The Independent Reading Level Assessment and Its Impact on Third Grade Reading Achievement

Amy Amato Jackson

Follow this and additional works at: https://pilotscholars.up.edu/etd

Part of the Educational Assessment, Evaluation, and Research Commons, and the Elementary Education Commons

Recommended Citation
Amato Jackson, Amy, "The Independent Reading Level Assessment and Its Impact on Third Grade Reading Achievement" (2019). Graduate Theses and Dissertations. 55.
https://pilotscholars.up.edu/etd/55

This Doctoral Dissertation is brought to you for free and open access by Pilot Scholars. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Pilot Scholars. For more information, please contact library@up.edu.
The Independent Reading Level Assessment and Its Impact on Third Grade Reading Achievement

by

Amy Amato Jackson

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Education in Neuroeducation

University of Portland School of Education

2019
The Independent Reading Level Assessment and Its Impact on Third Grade Reading Achievement

by

Amy Amato Jackson

This dissertation is completed as a partial requirement for the Doctor of Education (EdD) degree at the University of Portland in Portland, Oregon.

Approved: 

Chair of Committee: REDACTED

Date: 3/1/2019

March 1, 2019

Committee Member: REDACTED

Date: March 1, 2019

Committee Member: REDACTED

Date: March 1, 2019

If applicable:

Additional Committee Member

Date

Additional Committee Member

Date

Approved:

REDACTED

Date: 3/1/19

Graduate Program Director: REDACTED

Date: 3/1/19

Dean of the Unit: REDACTED

Date: 3/1/19

Graduate School Representative

Date
Abstract

This study focuses on the developmental reading taxonomy known as the Independent Reading Level Assessment (IRLA). The purposes of the study were to analyze the inclusion and timing of the elements required for learning to read, and to discover whether the use of the IRLA impacted results on summative third grade assessments (Smarter Balanced Assessments). A taxonomic structure was used to examine the content validity study of the IRLA. Findings were that while the developmental progressions were accurate and the elements were inclusive, the areas of executive functioning, phonological awareness, and vocabulary could be strengthened, and that spelling could be more pronounced to strengthen decoding and encoding of language. ANOVA analysis of standardized test scores in 49 IRLA schools over four years showed no statistically significant change. The wide range of scores from year one to year four indicated potential issues with implementation of the IRLA. This study yielded two conclusions: (1) the elements and timing for teaching reading as presented in the IRLA largely match the research base; 2) no evidence through standardized test scores of the impact of the IRLA was found.
Acknowledgements

Thank you, Jane, and the American Reading family for the ideas, space, and encouragement to see this through.

Thank you, mom, dad, and family for supporting my work and helping me find time to play. Your cultivation of and tolerance for my tenacity are among the greatest gifts of my life.

Thank you, John. A million times.
# Table of Contents

Abstract ........................................................................................................................................... iii
Acknowledgements ......................................................................................................................... iv
Table of Contents .......................................................................................................................... v
List of Tables ................................................................................................................................... xi

## Chapter 1: Introduction

- Consensus and Confusion ........................................................................................................... 2
- Implementation Challenges ......................................................................................................... 6
- Implementation Supports ............................................................................................................. 7
- Structure as Support .................................................................................................................... 9
- The Promise of Taxonomy as Organizational Structure .......................................................... 10
- Developmental Reading Taxonomy ............................................................................................. 12
- Purpose Statement ....................................................................................................................... 14
- Significance of the Study ............................................................................................................ 15
- Summary .................................................................................................................................... 15

## Chapter 2: Literature Review

- Comprehension of Complex Text ............................................................................................... 17
  - Definition ................................................................................................................................ 19
  - Empirical research .................................................................................................................... 20

  *The reader and the skills the reader brings to the text......................................................... 21
  *The text and its demands. ............................................................................................................ 23
The activity, made up of purpose, processing, and outcome, done by the reader with the text ............................................. 28

Executive Function ............................................................................................................. 31
Definition ......................................................................................................................... 31
Empirical research ............................................................................................................ 33

Print Concepts .................................................................................................................... 35
Definition .......................................................................................................................... 35
Empirical research .......................................................................................................... 35
Interaction with other elements ....................................................................................... 38

Phonological Awareness .................................................................................................... 38
Definition .......................................................................................................................... 38
Empirical research .......................................................................................................... 39
Interaction with other elements ....................................................................................... 43

Word Recognition .............................................................................................................. 44
Definition .......................................................................................................................... 44
Empirical research .......................................................................................................... 44
Interaction with other elements ....................................................................................... 47

Phonics ................................................................................................................................... 49
Definition .......................................................................................................................... 49
Empirical research .......................................................................................................... 49
Interaction with other elements ....................................................................................... 54

Vocabulary .......................................................................................................................... 54
Definition .......................................................................................................................... 54
Empirical research .......................................................................................................... 55
Interaction with other elements ....................................................................................... 59

Fluency .................................................................................................................................. 61
Definition .......................................................................................................................... 61
Empirical research .......................................................................................................... 62
Interaction with other elements ............................................................ 65

Formative Assessment .................................................................................... 66
Definition .............................................................................................. 66
Empirical research ................................................................. 68

One-to-One Instruction .................................................................................. 69
Definition .............................................................................................. 69
Empirical research ................................................................. 70

Timing of Skills ............................................................................................... 73

Inclusion and Timing of Taxonomic Elements of the IRLA ....................... 81
Comprehension of complex text in the IRLA. ..................................... 82
Executive functioning in the IRLA. ..................................................... 86
Print concepts in the IRLA ................................................................. 88
Phonological awareness in the IRLA .................................................... 90
Word recognition in the IRLA ............................................................. 95
Phonics in the IRLA ................................................................. 97
Vocabulary in the IRLA ................................................................. 102
Fluency in the IRLA ................................................................. 105
Formative assessment in the IRLA ........................................................ 107
One-to-one instruction in the IRLA ..................................................... 109

Summary ....................................................................................................... 109

Chapter 3: Methodology .......................................................................................... 111

Research Questions ........................................................................................ 111

Independent Reading Level Assessment ................................................... 111
Exploration of 1B ............................................................................... 113

Rationale for Methodology ............................................................................ 119

Sampling Strategy ....................................................................................... 121

Participants and Setting .............................................................................. 122
Instrumentation ............................................................................................................. 123
Procedures and Timeline ............................................................................................. 125
Data Analysis .................................................................................................................. 125
Disclosure Statement .................................................................................................... 126
Summary .......................................................................................................................... 126

Chapter 4: Results ........................................................................................................... 128

Research Question and Hypothesis .............................................................................. 128

Findings: Inclusion and Timing of Elements .............................................................. 128
  Comprehension of complex text .................................................................................. 129
  Executive functioning .................................................................................................. 130
  Print concepts .............................................................................................................. 130
  Phonological awareness .............................................................................................. 131
  Word recognition ........................................................................................................ 131
  Phonics ........................................................................................................................ 132
  Vocabulary .................................................................................................................. 132
  Fluency ........................................................................................................................ 133
  Formative assessment ................................................................................................. 134
  One-to-one instruction ............................................................................................... 135

Impact of the IRLA on Results of Third Grade Summative Assessments
........................................................................................................................................ 136
  Participants .................................................................................................................. 137
  Change in performance of third grade students ......................................................... 140
  Analysis of variance in performance of third grade students .................................... 142

Summary .......................................................................................................................... 144

Chapter 5: Discussion .................................................................................................... 145

Findings: Inclusion and Timing of Elements in the IRLA ........................................... 147
  Independent reading .................................................................................................... 152
Findings: Results on Third Grade Summative Assessments .................... 153
  SBA proficient ................................................................................. 153
  Variability ......................................................................................... 154

Implications Embedded in the Statistical Analysis ............................. 157

Limitations of the Study .................................................................... 158
  Unit of analysis .................................................................................. 158
  Pre-test information is not available ................................................ 159
  Quantity of data ................................................................................ 159
  Duration of data ................................................................................ 160
  Implementation variations ................................................................ 160

Limitations of IRLA Implementations ............................................... 161
  Educator knowledge and readiness .................................................... 161
  Accuracy of IRLA data ..................................................................... 162
  Teacher training ................................................................................ 163

Recommendations for Implementing the IRLA .................................... 165
  Conditions for using the IRLA .......................................................... 166

Implementation Science as a Tool for Effective Implementation ........ 167
  Implementation drivers ..................................................................... 168
    Competency: Selection .................................................................... 169
    Competency: Training .................................................................... 169
    Competency: Coaching ................................................................... 169
    Competency: Performance assessment .......................................... 170
    Organization: Decision support data systems ................................ 170
    Organization: Facilitative administration ....................................... 171
    Organization: Systems interventions .............................................. 171
    Leadership ....................................................................................... 172

Recommendations for Further Study .................................................. 172
  Annual replication of the current study .............................................. 173
  Student-by-student growth within systems ...................................... 173
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <em>K-3rd Grade Reading Instructional Ingredients in U.S. Policy and Position Papers</em></td>
<td>4</td>
</tr>
<tr>
<td>2. <em>Hambrick’s Taxonomic Structure of Elements of Learning to Read</em></td>
<td>11</td>
</tr>
<tr>
<td>3. <em>Combined Taxonomic Structure Realized in the IRLA</em></td>
<td>18</td>
</tr>
<tr>
<td>5. <em>Combined Taxonomic Structure Realized in the IRLA</em></td>
<td>81</td>
</tr>
<tr>
<td>6. <em>Combined Taxonomic Structure Realized in the IRLA</em></td>
<td>113</td>
</tr>
<tr>
<td>7. <em>1B Foundational Skills</em></td>
<td>114</td>
</tr>
<tr>
<td>8. <em>Formative Assessment Protocol</em></td>
<td>116</td>
</tr>
<tr>
<td>9. <em>Cluster for Analysis Including Baseline Year</em></td>
<td>121</td>
</tr>
<tr>
<td>10. <em>Total Districts, Schools, and Number of 3rd Grade Students Included in the Study</em></td>
<td>137</td>
</tr>
<tr>
<td>11. <em>Demographic Breakdown of Students in the Study</em></td>
<td>139</td>
</tr>
<tr>
<td>13. <em>Group Means and Standard Deviations for 3rd Grade Achievement on SBA</em></td>
<td>142</td>
</tr>
<tr>
<td>14. <em>Demographic Descriptors of the Two Schools Gaining More Than 20% from Year 1 to Year 4</em></td>
<td>154</td>
</tr>
</tbody>
</table>
15. Demographic Descriptors of the Three Schools Losing More Than 20% from Year 1 to Year 4 ................................................................. 156

16. Total Students Using IRLA by Grade in Schools Included in the Study by Year, as Reported in SchoolPace ........................................... 160

17. To-Scale Progression of Points from Kindergarten Through Second Grade ................................................................. 162

18. Professional Development Received by the Five Districts Over the Four Years in the Study ................................................................. 164
Chapter 1: Introduction

The ability to read supersedes every other academic skill. Without it, access to the vast bank of resources available through print on every known subject is blocked. Reading critically, thoughtfully, and evidentially enables participation in a functioning society. Yet, according to the National Assessment of Educational Progress (NAEP, 2017) students in the US at grades 4, 8, and 11 have not shown gains in their ability to read and comprehend grade level text over the past 30 years. NAEP reports 68% of fourth graders performing at basic reading levels in 2017, while 67% performed at that level in 2007, an increase of just 7% from 1998. Thirty-seven percent of fourth grade students achieved a proficient ranking in 2017, which is up 8% from 29% in 1998, but neither score suggests the majority of students are reading at grade level by the fourth grade.

Internationally, the Programme for International Student Assessment (PISA) which surveys for achievement triennially, shows that over the past 15 years the US has remained flat in all areas (http://www.oecd.org/pisa/data/). Mullis, Martin, Foy, and Hooper (2017) discuss that at the fourth-grade level, PISA shows a trend of lower than average achievement for the US when compared to 41 countries around the world. Similarly, the Progress in International Reading Study (PIRLS) has collected data on a five-year cycle and also reveals a flat trend for the US, echoing NAEP and PISA. The lack of improvement in reading achievement serves as a call to action for
researchers to discover what knowledge or conditions are necessary for teaching reading effectively.

**Consensus and Confusion**

The lack of progress shown on national and international standardized assessments suggests that the thousands of research studies conducted on early reading either have not been implemented or they are having no effect. Learning how to read and how to teach beginning readers is complex, as evidenced by countless studies trying to understand what takes place for a learner, and what decisions, materials, and conditions are necessary for a teacher to help children learn to read. In the past 20 years, numerous meta-analyses and policy changes focused on improving reading have included the National Research Council (Snow, Burns, & Griffin, 1998), the National Reading Panel (NRP, 2000), the No Child Left Behind Act (NCLB, 2001), the establishment of Early Reading First as a component of NCLB, the Education Sciences Reform Act (2002), and the establishment of the Institute of Education Sciences (IES, https://ies.ed.gov) in 2002. Further, in an attempt to unify curricular standards across the country, the National Governor’s Association Common Core State Standards (CCSS) were completed in 2010 (NGA, 2010). Drawing on the preponderance of reading research, each of the policies and meta-analyses named above provided well-researched recommendations to the field of reading instruction. Guided by the NRP and enacted in legislation, phonemic awareness, phonics, vocabulary development, reading fluency, including oral reading skills, and reading comprehension (NCLB, 2001) make up the elements that have permeated the teaching culture, as evidenced by teachers who are keenly aware of *the big five*. Widespread
adoption of common language is evidence that federally requested meta-analyses and policy changes entering the landscape of reading instruction over the past 20 years have taken root. However, a study of the elements recommended through the collected works cited above reveals that consensus is elusive. The 2001 NCLB legislation naming *the big five* provided direction, but its omission of many other skills named in the research makes it incomplete. The research base, as seen in the abovementioned meta-analyses and as will be reviewed in Chapter 2, is much more comprehensive than *the big five*. Table 1 provides an overview of the elements of reading each of the cited works recommends as a necessary ingredient.
Table 1

*K-3rd Grade Reading Instructional Ingredients in U.S. Policy and Position Papers*

<table>
<thead>
<tr>
<th>The Big Five as recommended by NRP and included in NCLB</th>
<th>Named Instructional Elements</th>
<th>NRC</th>
<th>NRP</th>
<th>NCLB</th>
<th>NELP</th>
<th>CCSS</th>
<th>IES: Comprehension</th>
<th>IES: Foundational Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phonemic Awareness</td>
<td>Phonemic Awareness</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Alphabetic Principle</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phonics</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rapid Naming</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reading Fluency, including oral reading skills</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Expressive and Receptive Oral Language</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluency</td>
<td>Vocabulary</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comprehension</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Text Comprehension</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comprehension Strategies</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Comprehension</td>
<td>Comprehension: Key Ideas/Details</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comprehension: Craft &amp; Structure</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Comprehension: Integration of knowledge and ideas</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Print Knowledge</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mapping Speech Sounds to Word Parts</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Explicit in the “Big Five”</td>
<td>Visual Memory</td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Visual Perceptual Abilities</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sight Words</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Invented Spelling</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*a*National Research Council (Snow, Burns, & Griffin, 1998);  *b*National Reading Panel Report (2000);  *c*No Child Left Behind (2001);  *d*National Early Learning Panel (Shanahan & Lonigan, 2010);  *e*Common Core State Standards (NGA, 2010);  *f*IES: Comprehension (Shanahan et al., 2010);  *g*IES: Foundational Skills (Foorman et al, 2016)
Notable in Table 1 are six elements not included in the NCLB’s identified *big five*: print knowledge, mapping speech sounds to word parts, visual memory, visual perceptual abilities, sight words and invented spelling. Also, the singularity of *vocabulary* stands in sharp contrast to the variations of *comprehension*. While CCSS lists three types of comprehension (key ideas/details, craft and structure, and integration of knowledge and ideas), the IES report on comprehension only denotes two of the three, the NRP specifies only comprehension, and the IES Foundational skills report names both comprehension and text comprehension. For teachers not versed in the literature distinguishing the underlying processes of comprehension and text complexity, this generates confusion (Reynolds & Daniel, 2018). Teachers’ ability to recite the NCLB elements does not indicate their ability to distinguish nuance within them, recognize when an unnamed key research-proven element is needed in instruction, and does not prepare them to navigate the intricacies among them (Vesay & Gischar, 2013; Zeece, 2010).

