	Percent
Valid	Never	184	1.0	1.8	1.8
	Once a year	192	1.1	1.9	3.7
	2 times a year	429	2.4	4.2	7.8
	3-4 times a year	2306	12.7	22.4	30.2
	5-8 times a year	1540	8.5	15.0	45.2
	1-2 times a month	3146	17.3	30.5	75.7
	1-2 times a week	2502	13.8	24.3	100.0
	Total	10299	56.7	100.0	
Missing	Not ascertained	128	.7		
	System	7747	42.6		
	Total	7875	43.3		
Total		18174	100.0		

Descriptive Statistics: Use of Standardized Tests to Identify Reading Deficits

A total of 10,299 teachers responded to how often they used standardized tests in reading

to identify reading deficits. Responses (30.5%) indicate they are used most 1-2 times a

month, with 1-2 times a week following with 24.3% of responses (see Table 9).

Table 10

Descriptive Statistics: Use of Standardized Tests to Monitor Progress of Students Below Benchmarks

					Cumulative
		Frequency	Percent	Valid Percent	Percent
Valid	Never	250	1.4	2.4	2.4
	Once a year	126	.7	1.2	3.6
	2 times a year	435	2.4	4.2	7.9
	3-4 times a year	2383	13.1	23.1	31.0
	5-8 times a year	1660	9.1	16.1	47.0
	1-2 times a month	3263	18.0	31.6	78.7
	1-2 times a week	2201	12.1	21.3	100.0
	Total	10318	56.8	100.0	
Missing	Not ascertained	109	.6		
	System	7747	42.6		
	Total	7856	43.2		
Total		18174	100.0		

A total of 10,318 teachers responded to how often they used standardized tests to monitor progress of students falling below benchmark. Responses (31.6%) indicate they are used most 1-2 times a month, with 3-4 times a year following with 23.1% of responses (see Table 10).

Table 11

Total

					Cumulative
		Frequency	Percent	Valid Percent	Percent
Valid	Never	736	4.0	7.1	7.1
	Once a year	530	2.9	5.1	12.3
	2 times a year	837	4.6	8.1	20.4
	3-4 times a year	2881	15.9	27.9	48.3
	5-8 times a year	1502	8.3	14.6	62.9
	1-2 times a month	2612	14.4	25.3	88.3
	1-2 times a week	1211	6.7	11.7	100.0
	Total	10309	56.7	100.0	
Missing	Not ascertained	118	.6		
	System	7747	42.6		
	Total	7865	43.3		

Descriptive Statistics: Use of Standardized Tests to Determine Placement

A total of 10,309 teachers responded to how often they used standardized tests to determine placement in more or less intensive levels of instruction. Responses (27.9%) indicate they are used most 3-4 times a year, with 1-2 times a month following with 25.3% of responses (see Table 11).

100.0

18174

Teachers' Assessment-Related Professional Development Participation

Teachers were also asked three questions, with five possible responses each, related to how often they had participated in assessment-related professional development activities for reading. Table 12 shows the sample size (N) per question, along with the means. Participation in professional development activities on to how to implement the reading curriculum (ImpmtRdngCurr) had the highest mean (M = 2.74), followed by activities on how to use assessment data to guide reading instruction (GuideRdngInst, M

= 2.72), and activities on how to identify struggling students (IdentStrgStdnts, M = 2.68).

Table 12

Descriptive Statistics: Teachers' Assessment-Related Professional Development Participation

						Std.
	Ν	Range	Minimum	Maximum	Mean	Deviation
IdentStrgStdnts	9909	4	1	5	2.68	1.313
GuideRdngInst	9904	4	1	5	2.72	1.301
ImpmtRdngCurr	9903	4	1	5	2.74	1.407
Valid N	9895					
(listwise)						

Table 13

Descriptive Statistics: Professional Development Activities on How to Use Data to Identify Struggling Students

					Cumulative
		Frequency	Percent	Valid Percent	Percent
Valid	Never	2249	12.4	22.7	22.7
	Once	2721	15.0	27.5	50.2
	2 times	2090	11.5	21.1	71.2
	3-4 times	1666	9.2	16.8	88.1
	More than 4 times	1183	6.5	11.9	100.0
	Total	9909	54.5	100.0	
Missing	Not ascertained	99	.5		
	Not applicable	419	2.3		
	System	7747	42.6		
	Total	8265	45.5		
Total		18174	100.0		

A total of 9,909 teachers responded to how often they participated in professional

development activities on how to use data to identify struggling students. Most teachers (27.5%) participated in these types of activities once, followed by never (22.7%), twice (21.1%), 3-4 times (16.8%), and more than 4 times (11.9%) (see Table 13).

					Cumulative
		Frequency	Percent	Valid Percent	Percent
Valid	Never	2048	11.3	20.7	20.7
	Once	2874	15.8	29.0	49.7
	2 times	2005	11.0	20.2	69.9
	3-4 times	1803	9.9	18.2	88.1
	More than 4 times	1174	6.5	11.9	100.0
	Total	9904	54.5	100.0	
Missing	Not ascertained	104	.6		
	Not applicable	419	2.3		
	System	7747	42.6		
	Total	8270	45.5		
Total		18174	100.0		

Descriptive Statistics: Professional Development Activities on How to Use Data to Guide Reading Instruction

A total of 9,904 teachers responded to how often they participated in professional development activities on how to use data to guide reading instruction. Most teachers (29%) participated in these types of activities once, followed by never (20.7%), twice (20.2%), 3-4 times (18.2%), and more than 4 times (11.9%) (see Table 14).

Table 15

Descriptive Statistics: Professional Development Activities on How to Implement the Reading Curriculum

					Cumulative
		Frequency	Percent	Valid Percent	Percent
Valid	Never	2488	13.7	25.1	25.1
	Once	2409	13.3	24.3	49.4
	2 times	1701	9.4	17.2	66.6
	3-4 times	1775	9.8	17.9	84.6
	More than 4 times	1530	8.4	15.4	100.0
	Total	9903	54.5	100.0	
Missing	Not ascertained	105	.6		
	Not applicable	419	2.3		
	System	7747	42.6		

Total	8271	45.5
Total	18174	100.0

A total of 9,903 teachers responded to how often they participated in professional development activities on how to implement the reading curriculum. Most teachers (25%) had not participated, followed by once (24.3%), 3-4 times (17.9%), twice (17.2%), and more than 4 times (15.4%) (see Table 15).

Teachers' Reading Instructional Practices

Teachers were also asked how often they taught specific reading skills related to literature and informational text from the beginning of the school year to the day the survey was administered (spring).

Literary Text

Table 16 shows the three questions asked regarding literary text. Understanding key ideas of literary text (LIT_IDEA) obtained the highest mean (M = 5.20), followed by understanding craft and structure (LIT_CRFT, M = 4.41), and integrating knowledge and ideas (LIT_KNOW, M = 4.13).

Table 16

Descriptive Statistics: Teach	hers' Reading Instruct	ional Practices with Literature

						Std.
	Ν	Range	Minimum	Maximum	Mean	Deviation
LIT_IDEA	10300	5	1	6	5.20	1.029
LIT_CRFT	10304	5	1	6	4.41	1.213
LIT_KNOW	10283	5	1	6	4.13	1.368
Valid N	10274					
(listwise)						

				Valid	Cumulative
		Frequency	Percent	Percent	Percent
Valid	Not yet taught/not	30	.2	.3	.3
	taught in this grade				
	level				
	1-10 days	194	1.1	1.9	2.2
	11-20 days	589	3.2	5.7	7.9
	21-40 days	1397	7.7	13.6	21.5
	41-80 days	2759	15.2	26.8	48.2
	More than 80 days	5331	29.3	51.8	100.0
	Total	10300	56.7	100.0	
Missing	Not ascertained	127	.7		
	System	7747	42.6		
	Total	7874	43.3		
Total		18174	100.0		

Descriptive Statistics: Teachers' Reading Instructional Practices with Literary Text: Understanding Key Ideas and details

A total of 10,300 teachers responded to how often they taught understanding key ideas and details in literary text. Most teachers (51.8%) had taught it on more than 80 days, followed by 41-80 days (26.8%), 21-40 days (13.6%), 11-20 days (5.7%), and 1-10 days (1.9%). The skill was not taught or applicable to the grade level for 0.3% of respondents

(see Table 17).

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not yet taught/not taught in this grade	124	.7	1.2	1.2
	level				
	1-10 days	597	3.3	5.8	7.0
	11-20 days	1641	9.0	15.9	22.9
	21-40 days	2716	14.9	26.4	49.3
	41-80 days	3059	16.8	29.7	79.0
	More than 80 days	2167	11.9	21.0	100.0
	Total	10304	56.7	100.0	
Missing	Not ascertained	123	.7		
	System	7747	42.6		
	Total	7870	43.3		
Total		18174	100.0		

Descriptive Statistics: Teachers' Reading Instructional Practices with Literary Text: Understanding Craft and Structure

A total of 10,304 teachers responded to how often they taught understanding craft and structure in literary text. Most teachers (29.7%) had taught it 41-80 days, followed by 21-40 days (26.6%), more than 80 days (21%), 11-20 days (15.9%), and 1-10 days (5.8%). The skill was not taught or applicable to the grade level for 1.2% of respondents (see Table 18).

				Valid	Cumulative
		Frequency	Percent	Percent	Percent
Valid	Not yet taught/not	239	1.3	2.3	2.3
	taught in this grade				
	level				
	1-10 days	1210	6.7	11.8	14.1
	11-20 days	1955	10.8	19.0	33.1
	21-40 days	2510	13.8	24.4	57.5
	41-80 days	2336	12.9	22.7	80.2
	More than 80 days	2033	11.2	19.8	100.0
	Total	10283	56.6	100.0	
Missing	Not ascertained	144	.8		
	System	7747	42.6		
	Total	7891	43.4		
Total		18174	100.0		

Descriptive Statistics: Teachers' Reading Instructional Practices with Literary Text: Integrating Knowledge and Ideas

A total of 10,283 teachers responded to how often they taught integrating knowledge and ideas in literary text. Most teachers (24.4%) had taught it 21-40 days, followed by 41-80 days (22.7%), more than 80 days (19.8%), 11-20 days (19%), and 1-10 days (11.8%). The skill was not taught or applicable to the grade level for 2.3% of respondents (see Table 19).

Informational Text

Table 20 shows the three questions asked regarding informational text.

Understanding key ideas of informational text (IT_IDEA) obtained the highest mean (M

= 5.08), and understanding craft and structure (IT_CRFT) along with integrating

knowledge and ideas (IT_KNOW) obtained the same mean (M = 4.40).

Table 20

Descriptive Statistics: Teachers' Reading Instructional Practices with Informational Texts

	Minimu					Std.
	Ν	Range	m	Maximum	Mean	Deviation
IT_IDEA	10304	5	1	6	5.08	1.060
IT_CRFT	10296	5	1	6	4.40	1.226
IT_KNOW	10303	5	1	6	4.40	1.283
Valid N	10293					
(listwise)						

Descriptive Statistics: Teachers' Reading Instructional Practices with Informational Text: Understanding Key Ideas and details

				Valid	Cumulative
		Frequency	Percent	Percent	Percent
Valid	Not yet taught/not	32	.2	.3	.3
	taught in this grade				
	level				
	1-10 days	221	1.2	2.1	2.5
	11-20 days	683	3.8	6.6	9.1
	21-40 days	1732	9.5	16.8	25.9
	41-80 days	2953	16.2	28.7	54.6
	More than 80 days	4683	25.8	45.4	100.0
	Total	10304	56.7	100.0	
Missing	Not ascertained	123	.7		
	System	7747	42.6		
	Total	7870	43.3		
Total		18174	100.0		

A total of 10,304 teachers responded to how often they taught understanding key ideas and details in informational text. Most teachers (45.4%) had taught it on more than 80 days, followed by 41-80 days (28.7%), 21-40 days (16.8%), 11-20 days (6.6%), and 1-10 days (2.1%). The skill was not taught or applicable to the grade level for 0.3% of respondents (see Table 21).

				Valid	Cumulative
		Frequency	Percent	Percent	Percent
Valid	Not yet taught/not taught in this grade level	144	.8	1.4	1.4
	1-10 days	590	3.2	5.7	7.1
	11-20 days	1654	9.1	16.1	23.2
	21-40 days	2728	15.0	26.5	49.7
	41-80 days	2973	16.4	28.9	78.6
	More than 80 days	2207	12.1	21.4	100.0
	Total	10296	56.7	100.0	
Missing	Not ascertained	131	.7		
	System	7747	42.6		
	Total	7878	43.3		
Total		18174	100.0		

Descriptive Statistics: Teachers' Reading Instructional Practices with Informational Text: Understanding Craft and Structure

A total of 10,296 teachers responded to how often they taught understanding craft and

structure in informational text. Most teachers (28.9%) had taught it 41-80 days, followed by 21-40 days (26.5%), more than 80 days (21.4%), 11-20 days (16.1%), and 1-10 days (5.7%). The skill was not taught or applicable to the grade level for 1.4% of respondents (see Table 22).