The impact of these meta-analyses and policies on reading instruction is seen and felt in public schools across the country through innovative materials adoptions, intervention structures, and standardized summative assessments, all informed by quantities of carefully analyzed research on the key instructional ingredients named in the studies and policies shown above. With these innovations being implemented by elementary reading teachers, improvements would likely be seen through NAEP, PISA and PIRLS, but they are not.
Implementation Challenges

Teachers cannot be expected to know all of the research, nor can they reasonably stay abreast of the deep policy changes and rationale behind them. Teachers have been shown to make instructional decisions based on judgement (McLeskey & Waldron, 2002), beliefs (Dewitz & Jones, 2013), and “just good teaching” (Griffith, Bauml, & Barksdale, 2015). It is therefore imperative that the foundational knowledge used to ground teacher judgement and beliefs is true, common, and supported. The establishment of “the big five” has helped teachers focus on evidence-based elements of reading but may have provided an additional challenge by researching, naming, and expounding on them in isolation.

Overwhelming research will be unpacked in Chapter 2 showing that a recurring theme throughout the literature on each ingredient is that they must be taught in concert, not in isolation (Adams, 1990; Bear, Negrete, & Cathey, 2012; Chall, 1967; Dooley, 2010; Ehri & McCormick, 1998; Foorman et al., 2016; Honig, Diamond, & Gutlohn, 2000; Jarmulowicz, Hay, Taran, & Ethington, 2008; Shanahan et al., 2010; Snow, Burns, & Griffin, 1998). However, practicable information for classroom teachers about how to do this remains elusive. The implicit message contained in the exposition of “the big five” is that these elements are the only ones necessary, and that they are to be taught independently. Given the stagnant results on NAEP, PISA and PIRLS, it is fair to speculate that this compartmentalized, incomplete view of instructional components for teaching reading has permeated the instructional ethos and may be shielding student growth.
Implementation Supports

Implementation of any method or strategy requires support. Gunn, Simmons, and Kame’enui (1995) discussed developmental patterns that teachers can use to help young children link new literacy information to existing knowledge and showed that teachers who use sequential teaching events produce more successful readers. If the ingredients required for teaching reading produce confusion and conflict, the need for teaching them sequentially compounds this confusion. In addition to understanding what to teach, teachers must understand research-based, developmentally appropriate sequencing for the ingredients named within the research. Materials that support teachers’ knowledge are widely available and attempt to lay out the components for learning to read in sequential strands. Two such resources are the CORE Sourcebook (Honig et al., 2000) and the Continuum of Literacy Learning (Pinnell & Fountas, 2007). The Consortium of Reading Excellence (CORE) provides a comprehensive collection, or sourcebook, of research-based reading strategies including instructional sequences. Each instructional ingredient is presented with a relative grade level for introduction and completion. The CORE Sourcebook resource presents a timeline with building blocks expecting mastery, but considerable overlap exists with no guidance given about which skills are mutually supportive, leaving a teacher to wonder what to prioritize, where to begin and what skill may benefit from being taught in conjunction with another. In the Pinnell and Fountas resource, continua of reading elements are presented in a comprehensive collection. Intending to “provide teachers with a conceptual tool that they can use to think constructively about their work” (Pinnell & Fountas, 2011, p. 10), the continuum provides sequences for
elements required for learning to read such as phonics, phonological awareness, letter knowledge, high-frequency words and vocabulary, but the continuum does not illuminate ways to link new and existing knowledge. Like the CORE Sourcebook, this list of skills organized as singular strands does not explicitly support teachers as they build confident and research-informed teaching practices to lean on when making in-the-moment instructional decisions. Zeece (2010) discussed that student motivation is provided by anchoring critical task features to knowledge already in place. An organizational structure that champions what a student already knows and can do, enables teachers to manage the landscape of instructional ingredients as they interact with one another over time for a variety of learners, all while providing the research-based instructional ingredients in a predictable sequence is needed.

In a broader attempt to support teachers as they implement the known research for teaching early reading, educational think tanks across the country—for example EdReports (edreports.org), The Council of Great City Schools (cgcs.org), Achieve the Core (achievethecore.org), Regional Educational Laboratory Southeast (Foorman, Smith, & Kosanovich, 2017)—have developed tools to evaluate published materials for alignment with standards, research, and the variety of needs each unique community presents. In spite of these efforts, outcomes have still not shifted. Students in the fourth-grade today perform nearly at the same levels as students in the fourth grade 20 years ago (NAEP, 2017). The lack of progress indicates that guidance provided through basal, or core reading programs should be further supported with a research-based organizational structure that comprehensively brings the known instructional ingredients in line with known sequential and combinational timings.
Structure as Support

Continua and lists attempt to support teachers, but the complex landscape of reading research needs an organizational structure with greater dimensionality. Mosenthal (1987) argued for the creation of a universal taxonomy of reading goals as a tool to organize and make actionable research findings in a simple, efficient, and instructionally informative system that supports teachers of early readers. Mosenthal asserted that the creation of such a tool would support the field of reading instruction by providing common goals and purpose. Travers (1980) undertook the idea of a taxonomy in the field of education and described the evolution of taxonomy as a classification system, offering critiques and suggestions. Travers reviewed and cited the centuries-long development of classification systems in biology for both plants (Linneaus) and animals (Aristotle and Lamarck), and chemistry (the periodic table, conceptualized by Lavoisier and advanced by Dalton). Travers noted that classification systems have also been developed in astronomy, geology, crystallography, and physics. Citing an 1874 work by Jevons, Travers agrees that classification systems are constructs of the intellect, and not categories existing in nature, as the first classification structures were conceptualized. In the logical building of a classification system, Travers considered what structures must be included to be useful. He cited Körner (1970) as laying out three things a categorical framework must do: make categories explicit, state the attributes and relationships that maximize the separations of classes, and state the logic underlying the thinking.

Hambrick (1984) proposed that because strategies are typically conceived as interdependent wholes, a taxonomic approach is an almost ideal way of studying them.
Hambrick wrote about using a taxonomic approach to classify elements in a complex system for closer study and declared classification a fundamental cognitive aid. Asserting that a classification scheme helps bring order to “cluttered conceptual landscape[s]” (p. 27), Hambrick’s classification structure first considers the environment, or uncontrollable attributes, and moves to strategic content, considering firm or fixed variables. Next, controllable strategic choices are placed, and finally the process by which all elements are directed is given.

**The Promise of Taxonomy as Organizational Structure**

The use of a classification structure such as a taxonomy to display the elements of reading holds promise (Mosenthal, 1987). Table 2 imposes Hambrick’s (1984) taxonomic approach on the complex system of learning to read. Hambrick’s *environment* is equated with the comprehension of complex text, as the ultimate goal of learning to read is the ability to read and comprehend anything in print, no matter the complexity. *Strategic content* as presented by Hambrick can be equated to NCLB elements of phonological awareness, phonics, fluency, vocabulary, and comprehension strategies. Strategic, controllable choices are equated to assessments and the processes that direct decision-making about selecting strategic areas to access and at which times they should be used in the instructional delivery model.

A classification system, as described by Travers (1980) must address Körner’s (1970) taxonomic elements. As envisioned by Hambrick (1984) and elaborated in Table 2, a taxonomy provides structure for making categories explicit by providing an organizational frame. Hambrick’s frame as shown in Table 2 organizes explicit categories, or elements, necessary for teaching and learning how to read.
### Table 2

*Hambrick’s Taxonomic Structure of Elements of Learning to Read*

<table>
<thead>
<tr>
<th>Hambrick’s structure</th>
<th>Elements of learning to read</th>
</tr>
</thead>
<tbody>
<tr>
<td>Environment/Uncontrollable Attribute</td>
<td>Comprehension of Complex Text</td>
</tr>
<tr>
<td></td>
<td>Phonological Awareness</td>
</tr>
<tr>
<td>Strategic Content or Fixed Variables</td>
<td>Phonics</td>
</tr>
<tr>
<td></td>
<td>Fluency</td>
</tr>
<tr>
<td></td>
<td>Vocabulary</td>
</tr>
<tr>
<td>Strategic choices</td>
<td>Assessment</td>
</tr>
<tr>
<td>Process</td>
<td>Instructional delivery</td>
</tr>
</tbody>
</table>

However, a developmental dimension considering increasing complexity through the categories over time must be added to ensure an appropriate sequencing of skills over time. As suggested by Travers (1980), relationships that separate classes, or categories divided by developmental stages, can be crossed with the categories to provide a stable and workable form for the ingredients of teaching reading and the optimal time in which to teach them. In its basest form, these categories can be represented by grade levels. A developmental taxonomy describing the teaching of reading along two axes, content and time, would bring direct benefits to school systems by identifying a path for teachers to impact the growth of their students as proficient and thoughtful readers by organizing instructional decision-making. In direct alignment with research-based best practices and a standards-informed
educational system, this type of organizational structure could positively and profoundly impact outcome measures of student reading (Mosenthal, 1987).

**Developmental Reading Taxonomy**

The tool being investigated in this study is a developmental reading taxonomy called the Independent Reading Level Assessment, or IRLA (Hileman & Zorzi Cline, 2017). Hambrick’s (1984) structure for categories combined with Travers’ (1980) structure for classifications provides a gridded frame for the taxonomic whole. Divisions, or separations of classes as laid out by Travers, are realized in the developmental sequencing provided in vertical groupings represented as color levels defining each element of reading to a developmental specificity. Relationships that maximize the separation of the classes, or categories, are the entry requirements given at the start of each color level. The logic underlying this thinking is provided for each color level as its research-based learning focus. The IRLA spans grades Pre-K through 12, though as this study seeks to measure the impact of beginning reading influenced by the IRLA as seen on third grade summative assessments, the focus will be on the foundational levels, grades Pre-K-2, or levels Read to Me (RTM) through 2 Red (2R).

The IRLA was developed by a reading teacher whose practice did not comply with imposed rigid curricula demanding decontextualized skill-based practice. Coining the phrase “100-Book Challenge,” students in the second grade were challenged by their teacher to read 100 books in that school year. Those second graders exceeded all known expectations (Bennett, 2004). Leveraging knowledge of the reading research base, conferences with students, and their independent reading of
many books, this teacher showed that when given authentic and engaging books and individualized instruction that links what a student knows with what a student needs next across a spectrum of skills, readers grow. The 100-Book Challenge gained notoriety and inspired a handful of studies on the impact of independent reading and knowledgeable coaching on student growth (DuCette, 1999, 2001; Florida Center for Reading Research, 2006; Offenberg, 2005; Public Citizens for Children and Youth, 2008). Student success prompted formalized development of a carefully organized leveling system that is “keyed to national standards and high-stakes tests, making it easier for educators to determine where students are in terms of their ability and what they need to work on next” (Hileman, in Baca, 2004, p.L1). As it exists today, the IRLA is a comprehensive standards-based resource supporting teachers as they navigate teaching reading in grades Pre-K -12.

Griswold and Bunch (2014) studied the validity and reliability of the IRLA as a formative assessment framework and found that it aligns to other assessments, measures what it sets out to measure, and is well aligned with the CCSS. Ralston, Waggoner, Tarasawa, and Jackson (2016) studied the concurrent validity of the IRLA with a summative state assessment. This first attempt to disaggregate validity evidence across grade, program, and race/ethnicity showed correlations above .75 for nearly all groups. However, there has been no known study done to date on the impact of the IRLA on summative state test scores, nor has there been a direct and comprehensive analysis of the elements in the IRLA to the research base. These gaps in the research are what this study will address.
Purpose Statement

The developmental reading taxonomy IRLA is in use in a variety of school systems across the US. The purpose of this research is two-fold: to determine whether the inclusion and timing of each element presented in the foundational levels of the IRLA is appropriate and justified, and to discover whether learning informed by the IRLA yields increased results on third grade summative assessments. Elements of the taxonomy include NCLB elements in addition to skills identified in the research that are not legislated through NCLB: print concepts, word recognition, range of reading and text complexity. This study, through the literature review provided in Chapter 2, explores the research base for the ingredients for reading instruction as provided in the IRLA and compares each element in the taxonomy to the research base on its developmental sequence and its connections to mutually supporting elements. The literature review is in service of answering these two questions: are the elements included in the IRLA appropriately inclusive of the research on teaching reading, and are these elements placed in a developmental sequence supported by research? Chapter 4 will unpack the IRLA in relation to each element in search of alignment—inclusion and timing—from the research. Additionally, a quantitative study, the results of which are compiled in Chapter 4, will examine the reading achievement scores for districts using the developmental reading taxonomy over time to help discover whether learning informed in this way increases third grade reading achievement.
Significance of the Study

By providing a clear and focused organizational structure to teachers of beginning reading, this study could make a significant contribution to education. A tool that contains and categorizes relevant and known instructional components for learning to read, while also providing a layered developmental sequence ensuring simultaneous mastery of all components within a category prior to moving to the next can support teachers as they work with a variety of skills, confident that student learning will be enhanced. The quantitative element of this study will show to what degree the use of the taxonomy has impacted reading achievement for third grade students. Increased understanding about how and why to employ a dynamic and synergistic—yet carefully organized—tool will enable teachers of beginning reading to confidently engage in robust, authentic literacy activities.

Summary

The quantity of information about teaching early readers is large, leading to a complex landscape for teachers to navigate. A clean, comprehensive organizational structure is required to harness the complexities for teachers, and one has been built and is in use in a variety of schools and districts around the country, called the IRLA. Chapter 2 of this dissertation will use a taxonomic structure as theoretical framework containing the elements and their developmental sequencing required for learning to read, and it will investigate empirical studies from the research base supporting the inclusion and timing for each of the key elements. It will further present the inclusion of each element as it appears in the IRLA. Chapter 3 will describe the IRLA and outline methodology for the analysis of elements included in the IRLA and for the
quantitative study to analyze the impact of third grade reading achievement scores in
districts using the IRLA at grades K-2. The analysis of the IRLA and results of the
quantitative study will be revealed in Chapter 4, and Chapter 5 will comprehensively
review findings from the review of the literature as well as from the quantitative study,
providing recommendations for next steps in the pursuit of using an organizational
structure to make the teaching of reading to beginners powerful enough to ensure a
successful trajectory towards becoming fully literate.
Chapter 2: Literature Review

The literature review for this dissertation will explore elements of reading as named in the developmental reading taxonomy called the IRLA, or the Independent Reading Level Assessment (Hileman & Zorzi Cline, 2017) as a purpose of this research is to determine whether the inclusion and timing of each element presented in the foundational levels of the IRLA is appropriate and justified. In addition to NCLB reading elements: phonemic awareness, phonics, fluency, vocabulary and comprehension, the IRLA includes print concepts, word recognition, range of reading and text complexity. It also addresses executive function, formative assessment, and one-to-one instruction. As discussed in Chapter 1, the quantity of elements and research describing and validating these elements combined with their individualized developmental sequences provides teachers with a cluttered conceptual landscape, and support is required so teachers can effectively teach students to read.

Chapter 1 presented Hambrick’s (1984) classification theory as a way to organize the elements necessary for learning to read, aligning with the suggestion of taxonomy as classification scheme to help bring order to cluttered conceptual landscapes. Also presented in Chapter 1 was Travers’ (1980) separation of classes, providing structure for the timing of instructional ingredients. When these two conceptions of taxonomies are combined, a gridded structure that lists the elements of learning to read as strategic content, further organized by consideration of environment, strategic choices, and process function as horizontal rows. The timing, or developmental sequencing for each, provides a vertical structure, with classifications separated by developmental time frame such as pre-kindergarten,
kindergarten, 1st grade and 2nd grade, containing a developmental slice of each element required for reading. The organizational structure of the IRLA follows this combined taxonomic structure but extends it by including elements not included in NCLB reading elements, namely executive function, print concepts, word recognition, formative assessment, and one-to-one instruction. The IRLA also provides classifications in smaller increments than whole grade level for beginning readers. This combined structure is represented in Table 3.