				Valid	Cumulative
		Frequency	Percent	Percent	Percent
Valid	Not yet taught/not	154	.8	1.5	1.5
	taught in this grade				
	level				
	1-10 days	726	4.0	7.0	8.5
	11-20 days	1635	9.0	15.9	24.4
	21-40 days	2670	14.7	25.9	50.3
	41-80 days	2612	14.4	25.4	75.7
	More than 80 days	2506	13.8	24.3	100.0
	Total	10303	56.7	100.0	
Missing	Not ascertained	124	.7		
	System	7747	42.6		
	Total	7871	43.3		
Total		18174	100.0		

Descriptive Statistics: Teachers' Reading Instructional Practices with Informational Text: Integrating Knowledge and Ideas

A total of 10,303 teachers responded to how often they taught integrating knowledge and ideas in informational text. Most teachers (25.9%) had taught it 21-40 days, followed by 41-80 days (25.4%), more than 80 days (24.3%), 11-20 days (15.9%), and 1-10 days (7%). The skill was not taught or applicable to the grade level for 1.5% of respondents (see Table 23).

Examining Correlations and Creating Constructs for Survey Items

To begin to answer the research questions, correlations amongst similar-topic survey items were assessed. A Pearson's Product-Moment Correlations test was first performed to determine the direction (positive or negative) and significance of the relationship amongst items. The closer the Pearson's correlation coefficient (r) to 1, the stronger the relationship (Chen & Krauss, 2004). This was proceeded by a Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) test, which assesses the adequacy of the sample for factor analysis (Field, 2005, p. 647). KMO values should, at a minimum, exceed 0.5 (Kaiser, 1974), with values exceeding 0.7 deemed "good" (Field, 2005). Barlett's test of Sphericity was also performed. This test examines redundancies between variables, providing a significant result when correlations amongst variables are significantly different from zero (Field, 2005). An Exploratory Factory Analysis (EFA) was then carried out to determine the number of constructs made up by these items. As per Kaiser's (1960) recommendation, factors with eigenvalues greater than 1 were retained. Last, a Cronbach's Alpha (α) test was performed to test reliability of the constructs; with values greater than .8 deemed reliable (Field, 2005). Significance, for the purpose of this study, will be established for *p* values smaller than .05, as Type II errors are more likely to occur with the nature of this study (Salkind, 2010).

Teachers' Frequency of Use of Standardized Tests Survey Items

The Pearson's Product-Moment Correlations test (see Table 24) indicated all variables have a positive relationship and are significantly correlated (p < 0.05, sig = 0.00). The highest correlations were between using standardized tests to (a) monitor the progress of students who fall below benchmarks and identify deficits in specific skills of struggling students (r = .688) and (b) monitor student progress on specific skills throughout the school year and identify deficits in specific skills of struggling students (r = .644). KMO's value (.806) exceeded the heuristic of .50 (Kaiser, 1974), indicating correlations were adequate to continue with factor analysis (see Table 25), and Bartlett's value indicated correlations amongst variables were significantly different from zero (Chi-square = 20453.960, p < 0.05, sig = .000). Communalities ranged from .464 to .759 (see Table 26) thus, all items were used to conduct the exploratory factor analysis. Table

27 shows five components were extracted in total, accounting for 100% of the variance; one factor (i.e., constructs) emerged with an eigenvalue greater than 1 (3.019), accounting for 60.386% of the variance. Lastly, the sub-scale of five items indicated a high internal reliability with a Cronbach's alpha of .829 (see Table 28). Thus, using the mean across survey items belonging to each component, one construct was created: frequency of use of standardized tests.

Table 24

					MntrPrg	
		RespCore	MntrPrg	IdentDf	BelBnch	DetPlace
		Ċur	rss	cts	mrk	ment
RespCoreCur	Pearson	1	.522**	.464**	.390**	.338**
	Correlation					
	Sig. (2-tailed)		.000	.000	.000	.000
	Ν	10304	10299	10282	10300	10291
MntrPrgrss	Pearson	.522**	1	.644**	.496**	.373**
	Correlation					
	Sig. (2-tailed)	.000		.000	.000	.000
	Ν	10299	10316	10294	10312	10303
IdentDfcts	Pearson	.464**	.644**	1	$.688^{**}$.529**
	Correlation					
	Sig. (2-tailed)	.000	.000		.000	.000
	Ν	10282	10294	10299	10295	10288
MntrPrgBelBn	Pearson	.390**	.496**	$.688^{**}$	1	.569**
chmrk	Correlation					
	Sig. (2-tailed)	.000	.000	.000		.000
	Ν	10300	10312	10295	10318	10307
DetPlacement	Pearson	.338**	.373**	.529**	.569**	1
	Correlation					
	Sig. (2-tailed)	.000	.000	.000	.000	
	Ν	10291	10303	10288	10307	10309

Correlations: Frequency of Use of Standardized Tests Survey Items

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

Table 25*KMO and Bartlett's Test: Frequency of Use of Standardized Tests Survey Items*

Kaiser-Meyer-Olkin	.806	
Adequacy.		
Bartlett's Test of	Approx. Chi-Square	20453.960
Sphericity	Df	10
	Sig.	.000

Communalities: Frequency of Use of Standardized Tests Survey Items

	Initial	Extraction
RespCoreCur	1.000	.464
MntrPrgrss	1.000	.614
IdentDfcts	1.000	.759
MntrPrgBelBnchmrk	1.000	.671
DetPlacement	1.000	.511

Extraction Method: Principal Component Analysis.

Table 27

Total Variance Explained: Frequency of Use of Standardized Tests Survey Items

	Initial Eigenvalues			Extraction Sums of Squared Loading		
		% of	Cumulative		% of	Cumulative
Component	Total	Variance	%	Total	Variance	%
1	3.019	60.386	60.386	3.019	60.386	60.386
2	.769	15.380	75.766			
3	.540	10.799	86.565			
4	.400	8.007	94.572			
5	.271	5.428	100.000			

Extraction Method: Principal Component Analysis.

Table 28

Reliability Statistics: Frequency of Use of Standardized Tests Survey Items

Cronbach's	
Alpha	N of Items
.829	5

Teachers' Assessment-Related Professional Development Participation Survey Items

Similarly to teachers' frequency of use of standardized tests survey items, the Pearson's Product-Moment Correlations test (see Table 29) indicated all variables have a positive relationship and are significantly correlated (p < 0.05, sig = 0.00). The highest correlation was between professional development on using data to identify struggling students and using data to guide reading instruction (r = .869). KMO's value (.669) exceeded the heuristic of .50 (Kaiser, 1974), indicating correlations were adequate to continue with factor analysis (see Table 30), and Bartlett's value indicated correlations amongst variables were significantly different from zero (Chi-square = 19940.101, p < 1000.05, sig = .000). Communalities ranged from .688 to .899 (see Table 31) thus, all items were used to conduct the exploratory factor analysis. Table 32 shows three components were extracted in total, accounting for 100% of the variance; one factor (i.e., constructs) emerged with an eigenvalue greater than 1 (2.433), accounting for 81.11% of the variance. Last, the sub-scale of three items indicated a high internal reliability with a Cronbach's alpha of .879 (see Table 33). Thus, using the mean across survey items belonging to each component, one construct was created: professional development.

Table 29

		IdentStrgStdn	GuideRdngIn	ImpmtRdngC
		ts	st	urr
IdentStrgStdnts	Pearson Correlation	1	$.869^{**}$.596**
	Sig. (2-tailed)		.000	.000
	Ν	9909	9902	9895
GuideRdngInst	Pearson Correlation	.869**	1	.675**
	Sig. (2-tailed)	.000		.000
	Ν	9902	9904	9897

Correlations	: Professional	Development	Survey Items
--------------	----------------	-------------	--------------

ImpmtRdngCurr	Pearson Correlation	.596**	.675**	1
	Sig. (2-tailed)	.000	.000	
	Ν	9895	9897	9903

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

Table 30

KMO and Bartlett's Test Professional Development Survey Items

Kaiser-Meyer-Olkin Measure	.669	
Bartlett's Test of Sphericity	Approx. Chi-Square	19940.101
	Df	3
	Sig.	.000

Table 31

Communalities: Professional Development Survey Items

	Initial	Extraction
IdentStrgStdnts	1.000	.846
GuideRdngInst	1.000	.899
ImpmtRdngCurr	1.000	.688

Extraction Method: Principal

Component Analysis.

Table 32

Total Variance Explained: Professional Development Survey Items

	Initial Eigenvalues		Extraction S	Sums of Squa	red Loadings	
		% of	Cumulative		% of	Cumulative
Component	Total	Variance	%	Total	Variance	%
1	2.433	81.111	81.111	2.433	81.111	81.111
2	.443	14.770	95.881			
3	.124	4.119	100.000			

Extraction Method: Principal Component Analysis.

Table 33

Reliability Statistics: Professional Development Survey Items

Cronbach's	
Alpha	N of Items
.879	3

Teachers' Reading Instructional Practices Survey Items

Survey items related to teachers' reading instructional practices were broken down into two different types of texts: literary and informational.

Teachers' Instructional Practices with Literary Texts

The Pearson's Product-Moment Correlations test (see Table 34) indicated all variables have a positive relationship and are significantly correlated (p < 0.05, sig = 0.00). The highest correlation was between understanding craft and structure and integrating knowledge and ideas (r = .696). KMO's value (.685) exceeded the heuristic of .50 (Kaiser, 1974), indicating correlations were adequate to continue with factor analysis (see Table 35), and Bartlett's value indicated correlations amongst variables were significantly different from zero (Chi-square = 12320.993, p < .05, sig = .000). Communalities ranged from .680 to .823 (see Table 36) thus, all items were used to conduct the exploratory factor analysis. Table 32 shows three components were extracted in total, accounting for 100% of the variance; one factor (i.e., constructs) emerged with an eigenvalue greater than 1 (2.235), accounting for 74.49% of the variance. Last, the sub-scale of three items indicated a high internal reliability with a Cronbach's alpha of .823 (see Table 38). Thus, using the mean across survey items belonging to each component, one construct was created: literary texts instructional practices.

Table 34 Correlations: Instructional Practices with Literary

		LIT_IDEA	LIT_CRFT	LIT_KNOW
LIT_IDEA	Pearson Correlation	1	.637**	.515**
	Sig. (2-tailed)		.000	.000
	Ν	10300	10298	10276
LIT_CRFT	Pearson Correlation	.637**	1	.696**
	Sig. (2-tailed)	.000		.000
	Ν	10298	10304	10278
LIT_KNOW	Pearson Correlation	.515**	.696**	1
	Sig. (2-tailed)	.000	.000	
	Ν	10276	10278	10283

Correlations: Instructional Practices with Literary Texts

Table 35

KMO and Bartlett's Test: Instructional Practices with Literary Texts

Kaiser-Meyer-Olkin Measure	.685	
Bartlett's Test of Sphericity Approx. Chi-Square		12320.993
	Df	3
	Sig.	.000

Table 36

Communalities: Instructional Practices with Literary Texts

	Initial	Extraction	
LIT_IDEA	1.000	.680	
LIT_CRFT	1.000	.823	
LIT_KNOW	1.000	.732	

Extraction Method: Principal Component Analysis.

_	Initial Eigenvalues		Extraction Sums of Squared Loadings			
_		% of	Cumulative		% of	Cumulative
Component	Total	Variance	%	Total	Variance	%
1	2.235	74.491	74.491	2.235	74.491	74.491
2	.490	16.339	90.831			
3	.275	9.169	100.000			

 Table 37

 Total Variance Explained: Instructional Practices with Literary Texts

Extraction Method: Principal Component Analysis.

Table 38

Reliability Statistics: Instructional Practices with Literary Texts

Cronbach's Alpha	N of Items	
.823		3

Teachers' Instructional Practices with Informational Texts

The Pearson's Product-Moment Correlations test (see Table 39) indicated all variables have a positive relationship and are significantly correlated (p < 0.05, sig = 0.00). Similar to literary texts, the highest correlation was between understanding craft and structure and integrating knowledge and ideas (r = .715). KMO's value (.718) exceeded the heuristic of .50 (Kaiser, 1974), indicating correlations were adequate to continue with factor analysis (see Table 40), and Bartlett's value indicated correlations amongst variables were significantly different from zero (Chi-square = 13854.632, p < .05, sig = .000). Communalities ranged from .727 to .818 (see Table 41) thus, all items were used to conduct the exploratory factor analysis. Table 42 shows three components were extracted in total, accounting for 100% of the variance; one factor (i.e., constructs) emerged with an eigenvalue greater than 1 (2.316), accounting for 77.210% of the variance. Last, the sub-scale of three items indicated a high internal reliability with a Cronbach's alpha of .850 (see Table 43). Thus, using the mean across survey items
belonging to each component, one construct was created: informational text instructional

practices.

Table 39

Correlations: Instructional Practices with Informational Texts

		IT_IDEA	IT_CRFT	IT_KNOW
IT_IDEA	Pearson Correlation	1	.659**	.599**
	Sig. (2-tailed)		.000	.000
	Ν	10304	10295	10302
IT_CRFT	Pearson Correlation	.659**	1	.715**
	Sig. (2-tailed)	.000		.000
	Ν	10295	10296	10294
IT_KNOW	Pearson Correlation	.599**	.715**	1
	Sig. (2-tailed)	.000	.000	
	Ν	10302	10294	10303

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

Table 40

KMO and Bartlett's Test: Instructional Practices with Informational Texts

Kaiser-Meyer-Olkin	.718	
Adequacy.		
Bartlett's Test of	Approx. Chi-Square	13854.632
Sphericity	df	3
	Sig.	.000

Table 41

Communalities: Instructional Practices with Informational Texts

	Initial	Extraction
IT_IDEA	1.000	.727
IT_CRFT	1.000	.818
IT_KNOW	1.000	.772

Extraction Method: Principal Component Analysis.