Table 3

*Combined Taxonomic Structure Realized in the IRLA*

<table>
<thead>
<tr>
<th>Environmental/Uncontrollable Attribute</th>
<th>Pre-K</th>
<th>Kindergarten</th>
<th>1st Grade</th>
<th>2nd Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>RTM</td>
<td>1Y</td>
<td>2Y</td>
<td>3Y</td>
<td>1G</td>
</tr>
<tr>
<td>Uncontrollable Attribute</td>
<td></td>
<td>2G</td>
<td>1B</td>
<td>2B</td>
</tr>
<tr>
<td>Attribute</td>
<td></td>
<td>1R</td>
<td>1R</td>
<td>2R</td>
</tr>
<tr>
<td>Comprehension of Complex Text</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Executive Function</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Print Concepts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonological Awareness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word Recognition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehension Strategies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formative Assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One-to-One Instruction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Given the above structure, within each level, or class according to Travers (1980), is a collection of developmentally aligned instructional ingredients that help teachers identify which skills students possess to mastery, which they need to practice, and which they need to learn. Students cannot advance to the next level without attaining mastery of all skills within their identified independent level, preventing any individual skill from under- or over-development. By working one-to-one with students informed by a continuous cycle of formative assessment, teachers become adept at knowing whether a student is ready for the skills of the next level.

Chapter 2 will explore each element of reading identified in the IRLA and represented in Table 3, providing a definition of the element, a brief discussion of empirical evidence for its inclusion, and consideration of its interaction or lack thereof with the other elements. Timing of skills requisite for learning to read as presented in a variety of empirical and theoretical studies will then be explored. Lastly, Chapter 2 will present the inclusion of the elements and their timing as included in the IRLA.

### Comprehension of Complex Text

**Definition.** The purpose of reading is to make sense of anything seen in print. Print pervades our culture as a primary method of conveying information. From dissertations, comics, and advertisements to essays, editorials, and stories, print assumes a variety of forms, requiring flexibility and adaptability by the reader. Learning to read means acquiring the ability to perceive print in any form and make meaning. An oddity in the research on comprehension is that little to no time is spent defining it, and assumptions of what it means to comprehend are often left to the

Defining reading comprehension as “the process of simultaneously extracting and constructing meaning through interaction and involvement with written language,” (Snow, 2002, p. xiii), Snow claimed that while the current knowledge base on comprehension is sizeable, it is also sketchy, unfocused, and inadequate. Speculating that the problem was not the researchers’ or the studies’, but instead a lack of unified definition, a model of reading grounded in context was presented. Within this model, three elements central to reading were given as 1) the reader and the skills the reader brings to the text, 2) the text and its demands, and 3) the activity, made up of purpose, processing, and outcome done by the reader with the text.

**Empirical research.** Common among the studies on comprehension are relationships between comprehension and decoding skills, comprehension and language skills, and comprehension and prior knowledge. It is clear that without the acquisition of decoding skills, students are unlikely to become comprehenders (Abbott, Wills, Miller, & Kaufman, 2012; Cain & Oakhill, 2011; Elwér et al., 2015; Kendeou, van den Broek, White, & Lynch, 2009; Kim, Petscher, Schatschneider, & Foorman, 2010; Perfetti, Landi, & Oakhill, 2007). The development of language plays a crucial role in comprehension, whether listening or reading (Elwér et al., 2015; Kendeou et al., 2009; Pearson, 2004; Perfetti et al., 2007; Snow, 2002). Last is the notion that knowledge itself is necessary for a reader to begin to comprehend, and by comprehending, a reader gains knowledge (Cain & Oakhill, 2011; Clay, 1991; Connor et al. 2016; Fielding & Pearson, 1994; Pearson & Lieben, 2018; Perfetti et al., 2007).
The following empirical exploration of comprehension is grounded in the model of comprehension made up of the three elements central to reading presented by Snow (2002) and later corroborated by Toyama, Hiebert, and Pearson (2017): 1) the reader and the skills the reader brings to the text, 2) the text and its demands, and 3) the activity, made up of purpose, processing, and outcome done by the reader with the text.

**The reader and the skills the reader brings to the text.** There is no question that without the skills necessary to decode text, there cannot be comprehension. Written text must be translated into meaningful language to be understood. Connor and Craig (2006) presented that the more language a student has, the more likely he or she is to become literate. Dickinson and Porsche (2011) found that the more a student is read aloud to in pre-kindergarten, the better he or she will perform on reading outcome measures in grade four. As will be discussed throughout this paper, there is unanimity in the research base about the skills a reader must bring to a text.

Shanahan et al. (2010) stated in the introduction of a meta-analysis on comprehension for early readers that word-level skills (phonemic awareness, word analysis strategies, and sight word vocabulary), vocabulary knowledge, and oral language skills are known to be necessary for readers to be able to access meaning in text. Kendeou et al. (2009) concluded that both oral language skills and decoding skills in the early grades are predictive of comprehension as children develop. Using a battery of measures for oral language (listening comprehension, television comprehension, and vocabulary), decoding skills (letter identification, word identification, and phonological awareness), and a reading comprehension measure
they devised, Kendeou et al. focused on how oral language skills and decoding skills develop in relation to one another from preschool through early elementary school. Kendeou et al. argue based on their results that early assessment of comprehension skills is useful because the risk of a child developing reading comprehension difficulties is smallest when he or she progresses appropriately along the trajectories of decoding skills (phonological awareness and letter and word identification) and oral language skills (vocabulary and discourse comprehension). In other words, if comprehension difficulties are observed but instructional attention is focused only on decoding, then there is a risk that comprehension skills will lag in development, asserting later challenges.

In two studies, the skills required for comprehension are acknowledged through their assumption of or their control for them. Kim et al. (2010), used data from a statewide data collection system in Florida, established and maintained under Florida’s Reading First initiative to conduct a longitudinal study of 12,536 students from 1st through 3rd grade to look at the relationship between a suite of skills (phonemic awareness, letter names, nonsense word reading, oral reading fluency and vocabulary) with comprehension. This study did not examine the inherent qualities of comprehension and how to teach, only whether a student gains the foundational skills to the degree necessary to comprehend. Cain and Oakhill (2011) removed students with decoding deficits from their sample to control for decoding challenges in their study.

Elwér et al. (2015) were concerned with a subset of readers who seemed to be developing age appropriate decoding skills but were not developing comprehension
skills at the same rate. Between the two groups, comprehenders and poor comprehenders, no differences were noted in the areas of phonological awareness or rapid automatized naming. The skills impacting comprehension had to do with language, including weaknesses in vocabulary, grammar, and verbal memory. Elwér et al.’s research brings focus to skills beyond decoding that are necessary for early readers learning to comprehend, an assertion reinforced by Kendeou et al. (2009), who presented that comprehension is not a unitary phenomenon, but a family of skills that develop simultaneously. The claim made by Elwér et al. and Kendeou et al. is that successful reading comprehension is the result of a confluence of elemental skills, each of which has its own developmental trajectory: a reader must bring word level, meaning level (vocabulary) and language skills to a text if there is to be hope of comprehension.

The text and its demands. Providing beginning readers leveled texts matched to their abilities to support the development of skills and growing comprehension is a commonly used practice. Levelling texts is not a new concept, but a wide variety of levelling systems contributes to a chaotic comprehension landscape. Publishers like Scholastic, Pearson, Heinemann (Fountas and Pinnell), American Reading Company, and Reading A-Z offer leveled books and, often, assessment tools to accompany them. The leveling systems are different, providing users whole number gradations, decimaled gradations, alphabetic, or color gradations with each system prioritizing different aspects of the text. A teacher will often find the same book to have a variety of levels ascribed to it more often than not, though this could be a proprietary necessity. For example, Charlotte’s Web (White, 1952) is leveled a 40 DRA, Q by
Fountas and Pinnell, R by Scholastic, 490 Lexile, and Orange by ARC. Taken collectively, this book is appropriate for a reader somewhere in the third to fifth grade range, though each system has rationale unique to itself justifying its placement. Levels are a place to begin when considering the text and its demands, but a true investigation of the complexity of a text is needed when considering the text and its demands.

A text is the what involved in comprehension. The CCSS (NGA, 2010) provided Appendix A as an overview of what is meant by text complexity and established reasons for why it is important to reading. The CCSS proclaim that all students must be able to comprehend texts of steadily increasing complexity as they progress through school. This proclamation is in response to a decline of text complexity in schools (Adams, 2010; NGA, Appendix A, 2010; Hayes, Wolfer, & Wolfe, 1996) and the rise of text complexity in college and real-world situations (NGA, Appendix A, 2010). Appendix A calls for levelling texts using quantitative, qualitative, and reader and task considerations.

Quantitative measures are determined by the frequency or length of words and the length of sentences. Commonly in use and discussed in Appendix A is the Lexile measurement system, a purely quantitative and computerized tool that considers length of words and sentences. Although quantitative tools provide some general information about the difficulty of a text and measure aspects that are nearly impossible for a human reader to evaluate, they are imperfect. A text given to the Lexile generator will come out with the same score whether put in frontwards or backwards, and it is incapable of determining the complexity of theme in stories.
Being able to qualitatively determine the complexity of a text by following the recommendations of Appendix A is an important step towards increasing the rigor and depth required by college texts. However, there are challenges, particularly for newer readers. Mesmer, Cunningham, and Hiebert (2012) pose that text complexity for early readers needs to be defined differently than it does for established readers and needs to be more deeply researched. Proposing four main areas for qualitatively determining complexity, Mesmer et al. bring the conversation to readers lower than grade two and provide concepts such as imageability, the idea that if a word is concrete enough to have a picture to go with it, then students will understand it faster, making the text a lower level of complexity. They also discuss word maturity which is a consideration of how understandings of words grow, develop and deepen as they are encountered in text. A comprehensive theoretical overview is presented in their article providing teachers with guidance to match reader skill to appropriate books. Matching reader skill to the appropriate complexity of text is an important job, and one that can help a reader advance. But making this match requires that teachers know their students’ skills and the challenges and opportunities present in texts.

Understanding text complexity and using texts of matched complexity levels with students is difficult. Hiebert (2013) presented an article introducing the Text-Complexity Multi Index (TCMI) as a remedy for the lack of guidance provided for qualitative and reader-task evaluation of texts. The process consists of a series of rubrics teachers can use to guide this work. Fisher and Frey (2015) similarly provide a model for text complexity in informational text and propose four areas to consider: levels of meaning and purpose, structure, language conventionality and clarity, and
knowledge demands. Each area is measured on a three-point scale and has subcomponents, i.e., levels of meaning and purpose includes density and complexity, figurative language, and purpose.

Though careful book-by-book analysis of many components of text complexity may be ideal for matching reader to level of complexity, the solutions proposed by Hiebert (2013) and Fisher and Frey (2015) are extensive, not user friendly, nor grade level specific. Morris (2014) mused about a proposal made by Cunningham, Koppenhaver, Erickson, and Spadorcia (2004) about leveled books. Cunningham et al. proposed that the ideal leveled books for beginning readers would have repetition of high frequency words, adequate repetition of decodable patterns, text that leads the beginner to anticipate upcoming words, and interesting story lines. Morris believed that such books could be written for beginning readers and pointed to examples in his article of publishers who are generating such books. The idea of leveled books is not new, but whether the leveling system is adequately aligned with grade levels, with the right skills at the right times, and with the text complexity demands of the common core is currently under study.

Nelson, Perfetti, Liben, and Liben (2012) acknowledged that quantitative measures of text complexity validated by research are needed to help stakeholders identify what makes texts complex. Nelson et al. worked with six difficulty measures proven to be valid, reliable, and calibrated with the CCSS demand for appropriate text complexity. They then used five sets of texts as reference measures taken from established leveled sources (exemplars from Appendix B of CCSS, standardized state test passages, Stanford Achievement Test, Gates-MacGinitie, and MetaMetrics Oasis),
and compared their grade level equivalents with the results from the difficulty measures. Because the reference texts included measures of grade level and measures of student performance, the findings have implications for teachers in the field working to quickly and effectively match students to appropriately complex text. Using Spearman’s Rho to rank order correlations, they found that measurements were more closely related at the lower levels than the higher ones, and also that metrics including a broader range of linguistic and text measures produced higher correlations than those that only used word difficulty and sentence length measures. Informational text correlated more closely than narrative text, and when predicting student performance, the measures of text complexity were strong. Nelson et al. called for further study to understand how the features that make texts difficult for readers change with grade levels but affirmed that objectively measuring text complexity is a good start for helping to determine the grade level demands a text contains.

Claiming that the definition of grade levelness differs across assessments, Toyama et al. (2017) investigated the complexity of leveled passages used in four classroom reading assessments using four analytical tools of text complexity. Considering informal reading inventories and curriculum-based measures, Toyama et al. studied the trajectories of text complexity across widely used classroom assessments, how the assessments compared in terms of grade level equivalency, whether newer analytic tools reveal additional information, and how well the text complexity progression within assessments is aligned with the text complexity expectations of the CCSS. Findings include that while measures are getting more sophisticated and closer to accurate, there are still differences in the grade levelness of
passages across assessments as measured by widely used text analysis systems.
Toyama et al.’s critique of the classroom assessment tools studied is that they fail to
provide a reasonable staircase of complexity as required by the CCSS. Toyama et al.
recommend that the human element considering context and background knowledge
remain a component of leveling practices.

As a second component of the model of comprehension put forward by Snow
(2002), measuring a text and the demand it puts on a reader is a complex and multi-
faceted collection of decisions. Although leveling systems have existed for many
years, each one is built on different decisions about what makes a text complex. A
clearly defined stair-stepped progression of text complexity seems not to be actualized
as theorized and presented in the CCSS. Given that every reader brings his or her own
collection of skills and experiences, it may be enough that a teacher is aware of the
skills a text demands to make a match that will enable a reader to have success and
gain additional skills at the same time.

The activity, made up of purpose, processing, and outcome, done by the
reader with the text. The activity of a reader, by this definition, equals purpose,
processing, and outcome (Snow, 2002). In Snow’s elaboration of this assertion, the
purpose for reading can be internally generated (wanting to build a Lego set) or
externally imposed (class assignment) and is influenced by the motivation of the
reader. Interest and prior knowledge set the degree to which the activity of reading is
more or less purposeful. DeNaeghel, Van Keer, Vansteenkiste, and Rosseel (2012)
assert that “if children consider reading as personally relevant or identify themselves
with the value of reading, their tendency to engage in reading activities has been
internalized” (p. 1007). If a reader is disinterested or unmotivated, comprehension will be compromised. The purpose for reading changes as the reader reads and grows. Whether for compliance, enjoyment, information extraction, or growth of skill, the reader ultimately sets the purpose.

Internally generated purpose, often called intrinsic motivation, is a force of nature. For a teacher to find what motivates a child and capitalize on it is a constant goal for teachers. Guthrie, Wigfield, and VonSecker (2000) reported that intrinsic motivation is a stronger predictor of reading achievement than extrinsic motivation for elementary and middle school students, but that intrinsic motivation often declines during elementary school. Findings from their study of whether a classroom intervention can influence students’ intrinsic reading motivation, Guthrie et al. found higher degrees of curiosity and higher usage of comprehension strategies where students perceived the classroom as supportive of autonomy, meaning students had choice and were supported in the growth of competence, primarily through setting learning goals.

DeNaeghel et al. (2012) further studied the phenomenon of motivation by investigating the association between autonomous and controlled reading motivation and reading behavior and performance. They found that children’s concept of themselves as readers positively related to their frequency of reading, their engagement, and their comprehension, and claim that in the academic setting reading motivation was associated with higher quantity of reading, higher engagement, and higher comprehension. In a different study, Saeed and Zyngier (2012) concluded that internally motivated students were more authentically engaged than students
extrinsically motivated and that an environment where student choice supported by the teacher was present, intrinsic motivation strengthened students’ ability to be authentically engaged in learning. The above studies included students from grades four and five, but there is some research extending motivation to younger students as well.