	Iı	nitial Eigenva	alues	Extraction Sums of Squared Loading				
		% of	Cumulative		% of	Cumulative		
Component	Total	Variance	%	Total	Variance	%		
1	2.316	77.210	77.210	2.316	77.210	77.210		
2	.409	13.640	90.850					
3	.274	9.150	100.000					

Table 42Total Variance Explained: Instructional Practices with Informational Texts

Extraction Method: Principal Component Analysis.

Table 43

Reliability Statistics: Instructional Practices with Informational Texts

Cronbach's	
Alpha	N of Items
.850	3

Results of the Study

During the past three decades, the use of standardized tests has increased due to global education demands (Vinovskis, 2019). However, the use of the data these tests yield to monitor learning, and its impact on elementary student achievement, has limited empirical evidence (Hao & Johnson, 2013) and is impacted by individual student and teacher characteristics (Black & William, 1998; Dee & Jacob, 2011). Hence, this study sought to answer the following research questions:

Question 1: Does fifth grade teachers' use of standardized tests to monitor learning relate to students' reading achievement when accounting for students' socioeconomic status and race/ethnicity and teachers' assessment-related professional development participation? Hypothesis 1: Fifth grade teachers' use of standardized tests to monitor learning will be corelated to students' reading achievement, when accounting for students' socioeconomic status and ethnicity and teachers' assessment-related professional development participation.

Question 2: Do fifth grade teachers' instructional practices relate to their use of standardized tests to monitor learning after controlling for the effects of students' socioeconomic status and race/ethnicity and teachers' assessment-related professional development participation?

Hypothesis 2: Fifth grade teachers' instructional practices will be correlated their use of standardized tests to monitor learning after controlling for the effects of students' socioeconomic status and race/ethnicity and teachers' assessment-related professional development participation.

Question 3: Do fifth grade teachers' instructional practices relate to student achievement after controlling for the effects of students' socioeconomic status and race/ethnicity and teachers' assessment-related professional development participation?

Hypothesis 3: Fifth grade teachers' instructional practices will be correlated to student achievement after controlling for the effects of students' socioeconomic status and race/ethnicity and teachers' assessment-related professional development participation.

Frequency of use of standardized tests and reading achievement

When analyzing the effects of teachers' frequency of use of standardized tests to monitor learning on student reading achievement, both student performance on standardized tests and teacher judgments were used, as they provide both formal and

60

informal data, respectively. Data yield by standardized tests, aside from serving as an accountability measure for multiple stakeholders, are aligned to grade-specific competencies students must acquire, whereas teacher judgments of students' reading levels go beyond grade-specific competencies and consider classroom dynamics and individual student characteristics. Controlling for student and teacher characteristics, such as socioeconomic status and teachers' assessment-related professional development participation, yield more meaningful results - teacher practices and individual student characteristics can impact achievement (Black & William, 1998; Hao & Johnson, 2013). As such, two hierarchical regression analyses were performed to answer the first research question, one for each dependent variable (IRT theta scores and teachers' judgement of students' reading level). In order to proceed with this analysis, assumption testing for normality was completed. Normality tests were implemented for both dependent variables. Kolmogorov-Smirnov test indicated they were not normally distributed (p < p.05, sig = .000) (see Tables 44-45); however, though this assumption was violated, normal distribution is not essential in studies with large sample sizes (Williams et al., 2013).

Table 44

Tests of Normality: IRT Theta Scores

	Kolmogorov-Smirnov ^a					
_	Statistic	df	Sig.			
IRT_THETA	.025	11427	.000			

a. Lilliefors Significance Correction

Table 45	
Tests of Normality: Teachers' Judgments of Student Reading Level	

	Kolmogorov-Smirnov ^a					
	Statistic	df	Sig.			
TEACHJUDG	.215	10381	.000			

a. Lilliefors Significance Correction

With IRT theta scores as the dependent variable, and following Field's (2005) suggestion that the first model in a hierarchical regression should be comprised of known predictors, students' socioeconomic status and their race/ethnicity were used and significantly explained 20.5% of variance in student achievement ($R^2 = .205$, F =258.263, p < .05, sig = .000). The second and third models added teachers' assessmentrelated professional development participation and their frequency of use of standardized tests to monitor learning, respectively. The addition of these variables did not significantly explain variance in student achievement, as the variance in IRT theta scores did not change ($R^2 = .205$, F = 1.222, p > .05, sig = .269 and $R^2 = .205$, F = .022, p > .05, sig = .883, respectively) (see Table 46). However, all three models, independently, were statistically significant (p < .05, sig = .000) (see Table 47). The model encompassing all independent variables (Model 3), though statistically significant (F = 206.715, p < .05, sig = .000), included the following variables that, with one-unit change, do not significantly change the mean of IRT theta scores (see Table 48): Asian, non-Hispanic (B = -.052, p >.05, sig = .429), American Indian or Alaska Native, non-Hispanic (B = .049, p > .05, sig = .429), two or more races, non-Hispanic (B = .014, p > .05, sig = .744), unknown race/ethnicity (B = .034, p > .05, sig = .061), teachers' assessment-related professional development (B = -.003, p > .05, sig = .267), and teachers' frequency of use of standardized tests to monitor learning (B = .000, p > .05, sig = .883). On the other hand, as the value of students' socioeconomic status and Native Hawaiian or Other Pacific Islander race/ethnicity category increases, so do IRT theta scores significantly (B = .164, p < .05, sig = .000 and B = .061, p < .05, sig = .000, respectively). Conversely, as the value of the Black, non-Hispanic (B = -.128, p < .05, sig = .000) and Hispanic (B = -.065, p < .05, sig = .000) variables increase, IRT theta scores significantly decrease.

Table 46

Model Summary: Frequency of Use of Standardized Tests and Reading Achievement (IRT Theta Scores)

				Std	Change Statistics				
				Error of		F			
Mod		R	Adjusted	the	R Square	Chang			Sig. F
el	R	Square	R Square	Estimate	Change	e	df1	df2	Change
1	.453 ^a	.205	.204	.3142015	.205	258.26	8	8003	.000
						3			
2	.453 ^b	.205	.204	.3141971	.000	1.222	1	8002	.269
3	.453 ^c	.205	.204	.3142164	.000	.022	1	8001	.883

a. Predictors: (Constant), Unknown, AsianNonHispanic, AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES, Hispanic

b. Predictors: (Constant), Unknown, AsianNonHispanic, AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES, Hispanic, ProfDev

c. Predictors: (Constant), Unknown, AsianNonHispanic, AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES, Hispanic, ProfDev, FreqUseStandTests

d. Dependent Variable: IRT_THETA

Table 47

ANOVA: Frequency of Use of Standardized Tests and Reading Achievement (IRT Theta Scores)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	203.971	8	25.496	258.263	.000 ^b
	Residual	790.077	8003	.099		
	Total	994.048	8011			
2	Regression	204.091	9	22.677	229.709	.000 ^c
	Residual	789.956	8002	.099		
	Total	994.048	8011			

P	1					
	Total	994.048	8011			
	Residual	789.954	8001	.099		
3	Regression	204.094	10	20.409	206.715	.000 ^d

a. Dependent Variable: IRT_THETA

b. Predictors: (Constant), Hispanic, AsianNonHispanic,

AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown,

NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES

c. Predictors: (Constant), Hispanic, AsianNonHispanic,

AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown, NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES, ProfDev d. Predictors: (Constant), Hispanic, AsianNonHispanic,

AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown, NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES, ProfDev, FreqUseStandTests

Table 48

Effects of Frequency of Use of Standardized Tests on Reading Achievement (IRT Theta Scores)

	_	Unstanda Coeffic	ardized cients	Standardi zed Coefficie nts			Co	rrelation	5	Colline Statis	earity tics
			Std.				Zero-	Partia		Toleran	
Mod	el	В	Error	Beta	t	Sig.	order	1	Part	ce	VIF
1	(Constant)	1.510	.005		301.50	.000					
					7						
	SES	.165	.005	.382	33.824	.000	.434	.354	.337	.778	1.285
	BlackNonHispan	129	.013	103	-9.757	.000	136	108	097	.893	1.120
	ic										
	AsianNonHispan	053	.066	008	801	.423	022	009	008	.992	1.008
	ic										
	NativeHawaiian	.061	.014	.045	4.359	.000	.104	.049	.043	.946	1.057
	OtherPacificIslan										
	derNonHispanic										
	AmericanIndian	.047	.054	.009	.877	.381	.011	.010	.009	.995	1.005
	OrAlaskaNative										
	NonHispanic										
	TwoOrMoreRace	.015	.041	.004	.351	.725	008	.004	.004	.986	1.014
	sNonHispanic										
	Unknown	.034	.018	.019	1.864	.062	.053	.021	.019	.966	1.035

	Hispanic	066	.009	083	-7.081	.000	233	079	071	.721	1.387
2	(Constant)	1.519	.009		165.45	.000					
					6						
	SES	.164	.005	.381	33.615	.000	.434	.352	.335	.773	1.294
	BlackNonHispan	128	.013	102	-9.641	.000	136	107	096	.887	1.127
	ic										
	AsianNonHispan	052	.066	008	789	.430	022	009	008	.992	1.008
	ic										
	NativeHawaiian	.061	.014	.045	4.372	.000	.104	.049	.044	.946	1.057
	OtherPacificIslan										
	derNonHispanic										
	AmericanIndian	.049	.054	.009	.898	.369	.011	.010	.009	.995	1.005
	OrAlaskaNative										
	NonHispanic										
	TwoOrMoreRace	.013	.041	.003	.325	.745	008	.004	.003	.986	1.015
	sNonHispanic										
	Unknown	.034	.018	.019	1.874	.061	.053	.021	.019	.966	1.035
	Hispanic	065	.009	082	-6.984	.000	233	078	070	.717	1.394
	ProfDev	003	.003	011	-1.106	.269	080	012	011	.972	1.029
3	(Constant)	1.516	.018		84.199	.000					
	SES	.164	.005	.381	33.574	.000	.434	.351	.335	.771	1.298
	BlackNonHispan	128	.013	102	-9.639	.000	136	107	096	.886	1.129
	ic										
	AsianNonHispan	052	.066	008	792	.429	022	009	008	.991	1.009
	ic										
	NativeHawaiian	.061	.014	.045	4.374	.000	.104	.049	.044	.945	1.058
	OtherPacificIslan										
	derNonHispanic										
	AmericanIndian	.049	.054	.009	.899	.369	.011	.010	.009	.995	1.005
	OrAlaskaNative										
	NonHispanic										
	TwoOrMoreRace	.014	.041	.003	.327	.744	008	.004	.003	.986	1.015
	sNonHispanic										
	Unknown	.034	.018	.019	1.875	.061	.053	.021	.019	.966	1.035

Hispanic	065	.009	082	-6.984	.000	233	078	070	.716	1.397
ProfDev	003	.003	012	-1.110	.267	080	012	011	.923	1.084
FreqUseStandTe	.000	.003	.002	.147	.883	061	.002	.001	.930	1.075
sts										

a. Dependent Variable: IRT_THETA

Similar to the first model, results when using teachers' judgement of students' reading levels as the dependent variable included students' socioeconomic status and all race/ethnicity categories. This model significantly explained 12.9% of the variance in student achievement ($R^2 = .129$, F = 148.465, p < .05, sig = .000). The addition of teachers' assessment-related professional development participation to the second model also yielded significant results ($R^2 = .130$, F = 14.658, p < .05, sig = .000), as well as the addition of teachers' frequency of use of standardized tests to monitor learning to the third model ($R^2 = .131$, F = 4.674, p < .05, sig = .031) (see Table 49). All three models, independently, were also statistically significant (p < .05, sig = .000) (see Table 50). The model encompassing all independent variables (Model 3), though statistically significant (F = 120.962, p < .05, sig = .000), included the following variables that, with one-unit change, do not significantly change the mean of teachers' judgement of students' reading level (see Table 51): Asian, non-Hispanic (B = .087, p > .05, sig = .547), American Indian or Alaska Native, non-Hispanic (B = .163, p > .05, sig = .179), two or more races, non-Hispanic (B = .072, p > .05, sig = .776), and unknown race/ethnicity (B = .023, p > .05, sig = .565). On the other hand, as the value of students' socioeconomic status (B = .308, p < .308) .05, sig = .000), Native Hawaiian or Other Pacific Islander race/ethnicity category (B =.129, p < .05, sig = .000), teachers' assessment-related professional development (B =.022, p < .05, sig = .001), and teachers' frequency of use of standardized tests to monitor learning (B = .016, p < .05, sig = .031) increases, so do teachers' judgement of their

reading level. Conversely, as the value of the Black, non-Hispanic (B = -.163, p < .05, sig

= .000) and Hispanic (B = -.049, p <.05, sig = .018) variables increase, teachers'

judgement of their reading level decreases.