Baroody and Diamond (2012) studied young children’s self-reported interest. Through a study of four-year old children, they found that children’s self-reported interest in literacy-related activities positively related to their code-related emergent literacy skills. In other words, the pre-reading children who reported interest in literacy activities had stronger emergent literacy skills than pre-reading children who reported being disinterested. Interest and success are connected and can be powerful drivers.

To further explore Snow’s (2002) definition of activity, processing involves the application of decoding skills, accessing higher level linguistic and semantic processing systems, and monitoring for comprehension. The process of reading is dependent on the reader’s purpose in conjunction with the reader’s background, thereby producing outcomes for the reader. Rounding out the definition, Snow refers to the outcomes of reading as the consequences of the activity of comprehending text. Whether increasing knowledge, solving a real-world problem or being engaged with a text, when the reader is through with the task of reading, there is a consequence, or outcome, that may or may not be related to the original purpose.

Content can drive purpose for reading, and even accelerate reading ability. Kuhn et al. (2006) found that where second graders spent time in a wide-reading
approach to instruction, reading a wide variety of texts, they grew as readers more rapidly. Choice, interest, and teacher scaffolding played a role in student growth. Gambrell (2015) pointed out that moderate challenge supports motivation, particularly where there is high topic interest. Referring to Fulmer and Frijters’ (2011) research finding that students’ interest in a story topic buffers the negative influence of difficulty level, Gambrell advocates self-selection of high interest text within a reasonable range of difficulty to encourage a love of reading and development of a reading habit. As an element of the comprehension model being explored, the activity made up of purpose done by a reader is informed and supported by intrinsic motivation, which is activated by a sense of autonomy and access to high interest text.

**Executive Function**

As children gain a sense of autonomy and interact with the environment, their capacity to respond, react, and adapt grows. The ability to make decisions about a variety of inputs is dependent upon three domain-general processes: working memory, interference (or inhibitory) control, and cognitive flexibility (Willoughby, Magnus, Vernon-Feagans, Blair, & Family Life Project Investigators, 2017). Known together as executive functioning, beginning readers rely on these processes working together to begin making meaning of the abstraction of text. Jones, Bailey, Barnes, and Partee (2016) emphasized that while the term executive functioning is used across the lifespan, it means different things based on the structure and complexity of the skill at different times in development.

**Definition.** Early childhood executive functioning enables children to learn how to learn and is defined as a synergistic interaction of working memory, inhibitory
control, and cognitive flexibility (or shifting) (Baddeley, 2003; Blair & Razza, 2007; Center for the Developing Child, 2018; Foy & Mann, 2013; Little, 2017; Rothlisberger, Neuenschwander, Cimeli, Michel, & Roebers, 2012).

Working memory is the ability to hold an idea in consciousness. The idea remains active or alive long enough to be either discarded or distributed to neural networks deeper in the brain as it seeks something to connect to: a prior experience or a similar idea. In work focused on working memory, Baddeley (2003) proposed that it is composed of subcomponents dependent upon an episodic buffer that determines whether an idea or concept is sent to long term storage or not. Baddeley’s theory is dependent upon the brain’s ability to focus on a memory long enough for crystallization to occur. DeBeni, Palladino, Pazzaglia, and Cornoldi (1998) conducted an experiment that focused on working memory. In a study of 17- and 18-year old students separated into good comprehender and poor comprehender groups, the poor comprehenders had a lower listening span and less efficient suppression ability regardless of the complexity of the listening or reading task. Their findings indicated an inextricable link between working memory and inhibitory control.

Inhibitory control is the brain’s directive to focus. Without focus, working memory is compromised because there is simply too much to hold onto. Without inhibitory control, people say and do random things and conforming to the environment presents many difficulties. Blair and Razza (2007) conducted an experiment with 141 3- to 5-year old children from low-income homes to consider the relationship between inhibitory control and academic outcomes. Their findings confirmed that for children with better inhibitory control, letter identification and
phonemic awareness skills were stronger. Foy and Mann (2013) stated that less skilled readers have more trouble inhibiting irrelevant information and doubted that it was due to inhibitory control, as poor readers have no trouble processing nonverbal sounds, spatial working memory, or focus, but they do have trouble with phonological awareness. Their study suggested that early reading skills might be better supported with instruction and practice in inhibiting irrelevant phonemic stimuli.

Cognitive flexibility refers to the ability to switch attention from one thing to another to another. Cognitive flexibility relies on inhibitory control, and without these two aspects functioning in synchrony, working memory becomes compromised. The ability to shift focus from one task or idea to another and back again is directly related to working memory and inhibitory control.

**Empirical research.** Many researchers have studied the effects of one element of executive functioning on another and have found them to be interdependent (Blair & Razza, 2007; DeBeni et al., 1998; Foy & Mann, 2013; Rothlisberger et al., 2012; Rueda, Rothbart, McCandliss, Saccomanno, & Posner, 2005; Willoughby et al., 2017). Other researchers have studied the whole of executive functioning on outcomes. Little (2017) used data from the Early Childhood Longitudinal Study, Kindergarten Class of 2010-2011 to examine the executive functioning skills of elementary students from a racial and socio-economic perspective. Significant gaps were found both racially and economically, with students from lower SES performing lower than their higher SES peers, and Black and Hispanic children underperforming their peers. There was convincing evidence, though, that the gap in executive
functioning closed more quickly than the perseverant gaps in academic performance between the same groups in the course of a typical school year.

Executive function develops, children are not born with it (Center for the Developing Child, 2018). Its development is thought to be enhanced by stable, loving, communicative and safe environments. Where chronic stress marks early childhood due to unpredictability, neglect, or a variety of other factors, executive function is slower to develop (Diamond & Lee, 2011; Fitzpatrick, McKinnon, Blair, & Willoughby, 2014; Hackman, Gallop, Evans, & Farah, 2015). While evidence is growing that children with stronger executive functioning skills perform better on academic tasks (Blair & Razza, 2007; Becker, Miao, Duncan, & McClelland, 2014; Little, 2017), it is also revealing that children from lower socio-economic status have less well-developed executive functioning skills upon arrival to school (Little, 2017; Willoughby et al., 2017).

Becker et al. (2014) conducted a study to investigate the interplay between executive functioning, self-regulation and visual motor skills. Citing Floyer-Lea and Matthews (2004), Becker et al. discussed that “children with better VMS [visual motor skills] are more likely to show a faster rate of automaticity, allowing for an easier translation of letters and numbers to paper” (p. 5). Further, they invoked research showing that visual motor skills in kindergarten predict third grade literacy (Grissmer, Grimm, Aiyer, Murrah, & Steele, 2010; McPhillips & Jordan-Black, 2007; Taylor, 1999). With 127 children from pre-kindergarten (49 children) and kindergarten (78 children), Becker et al. used known direct assessments of behavior self-regulation (Head-Toes-Knees-Shoulders task), inhibitory control (the Day-Night Stroop task),
working memory (the Woodcock-Johnson Auditory Working Memory subtest) and visuomotor skills (Beery Visual-Motor Integration). Academic assessments (subtests from the Woodcock-Johnson in math, emergent literacy, and picture vocabulary) were also administered. Comparisons between results were conducted and revealed that all domains (behavioral self-regulation, executive function and visual motor skills) were significantly associated with emergent literacy and in particular, vocabulary.

In their discussion, Becker et al. (2014) suggest that the “positive link between behavioral SR, EF, and VMS with literacy could point to a synergistic relationship as reading skills develop” (p. 30). Interventions for teaching executive functions are included in the literature. They include structures, scaffolding, routines, and high-interest opportunities to control and focus attention.

Print Concepts

**Definition.** Print concepts include knowledge of the functions of print, concept of letter, concept of word, directionality of print, and the general organizational schema of books (Meisels & Piker, 2001). Justice, Bowles, and Skibbe (2006) defined print knowledge as “an umbrella term that describes children’s maturing knowledge about the rule-governed system of orthography and written language” (p. 224).

**Empirical research.** In a study conducted to validate the use of a tool to measure preschoolers’ print concept knowledge, Justice et al. (2006) took a discrete look at print concept knowledge. They studied 128 typically developing 3- to 5-year old native English-speaking children from a variety of socioeconomic tiers using the Preschool Word and Print Awareness (PWPA) assessment. The PWPA tool measures
14 elements: front of book, title of book, purpose of title, function of narrative text, left to right organization, print versus pictures, organization of book, organization of print (top and bottom lines), purpose of contextualized print, letter as a print unit, first letter, capital letter, and meaning of contextualized print. They concluded that the PWPA proved a valid measure of these skills, and their discussion carries recommendations for early identification of deficit skills for preventive interventions. Additionally, they found that low-SES children had lower scores than those from middle-SES children, which is a theme seen throughout the literature on emergent literacy skills (Fitzpatrick et al., 2014; Foy & Mann, 2013; Hackman et al., 2015; Little, 2017).

Incorporating Ehri’s (1998) work on the developmental stages of reading, Flanigan’s (2007) study set out to replicate findings by Morris (1993), where reading acquisition including concept of word in text was found to be a central feature in learning to read. Fifty-six native-English speaking typically-developing kindergarten students were provided balanced literacy instruction, including systematic phonological awareness instruction. Five assessments of early literacy skills were administered during the first two weeks of May: beginning consonant awareness, concept of word in text, spelling, phoneme segmentation, and word recognition. Based on the results, Flanigan concluded that “mastery of three critical early literacy skills—full phoneme segmentation ability, phonemic spelling ability, and word recognition—occurs after a child has learned to match spoken words to printed words while reading text” (p. 56). Morris reached a similar conclusion: “a stable concept of word in text can actually facilitate a child’s awareness of the sequential sounds in
words” (p. 149). The recurring theme that acquiring the concept of a word as an orienting concept is pivotal in learning to read and provides insight and depth to the established reciprocating relationship between phonological awareness and early decoding.

Wondering whether print concepts develop sequentially, Mesmer and Williams (2015) conducted a study that sequenced abilities within print concepts. In the phonological awareness literature, children identify from large to small: word, then onset/rime, then phoneme. Mesmer and Williams argued that the word is the context within which letter and phoneme knowledge becomes useful. Because the speech stream contains those units but has no spaces between successive spoken words (Tunmer, Bowey, & Grieve, 1983), the convention of printed language presents an invisible mystery to beginning readers (Clay, 1979, 1991; Flanagan, 2007; Holden & MacGinite, 1972; Mesmer & Williams, 2015). But, when learners understand that words can be seen and heard, they are ready to understand that words are separated by spaces on a page. By studying 101 preschoolers and comparing the interplay between concept of word, beginning sound awareness, letter naming, and phonological awareness, they concluded that for print concepts, the same progression is true, first students see print in lines, then they see spaces and words, then letters and sounds. Mesmer and Williams went one step further to discuss that syllable knowledge happens next. Implications from this last piece, that syllable knowledge is an element of print concept, are that the print concept is not consolidated once a child can handle and interact in a conventional way with a book—there is a further step. Mesmer and Williams’ work stretches print concept beyond the preschool and kindergarten year in
a way that Clay (1979), Morris (1993), Justice et al. (2006), and Flanigan (2007) did not do.

**Interaction with other elements.** Emergent literacy skills, including print concepts, prepare children to learn to read. In addition to print concepts, emergent literacy skills encompass phonological and orthographic awareness. Children with emerging awareness of the separable sounds in language and a general understanding that orthographic symbols (letters) represent sounds make more rapid progress than those who do not (Flanigan, 2007; Justice et al., 2006; Morris, Bloodgood, Lomax, & Perney, 2003). Phonological awareness has a deep and robust literature base expounding on it as an emergent literacy skill. As discussed in the phonological awareness section of this paper, there is a known reciprocal relationship between hearing and seeing sounds and compounding phonological awareness with orthographic awareness (Hohn & Ehri, 1983). The exploitation of this reciprocity is what moves children from comprehending spoken language to deciphering and comprehending written language. In between speech and text are conventional tools that are key to growing understanding about printed language (Justice et al. 2006). The relationships between print concepts, phonological awareness and phonics are clear, deep and intertwined.

**Phonological Awareness**

**Definition.** Phonological awareness is defined by Foorman et al., (2016) as recognizing and manipulating segments of sound in words. Bentin and Leschem (1993) defined it as a child’s ability to consciously manipulate syllabic, subsyllabic and phonemic segments of spoken words. Torgesen, Morgan, and Davis (1992),
defined it as “one’s sensitivity to, or explicit awareness of the phonological structure of the words in one’s language” (p. 364). Though the term phonological awareness is sometimes used synonymously with the term phonemic awareness (Ball & Blachman, 1991; Lewkowicz, 1980), phonemic awareness is a subset of phonological awareness and refers to the smallest grain size of sound in a word (/h/ /i/ /m/). Phonological awareness refers to awareness of phonemes, syllables and words (Ziegler & Goswami, 2005). The overview of phonological awareness presented here will include a look at phonemic awareness, the smallest sound bits of words, and phonological awareness, the larger sound structure of language in consideration of how the sounds of language support the translation of print to meaning.

Empirical research. Ehri and Wilch (1987) studied whether phonemic awareness and spelling have a relationship as suggested by Bradley and Bryant (1983). It was clear that knowing the sounds in words, as well as the symbols representing the sounds made it easier for children to spell words with accuracy. Further, in a discussion about methodology for teaching children to spell, they speculated that when students learn spelling by rote memorization, they miss the opportunity to acquire or consolidate phonemic and phonetic skills, further emphasizing the connection between the sounds of the language and its written representation.

Many researchers have worked to determine which sub-skills are prerequisite for learning to read, and when skills were most ripe for teaching. Ball and Blachman (1991) found that explicit teaching of segmenting and mapping sounds to letters was advantageous to kindergarten students in reading and spelling. Torgesen et al. (1992) concluded that teaching the blending of sounds alone is not sufficient, and that while
segmenting and blending is necessary, which should be taught first was not discovered as the results were virtually the same either way. Similarly, Oudeans (2003) concluded that the instructional sequencing of phonological skills is not what produces gains in reading performance. Iversen and Tunmer (1993) introduced a systematic and explicit training of phonological recoding to augment the Reading Recovery model and concluded that phonemic awareness is not an incidental occurrence for those students selected for the program. They concluded that this skill needs to be taught. Their argument hinged on evidence found that deficient phonological awareness is widely regarded as a major cause of reading disability.

These studies, along with countless others discussed in collective literature reviews influencing policy (NRP, 2000; Snow et al., 1998; Foorman et al., 2016) affirm the importance of phonemic awareness for beginning readers and prove that explicit teaching of the discrete skills that make up phonemic awareness (blending, segmenting, substituting) results in children learning them. However, a look at the research not included in those literature reviews reveals a different line of thinking.

A study conducted in both English and Hebrew by Bentin and Leschem (1993) found that for students who come from homes with literate backgrounds, the introduction of the alphabetic system triggers their awareness of sounds and they do not need to be taught that language is made of composite sounds. However, if children do not come from linguistically rich homes, the skill is deficit and must be taught. Included in their findings is the idea that reading increases phonological awareness. This corroborates the Iversen and Tunmer (1993) finding in that the children chosen for Reading Recovery come from disadvantaged homes and seem to
have greater need for explicit instruction in phonological awareness. To drive the point further, the author of Reading Recovery states in Clay (1991): “it is by no means clear from many research papers whether phonological processes (which are cognitive processes) precede and contribute to reading acquisition, or whether they could be cognitive outcomes of being a reader” (p. 15). Notable here is Clay’s consideration of being a reader. Absent from the aforementioned studies are discussion of comprehension, connection, enjoyment, or learning from the concepts embedded in print. While these aspects may be reasonably inferred, it is notable that the studies become more granular in their quest to determine the smallest, most isolated aspect of phonological awareness to teach first, almost as an anchor point to the monumental task of learning to read.