Table 49

Model Summary: Frequency of Use of Standardized Tests and Reading Achievement (Teacher Judgements)

				Ctd. Emer	Change Statistics						
Mod		R	Adjusted	of the	R Square	F			Sig. F		
el	R	Square	R Square	Estimate	Change	Change	df1	df2	Change		
1	.359ª	.129	.128	.702	.129	148.46	8	8029	.000		
						5					
2	.361 ^b	.130	.129	.702	.002	14.658	1	8028	.000		
3	.362°	.131	.130	.701	.001	4.674	1	8027	.031		

a. Predictors: (Constant), Hispanic, AsianNonHispanic, AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown, NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES

b. Predictors: (Constant), Hispanic, AsianNonHispanic, AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown, NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES, ProfDev

c. Predictors: (Constant), Hispanic, AsianNonHispanic, AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown, NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES, ProfDev, FreqUseStandTests

d. Dependent Variable: TEACHJUDG

Table 50

ANOVA: Frequency of Use of Standardized Tests and Reading Achievement (Teacher Judgements)

		Sum of				
Model		Squares	df	Mean Square	F	Sig.
1	Regression	585.700	8	73.213	148.465	.000 ^b
	Residual	3959.351	8029	.493		
	Total	4545.051	8037			
2	Regression	592.916	9	65.880	133.822	.000 ^c
	Residual	3952.135	8028	.492		
	Total	4545.051	8037			
3	Regression	595.216	10	59.522	120.962	.000 ^d
	Residual	3949.835	8027	.492		
	Total	4545.051	8037			
a Danar	dont Variables T	EACHILIDC				

a. Dependent Variable: TEACHJUDG

b. Predictors: (Constant), Hispanic, AsianNonHispanic,
AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown,
NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES
c. Predictors: (Constant), Hispanic, AsianNonHispanic,
AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown,
NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES, ProfDev
d. Predictors: (Constant), Hispanic, AsianNonHispanic, BlackNonHispanic, SES, ProfDev
d. Predictors: (Constant), Hispanic, AsianNonHispanic,
AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown,
NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES, ProfDev,
FreqUseStandTests

Table 51

Effects of Frequency of Use of Standardized Tests on Reading Achievement (Teacher Judgements)

	Unstandardized Coefficients		Standardi zed Coefficie nts			Co	rrelation	s	Colline Statis	earity tics	
Mod	del	В	Std. Error	Beta	t	Sig.	Zero- order	Partia 1	Part	Tolera nce	VIF
1	(Constant)	2.135	.011		191.2	.000					
	SES	.303	.011	.330	19 27.94 7	.000	.350	.298	.291	.780	1.283
	BlackNonHispan	151	.030	056	-5.123	.000	094	057	053	.894	1.119
	R AsianNonHispan ic	.098	.144	.007	.683	.495	008	.008	.007	.992	1.008
	NativeHawaiian OtherPacificIsla nderNonHispani	.129	.031	.044	4.150	.000	.085	.046	.043	.946	1.057
	C AmericanIndian OrAlaskaNative NonHispanic	.170	.121	.015	1.408	.159	.014	.016	.015	.995	1.005
	TwoOrMoreRac esNonHispanic	.061	.092	.007	.657	.511	006	.007	.007	.986	1.014
	Unknown	.024	.040	.006	.586	.558	.028	.007	.006	.966	1.035

	Hispanic	042	.021	025	-2.008	.045	160	022	021	.723	1.384
2	(Constant)	2.069	.020		101.2	.000					
					56						
	SES	.307	.011	.333	28.18	.000	.350	.300	.293	.775	1.291
					9						
	BlackNonHispan	160	.030	060	-5.402	.000	094	060	056	.888	1.126
	ic										
	AsianNonHispan	.091	.144	.007	.634	.526	008	.007	.007	.991	1.009
	ic										
	NativeHawaiian	.128	.031	.044	4.109	.000	.085	.046	.043	.946	1.057
	OtherPacificIsla										
	nderNonHispani										
	с										
	AmericanIndian	.161	.121	.014	1.333	.182	.014	.015	.014	.995	1.005
	OrAlaskaNative										
	NonHispanic										
	TwoOrMoreRac	.069	.092	.008	.748	.455	006	.008	.008	.986	1.014
	esNonHispanic										
	Unknown	.022	.040	.006	.560	.575	.028	.006	.006	.966	1.035
	Hispanic	047	.021	028	-2.275	.023	160	025	024	.719	1.391
	ProfDev	.025	.007	.040	3.829	.000	012	.043	.040	.973	1.028
3	(Constant)	1.995	.040		50.11	.000					
					5						
	SES	.308	.011	.335	28.26	.000	.350	.301	.294	.773	1.294
					9						
	BlackNonHispan	163	.030	061	-5.487	.000	094	061	057	.887	1.127
	ic										
	AsianNonHispan	.087	.144	.006	.602	.547	008	.007	.006	.991	1.009
	ic										
	NativeHawaiian	.129	.031	.044	4,154	.000	.085	.046	.043	.946	1.057
	OtherPacificIsla	/		.011			.005	.515		.210	1.557
	nderNonHispani										

____ c

AmericanIndian	.163	.121	.014	1.345	.179	.014	.015	.014	.995	1.005
OrAlaskaNative										
NonHispanic										
TwoOrMoreRac	.072	.092	.008	.776	.438	006	.009	.008	.986	1.014
esNonHispanic										
Unknown	.023	.040	.006	.576	.565	.028	.006	.006	.966	1.035
Hispanic	049	.021	029	-2.360	.018	160	026	025	.718	1.393
ProfDev	.022	.007	.035	3.242	.001	012	.036	.034	.923	1.083
FreqUseStandTe	.016	.007	.023	2.162	.031	013	.024	.022	.931	1.074
sts										

a. Dependent Variable: TEACHJUDG

Thus, the hypothesis that fifth grade teachers' use of standardized tests to monitor learning is significantly correlated to students' reading achievement (as measured by students' IRT theta scores on a direct cognitive reading assessment and teachers' judgement of students' reading levels), when accounting for students' socioeconomic status and race/ethnicity and teachers' assessment-related professional development participation, was accepted. Using data to monitor learning, however, is just the first step in ensuring student achievement – to effectively use these data and help students succeed, teachers must implement appropriate instructional practices.

Instructional Practices and Frequency of Use of Standardized Tests

The implementation of effective instructional practices should work, simultaneously, with teachers' use of data to monitor learning (Black et al., 2011; DeLuca & Johnson, 2017; Gómez-Monarrez et al., 2019; Livingston & Hutchinson, 2017; Mandinach, 2012; Reutzel & Cooter, 2019). Based on data gathered, teachers implement instructional practices appropriate to individual student needs. When looking at the effects of teachers' instructional practices with literary and informational text on their frequency of use of standardized tests to monitor learning, one must again account for student characteristics and teachers' assessment-related professional development participation, as these can impact student achievement (Curry, 2014; Hao & Johnson, 2013; Im, 2017). Thus, another hierarchical regression analysis was performed to answer the second research question. Kolmogorov-Smirnov test indicated the dependent variable, teachers' frequency of use of standardized tests, was not normally distributed (p < .05, sig = .000) (see Table 52); however, as previously mentioned, normal distribution is not essential in studies with large sample sizes (Williams et al., 2013).

Table 52

Tests of Normality: Frequency of Use of Standardized Tests

	Kolmogorov-Smirnov ^a							
	Statistic	df	Sig.					
FreqUseStandTests	.086	10322	.000					
T 111 0 01 10	<u> </u>							

a. Lilliefors Significance Correction

Table 53 shows the three models used for the hierarchical regression. The first model included students' socioeconomic status and all race/ethnicity categories. This model significantly explained 2% of the variance in teachers' frequency of use of standardized tests ($R^2 = .019$, F = 19.92, p < .05, sig = .000). Adding teachers' assessment-related professional development participation (model 2) significantly explained 6.9% of the variance ($R^2 = .069$, F = 66.261, p < .05, sig = .000), and adding teachers' instructional practices with literary and informational text (model 3) significantly explained 10.1% of the variance ($R^2 = .101$, F = 81.869, p < .05, sig = .000) (see Table 55). The third model, though statistically significant (F = 81.869, p < .05, sig = .000), included the following variables that, with one-unit change, do not significantly change the mean of teachers' frequency of use of standardized tests

to monitor learning (see Table 55): Asian, non-Hispanic (B = .251, p > .05, sig = .240), American Indian or Alaska Native, non-Hispanic (B = -.100, p > .05, sig = .581), two or more races, non-Hispanic (B = -.184, p > .05, sig = .176), unknown race/ethnicity (B = -.184, p > .05, sig = .176, p > .056, sig = .176, p > .056, p > .05.050, p > .05, sig = .400), and instructional practices with literary texts (B = .011, p > .05, sig = .587). On the other hand, as the value of the Black, non-Hispanic (B = .169, p < .169, p.05, sig = .000), Hispanic (B = .080, p < .05, sig = .009), teachers' assessment-related professional development (B = .173, p < .05, sig = .000), and informational text instruction (B = .185, p < .05, sig = .000) variables increases, teachers' use of standardized tests to monitor learning increases. Conversely, as the value of students' socioeconomic status (B = -.086, p < .05, sig = .000) and Native Hawaiian or Other Pacific Islander race/ethnicity category (B = -.100, p < .05, sig = .031) increases, teachers' frequency of use of standardized tests decreases. As such, the hypothesis that fifth grade teachers' instructional practices will be correlated with their use of standardized tests to monitor learning, after controlling for the effects of students' socioeconomic status and race/ethnicity and teachers' assessment-related professional development participation, was also accepted.

				Std Emon	Change Statistics							
Mod		R	Adjusted	of the	R Square	F			Sig. F			
el	R	Square	R Square	Estimate	Change	Change	df1	df2	Change			
1	.140ª	.020	.019	1.08580	.020	19.992	8	8040	.000			
2	.263 ^b	.069	.068	1.05807	.050	427.91	1	8039	.000			
						8						
3	.317°	.101	.100	1.04003	.032	141.67	2	8037	.000			
						0						

Table 53Model Summary: Instructional Practices and Frequency of Use of Standardized Tests

a. Predictors: (Constant), Hispanic, AsianNonHispanic, AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown, NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES

b. Predictors: (Constant), Hispanic, AsianNonHispanic, AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown, NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES, ProfDev

c. Predictors: (Constant), Hispanic, AsianNonHispanic, AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown, NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES, ProfDev, InstPracLitText, InstPracInfoText

d. Dependent Variable: FreqUseStandTests

Table 54

ANOVA: Instructional Practices and	Frequency of	^c Use of	Standardized	Tests
------------------------------------	--------------	---------------------	--------------	-------

		Sum of				
Mode	1	Squares	df	Mean Square	F	Sig.
1	Regression	188.560	8	23.570	19.992	.000 ^b
	Residual	9478.853	8040	1.179		
	Total	9667.413	8048			
2	Regression	667.621	9	74.180	66.261	.000 ^c
	Residual	8999.792	8039	1.120		
	Total	9667.413	8048			
3	Regression	974.100	11	88.555	81.869	.000 ^d
	Residual	8693.314	8037	1.082		
	Total	9667.413	8048			

a. Dependent Variable: FreqUseStandTests

b. Predictors: (Constant), Hispanic, AsianNonHispanic,

 $American Indian Or Alaska Native Non Hispanic, \ Two Or More Races Non Hispanic, \ Unknown,$

 $Native Hawaiian Other Pacific Islander Non Hispanic, \ Black Non Hispanic, \ SES$

c. Predictors: (Constant), Hispanic, AsianNonHispanic,

AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown, NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES, ProfDev

d. Predictors: (Constant), Hispanic, AsianNonHispanic, AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown, NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES, ProfDev, InstPracLitText, InstPracInfoText

Table 55

Effects of Instructional Practices on Frequency of Use of Standardized Tests

	-	Unstandardized Coefficients		Standardi zed Coefficie nts			Со	rrelation	S	Colline Statis	earity stics
Mod	del	В	Std. Error	Beta	t	Sig.	Zero- order	Partia 1	Part	Tolera nce	VIF
1	(Constant)	5.162	.017		299.1	.000					
	SES	101	.017	076	10 -6.042	.000	112	067	067	.779	1.283
	BlackNonHispa nic	.242	.046	.062	5.299	.000	.061	.059	.059	.893	1.120
	AsianNonHispa nic	.344	.223	.017	1.543	.123	.018	.017	.017	.992	1.008
	NativeHawaiian OtherPacificIsla nderNonHispani	075	.048	018	-1.549	.121	041	017	017	.946	1.057
	c AmericanIndian OrAlaskaNative NonHispanic	.027	.190	.002	.142	.887	001	.002	.002	.995	1.005
	TwoOrMoreRac esNonHispanic	207	.142	016	-1.462	.144	017	016	016	.986	1.014
	Unknown	040	.062	007	639	.523	022	007	007	.966	1.035
	Hispanic	.162	.032	.065	5.027	.000	.089	.056	.056	.723	1.384
2	(Constant)	4.628	.031		150.2 52	.000					

	SES	074	.016	055	-4.489	.000	112	050	048	.774	1.292
	BlackNonHispa nic	.171	.045	.044	3.821	.000	.061	.043	.041	.888	1.127
	AsianNonHispa	.285	.217	.014	1.314	.189	.018	.015	.014	.991	1.009
	nic										
	NativeHawaiian	088	.047	021	-1.870	.061	041	021	020	.946	1.057
	OtherPacificIsla										
	nderNonHispani										
	c										
	AmericanIndian	042	.185	002	228	.820	001	003	002	.995	1.005
	OrAlaskaNative										
	NonHispanic	129	120	011	1.002	217	017	011	011	0.97	1.015
	I woOrMoreKac	138	.138	011	-1.002	.317	017	011	011	.986	1.015
	Unknown	- 048	061	- 009	- 787	132	- 022	- 009	- 008	966	1.035
	Chkhown	048	.001	007	/0/	.432	022	007	008	.700	1.055
	Hispanic	.115	.031	.046	3.654	.000	.089	.041	.039	.719	1.391
	-										
	ProfDev	.205	.010	.226	20.68	.000	.242	.225	.223	.972	1.028
					6						
3	(Constant)	3.822	.058		65.59	.000					
					1						
					1						
	SES	086	.016	064	1 -5.356	.000	112	060	057	.772	1.295
	SES BlackNonHispa	086 .169	.016 .044	064 .043	-5.356 3.838	.000 .000	112 .061	060 .043	057 .041	.772 .886	1.295 1.129
	SES BlackNonHispa nic	086 .169	.016 .044	064 .043	-5.356 3.838	.000 .000	112 .061	060 .043	057 .041	.772 .886	1.295 1.129
	SES BlackNonHispa nic AsianNonHispa	086 .169 .251	.016 .044 .214	064 .043 .012	-5.356 3.838 1.176	.000 .000 .240	112 .061 .018	060 .043 .013	057 .041 .012	.772 .886 .991	1.295 1.129 1.009
	SES BlackNonHispa nic AsianNonHispa nic	086 .169 .251	.016 .044 .214	064 .043 .012	-5.356 3.838 1.176	.000 .000 .240	112 .061 .018	060 .043 .013	057 .041 .012	.772 .886 .991	1.295 1.129 1.009
	SES BlackNonHispa nic AsianNonHispa nic NativeHawaiian	086 .169 .251 100	.016 .044 .214 .046	064 .043 .012 024	-5.356 3.838 1.176 -2.164	.000 .000 .240 .031	112 .061 .018 041	060 .043 .013 024	057 .041 .012 023	.772 .886 .991 .946	 1.295 1.129 1.009 1.057
	SES BlackNonHispa nic AsianNonHispa nic NativeHawaiian OtherPacificIsla	086 .169 .251 100	.016 .044 .214 .046	064 .043 .012 024	1 -5.356 3.838 1.176 -2.164	.000 .000 .240 .031	112 .061 .018 041	060 .043 .013 024	057 .041 .012 023	.772 .886 .991 .946	1.295 1.129 1.009 1.057
	SES BlackNonHispa nic AsianNonHispa nic NativeHawaiian OtherPacificIsla nderNonHispani	086 .169 .251 100	.016 .044 .214 .046	064 .043 .012 024	1 -5.356 3.838 1.176 -2.164	.000 .000 .240 .031	112 .061 .018 041	060 .043 .013 024	057 .041 .012 023	.772 .886 .991 .946	1.295 1.129 1.009 1.057
	SES BlackNonHispa nic AsianNonHispa nic NativeHawaiian OtherPacificIsla nderNonHispani c	086 .169 .251 100	.016 .044 .214 .046	064 .043 .012 024	1 -5.356 3.838 1.176 -2.164	.000 .000 .240 .031	112 .061 .018 041	060 .043 .013 024	057 .041 .012 023	.772 .886 .991 .946	1.295 1.129 1.009 1.057
	SES BlackNonHispa nic AsianNonHispa nic NativeHawaiian OtherPacificIsla nderNonHispani c AmericanIndian OrAlaskaNative	086 .169 .251 100	.016 .044 .214 .046	064 .043 .012 024	-5.356 3.838 1.176 -2.164	.000 .000 .240 .031	112 .061 .018 041	060 .043 .013 024	057 .041 .012 023	.772 .886 .991 .946	1.295 1.129 1.009 1.057
	SES BlackNonHispa nic AsianNonHispa nic NativeHawaiian OtherPacificIsla nderNonHispani c AmericanIndian OrAlaskaNative NonHispanic	086 .169 .251 100	.016 .044 .214 .046	064 .043 .012 024	-5.356 3.838 1.176 -2.164 552	.000 .000 .240 .031	112 .061 .018 041	060 .043 .013 024	057 .041 .012 023	.772 .886 .991 .946	1.295 1.129 1.009 1.057 1.005
	SES BlackNonHispa nic AsianNonHispa nic NativeHawaiian OtherPacificIsla nderNonHispani c AmericanIndian OrAlaskaNative NonHispanic	086 .169 .251 100 100	.016 .044 .214 .046 .182	064 .043 .012 024 006	-5.356 3.838 1.176 -2.164 552 -1.354	.000 .000 .240 .031 .581	112 .061 .018 041 001	060 .043 .013 024 006	057 .041 .012 023 006	.772 .886 .991 .946 .995	1.295 1.129 1.009 1.057 1.005
	SES BlackNonHispa nic AsianNonHispa nic NativeHawaiian OtherPacificIsla nderNonHispani c AmericanIndian OrAlaskaNative NonHispanic	086 .169 .251 100 100	.016 .044 .214 .046 .182 .136	064 .043 .012 024 006	-5.356 3.838 1.176 -2.164 552 -1.354	.000 .000 .240 .031 .581 .176	112 .061 .018 041 001	060 .043 .013 024 006 015	057 .041 .012 023 006	.772 .886 .991 .946 .995 .983	1.295 1.129 1.009 1.057 1.005
	SES BlackNonHispa nic AsianNonHispa nic NativeHawaiian OtherPacificIsla nderNonHispani c AmericanIndian OrAlaskaNative NonHispanic TwoOrMoreRac esNonHispanic	086 .169 .251 100 100 184 050	.016 .044 .214 .046 .182 .136 .060	064 .043 .012 024 006 014 009	-5.356 3.838 1.176 -2.164 552 -1.354 841	.000 .000 .240 .031 .581 .176 .400	112 .061 .018 041 001 017 022	060 .043 .013 024 006 015 009	057 .041 .012 023 006 014 009	.772 .886 .991 .946 .995 .983 .983	1.295 1.129 1.009 1.057 1.005 1.017 1.036

ProfDev	.173	.010	.190	17.38	.000	.242	.190	.184	.934	1.071
				9						
InstPracLitText	.011	.020	.010	.543	.587	.187	.006	.006	.324	3.090
InstPracInfoText	.185	.020	.174	9.336	.000	.221	.104	.099	.323	3.094

a. Dependent Variable: FreqUseStandTests

Instructional Practices and Student Achievement

The significance of the relationship between teachers' use of standardized tests and (a) student achievement and (b) teachers' instructional practices has already been established. What remains to be explored amongst these three variables, and which will fully validate the model presented in *Figure* 1, is the effect of teachers' instructional practices on student reading achievement. This is aligned with the second level of Xu and Brown's (2016) framework, which focuses on the interrelationship amongst assessment, instructional practices, and student achievement. Thus, to answer the last research question, two other hierarchical regression analyses (with three models each) were performed, one per dependent variable of student achievement. Normality testing for these two variables was presented during the discussion of the first research question.

The first hierarchical regression used students' IRT theta scores as the dependent variable. The first model included students' socioeconomic status and all race/ethnicity categories. Except for White, non-Hispanic, which was excluded, this model significantly explained 20.5% of the variance in students' IRT theta scores ($R^2 = .205$, F = 256.759, p < .05, sig = .000), whereas adding teachers' assessment-related professional development participation (model 2) was not significant, as the variance percentage does not change ($R^2 = .205$, F = 1.275, p > .05, sig = .259). On the other hand, adding teachers' instructional practices with literary and informational texts (model 3) significantly explained 20.9% of the variance in students' IRT theta scores ($R^2 = .205$, F = 22.562, p <

.05, sig = .000) (see Table 56). As with the first two research questions, all three models, independently, were statistically significant (p < .05, sig = .000) (see Table 57). The full model, though statistically significant (F = 191.968, p < .05, sig = .000), included the following variables that, with one-unit change, do not significantly change the mean of students' IRT theta scores (see Table 58): Asian, non-Hispanic (B = -.057, p > .05, sig = .387), American Indian or Alaska Native, non-Hispanic (B = .031, p > .05, sig = .572), two or more races, non-Hispanic (B = .003, p > .05, sig = .933), unknown race/ethnicity (B = .032, p > .05, sig = .074), and instructional practices with informational texts (B = .032, p > .05, sig = .074).003, p > .05, sig = .635). On the other hand, as the value of students' socioeconomic status (B = .163, p < .05, sig = .000), and Native Hawaiian or Other Pacific Islander race/ethnicity category (B = .058, p < .05, sig = .000), and teachers' literary text instruction (B = .021, p < .05, sig = .000) variables increases, IRT theta scores increase. Conversely, as the value of the Black, non-Hispanic (B = -.127, p < .05, sig = .000), Hispanic (B = -.069, p < .05, sig = .000), and teachers' assessment related-professional development (B = -.007, p < .05, sig = .017) increases IRT theta scores tests decrease.

Table 56

Model	Summary:	Instructional	Practices	and Stud	lent Achieveme	ent (IRT	' Theta S	lcores)
-------	----------	---------------	-----------	----------	----------------	----------	-----------	---------

				C(1 Emer					
Mod		R	Adjusted	of the	R Square	F			Sig. F
el	R	Square	R Square	Estimate	Change	Change	df1	df2	Change
1	.452ª	.205	.204	.3144094	.205	256.75	8	7983	.000
						9			
2	.453 ^b	.205	.204	.3144040	.000	1.275	1	7982	.259
3	.457°	.209	.208	.3135582	.004	22.562	2	7980	.000

a. Predictors: (Constant), Hispanic, AsianNonHispanic, AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown, NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES

b. Predictors: (Constant), Hispanic, AsianNonHispanic, AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown, NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES, ProfDev

c. Predictors: (Constant), Hispanic, AsianNonHispanic, AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown, NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES, ProfDev, InstPracLitText, InstPracInfoText d. Dependent Variable: IRT_THETA

Table 57

Effects of Instructional Practices on Student Achievement (IRT Theta Scores)

		Sum of				
Model		Squares	df	Mean Square	F	Sig.
1	Regression	203.051	8	25.381	256.759	.000 ^b
	Residual	789.146	7983	.099		
	Total	992.197	7991			
2	Regression	203.178	9	22.575	228.379	.000 ^c
	Residual	789.020	7982	.099		
	Total	992.197	7991			
3	Regression	207.614	11	18.874	191.968	.000 ^d
	Residual	784.583	7980	.098		
	Total	992.197	7991			

a. Dependent Variable: IRT_THETA

b. Predictors: (Constant), Hispanic, AsianNonHispanic,

 $American Indian Or Alaska Native Non Hispanic, \ Two Or More Races Non Hispanic, \ Unknown,$

 $Native Hawaiian Other Pacific Islander Non Hispanic, \,Black Non Hispanic, \,SES$

c. Predictors: (Constant), Hispanic, AsianNonHispanic,

AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown, NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES, ProfDev d. Predictors: (Constant), Hispanic, AsianNonHispanic,

AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown, NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES, ProfDev, InstPracLitText, InstPracInfoText

Table 58

Effect of Instructional Practices on Student Achievement (IRT Theta Scores)

		Unstanda Coeffic	Standardi zed Unstandardized Coefficie Coefficients nts				Co	rrelation	s	Colline Statis	earity stics
Std.					Zero-	Partia	_	Tolera			
Mo	del	В	Error	Beta	t	Sig.	order	1	Part	nce	VIF
1	(Constant)	1.510	.005		300.9	.000					
					43						
	SES	.165	.005	.382	33.73	.000	.434	.353	.337	.778	1.286
					9						

	BlackNonHispan	130	.013	103	-9.750	.000	137	108	097	.893	1.120	
	ic											
	AsianNonHispan	053	.066	008	805	.421	022	009	008	.992	1.008	
	ic											
	NativeHawaiian	.060	.014	.044	4.287	.000	.104	.048	.043	.946	1.057	
	OtherPacificIsla											
	nderNonHispani											
	с											
	AmericanIndian	.035	.055	.006	.641	.522	.009	.007	.006	.995	1.005	
	OrAlaskaNative											
	NonHispanic											
	TwoOrMoreRac	.014	.041	.003	.344	.731	008	.004	.003	.986	1.014	
	esNonHispanic											
	Unknown	.034	.018	.019	1.858	.063	.053	.021	.019	.966	1.035	
	Hispanic	066	.009	083	-7.030	.000	232	078	070	.721	1.388	
2	(Constant)	1.519	.009		165.2	.000						
					40							
	SES	.164	.005	.381	33.53	.000	.434	.351	.335	.773	1.294	
					2							
	BlackNonHispan	128	.013	102	-9.630	.000	137	107	096	.887	1.127	
	ic											
	AsianNonHispan	052	.066	008	792	.428	022	009	008	.992	1.008	
	ic											
	NativeHawaiian	.060	.014	.044	4.301	.000	.104	.048	.043	.946	1.057	
	OtherPacificIsla											
	nderNonHispani											
	c											
	AmericanIndian	.036	.055	.007	.661	.508	.009	.007	.007	.995	1.005	
	OrAlaskaNative											
	NonHispanic											
	TwoOrMoreRac	.013	.041	.003	.318	.751	008	.004	.003	.986	1.015	
	esNonHispanic											
	Unknown	.034	.018	.019	1.866	.062	.053	.021	.019	.966	1.035	
	Hispanic	065	.009	082	-6.933	.000	232	077	069	.717	1.395	