The empirically supported theory of phonemic awareness and the growing body of evidence proclaiming that phonological deficits predict later reading deficits, that phonological deficits are more pronounced in students with learning disabilities, and that poor phonological awareness is a neuropsychological and genetic issue inspired Wise, Ring, and Olson’s (2000) study. Their study applied teaching techniques to students identified as lower- and higher-level readers and proved that the students who started out lower gained more than those who started out higher. Wise et al.’s. findings align with Torgesen, Wagner, Rashotte, Burgess, and Hecht (1997) and Bentin and Leshem (1997) in that students who come to school from literate backgrounds do not need instruction in phonemic awareness to the same degree as those from less literate backgrounds. Hagans and Good (2013) looked at the effect phonemic awareness training had on students from low SES backgrounds as compared
to those from high SES backgrounds and speculated that perhaps a more comprehensive approach to teaching this skill would produce more lasting effects. Wise et al. (2000) also began to bring the discussion of the literature back around to real world reading ability and application. Rather than an emphasis on how or what to teach, the recommendation shifted to who to teach.

Walton, Walton, and Felton (2001) conducted a study with grade one prereaders with weak letter-sound and phonological skills and discovered that they develop word reading ability relatively quickly if given experience with rime analogy. They showed that rather than teaching each phoneme individually, teaching word parts is sufficient in helping students read words. The rime analogy strategy involves using a shared spelling sequence to predict a shared pronunciation. A minimum level of competence in phonological and letter-sound skills is required. This idea, that the grain size of the phoneme and the ability of a reader to identify it before beginning to read, is seen throughout the literature.

Advising that teachers begin phonological instruction as soon as possible, Foorman et al. (2016) based their recommendation on a review of the literature from 2000. Developing awareness of the segments of sound in speech and how they link to letters should be taught from kindergarten through the first-grade year. Their instructional recommendation begins with introducing students to words and drawing their attention to smaller and smaller sound segments. Next in Foorman et al.’s instructional sequence is the recognition of onset and rime (or initial sound and ending sound segment of a word), and finally the decomposition of words into their individual phonemes. After students can isolate phonemes, then teachers should help them map
phonemes onto letters. The sequencing of these skills begins with single letter consonants and short vowels, then consonant blends, and next consonant digraphs (two letters making one sound: /wh/), then vowel digraphs, and eventually building words with letters. Consideration for assigning meaning to the sounds comes at the end when students are building words.

Grain size of phoneme across languages was studied in an extensive review of literature done by Ziegler and Goswami (2006). They found full access to phonemes only develops when children are taught to read and spell. It did not seem to matter whether the sounds of the language were individually discernable before the shapes of the written language were assigned individual sounds. They suggest that perhaps a more holistic view of the sound-symbol relationship be considered. This theory is further incorporated in the Goswami and Bryant (2016) research looking at phonological impairment for dyslexic students. Impaired elements are given as difficulty counting syllables, recognizing rhymes and phoneme substitution, phonological short-term memory and rapid automatized naming of object and color. These elements are not all directly related to the sounds of language and mapping those sounds onto objects, which begins a deeper consideration for a more holistic view of what children are capable of and ways in which those capabilities are mutually supportive.

**Interaction with other elements.** Phonological awareness interacts with other elements of reading in a variety of ways. Developing in tandem with print concepts, beginning readers hear the words as they see them, mapping sounds to letters. Hearing a word and mapping it to the symbology of the word taps into decoding
strategies. Fluency develops through the connections built between spoken language
and print—without awareness of how sounds work in spoken language, a reader
cannot bring the print to life in a prosaic way. Elwér et al. (2015) found evidence that
phonological awareness as measured in preschool develops independently from word
recognition but seems to be deeply related to comprehension and language skill.

**Word Recognition**

**Definition.** The ability to see a word and say it characterized the act of
reading in the early- to mid-twentieth century (Chall, 1967). Dolch (1936) used the
term *sight word*, LaBerge and Samuels (1974) used *automaticity*, and Perfetti and
Hogaboam (1975), referred to *rapid word decoding*, all in service of a see-it say-it
strategy. In the 1980’s and 1990’s several studies were conducted to consider the
impact of rapid naming, or word naming speed on reading. This reflects a change in
the terminology used for the phenomenon of sight word reading as the study of it
matured. The IRLA provides the term *word recognition* for this paper, and the skill it
promotes is a conglomerate of sight words, automaticity, and rapid naming.

**Empirical research.** Given the known importance of phonological awareness
and the sound-sight connections made as children decode (Ehri, Nunes, Stahl, &
Willows, 2001; Landerl, 2000; Shapiro & Solity, 2016; Torgeson, Brooks, & Hall,
2006), Ehri and Saltmarsh (1995) provide a discussion about the necessity of reading
by sight because English spellings are not perfectly decodable. They presented that
the acquisition of a sight vocabulary is the most important type of word learning to
understand. Ehri and Saltmarsh theorized that readers access connections that they
have formed between letters in the spellings of specific words and phonemes detected
in pronunciations. Their study did not consider sight words in a see-it say-it way. Instead, they looked at sight words as mapped because of graphophonemic regularities or predictabilities. Gauging what elements of the words the students in the study used supported their hypothesis that readers access connections that they have formed between letters in the spellings of specific words and phonemes detected in pronunciations.

Ehri’s (2005) research explored the intersection of phonological awareness and phonics, phonics and spelling, phonological awareness and spelling, and the nexus of phonological awareness, phonics, spelling, and a sight vocabulary. Asserting that spellings are phonemic maps that visually lay out elements of their pronunciations, Ehri claims that pre-alphabetic phase readers must learn whole words because they do not have the knowledge of letters and their sounds. As children gain knowledge of the alphabet, they make connections between words they have learned by sight and the spellings that are presented within them. By bonding the symbology of a word to the sound of the word, a child truly knows the word. With practice, all words come to be read by sight, which supports the automaticity required for fluent reading and processing freedom.

The National Reading Panel (2000) did not discretely address sight words in their report. However, within the section on fluency, they discuss the rapid mapping of language and expound on building automaticity. Through their meta-analysis, they found that skilled readers read words accurately, rapidly, and efficiently. This thread continues in the National Early Literacy Report (Lonigan & Shanahan, 2009), where two of six variables showing correlations with later literacy included rapid automatic
naming (RAN). RAN is not isolated to letters, sounds, words or even shapes; the ability to rapidly name a sequence of repeating random sets of pictures of objects indicates later reading ability.

Schwanenflugel, Morris, Kuhn, Strauss, and Sieczko (2007) contributed to the word recognition literature by presenting that automatic word recognition, the quick and accurate reading of words, is a key step in becoming a skilled reader. Using measures that rely on the Stroop effect (which considers that inhibitory abilities are more difficult when a skill is automatized), Schwanenflugel et al. tested a variety of unit sizes to determine whether reading is done by whole word or known word parts (regular grapheme-phoneme units). Testing 118 first graders, 84 second graders, and 79 third graders, they found that the methods readers apply to reading words are dependent upon their personal lexicon. Where students were older and more practiced readers, larger units were used to read words. Younger and less experienced readers used smaller units to read words. Schwanenflugel et al. presents that flexibility with the use of known word units and the ability to recognize and piece them together with automaticity supports an analogy-focused phonics approach, and that a sizable sight word vocabulary is necessary to best support readers.

Intensely teaching high frequency words by sight is an effective strategy. Watts and Gardner (2013) studied a small sample (8 students) of 8-year-old students and found that through intense teaching of high-frequency words, regardless of whether the predominant phonics instruction had been synthetic or analytic, most children learned high-frequency words by sight providing a positive effect on their fluency and comprehension. They recommend teaching high-frequency words as an
integral part of literacy instruction and ground this recommendation not only in their results, but in defense of the presence of irregular words in English, which they claim renders an exclusive synthetic phonic method problematic. January, Lovelace, Foster, and Ardoin (2017) tested a new technique in the instruction and practice of sight words to students. Arguing that the more sight words a student knows the less effort he or she needs to expend on decoding, which allows more energy to be spent on comprehension, they modified a pure direct instruction method to include more flexible and incremental introduction of new words. Most students made gains.

Recognizing common words increases efficiencies in reading, Clark (2016) argued that children should be encouraged to amass sight words. Relatively few words in the English language are high-use words, and many of these are not regularly spelled. An efficiency for an early reader, therefore, is to teach her to read these words by sight. This brings us back to Dolch and his list of sight words. Dolch presented a list of the most commonly used words, which were pulled out of primers of the day. Fry (1957) expanded on Dolch’s work and presented instant words, six lists of 100 words ranked by frequency values to the best ability of the technology of his time. He encouraged that these lists be used for remediating reading. The list was updated (Fry, 1980) based on a newer frequency count of five million words, and again in 2000, when they were released in a book (Fry, 2000).

**Interaction with other elements.** As a disembodied element of learning to read, word recognition is an historically assumed necessity. As a synergistic component of a connected process, word recognition supports fluency and comprehension and is supported by phonological awareness, phonics, and print
concepts. The question of whether learning words by sight is important is therefore moot. Instead, the question is which words are crucial to learn by sight and when.

Bowers (1995), looked at the naming speeds of poor and average readers from grades 2, 3, and 4 and found a direct relationship between speed and reading ability. Cornwall (1992) studied 54 9-year-old children with severe reading disabilities and found not only that rapid letter naming improved passage speed and accuracy, but also that several independent processes interact to determine the quality of reading and the extent of disability. A key finding in Cornwall’s study was that phonological awareness is a significant predictor of other reading skills. Torgeson et al. (1997) conducted a longitudinal study considering the relationship between rapid naming and reading skill. They concluded that while important, the isolation of word reading was not enough to determine the quality of reading. Multiple mutually supportive skills must come together for beginning readers, a finding that echoes Perfetti and Hogaboam (1975). Each of these studies were conducted with children older than second grade, limiting the understanding about early readers, the ways in which learning words by sight are efficient, and which words are the most beneficial to be learned this way.

Dolch (1936) and Fry (2000) have weighed in about which words, making the determination based on frequency. Where they differ is where the frequencies are derived; for Dolch, the materials children would be reading in school provided the words for frequency count, and Fry looked to the wider world of text. However, of the first 50 Fry words, there are only three words not in common: had, word, and each. The discovery in Schwanenflugel et al., (2007) that children use the word parts they
know to figure out words they don’t, considered with Ehri (2005) who discussed the ways in which pre-alphabetic stage readers can use sight words to enhance their alphabetic learning and Torgeson et al.’s (1997) assertion that the use of phonological and analytic strategies must be applied in extensive exposure to print make plain that the important part of word recognition is that children do it as mastering sight words will enable students to access every higher levels (Graves, Watts-Taffe & Graves, 1999).

**Phonics**

**Definition.** Phonics refers to the sound-symbol system of a language. Children learning to read rely on the orthography, or spelling, of sounds and blend them to read words. Morris (2014) stated that the purpose of phonics is automatic recognition of basic spelling patterns. Goswami and Bryant (2016) discuss the orthography of languages: some are transparent, meaning there is one spelling matching one symbol (Spanish is an example of an orthographically transparent language), and other languages are more complex, or opaque orthographically (English is an example of this type of orthography). The orthography makes the code of written language, and teaching that code is referred to as phonics. The term ‘systematic phonics instruction’ refers to the explicit, organized, and intentionally sequenced teaching of letter-sound relationships (Henbest & Apel, 2017; Rose, 2006; Shapiro & Solity, 2016; Torgerson et al., 2006).

**Empirical research.** In literature reviews of phonics instruction and its effects on early readers, Gough and Hillinger (1980), praised phonics as a virtually indispensable tool for discovering what spoken word goes with an unfamiliar written
Clay (1991) discussed the complexity of the coded sounds of speech and the usefulness of the alphabet as a substitution cipher. Henbest and Apel (2017) sum the research by stating that what is known is that systematic phonics must be explicitly taught to young children in the early stages of learning to read.

A meta-analysis of 66 high-quality treatment-control comparisons conducted by Ehri et al. (2001), found that systematic phonics instruction helped children learn to read better than all forms of control group instruction, including whole language. Torgerson et al. (2006) conducted a review of 20 randomized control trials (14 were in common with the Ehri et al., 2001 study, the remaining 6 took place between 2000 and 2005) and found no effect of systematic phonics on comprehension but affirmed that systematic phonics improved accuracy for typically developing students as well as students at-risk of reading failure. Foorman et al.’s (2016) meta-analysis expanded the National Reading Panel’s (2000) work by reviewing 235 studies published between 2000 and 2014, culling to 56 meeting the team’s design standards. Foorman et al.’s report directs teachers to systematically “teach students to decode words, analyze word parts, and write and recognize words” (p. 22). They conclude that children’s recognition of letter patterns and word parts and understanding that sounds relate to letters in predictable and unpredictable ways is crucial for learning to read. While the question of teaching phonics versus not teaching phonics has been answered with a resounding yes (Clark, 2016; Landerl, 2000; Nelson, Benner & Gonzales, 2005; Savage, Carless, & Stuart, 2003; Simmons et al., 2011; Vadasz, Sanders, and Peyton, 2006; Walton et al., 2001; Wright & Jacobs, 2003), there remains contention in the literature.
Using the preponderance of evidence regarding phonics instruction, and guided by a commissioned study (Rose, 2006), England mandated systematic synthetic phonics instruction for all students beginning school, which called into light a new line of debate about phonics instruction: synthetic versus analytic phonics. Chall (1967) discussed a synthetic approach as one in which parts to wholes was emphasized: first, teach letter-sound relationships, then blend sounds and letters into whole words. Henbest and Apel (2017) defined analytic phonics as instruction where the child is likely to be taught to manipulate the onset and rime of a word rather than individual letters and sounds. Little to no attention is given to blending the individual sounds in words. Glazzard (2017) described analytic phonics as being taught to analyze common phonemes in a set of words, and processing text from whole to part rather than part to whole. The National Reading Panel Report (2000) emphasized that what mattered was that phonics instruction be systematic; it went out of its way to clarify that synthetic and analytic phonics are shown to be equally effective (NRP, 2000). The National Reading Panel’s declaration is corroborated by many studies and is echoed throughout other meta-analyses arriving at the same conclusion (Ehri et al., 2001; Foorman et al., 2016; Torgerson et al., 2006).

In an effort to see whether differential effects in phoneme awareness and nonword reading were produced via synthetic or analytic approach, Savage et al. (2003) studied the lowest 108 4- and 5-year old readers from an initial pool of 414 children. They were randomly assigned to four groups for nine weeks of instruction. Group 1 was provided instruction in a synthetic program moving from phonemes (vowels and consonants) to building words with blocks and eventually writing them
Group 2 was treated with an analytic program using word parts (e.g., ‘-at’) and known alphabet letters to build words supported with pictures (rime-intervention). They built word families, wrote words, and sorted words. Group 3 was a mixed onset-rime and phoneme program following procedures from the analytic program but breaking the rime into its component vowel and consonants. Group 4 was a control group and followed the National Literacy Strategy lessons. Grouped by ability, the controls worked on initial, medial, and final sounds, consonant blends, and high-frequency words. Through a pre- and post-test protocol using 10 early literacy measures, Savage et al. confirmed that all three interventions provided by instructional assistants had been more successful than normal classroom instruction in improving children’s performance in onset-rime segmentation and blending and in letter-sound knowledge. Additionally, they analyzed data from the three interventions (not the control) to discover which had been most successful. There was no statistical difference. All three methods produced overall growth to the same levels. However, in the analytic (onset-rime) group, improvements in phoneme blending were made even though they were not explicitly taught. The transfer of skill was only one-way, as the synthetic group did not seem to have gained the flexibility of sound manipulation necessary to perform onset-rime tasks.

Landerl (2000) considered the complexity of the phonological code of English orthography. Many sounds in English are produced by a variety of spellings, and several letters represent more than one sound. Landerl’s study comparing English and German children’s acquisition of early reading and sound spellings found that English speaking children acquired phonological assembly skills (the ability to sound out)
more slowly than German speaking children and attributed this difference to the more consistent graphemic mapping of the German language than that of English.

Understanding the orthographical variety of languages as a factor in the efficacy of teaching synthetic phonics was undertaken by Goswami (2005). While phonological awareness plays a critical role in the acquisition of reading, analysis of phonological and phonic makeup of different languages is important. Showing that where there is a near 1:1 match of sound to letter in languages such as Greek, Finnish, German, Italian, and Spanish, students decode and blend letters quickly and accurately. In languages such as English and French there are more variations in mapping a sound to a letter. Where synthetic phonics is elevated as the only way to teach English, Goswami argues that confusions present themselves to students in the form of rimes that cannot be sounded out such as “ight,” “eild,” and “-tain” in captain versus “-tain” in fountain. Synthetic phonics instruction maps grapheme-phoneme correspondences, but in English there are approximately 600 different orthographic patterns, rather than 26 mapped to 26 letters. Noted in Gough and Hillinger (1980), phonics pairs each letter or letters with a syllable, not a phoneme. Offering a theoretical solution called psycholinguistic grain size (Ziegler & Goswami, 2005), a suggestion to teach students sound spellings based on their grain size is posed. This theory takes advantage of what is known about teaching initial reading in alphabetic orthographies with consistent spelling systems (synthetic phonics) and compounds it with phonological awareness by teaching correspondences for large units of sound, such as rimes or syllables.
Each empirical study reviewed for the phonics section of this paper claim, and prove, that breaking the code, or the ability to map sounds onto the symbols of the English language, is necessary for the act of reading to take place (Baumann, Hoffman, Duffy-Hester, & Ro, 2000; Comasky, Savage, & Abrami, 2009; Coyne, Kame’enui, Simmons, & Harn, 2004; Ehri & Robbins, 1992; Johnston & Watson, 2004; Landerl, 2000; McGown, Johnston, & Medford, 2012; Savage et al., 2003; Shapiro & Solity, 2016; Simmons et al., 2011; Wright & Jacobs, 2003).

**Interaction with other elements.** Throughout the literature on phonics instruction are two commonalities: phonics is not in and of itself sufficient for meaning making, and phonics is co-dependent with a variety of other reading skills: phonemic awareness, print concepts, sight word reading, morphemic awareness, command of language, and comprehension. Clark (2016) included in summation of a review of the literature that there is no evidence to support phonics instruction in isolation.

**Vocabulary**

**Definition.** Vocabulary, the body of words that make up a language, is a commonly known term. The number of words that make up an individual’s vocabulary, and the number of words that make up the entire body of a language’s vocabulary is speculated and calculated, but not conclusive. While the 2016 edition of the Oxford Dictionary contains 171,476 unique words in English, it also explains that it is an impossible task to count the number of words in English because of the variety of ways words are used (the dog is my pet, I’m dog tired, that hot dog is delicious,
etc.), words invented in slang, and words borrowed from other languages (French words in cooking, Latin words in law, Japanese words in martial arts).

Nagy and Anderson (1984) set out to discover the number of words in printed school English, and while their best estimate is about 88,500 distinct words, they also discuss at length that a word is not just a word. Meanings shift and change depending on context and syntax. Moreover, words are made up of units of meaning, called morphemes. Some words are monomorphemic, containing a single meaning, and others are multimorphemic. Nagy and Anderson claim that multimorphemic words outnumber monomorphemic words by about four to one. Given any word count for English, vocabulary instruction that teaches one word at a time is futile. There are simply too many words. Recent research in vocabulary acquisition and instruction places focus on strategies involving morphological awareness, as this section will show.

**Empirical research.** Measuring vocabulary growth and studying the phenomena of its development is noted as challenging throughout the literature (Adams, 1990; Baker et al., 2013; Justice, Meier, & Walpole, 2005; Nagy, Anderson, & Herman, 1987). The variables inherent in the simple question ‘when is a word known?’ are enough to make the measurement of vocabulary growth difficult. The measurement challenge may explain the fact that only seven studies met the standards for vocabulary and language instruction in Foorman et al.’s (2016) IES report focusing on foundational skills to support reading for understanding from kindergarten through third grade. Of those seven studies, only two found positive effects for interventions on vocabulary outcomes for children in this grade range.
From the studies included in the IES report, Foorman et al. (2016) recommended that students be taught “academic language skills, including the use of inferential and narrative language, and vocabulary knowledge” (p. 6). Academic language skills include more than just vocabulary. The ability to articulate ideas beyond their immediate context, to clearly relate a series of events, and to comprehend and use a wide range of academic vocabulary and grammatical structures are included in the IES definition of academic language. The studies included in the IES report focused more on singular vocabulary strategies than on a broader consideration of academic language.

Justice et al.’s (2005) study was grounded in three theoretical perspectives: incidental exposure to novel words is critical, word learning is gradual, and adult input can influence the acquisition of new words. Justice et al. studied 57 kindergarten students identified as at-risk due to past performance and socioeconomic level, in randomly assigned treatment or comparison groups. Students in the treatment group received 20 20-minute small-group storybook reading sessions over a 10-week period. There were 10 books used in the study and across the duration of the study each book was read four times. Sixty words were selected as new vocabulary and were randomly assigned to elaboration or non-elaboration conditions. There were three words per book. Teachers were only aware of the 30 words chosen for elaboration. The elaborated words produced a large effect for all children and non-elaborated words produced a large effect for the students who began the study with high vocabularies as determined by the pre-test, but students in the control group had negligible gains. Justice et al. used the context of language to build the number of words learned.
Teachers can conclude that when reading with children, words that are intentionally elaborated upon are more easily assimilated into a child’s working vocabulary.

Similar to Justice et al.’s (2005) study, Baker et al. (2013) studied the impact of an explicit read aloud intervention taught in a whole group format by categorizing words for instruction and teaching some explicitly and some implicitly. They found that the explicit noticing and teaching of vocabulary and thinking skills while reading aloud to a whole class works best to teach students words. The focus of this study was on listening comprehension, but the results yielded information about specific words learned.

Simmons et al. (2007) studied the impact of targeted vocabulary lessons. In 108 30-minute lessons conducted daily, vocabulary words were taught through repeated reading of story books, targeted vocabulary lessons, and exposing students to vocabulary words multiple times. There was a very small effect (Cohen’s d effect size on the Peabody Picture Vocabulary Test ranged from 0.01 to 0.33) on the acceleration of vocabulary word learning seen in this study.

Hypothesizing that primary students can benefit from structural knowledge of expository texts, Williams, Stafford, Lauer, Hall, & Pollini (2009) studied the implementation of strategies directly teaching students how expository text works, including strategies for learning content-specific words. Vocabulary concepts were related to critical features of animal classifications through simple definitions in everyday language. Although Williams et al. did not set out to investigate vocabulary growth, it found that by teaching students within a greater context, there were significant vocabulary gains.
As seen in the preceding paragraphs, research that informed the Foorman et al. (2016) IES study demonstrated approaches for teaching individual words (read alouds and authentic texts featured in all of them), but there is evidence beyond the studies Foorman et al. considered in the research of other strategies that may be as powerful, if not more so, for increasing students’ vocabulary and for its rapid expansion.

Morphological awareness refers to the ability to consciously consider and manipulate the smallest units of meaning in language, including base words and affixes. Kirby et al. (2012) studied whether the skill of morphological awareness should be considered a stand-alone, or if it is too deeply embedded in the other cognitive skills of reading. In their longitudinal study of 103 1st, 2nd, and 3rd graders, Kirby et al. considered the relationship of morphological awareness to reading development and concluded with a recommendation to assess and teach morphological awareness from the first grade. Significant relationships between word analogy and morphological awareness, word reading and morphological awareness, and overall comprehension and morphological awareness were found. While weakest in 1st graders, morphological awareness was found to be taking root at 1st grade and developing noticeably through the 2nd and 3rd grade.

Apel, Brimo, Diehm, and Apel (2013) conducted a feasibility study to determine whether morphological awareness instruction would impact the vocabulary of low socioeconomic students in kindergarten, 1st, and 2nd grade. They spent nine weeks working with 19 kindergarten, 21 1st grade and 21 2nd grade students four times each week for 25 minutes a time to increase awareness of affixes, inflected, and derived words. They found that students at kindergarten and 1st grade were aware of
the relationship of base words and their inflected and derived forms, and that their level of morphological awareness grew across the grades. Apel et al. found the most pronounced growth in the first three grades, but also that there continued to be gains across the grades as well. There were larger effect sizes than expected across all groups. This was the first study of its kind, and although small, it was successful enough that additional studies have been done to understand the impact of instruction in morphological awareness.

The use of morphological problem-solving strategies to read multimorphemic words successfully is a viable way for children to solve words for their meanings (Apel & Henbest, 2016). Stating that children begin developing morphological awareness as early as kindergarten and continuing to grow through high school, Apel and Henbest developed an affix meaning task to investigate the explicit knowledge of affix meanings of 1st, 2nd, and 3rd graders, and to discover whether affix meaning knowledge predicts reading abilities more than other skills. Additionally, they sought to understand when children develop affix naming knowledge to guide teachers and specialists to provide on-time instruction in morphological awareness for children not developing this skill when appropriate. They found that affix naming is autonomously learned through first grade, but continues to grow from there, requiring more deliberate instruction.

**Interaction with other elements.** The NRP (2000) included vocabulary within their analysis of comprehension strategies because both involve the meaning of text: vocabulary tied closely to individual words and comprehension more in larger units. They claimed that separating the two processes is difficult, if not impossible.
The relationship between vocabulary and comprehension is often discussed in the vocabulary literature (Adams, 1990; Beck, Perfetti, & McKeown, 1982; NRP, 2000), where speed or automaticity of access to word meaning is emphasized. For example, Beck, Perfetti, and McKeown (1982) asserted that comprehension depends in part on easy access to word meanings in text and focused their work on the reciprocal relationship between word meanings and semantic knowledge. In their meta-analysis, Stahl and Fairbanks (1986) found that through vocabulary instruction, growth in reading comprehension is facilitated and that quick knowledge of a word’s meaning prevents a cognitive bottleneck. Nagy and Anderson (1984) were concerned with the size of an individual’s vocabulary and determining efficiencies for helping students make meaning from words in service of comprehending connected text through the grades.

Kirby et al.’s (2012) study probed the relationship between morphological awareness and word analogizing. Their finding was that of the five component relationships studied, this was the most dramatic. Students who can solve a word’s meaning through analogy using its morphological components have a higher overall reading ability. Considering the connection between analytic phonics and morphological awareness is a next logical step in determining the most powerful way to teach early reading. If morphological awareness is a key to greater comprehension through a more facile realization of word meaning, a question to consider is whether learning to read words by analogy promotes morphological awareness, thereby improving comprehension. Though logical, research to explore this relationship has not yet been done.
The relationship between vocabulary and word recognition is called out in literature on both, and the connector here is fluency. Where readers can smoothly and accurately read connected text, comprehension is bolstered particularly when the meanings of individual words and phrases are known without having to stop. The fluency research elevates reading with expression as impactful to comprehension (Hudson et al., 2005), and knowledge of words enables that process.

Vocabulary is connected to phonological awareness as a child must have a receptive storage for the sound of the word, no matter how morphologically complex. When considering morphological awareness in tandem with phonological awareness, there are obvious links. Children must develop facility with small units, but this often happens after they gain facility with larger units. The same mechanisms are at play with the acquisition of literacy skills, signaling that leveraging the elements to support one another will enhance and potentially accelerate the growth of each.

**Fluency**

**Definition.** At its essence, reading fluency is equated with reading speed, or rate, generally measured by number of words read per minute (Adams, 1990; Allington, 1983; Foorman et al., 2016; Hudson, Lane, & Pullen, 2005; O’Connor, Swanson, & Geraghty, 2010; Pinnell, 1995; Samuels, 1979). Samuels (1979) operationalized 85 words per minute as the goal of a repeated reading intervention to promote the development of fluency. There are, however, a variety of nuanced definitions present in the literature, including automaticity (Harris & Hodges, 1995; NRP, 2000), accuracy (Foorman et al., 2016; Samuels, 1979), freedom from word identification problems (Harris & Hodges, 1995), expression (NRP, 2000; Foorman et
al., 2016; Pinnell, 1995), and smoothness (Samuels, 1979). At its core, reading fluency functions as an indicator that word recognition has become automatic. The ability to measure fluency quantitatively, by words per minute, makes it an element of reading easy to monitor and, as seen in the research, relatively easy to influence.

A consensus exists throughout the literature, claimed by Kuhn and Stahl (2003), regarding the primary components of fluency: accuracy, automaticity, and “appropriate use of prosodic features such as stress, pitch, and appropriate text phrasing” (p. 5). While expression features in some of the above definitions, Kuhn and Stahl name a more robust term: prosody. Spoken language is prosaic. Inflections, volume changes, modifications in voice tone, and fluctuations in the speed of language depending on what is being communicated enhance the comprehension of transmitted ideas (Pinnell, 1995). Written language should be equally prosaic, but with limited supporting text features in written language (Allington, 1983), beginning readers have trouble bringing written language to life. Hudson et al. (2005) describe prosodic readers as those who understand what they read, and when reading aloud, make it easier for others to comprehend as well.

**Empirical research.** Samuels (1979) was able to prove that repeated reading of a text increased fluency, but Kuhn and Stahl (2003) showed that repeated reading did not generalize fluency skills. Surprisingly, there are few studies to verify Kuhn and Stahl’s work regarding the generalizability of fluency skills. O’Connor et al. (2010) showed that extensive practice in reading connected text improved the reading rate for average readers in the primary grades. With opportunities to read aloud to an adult for 15 minutes two times a week over a 20-week period, students in Grades 2
and 4 made significant growth, regardless of the difficulty of text. Children in the control group did not read aloud to an adult and did not make noticeable growth. O’Connor et al. discussed that in studies where reading rate improved through repeated reading, generalization to unique text was not observed. Instruction in fluency has been shown to improve fluency but understanding whether fluency truly enhances or supports comprehension is still thin. O’Connor et al.’s study revealed that fluency had a positive impact on word reading and comprehension, but not on decoding or vocabulary. This study called for further consideration of the interplay of literacy skills as a whole.

Supplemental programs that home in on literacy skills have provided some of the evidence contributing to the empirical research base, but Begeny et al. (2010) point out that though many programs have a theoretical and research basis driving their development, very few if any studies are done to determine effectiveness of the developed and implemented product. Begeny et al. designed and carried out a study on their theory-based fluency program. Using a randomized control design with 2nd graders, Begeny et al. compared their program to a similar program and a control. From February through April of the 2nd grade year, students spent 10 minutes on Mondays, Wednesdays and Fridays in one-on-one sessions with a tutor. Using passages of relatively similar difficulty levels, students read an unfamiliar passage aloud, received some form of corrective feedback, received performance feedback aligned with reading goals, listened to an adult model the same passage, and received rewards for improved reading behavior. The other control program included repeated reading of the passage with suggested prompts (read for fluency and for
comprehension), systematic error-correction procedures and opportunity for students to orally retell the content. Both experimental groups improved in fluency, with the program containing prompts and repeated reading improving the most. Additionally, improvements were seen in word reading and decoding on the prompting program. What did not improve in either experimental condition was comprehension. Though the comprehension measure produced a medium effect size, the $p$-value was too great for the researchers to claim significance. The variety of variables present between the two experimental groups may have made understanding the instructional behaviors and the experiences of the learners difficult to interpret if significant differences had appeared.

Martens et al. (2007) found mean increases of two to three grade levels in passages mastered in their study of a fluency-based reading program. This study chose participants based on a phonemic awareness score indicating competence with phonemic awareness, but who were not yet reading grade-level text. Martens et al. divided the pool of participants into two groups: an experimental and a control. The experimental group participated in training sessions one-on-one with trained tutors. Each session lasted 30 minutes and was held three times each week for five weeks (for second graders) or six weeks (for third graders). During the session, students were given a passage, a reward ticket to potentially exchange for a prize, and asked to state the goal of reading 100 words per minute. They then read the passage. If the goal was met, they were able to choose a prize, receive another ticket and another passage. If the goal was not met, they were given feedback, provided with a model, and given another opportunity. This classic behaviorist model yielded increases in the number of
words read per minute. The statistical improvements were strong, showing an increase of words read per minute on increasingly difficult text over time for both grades.