	ProfDev	003	.003	011	-1.129	.259	080	013	011	.972	1.029
3	(Constant)	1.419	.018		79.60	.000					
					3						
	SES	.163	.005	.377	33.24	.000	.434	.349	.331	.771	1.298
					1						
	BlackNonHispan	127	.013	101	-9.519	.000	137	106	095	.885	1.129
	ic										
	AsianNonHispan	057	.066	009	866	.387	022	010	009	.992	1.009
	ic										
	NativeHawaiian	.058	.014	.043	4.202	.000	.104	.047	.042	.945	1.058
	OtherPacificIsla										
	nderNonHispani										
	c										
	AmericanIndian	.031	.055	.006	.565	.572	.009	.006	.006	.995	1.005
	OrAlaskaNative										
	NonHispanic										
	TwoOrMoreRac	.003	.041	.001	.084	.933	008	.001	.001	.984	1.017
	esNonHispanic										
	Unknown	.032	.018	.018	1.786	.074	.053	.020	.018	.966	1.036
	Hispanic	069	.009	086	-7.314	.000	232	082	073	.714	1.401
	ProfDev	007	.003	024	-2.378	.017	080	027	024	.935	1.069
	InstPracLitText	.021	.006	.061	3.533	.000	.066	.040	.035	.328	3.048
	InstPracInfoText	.003	.006	.008	.475	.635	.043	.005	.005	.328	3.052

a. Dependent Variable: IRT_THETA

The second hierarchical regression used teachers' judgement of students' reading level as the dependent variable. The first model included students' socioeconomic status and all race/ethnicity categories. Except for White, non-Hispanic, which was excluded, this model significantly explained 12.8% of the variance in teachers' judgment of students' reading level ($R^2 = .128$, F = 147.532, p < .05, sig = .000). Adding teachers' assessment-related professional development participation (model 2) and adding teachers' instructional practices with literary and informational text (model 3) also yielded significant results ($R^2 = .130$, F = 14.379, p < .05, sig = .000 and $R^2 = .135$, F =

20.934, p < .05, sig = .000, respectively; see Table 59). As with the previous hierarchical regression analysis using students' IRT theta scores, all three models, independently, were statistically significant (p < .05, sig = .000; see Table 60). However, the full model, though statistically significant (F = 113.130, p < .05, sig = .000), included the following variables that, with one-unit change, do not significantly change the mean of teachers' judgement of students' reading level (see Table 61): Asian, non-Hispanic (B = .083, p >.05, sig = .565), American Indian or Alaska Native, non-Hispanic (B = .119, p > .05, sig = .330), two or more races, non-Hispanic (B = .046, p > .05, sig = .616), unknown race/ethnicity (B = .021, p > .05, sig = .603), and instructional practices with informational texts (B = .019, p > .05, sig = .164). On the other hand, as the value of students' socioeconomic status (B = .302, p < .05, sig = .000), Native Hawaiian or Other Pacific Islander race/ethnicity category (B = .123, p < .05, sig = .000), teachers' assessment-related professional development (B = .016, p < .05, sig = .014), and teachers' literary text instruction (B = .034, p < .05, sig = .011) variables increases, teachers' judgement of students' reading levels increases. Conversely, as the value of the Black, non-Hispanic (B = -.158, p < .05, sig = .000) and Hispanic (B = -.056, p < .05, sig = .008) variables increases, teachers' judgement of students' reading level decreases.

 Table 59

 Model Summary: Instructional Practices and Student Achievement (Teacher Judgment)

				Std. Error		Chang	ge Statist	tics	
Mod		R	Adjusted	of the	R Square	F			Sig. F
el	R	Square	R Square	Estimate	Change	Change	df1	df2	Change
1	.358ª	.128	.128	.702	.128	147.53	8	8010	.000
						2			
2	.361 ^b	.130	.129	.702	.002	14.379	1	8009	.000
3	.367°	.135	.133	.700	.005	20.934	2	8007	.000

a. Predictors: (Constant), Hispanic, AsianNonHispanic, AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown, NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES

b. Predictors: (Constant), Hispanic, AsianNonHispanic, AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown, NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES, ProfDev

c. Predictors: (Constant), Hispanic, AsianNonHispanic, AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown, NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES, ProfDev, InstPracLitText, InstPracInfoText d. Dependent Variable: TEACHJUDG

d. Dependent variable: TEACHJU

Table 60

ANOVA: Instructional Practices and Student Achievement (Teacher Judgment)

		Sum of				
Model		Squares	df	Mean Square	F	Sig.
1	Regression	582.024	8	72.753	147.532	.000 ^b
	Residual	3949.985	8010	.493		
	Total	4532.009	8018			
2	Regression	589.102	9	65.456	132.957	.000 ^c
	Residual	3942.907	8009	.492		
	Total	4532.009	8018			
3	Regression	609.612	11	55.419	113.130	.000 ^d
	Residual	3922.397	8007	.490		
	Total	4532.009	8018			

a. Dependent Variable: TEACHJUDG

b. Predictors: (Constant), Hispanic, AsianNonHispanic,

AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown,

NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES

c. Predictors: (Constant), Hispanic, AsianNonHispanic,

AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown, NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES, ProfDev d. Predictors: (Constant), Hispanic, AsianNonHispanic,

AmericanIndianOrAlaskaNativeNonHispanic, TwoOrMoreRacesNonHispanic, Unknown,

NativeHawaiianOtherPacificIslanderNonHispanic, BlackNonHispanic, SES, ProfDev, InstPracLitText, InstPracInfoText

		Unstandardized Coefficients		Standardi zed Coefficie nts			Cor	rrelation	8	Colline Statis	earity tics
			Std.				Zero-	Partia		Tolera	
Mo	del	В	Error	Beta	t	Sig.	order	1	Part	nce	VIF
1	(Constant)	2.136	.011		191.0 48	.000					
	SES	.303	.011	.329	27.85 1	.000	.350	.297	.291	.779	1.283
	BlackNonHispan ic	151	.030	056	-5.112	.000	094	057	053	.893	1.119
	AsianNonHispan ic	.097	.144	.007	.673	.501	008	.008	.007	.992	1.008
	NativeHawaiian OtherPacificIsla nderNonHispani	.128	.031	.044	4.119	.000	.086	.046	.043	.946	1.057
	c AmericanIndian OrAlaskaNative NonHispanic	.141	.123	.012	1.148	.251	.012	.013	.012	.995	1.005
	TwoOrMoreRac esNonHispanic	.059	.092	.007	.643	.520	006	.007	.007	.986	1.014
	Unknown	.024	.040	.006	.608	.543	.028	.007	.006	.966	1.035
	Hispanic	042	.021	025	-2.009	.045	160	022	021	.722	1.384
2	(Constant)	2.071	.020		101.2 42	.000					
	SES	.306	.011	.333	28.09 0	.000	.350	.299	.293	.774	1.292
	BlackNonHispan ic	160	.030	060	-5.392	.000	094	060	056	.888	1.126
	AsianNonHispan ic	.090	.144	.007	.624	.533	008	.007	.007	.991	1.009

Table 61Effects of Instructional Practices on Student Achievement (Teacher Judgment)

	NativeHawaiian OtherPacificIsla	.127	.031	.044	4.074	.000	.086	.045	.042	.946	1.057
	nderNonHispani										
	с										
	AmericanIndian	.133	.123	.011	1.080	.280	.012	.012	.011	.995	1.005
	OrAlaskaNative										
	NonHispanic										
	TwoOrMoreRac	.068	.092	.008	.733	.464	006	.008	.008	.986	1.014
	esNonHispanic										
	Unknown	.024	.040	.006	.587	.557	.028	.007	.006	.966	1.035
	Hispanic	047	.021	028	-2.271	.023	160	025	024	.719	1.391
	ProfDev	.025	.007	.040	3.792	.000	012	.042	.040	.973	1.028
3	(Constant)	1.854	.039		47.14	.000					
					6						
	SES	.302	.011	.328	27.74	.000	.350	.296	.288	.772	1.296
					7						
	BlackNonHispan	158	.030	059	-5.319	.000	094	059	055	.886	1.128
	ic										
	AsianNonHispan	.083	.144	.006	.575	.565	008	.006	.006	.991	1.009
	ic										
	NativeHawaiian	.123	.031	.042	3.961	.000	.086	.044	.041	.946	1.057
	OtherPacificIsla										
	C										
	AmericanIndian	119	122	010	975	330	012	011	010	995	1.005
	OrAlaskaNative	,	.122	.010	.915	.550	.012	.011	.010	.,,,,	1.005
	NonHispanic										
	TwoOrMoreRac	.046	.092	.005	.501	.616	006	.006	.005	.984	1.016
	esNonHispanic										
	Unknown	.021	.040	.005	.519	.603	.028	.006	.005	.966	1.036
	Hispanic	056	.021	033	-2.668	.008	160	030	028	.716	1.397
	ProfDev	.016	.007	.026	2.447	.014	012	.027	.025	.935	1.070
	InstPracLitText	.034	.013	.046	2.547	.011	.078	.028	.026	.325	3.082
	InstPracInfoText	.019	.013	.025	1.390	.164	.066	.016	.014	.324	3.085

a. Dependent Variable: TEACHJUDG

As both full models using IRT theta scores and teachers' judgment of students' reading level, while controlling for the effects of students' socioeconomic status and race/ethnicity and teachers' assessment-related professional development participation, were statistically significant (p < .05, sig = .000), the hypothesis that teachers' instructional practices are correlated to student achievement was accepted.

Summary

Chapter IV provided an overall description of variables used and discussed the data analyses implemented and results of the study. As per the hierarchical regression analyses conducted, all three hypotheses were accepted, thus supporting the model presented in *Figure 1*. While controlling for students' socioeconomic status and race/ethnicity as well as teachers' assessment-related professional development participation, results of the study indicate that fifth grade teachers' (a) frequency of use of standardized tests to monitor learning is correlated to student reading achievement, (b) instructional practices are correlated with their frequency of use of standardized tests to monitor learning, and (c) instructional practices correlate with student achievement.

CHAPTER V

DISCUSSION

The purpose of this study was to explore, through the lens of data use theory (Mandinach et al., 2006) and by employing hierarchical regression analyses, the relationship between fifth grade teachers' use of standardized tests to monitor learning, instructional practices, and student achievement, while accounting for teacher and student characteristics. As Hutchins (1995) and Spillane (2012) argue, interactions amongst these variables play a key role in teachers' decision-making process and, in turn, on student achievement. This chapter includes a summary of the findings and discusses the implications and recommendations for future research.

Summary of Findings

The fourth-grade slump (Chall & Jacobs, 2003; Hirsch, 2003) refers to a decline in reading comprehension, particularly in low-income students. To help overcome it, targeted interventions have been recommended (Best et al., 2004; Palacios, 2017). These interventions effectively occur if teachers use data to monitor learning and implement appropriate instructional practices, in turn leading to student success. However, teacher preparation and individual student characteristics must be considered, as they also play an important role in student achievement (Dee & Jacob, 2011). The model presented in *Figure 1* illustrates the interconnectedness between the use of data to monitor learning, teachers' instructional practices, student achievement, teachers' assessment-related professional development, and student characteristics, all with the goal of better understanding the relationship between the use of assessment data and teachers' instructional practices and their impact on elementary student achievement, a need

86

identified in the literature (DeLuca & Johnson, 2017; DeLuca & Klinger, 2010; Hao & Johnson, 2013; Hill et al., 2017; Livingston & Hutchinson, 2017; Werts et al., 2014). Results of this study support this model – the effects of (a) teachers' frequency of use of standardized tests on reading achievement (both formal and informal), (b) teachers' instructional practices on their use of standardized tests, and (c) teachers' instructional practices on student achievement, were all positive and significant (p < .05, sig = .000), when accounting for relevant teacher and student characteristics (teachers' assessment-related professional development participation and students' socioeconomic status and race/ethnicity).