While Martins et al. (2007) noted limitations including the number of students (small) and the number of adults required for this study (large), and the isolation of only 2nd and 3rd grade students, they did not discuss comprehension at all. If, as the research discussed throughout this section shows, fluency and comprehension are related, it seems a more apt question is how they are related. Martins et al. assumes that fluency is the cause of comprehension, not the effect of it.

**Interaction with other elements.** Cause and effect between fluency and comprehension runs through the literature. “Fluency depends upon well-developed word recognition skills, but such skills do not inevitably lead to fluency” (NRP, 2000, p. 3-1). Each study of fluency used measures of phonological awareness and word recognition to determine readiness for the study. Meta-analyses called out repeated relationships between fluency and other elements. It is widely agreed that fluency is dependent upon automatic word reading, phonological awareness, decoding skills, and word meaning (Eldredge, 2005; Hudson et al., 2005; Kuhn & Stahl, 2003; Snow et al., 1998). If any of these components is too weak, fluency cannot be achieved. Further, comprehension is theorized to not occur without fluency and word meaning skills being in place. Of all the areas examined in this paper so far, fluency holds markers of being a gateway between early reading acquisition skills, or competently naming words in print, to comprehension, or arriving at meaning in print.
Formative Assessment

**Definition.** Although most articles featuring formative assessment spend time discussing and contemplating the definition of formative assessment (Bennett, 2011; Black & Wiliam, 1998; Bloom, 1969; Dunn & Mulvenon, 2009; Melmer, Burmaster, & James, 2008; Sadler, 1989; Stiggins, 2002; Taras, 2005), a common definition has yet to be agreed upon. In a seminal piece laying out the theory of formative assessment, Sadler claims that formative assessment can be used to shape and improve students’ competence by short-circuiting the randomness and inefficiency of trial and error learning. Black and Wiliam state that there is no consistent definition of formative assessment and attempt to provide clarity by calling “all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged” (p. 7-8) their interpretation of formative assessment. Black and Wiliam’s definition is so wide and broad that rather than providing clarity, it reinforces the nature of teaching, wherein teacher interactions with students inform next teaching behaviors in a continuous loop. The idea of a continuous loop is carried throughout the literature focusing on specific areas of learning exemplified by Afflerbach et al.’s (2008) comprehension article explaining that reading instruction follows a “regular cycle of modeling, explaining, and guiding…that leads to independent practice and fluency” (p. 370).

A much heralded and somewhat criticized meta-analysis conducted by Black and Wiliam (1998) (cited in 8,740 works, according to Google Scholar, August 2018) is a natural starting place for the examination of empirical evidence of the impact of
formative assessment. The definition of formative assessment has been somewhat mercurial, which leads to many difficulties in the evidentiary work. Black and Wiliam set a definition and collected research papers fitting it. Providing synopses of each of eight papers based on quantitative comparisons of learning gains, Black and Wiliam asserted that formative assessment accelerates learning, changes attitudes about learning and self, cuts prejudicial teaching and increases motivation. “We have not come across any report of negative effects following on an enhancement of formative practice” (p. 17). Stiggins (2002) heralded this work and furthered the claim by hypothesizing that if formative assessment as described in Black and Wiliam became standard classroom practice, achievement gaps would be erased.

The definition used by Black and Wiliam was not precise and did not explicitly match the definitions in the papers analyzed. Dunn and Mulvenon (2009) provided a critical analysis of Black and Wiliam, systematically refuting the empirical evidence they provided. First by undercutting the vastness of the definition used by Black and Wiliam, then discussing the nature of formative assessment as the way a teacher uses any assessment tool, Dunn and Mulvenon move through each of the eight papers presented to illuminate methodological issues. From poor interpretation of effect size calculations, skewed sample sizes, the use of technically inadequate dependent measures, inadequate teacher training, and inappropriate application of statistical measures, Dunn and Mulvenon dismantled Black and Wiliam’s findings. However, they found similar methodological issues in nine more recent articles on formative assessment reporting similarly positive impacts. Concluding that there is limited scientific evidence to support that formative assessment directly contributes to positive
educational outcomes, Dunn and Mulvenon call for a clear and shared lexicon for formative assessment and a sound, research-validated framework for best practices in formative assessment.

The Council of Chief State School Officers (CCSSO) formed a collaborative to define formative assessment in 2006, and the definition was refined in 2018 to embody ways in which the term is used in research, practice, and policy (CCSSO, 2018). Key elements in their definition are that formative assessment is a planned and ongoing process, and that student learning is elicited and evidenced to improve outcomes. The process includes setting clear goals, analyzing evidence of student learning, self-assessment and actionable feedback. The CCSSO paper carefully unpacks and elaborates on each element within its definition but does not provide empirical evidence upon which the definition was built. CCSSO does stress that formative assessment is a process that unless integrated in full, carries no promise of efficacy. However, by incorporating goal setting, evidence of knowledge or skill, self-assessment, and actionable feedback, students and teachers engaged in a formative assessment process can positively impact student learning.

**Empirical research.** A limited body of scientifically based empirical evidence exists about formative assessment (Bennett, 2011; Black & Wiliam, 1998, 2009; Dunn & Mulvenon, 2009; Shepard, 2000; Taras, 2005). This may be due to a lack of common definition, or it may be due to the individualized nature of formative assessment. Dunn and Mulvenon (2009) refer to formative assessment as an ethereal construct and make a case that further clarity be provided to enable more quantitative analysis of its impact on education. Where there is a social construct that is flexible
and responsive, there is difficulty providing quantitative, empirical evidence. The quantification of formative assessment is difficult because the separation of its component parts is virtually impossible. Throughout this dissertation, the component parts of reading (phonological awareness, phonics, word identification, etc.) have been separated, studied as singular entities, and reveal that learning them in isolation is not as powerful as learning them synergistically. Formative assessment is about relationships, which are unique between each child and teacher, about closing gaps in learning, which are also unique, and about setting next steps for each individual, another unique proposition. The permutations are infinite, requiring flexible, knowledgeable individuals to engage in cooperative discourse pointed at known goals. When this type of work is being done, and students grow as learners, it is impossible to point to which of the permutations provided the growth. Likewise, when learning is stalled, determining which element is lacking is hampered by the sheer quantities of variables. Bennett presents that although generally facilitative of learning, definitions of formative assessment have been varied throughout the literature, as have implementations, making the naming and evaluation of formative assessment difficult if not impossible to conclusively research. The inability to cleanly and quantitatively dissect and study the elements of formative assessment should not disqualify it as a valid and valuable tool for learning.

**One-to-One Instruction**

**Definition.** One-to-one instruction, individualized instruction, and tutoring are used synonymously throughout the literature and refer to the fact that a teacher and a child are working together without others. Frey (2006) draws a distinction, citing
special education literature, that one-to-one instruction is truly one-to-one, while individualized instruction also refers to small groups or clustered instruction. Ehri et al. (2001) claim that one-to-one tutoring is the preferred form of instruction for students who are having difficulties because lessons can be tailored to individual needs.

**Empirical research.** The Reading Recovery program, developed by Clay (1993), is an approved intervention program on the What Works Clearinghouse website (WWC, 2013), and has been the focus of several studies to determine its effectiveness (Iversen & Tunmer, 1993; Pinnell, DeFord, & Lyons, 1988; Schwartz, 2005). Pinnell’s (1989) study of the principles involved in Reading Recovery led to the conclusion that individualized instruction increases learning for at-risk readers, and that the key factor is not the Reading Recovery program itself, but the teacher’s ability to make good decisions based on observations of the child. Like prior studies, this work focused on children identified as behind and the economics of implementation.

Citing Bloom’s (1984) study and claiming that research has repeatedly demonstrated one-to-one instruction to be more effective than either whole-class or small-group instruction, Iversen and Tunmer (1993) found an effect size on reading level measures of more than eight standard deviations between students receiving one-to-one instruction in a modified Reading Recovery model over students in matched schools receiving support in groups of four to six. Iversen and Tunmer caution that their results may be more due to the focus of the instruction rather than group size.

Based on converging evidence that when 3 dimensions (grouping strategies, content needs, and management) are used to individualize instruction based on
students’ language and literacy skills in the early years, students achieve stronger reading performance (Al Otaiba et al., 2011; Connor et al., 2009; Connor et al., 2013), Connor designed Individualized Student Instruction (ISI) intervention software to make algorithmic recommendations to teachers based on data from assessments and classroom observation and predetermined end-of-year outcomes. Using the algorithmic recommendations, teachers engaged in individualized instruction.

Although Connor et al. (2009) and Connor et al. (2013) had shown this strategy to be effective with 1st and 2nd graders, Al Otaiba et al. studied the effectiveness of algorithmically recommended time and type of instruction on kindergarten students. Through a cluster-randomized control field trial, this study examined the effectiveness of two types of professional development conditions to support kindergarten teachers’ ability to differentiate. Students receiving individualized instruction designed on the three dimensions outperformed students in the contrast classrooms on assessments of word reading, decoding, alphabetic knowledge, and phonological awareness.

Discussion from this Al Otaiba et al.’s study includes qualitative information about the difficulties teachers had in incorporating individualized and differentiated instruction to their classroom environments, but with training, they provided more effective instruction, enabling teachers to better differentiate based on students’ identified language and literacy skills and ongoing needs as determined through assessments.

Effective instruction is highly complex. Connor et al. (2009) concluded that optimal patterns of instruction differ for each child following a study to determine the most effective type of instruction in first grade classrooms for schools requiring extra services. They observed the amounts and types of literacy instruction provided in
control (business as usual) and treatment (teachers trained in and using individualized instruction) classrooms. Children receiving individualized student instruction grew considerably in both comprehension and foundational skills. The trouble with this study was the variability across the treatment groups, rendering full implementation inconsistent. The amount of time each teacher allocated varied, group instruction provided outside of the individualized treatment varied, and the lessons required by student need varied. Among Connor et al.’s findings was that students with higher needs who received more individualized instruction made greater gains, but that the needs required proper identification for the algorithm to provide appropriate instructional direction, meaning focused or code focused, and broken further into comprehension and vocabulary, or phonics and word recognition.

There are two veins of research on individualized instruction. Al Otaiba et al. (2011), Connor, Morrison, and Slominski (2006), Connor et al. (2009), Connor et al. (2013), Juel and Minden-Cupp (2000), and Slavin, Madden, Karweit, Dolan, and Wasik (1991) discuss the use of individualized instruction for teaching foundational, or code-based skills: phonemic awareness, phonics and occasionally vocabulary. Clay (1993), Compton-Lilly (2009), Schwartz, Schmitt, and Lose (2012), Pinnell et al. (1988) and Pinnell, Lyons, DeFord, Bryk, and Seltzer (1994) focus on using the one-to-one arena as a place to coach meaning-making as students engage in using and growing reading strategies. In common between all of these researchers is the focus on readers who are behind or at risk of falling behind. Also in common is the idea that individualized instruction is the best possible instructional application (Bloom, 1984; Clay, 1993; Ehri et al., 2001; Pinnell, 1989).
Missing from the research and discussion are clarity regarding the amount of time recommended for working one-to-one with students. Reading Recovery explicitly states 30 minutes per day for a finite duration. The aim of Reading Recovery is to target students who are behind and bring them up to speed. Success for All (Slavin et al., 1991) is another program that touts one-to-one instruction, and while its focus is more on the explicit teaching of code-based skills in a whole group setting, it augments whole group instruction with after-school individualized tutoring, but the specific time for them is not given. In the multiple studies done by Connor (Connor et al., 2006; Connor et al., 2009; Connor et al., 2013; Connor, Alberto, Compton, and O’Connor, 2014), the impact of Individualized Student Instruction protocol is discussed and shown to be most effective when following the recommendations for time and mode of instruction provided by the algorithms built into the A2i software. However, discussion of the time specifications is not provided. From their writings about the proprietary algorithmic ISI software, it is unclear whether one-to-one interventions take place over five minutes or 30.

Timing of Skills

An exploration of the literature on developmental reading sequences reveals that while there are many elements in general agreement, evolving theories and frameworks do not provide consensus for precisely which skill should be taught at precisely what point in development.

Chall (1976) described the evolution of learning to read from its primitive beginnings to its most mature form. Coined stage theory, Chall presents each stage as having a definite structure different from the others in qualitative and characteristic
ways that follow in a hierarchical progression. In the pre-reading stage, students are learning about print, mastering the syntactic and semantic features of language. This stage is designated from birth through first grade. Stage 1 is presented as the decoding or initial reading stage. It spans grades 1 and 2. Stage 2, encompassing grades 2 and 3, is one of practice and consolidation of the skills gained in Stage 1. At this point, fluency and familiarity take hold for the early reader. Stage 3, from grades 4-6, sees the shift to reading for information and knowledge growth. These stages are broad and do not give way for meaning making at the early stages, nor do they specify the work a teacher and/or a student must do to grow through the stages.

Ehri and McCormick (1998) make a case for the sequential accumulation of word reading skills through distinct developmental phases. They argue that phases may overlap and that lack of mastery does not preclude advancement to the next phase. They layout the phases as follows: pre-alphabetic, partial alphabetic, full alphabetic, consolidated alphabetic, and automatic alphabetic. Although clear delineations of the phases aligned with the early grades are not given, there is mention that the pre-alphabetic phase is typical of pre-school children and early kindergarten students. The partial alphabetic phase takes students into first grade, the full alphabetic phase is fully active throughout first grade. The consolidated alphabetic phase is a common expectation at about second grade. By late second grade students are expected to be in the automatic phase, characterized by the ability to read with a high degree of automaticity.

The Consortium of Reading Excellence (CORE), under Honig’s supervision, published a comprehensive collection of research-based reading strategies including
instructional sequences (Honig et al., 2000). It is focused on explicit instruction and
presents skills and strategies organized around sound/print connection (print concepts,
alphabet recognition, and phonemic awareness), decoding (phonics, high-frequency
words, multisyllabic words, and reading fluency), spelling, vocabulary development
(specific word instruction and word-learning strategies), comprehension (strategic
reading, narrative text, and expository text), reading and responding (independent,
wide reading, and book discussions), and differentiated instruction (assessment and
instructional organization). Each of these instructional ingredients is presented with a
relative grade level for introduction and completion. For example, alphabet
recognition is given to begin in early kindergarten and conclude in early first grade.
This resource presents a timeline with building blocks expecting mastery, but also
with considerable overlap.

In a study by Morris et al. (2003), a theoretical model of early reading
development was tested; an interactive relationship between beginning readers’
concept of word in text and phoneme awareness was investigated. To do so, the study
considered the placement of skills in the following sequence from kindergarten
through first grade: alphabet knowledge, beginning consonant awareness, concept of
word in print, spelling with beginning and ending consonants, and finally phoneme
segmentation, word recognition, and contextual reading ability. They found an
interactive relationship between beginning readers’ concept of word in text and
phoneme awareness but brought additional concepts to the sequencing of early reading
skills: spelling using initial and ending consonants, and finger-point reading. By
teaching concept of word in text and having students isolate words in a one-to-one manner, the bridge from word to phoneme awareness was strengthened.