Though the three models overall were significant, the effect of the independent variables on the dependent variable, individually, varied. Students' socioeconomic status had significant positive effects on all dependent variables except on teachers' frequency of use of standardized tests (B = -.086, p < .005, sig = .000). Conversely, the Black, non-Hispanic and Hispanic variables had significant negative effects on all dependent variables except on teachers' frequency of use of standardized tests (B = .169, p < .005, sig = .000 and B = .080, p < .005, sig = .009, respectively). This finding supports the National Center for Education Statistics (2019) report on the reading achievement gap amongst fourth grade students and the fourth-grade slump phenomenon (Chall & Jacobs, 2003; Hirsch, 2003) – students with these demographics do seem to perform lower on standardized tests. Teachers' assessment-related professional development participation had a positive and significant effect on all dependent variables except on IRT theta scores when measuring the effect of teachers' frequency of use of standardized tests to monitor learning on reading achievement (B = .003, p > .005, sig = .267). When measuring the

87

effects of teachers' instructional practices on reading achievement, this same variable had a negative significant effect on students' IRT theta scores (B = -.007, p < .005, sig = .017). This further demonstrates the importance of teachers' assessment-related professional development, as argued in the literature (Black & William, 1998; Black et al., 2011; DeLuca & Johnson, 2017; Gómez-et al., 2019; Livingston & Hutchinson, 2017; Mandinach, 2012; Reutzel & Cooter, 2019). Furthermore, teachers' frequency of use of standardized test data to monitor learning was only significant (B = .016, p < .05, sig = .031) when using teachers' judgement of student reading levels as the dependent variable. Teachers' instructional practices with informational text variable yielded significant positive results (B = .185, p < .05, sig = .000) on its effect on teachers' frequency of use of standardized tests, indicating teachers' instructional practices with these types of texts might be intentional and based on standardized test data. Teachers' instructional practices with literary text, on the other hand, was the only instructional practice to yield significant results when examining its effect on students' IRT theta scores (B = .003, p < .003.05, sig = .000) and teachers' judgement of students' reading levels (B = .019, p < .05, sig = .011).

Implications

Results of this study pose several implications for the field of elementary reading education. First, teachers' instructional practices having a positive effect on their use of standardized tests to monitor learning when students are Black, non-Hispanic (B = .169, p< .05) or Hispanic (B = .080, p < .05, sig = .009) suggests the merging of instructional practices and use of data could play a significant role for these historically lowperforming students – as the value increases so does teachers' use of data to monitor learning. This is aligned with results for students' socioeconomic status - as the value increases, teachers' use of data to monitor learning decreases (B = -.086, p < .05, sig = .000). Students with a high socioeconomic status have been found to be performing above minority groups (National Center for Education Statistics, 2019), which could explain this result. However, despite these results, Black, non-Hispanic, Hispanic, and low socioeconomic status students continue to struggle, putting into question the appropriateness of instructional practices being implemented with these students. Implementing instructional practices while using data to monitor learning is a great start, but it does not guarantee the effectiveness or appropriateness of the instruction being provided. On the other hand, the sole use of standardized assessment data to monitor learning has a negative effect on their performance -- teachers' use of these had a negative, significant effect (p < .05) on the achievement of Black, non-Hispanic and Hispanic students – as teachers increase their use of standardized tests to monitor data, students' IRT theta scores and teachers' judgment of their reading performance decreases for Black, non-Hispanic and Hispanic students (B = -.163, p < .05, sig = .000 and B = -.049, p < .05, sig = .018, respectively). This demonstrates the importance of preparing both in-service and preservice teachers to accurately interpret data standardized tests yield and supports the need expressed in the literature for assessment-related teacher professional development (e.g, Black et al., 2011; DeLuca & Johnson, 2017; Gómez-Monarrez et al., 2019; Livingston & Hutchinson, 2017; Mandinach, 2012; Reutzel & Cooter, 2019).

Secondly, teachers' assessment-related professional development participation's positive significant effect on their judgment of students' reading levels (B = .022, p < .05, sig = .001) and on their frequency of use of standardized tests to monitor learning (B = .173, p < .05, sig = .000) suggests there is a benefit to targeted, assessment-related professional development, though this benefit is not yet apparent in student performance on formal assessments.

Moreover, the significant positive effect (B = .185, p < .05, sig = .000) of teachers' instruction with informational text on their use of standardized tests to monitor learning suggests teachers are aware of the importance of informational text instruction in the upper elementary grades. As their instruction with informational text increases, so does their use of standardized tests to monitor learning. This is aligned with Palacios's (2017) and Toste and Ciullo's (2017) discussion that this is the predominant type of text in third grade and above, and a possible contributor to the decrease in reading achievement scores in upper elementary students -- most of the texts used in standardized tests is informational. Similarly, teachers' instruction with literary texts having a significant effect on students' direct cognitive scores (B = .021, p < .05, sig = .000) and on teacher's judgment of student reading levels (B = .034, p < .05, sig = .011) suggests this type of text still plays a major role during classroom instruction – as instruction using narrative text increases, students' direct cognitive assessment score and teachers' judgment of reading level increases. Most importantly, however, when looking at the three models overall, their positive and significant (p < .05) results demonstrate that the path to student achievement is not linear, but rather complex with the intricacies of student and teacher characteristics playing a major role. Thus, instruction needs to be tailored to each individual student, just as much as professional development and teacher preparation programs need to be tailored to each of their teachers and the population of students they will serve.

Limitations and Recommendations for Future Research

Since this study used existing data to answer the research questions, several relevant questions remain unanswered due to its limitations. First, data used were from a representative sample of students across the country; however, specific regions of interest (e.g., minority-predominant areas) were not isolated. Therefore, exploring programs like the Master of Teaching in New Zealand (Hill et al., 2017) and the Targeted Reading Intervention (Vernon-Feagans et al., 2018) is recommended. In these programs, teachers are purposefully prepared to meet the needs of historically underachieving students by, in part, learning about effective assessment and instructional practices with these groups. Curricula from programs like these can be used in both teacher preparation programs and in professional development opportunities for in-service teachers. As Gómez-Monarrez et al. (2019) found, there seems to be limited training on how to adjust instruction within our teacher education programs, and Hispanics and Black students continue to struggle with reading (National Center for Education Statistics, 2019). A guiding question could be: what characteristics make such a program successful and how are they transferable to the diverse preservice teacher and student population in different regions of the United States?

91

The other limitation was the researcher's inability to manipulate teachers' survey questions – only questions related to teachers' pedagogical practices with informational and literary text were asked. Collecting data regarding teachers' content and pedagogical content knowledge would have provided a more thorough analysis, as teachers draw from both during instruction (Jordan et al., 2018). The second recommendation, thus, is to explore meaningful assessment-related professional development opportunities for teachers that specifically impact student achievement on formal assessments. Rather than "teaching to the test", do teachers have an in-depth understanding of the skills students are to demonstrate mastery of in standardized tests, and what pedagogical and content knowledge should they possess to help students improve their performance on these tests? The model presented in *Figure 1* can be used as the foundation to begin to answer this question.

The third recommendation, and perhaps the most important, is to deviate from the "standardized tests are bad" discourse and instead focus efforts on DeLuca and Johnson's (2017) advise to reconceptualize and integrate (rather than in disconnected fragments) assessment instruction at both the preservice and in-service teacher levels. To do so, both formal and informal assessments must be presented as tools that, when used appropriately, can yield important information about student performance.

Summary

Chapter V provided a summary of the findings, implications of the study, and recommendations for future research. Gaps in the literature regarding the effects of teachers' use of standardized tests to monitor learning and their assessment-related instructional practices on upper elementary student reading achievement were filled.

92

When accounting for students' socioeconomic status and race/ethnicity and teachers' assessment-related professional development participation, teachers' use of standardized test data to monitor learning, as well as their instructional practices, have significant positive effects on student reading achievement, as measured by performance on a standardized reading assessment and teachers' judgment of student reading level. Results suggest teachers are using data to differentiate instruction for historically low-performing students and supports the benefits described in the literature (e.g., Hao & Johnson, 2013) to targeted assessment-related professional development. However, there certainly is room for additional research since more than half of our states continue to use standardized tests (Ross & Walsh, 2019) on our diverse student population. Recommendations for future research were to explore programs that prepare teachers to help struggling students from specific populations, determine the pedagogical and content knowledge required for teachers to effectively help students improve performance on standardized tests, and reconceptualizing assessment instruction at both the preservice and in-service teacher levels.
REFERENCES

Adams, M. J. (2001). Alphabetic anxiety and explicit, systematic phonics instruction: A cognitive science perspective. *Handbook of early literacy research*, 1, 66-80.

Afflerbach, P. P., & Cho, B. (2011). The Classroom Assessment of Reading. In M. L. Kamil, P. D. Pearson, E. B. Moje, & P. P. Afflerbach (Eds.), *Handbook of Reading Research, Volume IV* (pp. 487-514). Routledge Taylor & Francis Group.

Afflerbach, P., Cho, B., Crassas, M. E., Kim, J. (2019) Best Practices in Reading Assessment. In L. M. Morrow & L. B. Gambrell (Eds), *Best Practices in Literacy Instruction* (6th ed.) (pp. 309-336). Guilford Press.

American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). *Standards for teacher competence in educational assessment of students*. U.S. Department of Education. https://files.eric.ed.gov/fulltext/ED323186.pdf

Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, 10 (18).

Andersson, C., & Palm, T. (2017). Characteristics of improved formative assessment practice. *Education Inquiry*, 8(2), 104-122.

Anderson, S. & Maxwell, S. (2018). Sample size. In B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation* (Vol. 1, pp. 1441-1444). SAGE Publications, Inc.

Assessment and Accountability. Fla. Stat. tit.48, § 1008.02-1008.30 (2019). http://www.leg.state.fl.us/statutes/index.cfm?App_mode=Display_Statute&URL=1000-1099/1008/1008.html

Au, W. (2011). Teaching under the new Taylorism: High-stakes testing and the standardization of the 21st century curriculum. *Journal of Curriculum Studies*, 43(1), 25-45.

Banta, T. W., Jones, E. A., & Black, K. E. (2009). *Designing effective assessment: Principles and profiles of good practice*. John Wiley & Sons.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6), 1173.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: principles, policy & practice*, 18(1), 5-25.

Best, R., Floyd, R. G., & McNamara, D. S. (2004). Understanding the fourth-grade slump: Comprehension difficulties as a function of reader aptitudes and text genre. In *85th Annual Meeting of the American Educational Research Association*. https://www.researchgate.net/profile/Danielle_Mcnamara/publication/251773068_Unders tanding_the_FourthGrade_Slump_Comprehension_Difficulties_as_a_Function_of_Reade r_Aptitudes_and_Text_Genre/links/5dc1b6d8299bf1a47b18f482/Understanding-the-Fourth-Grade-Slump-Comprehension-Difficulties-as-a-Function-of-Reader-Aptitudes-and-Text-Genre.pdf

Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2011). Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education: Principles, Policy & Practice*, 18(4), 451-469.

Black, P., & Wiliam, D., (1998). Inside the Black Box: Raising Standards through Classroom Assessment. *The Phi Delta Kappan*, 80(2), 139.

Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on formative and summative evaluation*. McGaw-Hill.

Brookhart, S. M. (2013). The use of teacher judgement for summative assessment in the USA. *Assessment in Education: Principles, Policy & Practice*, 20(1), 69-90.

Brown, G. T. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, 11(3), 301-318.

Brown, J., & Rolfe, S. A. (2005). Use of child development assessment in early childhood education: Early childhood practitioner and student attitudes toward formal and informal testing. *Early child development and care*, 175(3), 193-202.

Chall, J. S., & Jacobs, V. A. (2003). The classic study on poor children's fourth-grade slump. *American educator*, 27(1), 14-15.

Chen, P. & Krauss, A. (2004). Pearson's correlation coefficient. In M. S. Lewis-BeckA. Bryman & T. F. Liao (Eds.), *The SAGE encyclopedia of social science research methods* (Vol. 1, pp. 808-810). SAGE Publications, Inc.

Cheng, H. G., & Phillips, M. R. (2014). Secondary analysis of existing data: opportunities and implementation. *Shanghai archives of psychiatry*, 26(6), 371.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. (1966). Equality of educational opportunity. *U.S. Government Printing Office*. https://files.eric.ed.gov/fulltext/ED012275.pdf

Cooter, R. B. Jr., & Perkins, J. H. (2007). Looking to the future with *The Reading Teacher*: A 900-year-old sheep and *Papa na come! The Reading Teacher*, 61, 4-7.

Curry, D. L. (2014). The impact of teacher quality on reading achievement of fourth grade students: An analysis of the 2007, 2009, 2011, and 2013 National Assessment of Educational Progress (NAEP). University of North Texas.

Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, 30(3), 418-446.

DeLuca, C., & Johnson, S. (2017). Developing assessment capable teachers in this age of accountability. *Assessment in Education: Principles, Policy & Practice*, 24(2), 121-126.

DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17(4), 419-438.

Drake, G. and Walsh, K. (2020). 2020 Teacher Prep Review: Program Performance in Early Reading Instruction. National Council on Teacher Quality. www.nctq.org/publications/2020-Teacher-Prep-Review:-Program-Performance-in-Early-Reading-Instruction

Ercan, I., Yazici, B., Sigirli, D., Ediz, B., & Kan, I. (2007). Examining Cronbach alpha, theta, omega reliability coefficients according to sample size. *Journal of modern applied statistical methods*, 6(1), 27.

Every Student Succeeds Act, 20 U.S.C. § 6301 (2015). https://www.congress.gov/114/plaws/publ95/PLAW-114publ95.pdf

Field, A. (2005). *Discovering Statistics Using SPSS (Introducing Statistical Methods)*. Sage Publications.

Gómez-Monarrez, C., Caso-López, A. A. C., & Gutiérrez-Anguiano, N. N. (2019). Effective Teaching practices reported in empirical research in the 21st Century. *Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality, Association for Computing Machinery*, 158-162.

Hamilton, M. (2017). Path analysis. In M. Allen (Ed.), *The sage encyclopedia of communication research methods* (Vol. 3, pp. 1194-1197). SAGE Publications, Inc.

Hao, S., & Johnson, R. L. (2013). Teachers' classroom assessment practices and fourthgraders' reading literacy achievements: An international study. *Teaching and Teacher Education*, 29, 53-63.

Hayasbi, K. & Yuan, K. (2010). Exploratory factor analysis. In N. J. Salkind (Ed.), *Encyclopedia of research design* (pp. 459-465). SAGE Publications, Inc.

Hill, M. F., Ell, F., Grudnoff, L., Haigh, M., Cochran-Smith, M., Chang, W. C., & Ludlow, L. (2017). Assessment for equity: Learning how to use evidence to scaffold learning and improve teaching. *Assessment in Education: Principles, Policy & Practice*, 24(2), 185-204.

Hirsch, E. D. (2003). Reading comprehension requires knowledge of words and the world. *American Educator*, 27(1), 10-13.

Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of educational research*, 59(3), 297-313.

Hutchins, E. (1995). Cognition in the Wild (No. 1995). MIT press.

Im, H. (2017). Kindergarten standardized testing and reading achievement in the US: Evidence from the early childhood longitudinal study. *Studies in Educational Evaluation*, 55, 9-18.

International Literacy Association. (2017). Standards for the preparation of literacy professionals. https://literacyworldwide.org/get-resources/standards/standards-2017

Jordan, R. L., Bratsch-Hines, M., & Vernon-Feagans, L. (2018). Kindergarten and first grade teachers' content and pedagogical content knowledge of reading and associations with teacher characteristics at rural low-wealth schools. *Teaching and Teacher Education*, 74, 190-204.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and psychological measurement*, 20(1), 141-151.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31-36.

Knoester, M., & Au, W. (2017). Standardized testing and school segregation: like tinder for fire? *Race Ethnicity and Education*, 20(1), 1-14.

Klinger, D.A., McDivitt, P.R., Howard, B.B., Munoz, M.A., Rogers, W.T., & Wylie, E.C. (2015). *The Classroom Assessment Standards for PreK-12 Teachers*. Kindle Direct Press.

Li, H. (2016). How is formative assessment related to students' reading achievement? Findings from PISA 2009. Assessment in Education: *Principles, Policy & Practice*, 23(4), 473-494.

Livingston, K., & Hutchinson, C. (2017). Developing teachers' capacities in assessment through career-long professional learning. *Assessment in Education: Principles, Policy & Practice*, 24(2), 290-307.

Lysynchuk, L. M., Pressley, M., & Vye, N. J. (1990). Reciprocal teaching improves standardized reading-comprehension performance in poor comprehenders. *The Elementary School Journal*, 90(5), 469-484.

Mandinach, E. B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist*, 47(2), 71-85.

Mandinach, E. B., Honey, M., & Light, D. (2006). A theoretical framework for datadriven decision making. *In annual meeting of the American Educational Research Association*, San Francisco, CA.

McCombes-Tolis, J., & Spear-Swerling, L. (2012). The preparation of preservice elementary educators in understanding and applying the terms, concepts, and practices associated with response to intervention in early reading contexts. *Journal of School Leadership*, 21(3), 360-389.

McFarland, J., Hussar, B., Zhang, J., Wang, X., Wang, K., Hein, S., Diliberti, M., Forrest Cataldi, E., Bullock Mann, F., and Barmer, A. (2019). *The Condition of Education 2019* (NCES 2019-144). U.S. Department of Education. National Center for Education Statistics. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2019144

Meissel, K., Meyer, F., Yao, E. S., & Rubie-Davies, C. M. (2017). Subjectivity of teacher judgments: Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education*, 65, 48-60.

Meister, D. G., & Jenks, C. (2000). Making the transition from preservice to inservice teaching: Beginning teachers' reflections. *Action in Teacher Education*, 22(3), 1-11.

Melnick, S. A., & Meister, D. G. (2008). A comparison of beginning and experienced teachers' concerns. *Educational Research Quarterly*, 31(3), 39-56.

Mulligan, G.M., McCarroll, J.C., Flanagan, K.D., and McPhee, C. (2019). *Findings From the Fifth-Grade Round of the Early Childhood Longitudinal Study, Kindergarten Class of* 2010–11 (ECLS-K:2011) (NCES 2019-130). U.S. Department of Education: National Center for Education Statistics. https://nces.ed.gov/pubsearch.

National Center for Education Statistics (NCES). (2019). Reading 2019: *The nation's report card: National assessment of educational progress at grades 4 and 8*. U.S. Department of Education.

National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform. *The Elementary School Journal*, 84(2), 113-130.

No Child Left Behind Act of 2001. (2002). Pub. L. No.107-110,115 SAT/ 1425. https://files.eric.ed.gov/fulltext/ED556108.pdf Navarrete, C., Wilde, J., Nelson, C., Martinez, R., & Hargett, G. (1990). *Informal Assessment in Educational Evaluation: Implications for Bilingual Education Programs*. National Clearinghouse for Bilingual Education.

Osborne, J. W., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical assessment, research, and evaluation*, 8(1), 2.

Palacios, N. (2017). Why all teachers matter: The relationship between long-term teacher and classroom quality and children's reading achievement. *Journal of Research in Childhood Education*, 31(2), 178-198.

Paris, S. G., & Hamilton, E. E. (2014). The development of children's reading comprehension. In *Handbook of research on reading comprehension* (pp. 56-77). Routledge.

Pennell, S. (2017). *EXPLORING THE BALANCE: A Path Analysis Examination of the Maintaining the Balance Model*.

Popham, W. J. (2001). Teaching to the Test? *Educational leadership*, 58(6), 16-21.

Pratt, N. (2018). Playing the levelling field: teachers' management of assessment in English primary schools. *Assessment in Education: Principles, Policy & Practice*, 25(5), 504-518.

Reio Jr, T. G., & Shuck, B. (2015). Exploratory factor analysis: implications for theory, research, and practice. *Advances in Developing Human Resources*, 17(1), 12-25.

Reutzel, D. R., & Cooter, R. B. (2019). *Teaching children to read: The teacher makes the difference*. Pearson.

Ross, E. & Walsh, K. (2019). State of the States 2019: *Teacher and Principal Evaluation Policy*. National Council on Teacher Quality.

Ruff, R. R. (2019). State-Level Autonomy in the Era of Accountability: A Comparative Analysis of Virginia and Nebraska Education Policy through No Child Left Behind. *Education policy analysis archives*, 27(6), n6.

Ryan, J. E. (2004). The perverse incentives of the no child left behind act. *NYUL Rev.*,79, 932.

Salkind, N. J. (Ed.). (2010). Encyclopedia of research design (Vol. 1). Sage.

Segrin, C. (2010). Multiple regression. In N. J. Salkind (Ed.), *Encyclopedia of research design* (pp. 845-849). SAGE Publications.

Shanahan, T. (2014). Educational policy and literacy instruction: Worlds apart? *The Reading Teacher*, 68(1), 7-12.

Sheskin, D. (2010). Correlation. In N. J. Salkind (Ed.), *Encyclopedia of research design* (pp. 265-267). SAGE Publications.

Silva, C. N. (2010). Ex post facto study. Encyclopedia of research design, 3, 465-466.

Spear-Swerling, L., & Cheesman, E. (2012). Teachers' knowledge base for implementing response-to-intervention models in reading. *Reading and Writing*, 25(7), 1691-1723.

Spillane, J. P. (2012). Data in practice: Conceptualizing the data-based decision-making phenomena. *American Journal of Education*, 118(2), 113-141.

Stiggins, R. J., (1991). Assessment Literacy. The Phi Delta Kappan, 72(7), 534-539.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-765.

Streiner, D. L. (2005). Finding our way: an introduction to path analysis. *The Canadian Journal of Psychiatry*, 50(2), 115-122.

Temple, C., Ogle, D., Crawford, A., Temple, C. (2018). *All Children Read: Teaching for Literacy in Today's Diverse Classrooms* (5th ed.). Pearson.

Timperley, H., Wilson, A., Barrar, H., Fung, I. (2007). *Teacher professional learning and development: Best evidence synthesis iteration*. Ministry of Education.

Tompkins, G. E., (2017). Literacy for the 21st century: A Balanced Approach. Pearson.

Toste, J. R., & Ciullo, S. (2017). Reading and writing instruction in the upper elementary grades. *Intervention in School and Clinic*, 52(5), 259-261.

Tourangeau, K., Nord, C., Lê, T., Wallner-Allen, K., Vaden-Kiernan, N., Blaker, L. and Najarian, M. (2019). *Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011) User's Manual for the ECLS-K:2011 Kindergarten–Fifth Grade Data File and Electronic Codebook, Public Version (NCES 2019-051).* U.S. Department of Education. National Center for Education Statistics.

U.S. Department of Education. (2009). Race to the Top Program: Executive Summary.

Vernon-Feagans, L., Bratsch-Hines, M., Varghese, C., Cutrer, E. A., & Garwood, J. D. (2018). Improving struggling readers' early literacy skills through a tier 2 professional development program for rural classroom teachers: The Targeted Reading Intervention. *The Elementary School Journal*, 118(4), 525-548.

Vinovskis, M. A. (2019). History of testing in the United States: PK–12 education. *The ANNALS of the American Academy of Political and Social Science*, 683(1), 22-37.

Walk, M. & Rupp, A. (2010). Pearson product-moment correlation coefficient. In N. J. Salkind (Ed.), *Encyclopedia of research design* (pp. 1023-1026). SAGE Publications, Inc.

Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, 44(3), 219-246.

Werts, M. G., Carpenter, E. S., & Fewell, C. (2014). Barriers and benefits to response to intervention: Perceptions of special education teachers. *Rural Special Education Quarterly*, 33(2), 3-11.

Williams, M. N., Grajales, C. A. G., & Kurkiewicz, D. (2013). Assumptions of multiple regression: Correcting two misconceptions. *Practical Assessment, Research, and Evaluation*, 18(1), 11.

Xu, Y., & Brown, G. T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149-162.

VITA

EILYN SANABRIA

Born, Havana, Cuba

2008	B.S., Early Childhood Education Florida International University Miami, Florida
2009-2013	Elementary Teacher Miami-Dade County Miami, Florida
2010	M.S., Reading Education K-12 Florida International University Miami, Florida
2013-2015	Teaching Assistant Florida International University Miami, Florida
2015-2018	Coordinator, Institutional Effectiveness Florida International University Miami, Florida
2015-Current	Adjunct Instructor Florida International University Miami, Florida
2018-2020	Senior Coordinator, Institutional Effectiveness Florida International University Miami, Florida
2020-Current	Assistant Director of Assessment and Academic Improvement, Institutional Effectiveness Florida International University Miami, Florida
2020	Doctoral Candidate Florida International University Miami, Florida

PUBLICATIONS AND PRESENTATIONS

- Davis, M., Fine, J. C., & Sanabria, E., (2014). *General Education Teachers' Experiences with Inclusion*. Presented at Council for Exceptional Children, Philadelphia, PA.
- Fine, J. C., & Sanabria, E., (2014). A History of Reading as Viewed through the Archival Lens of the American Reading Forum Yearbooks: 1981-Present. Presented at International Literacy Association, New Orleans, LA.
- Fine, J. C., & Sanabria, E., (2013). Experienced Primary Teachers' Perceptions from Implementing the Common Core Standards. Paper Presented at American Reading Forum, Sanibel, FL.
- Fine, J. C., Sanabria, E., & O'Gorman-Fazzolari, C. (2014). How Do I Teach ELs with the Common Core? *American Reading Forum Annual Yearbook, 34*.
- Fine, J. C., & Sanabria, E., (2013). Teacher Preparation: What Can We Learn from the Voices of 4th and 5th Grade Students on Attitude Scales about their Student Teachers? Presented at International Literacy Association, St. Louis, MO.
- Perez, K., Sanabria, E., Lebin, S., Doherty-Restrepo, J. (2019) *IE Certificate: Effects of Teaching Faculty Assessment Best Practices.* Paper presented at the Association for Institutional Research, Denver, CO.
- Perez, K., Sanabria, E., Lebin, S., Doherty-Restrepo, J. (2019) *The Impact of an Assessment Certificate on Faculty Perceptions and Knowledge*. Presented at the University of Florida's Assessment in Higher Education annual conference, Gainesville, FL.
- Perez, K., Sanabria, E., Lebin, S., Doherty-Restrepo, J. (2020). The Impact of an Assessment Certificate on Faculty Perceptions and Knowledge. *The Journal of Assessment in Higher Education*, Gainesville, FL.
- Sanabria, E. (2013). An In-Depth Look at RtI: Using Current Research to Examine Its Effectiveness and Areas for Growth. *National Association of Special Education Teachers, Special Educator E-Journal,* September.
- Sanabria, E. (2013). Exploring Autism Through a Whole-Body Approach. National Association of Special Education Teachers, Autism Spectrum Disorder Series, September 2013.