Jarmulowicz et al. (2008) conducted a study to investigate whether phonological awareness and morphological awareness combined to create a new term: morphophonological accuracy. This refers to an oral task that manipulates phonological (sound-based) and morphographical (meaning-based) aspects of derived words, bringing meaning-based elements of early reading acquisition into early literacy classrooms. Due to the findings in previous research noting that morphological and morphophonological accuracy are important in mid- to late-elementary education, they looked at the impact on sequential skills by those immediately preceding them to see if work with these skills earlier would have a positive impact on reading development. They found that the developmental sequence of skills, or pathway, was impacted, and additionally, that morphophonological instruction had a significantly positive impact on subsequent decoding skill. They cautioned that a unidirectional model of development isn’t true in the strictest sense, and emphasized that throughout their model, earlier developing skills support later developing skills. The developmental sequence their study led them to propose is: receptive language, phonological awareness, morphological awareness, morphophonological accuracy, decoding, and finally comprehension. Specificity regarding grade level is present in their work, beginning pre-kindergarten with receptive language, growing phonological and morphological awareness through the kindergarten year, and proceeding through morphophonological accuracy, decoding, and comprehension during the first-grade year.
Discussing an emergent literacy model, Dooley (2010) proposed that comprehension comes before decoding. When preschool children play with books, Dooley claims they are developing through four stages of emergent literacy: book as prop, book as invitation, book as script, and finally, book as text. Preschool and preliterate children moving through these stages build an understanding that books hold meaning, and as they become more aware of meaning in text, they become more autonomous with the books and develop a desire to learn the processes of decoding the text that tells the story more truly than the interpretation of illustrations and the act of turning pages can satisfy.

Bear et al. (2012) explored developmental literacy instruction across three stages. In the emergent stage, students use word concepts to acquire sight words and early phonics skills. At the beginning stage students learn single syllable words using beginning consonants, digraphs, blends, short vowel families and consonant-vowel-consonant patterns. There is discussion that during the beginning stage, the way the sounds are made in the mouth gives readers important information about phonics and that by hearing and articulating the sounds of English, students acquire decoding skills more quickly. Finally, the transitional stage sees students reading in a way that sounds like language. It is accurate and fluent, and vowel patterns do not interfere, allowing expression to inform comprehension because the reading is automatic and prosaic. In this view of learning to read, however, students do not read for meaning until the transitional stage. This is contrary to what Dooley (2010) proposed and contributes to a lack of consensus regarding the timing of instructional ingredients.

Table 4 presents an assembly of the aforementioned instructional sequences.
Table 4

Timing of Instructional Ingredients for Reading Presented in Nine Studies Between 1976-2016

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-K</td>
<td>Pre-Reading (birth through 1st): Knows letters, words, books. Visual, visual-motor and auditory perceptual skills. Control over syntactic and semantic language. Has concepts of print.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early K</td>
<td>Stage 1 (grades 1 and 2): Initial reading, decoding</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mid K</td>
<td>Stage 2 (grades 2 and 3): Confirmation, fluency, ungluing from print</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Late K</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early 1st</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mid 1st</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Late 1st</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Early 2nd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mid 2nd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Late 2nd</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-alphabetic phase: limited to reading words from memory and guessing from context – reading the environment. Letters exert little influence on the guessed word</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partial-alphabetic phase: Use initial sounds, don’t have strong left-right reading orientation, starting to learn how to learn sight words, know consonants but not digraphs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-alphabetic phase: mastery required of phonemic awareness, decoding is painstaking, sight word vocabulary is growing, lots of text is being consumed</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consolidated-alphabetic phase: focus is on spelling patterns, chunking, growth of sight vocabulary</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automatic phase: highly developed automaticity and several strategies at disposal for identifying a word</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonemic Awareness: Pre-K – 3rd Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alphabet Recognition: K – 1st Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Print Concepts: K – 1st Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonics: K – 2nd Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-Frequency Words: K – 3rd Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specific Word Instruction: K – 4th Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Narrative Text Comprehension: K – 8th Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multi-Syllabic Words: 1st – 4th Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluency: 1st Grade and Above</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spelling: 1st Grade and Above</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word Learning Strategies: 2nd Grade and Above</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-K</td>
<td>Early K</td>
<td>Mid K</td>
<td>Late K</td>
<td>Early 1st</td>
<td>Mid 1st</td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>-------</td>
<td>--------</td>
<td>-----------</td>
<td>---------</td>
</tr>
<tr>
<td></td>
<td>Alphabet knowledge</td>
<td>Concept of word</td>
<td>Beginning consonant awareness</td>
<td>Spelling with beginning and ending consonants</td>
<td>Phoneme segmentation</td>
</tr>
<tr>
<td>Morris et al., 2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Receptive language</td>
<td>Phonological awareness</td>
<td>Morphological awareness</td>
<td>Morphophonological accuracy</td>
<td>Decoding</td>
</tr>
<tr>
<td>Jarmulowicz, 2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dooley, 2010</td>
<td><strong>Emergent Literacy 1:</strong> book as prop</td>
<td><strong>Emergent Literacy 2:</strong> book as invitation – attention to topic, images</td>
<td><strong>Emergent Literacy 3:</strong> book as script – images prompt script-sounding reading</td>
<td><strong>Emergent Literacy 4:</strong> book as text</td>
<td></td>
</tr>
<tr>
<td>Bear, 2012</td>
<td><strong>Emergent:</strong> Concept of word allows student to acquire sight words. Patterns and rhythms support this acquisition.</td>
<td><strong>Beginning:</strong> Just learning single syllable words, beginning consonants, digraphs, blends, short vowel families, CVC words</td>
<td><strong>Transitional:</strong> reading sounds like language – accurate, fluent, vowel patterns are learned, automatic and prosaic expression carries comprehension</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foorman et al., 2016</td>
<td><strong>Teach academic language skills:</strong> K – 3rd Grade. Conversation, narrative language skills, academic vocabulary in context of reading activities</td>
<td><strong>Develop awareness of segments of sounds in speech and how they map to letters:</strong> mid-K – 1st Grade</td>
<td><strong>Teach students to decode words, analyze word parts, write and recognize words:</strong> blend letter sounds and sound-spellings, instruct in sound-spellings, recognize common word parts, read decodable words, high frequency words, introduce non-decodable text</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pre-K</td>
<td>Early K</td>
<td>Mid K</td>
<td>Late K</td>
<td>Early 1st</td>
</tr>
<tr>
<td>--------</td>
<td>-------</td>
<td>---------</td>
<td>-------</td>
<td>--------</td>
<td>-----------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ensure student reads connected text every day to support accuracy, fluency and comprehension: model, scaffold, feedback, self-monitoring, practice and receive feedback.
**Inclusion and Timing of Taxonomic Elements of the IRLA**

Chapter 2 has so far presented a combined taxonomic structure as theoretical framework, probed the research base for empirical evidence on the strategic elements required for learning to read, and explored developmental timelines informing early reading instruction. The following section analyzes each element as it appears within the context of the IRLA. Findings to support or challenge the inclusion and timing of each element will be presented in Chapter 4. Table 5 is provided to remind the reader of the taxonomic structure organizing the concepts of reading element over time, and to report where each strategic content is explicitly presented in the IRLA, denoted with an x. Environmental attributes, strategic choices and process are employed throughout and are denoted with a -.
Table 5

Combined Taxonomic Structure Realized in the IRLA

<table>
<thead>
<tr>
<th></th>
<th>Pre-K</th>
<th>Kindergarten</th>
<th>1st Grade</th>
<th>2nd Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RTM</td>
<td>1Y</td>
<td>2Y</td>
<td>3Y</td>
</tr>
<tr>
<td>Environment/Uncontrollable Attribute</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehension of Complex Text</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Executive Function</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Strategic Content or Fixed Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Print Concepts</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Phonological Awareness</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Word Recognition</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Phonics</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Fluency</td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strategic Choices</td>
<td>Formative Assessment</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Process</td>
<td>One-to-One Instruction</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. Hyphens in the table represent elements that occur throughout the IRLA. xs indicate where an element of reading appears explicitly in a level of the IRLA.

**Comprehension of complex text in the IRLA.** To best prevent later reading comprehension issues, Kendeou et al. (2009) advised that comprehension be taught from the very beginning of learning to read. Comprehension is endemic to the IRLA—basic comprehension of text at each level is a required component for entry to that level, which underscores its place in the taxonomic structure as the environment serviced by interaction of and with the other reading skills.

At RTM, the pre-K level of the IRLA, students are expected to have enough background knowledge to name and talk about pictures common to beginning books. Included in the work done by Fielding and Pearson (1994), Perfetti et al. (2007), and
Connor et al. (2016), knowledge is an element the reader must bring to text if there is any hope that comprehension will occur. Kendeou et al. (2009) showed the benefits of starting early with comprehension of oral language and linking meaning to the development of decoding skills. The inclusion of background knowledge at RTM is in alignment with these studies.

The yellow levels, from the start through the middle of the kindergarten year, bring students through early foundational skills as shown throughout other sections. Comprehension is not an explicit entry requirement or point of practice of the IRLA at the yellow levels, but it is stressed that students make sense of yellow level books. At 3Y, when students are working to produce the initial sound of the unique word on the page (a decoding skill), meaningfully connecting the picture to the word and its beginning sound is emphasized by the teacher (Cunningham et al., 2004; Mesmer & Williams, 2012).

Within the IRLA, entry requirements for each level must be passed to ensure the student has amassed enough skill to take on the learning challenges of the level they are entering. From 1G on, the first entry requirement checked in every level is “Read and Comprehend Unfamiliar ___ Text”. To check for whether a student is ready to enter a level in the IRLA, the teacher has the student do a ‘cold read’ on a text at that level. For example, a student potentially ready to enter 1B will read and comprehend unfamiliar 1B text. A ‘cold read’ is a text the student has not read before at the level he or she is attempting to enter. Cold reads are calibrated to the text complexity appropriate to the level being entered (CCSS, Appendix A, 2010). The student reads the text aloud and the teacher listens for active reading strategies,
including “…read with purpose and understanding. Stop and try again if something doesn’t look right, sound right, or make sense. Self-correct ALL errors that interfere with meaning” (Hileman & Zorzi Cline, 2017, p. 29). The teacher presents general questions to determine whether the student has basic comprehension and can make inferences with that text. Connecting with the research on student motivation, four cold reads (two fiction and two non-fiction) are given in the IRLA for each level so the student has the opportunity to choose the text for him or herself (deNaeghel et al., 2012; Gambrell, 2015; Guthrie et al., 2000; Saeed & Zyngier, 2012). The basic comprehension question for fiction text is “What is happening so far in this story?”. The inferential question asks the student to consider something about the character and use evidence from the text to support the answer. For the non-fiction selections, the basic question asks the student to tell three things the text says about the topic, and the inferential question asks the student about some fact in the text and has them use evidence from the text to support the answer. Points are awarded for the strength of response on a scale from one to four, where one is weak and four is strong. Active reading strategies are scored ‘yes’ or ‘no’ with any ‘no’ response signaling that the student is not yet ready for the demands of text at that level.

At 1R, fluency is added to the expectations for active reading strategies, “read fluently and with expression, using punctuation.” If at any level, a student makes more than two uncorrected errors, does not self-correct all errors that interfere with meaning, or does not read fluently, entrance to that level is not granted. Notable in these requirements are the conditions. *More than two uncorrected errors* signals the teacher to listen for where the student gets hung up on phonic issues or sight words.
The skills from the prior levels must be strong enough to support the work at the incoming level, and the prompt cues the teacher to listen for foundational skills issues. *Does not self-correct all errors that interfere with meaning* prompts the teacher to listen to the semantic background a child brings to the text. If a student says the word *like* instead of the word *love*, meaning may not be impacted and the omission of a self-correction may be an indication that the student is relying more on meaning than on the phonic, or alphabetic principal. However, if the child substitutes the word *loaf* for *love* and does not pick up on the change in meaning, comprehension is implicated as the skill that needs work. The IRLA is used to assess readiness for reading text at the next level of complexity, listening to the child read aloud provides the teacher with information to help determine what next skill the student needs to learn (Afflerbach et al., 2008).

The IRLA is an independent reading level assessment. The identification of a color level for a student is dependent upon whether the student can read any book at that color level without support from the teacher. Independent does not indicate perfect, nor does it guarantee employment of deep comprehension strategies—yet. Independent means that a child can select any book at their identified color level and can apply appropriate reading skills to extract meaning from that text without feeling frustrated or defeated by the challenges presented by that color level. There will be challenges at that color level, though, and those challenges present themselves to the reader and the teacher as the next skill requiring instruction and practice.

Companion to the IRLA are books keyed to each color level identifying the collection of skills at that level. Because of the specificity of skills within each level,
books are carefully considered using the three parts of text complexity identified by CCSS Appendix A: quantitative, qualitative, and reader and task. Each book is analyzed quantitatively for word count and sentence length, provided by a search from Accelerated Reader and/or Lexile. This analysis is kept hidden from the people who do the qualitative review. The qualitative reviewers study each book against a backdrop of the skills denoted in the IRLA. They analyze the book for sight words and vocabulary, syllables and word families, frequency and type of vowel pairs and endings, irregularly spelled non-frequent words, and other skills spelled out in the IRLA and assign a color level. Two people do this analysis for each book, and if their decision about the level does not match, a third person analyzes it. With discussion, the color level is finally designated. Next, every book is evaluated for the appropriateness of audience. It is also evaluated to determine whether it would likely be selected by a student for wide-reading. If the topic is too content heavy or is not something that would appeal to a young reader, it is not included in the baskets of leveled books built by American Reading Company (ARC) intended to support the implementation of the IRLA (M. Lynch & M. Wiel, personal communication, July 31, 2018).

**Executive functioning in the IRLA.** The introduction of the IRLA states that the levels Read to Me (RTM) through 2Y are grounded in executive function & language. Quoting a chapter title from Sulla (2015), “Executive Function is Foundational for All Learning” (Hileman & Zorzi Cline, 2017, p. iv), the introduction includes that the RTM level represents the 2,000 hours of reading experiences typical of successful readiness, needed to provide groundwork for building reading skills.
While working memory (Baddeley, 2003), inhibition control (Blair & Razza, 2007; Foy & Mann, 2013), or cognitive flexibility (Jones et al., 2016) are not directly discussed, attention span has direct roots in executive function.

In the RTM level, foundational skills which are indicators of the elements of executive functioning are given for teachers to look for: “be attentive during group read-aloud” (Hileman & Zorzi Cline, 2017, p. 2) indicates inhibition control and working memory, “interactive language” (Hileman & Zorzi Cline, 2017, p. 2) indicates cognitive flexibility and inhibition control, and “story-making” indicates all three areas (Hileman & Zorzi Cline, 2017, p. 2). Not present are explicit descriptions of these elements of executive functioning.

There are three steps in the yellow level, 1Y, 2Y and 3Y. 1Y and 2Y are pre-reading levels, and 3Y introduces explicit phonic elements. The header for the yellow levels states that the big idea is executive function (Hileman & Zorzi Cline, 2017, p. 3). The learning focus states that children are sustaining concentration and monitoring comprehension, which are directly linked to the three elements of executive function (Jones et al., 2016). Students must remember a repeated sentence stem (working memory), read the main idea of the picture (inhibition control, especially if the student has something not related to the text to share), and say a word that matches the picture (cognitive flexibility). The expectation for children to be active and focused throughout the steps of this level is clear, and without strong executive functioning skills, this expectation may be difficult to meet. There is no further mention of executive function throughout the IRLA.
Print concepts in the IRLA. The print concepts strand stretches across the IRLA from RTM (pre-kindergarten) through 2B (last third of first grade). In RTM, the foundational skill “text awareness” provides a clear description of the behaviors that herald a development of the concept of print: “when ‘reading’ alone, point to the words and pretend to read them. Know that the printed words are what people read” (Hileman & Zorzi Cline, 2017, p. 2). Unpacking this description reveals several skills and awarenesses. “When ‘reading’ alone” (Hileman & Zorzi Cline, 2017, p. 2) invokes that a child understands what a book is and knows the outward behaviors of reading: looking at the book and engaging with it as something that holds meaning or has something to say (Meisels & Piker, 2001). “…point to the words” (IRLA, p. 2) assumes understanding the difference between print and pictures (Justice et al., 2006). “…pretend to read them” (IRLA, p. 2) assumes concept of narrative text and purpose of contextualized print (Clay 1979). “Book handling” (Hileman & Zorzi Cline, 2017, p. 2) is another RTM foundational skill, expecting children hold the book correctly and turn pages right to left. This print concept requires knowledge of the left to right organization of print (Justice et al., 2006).

At 1Y the IRLA is clear that children do not actually read any words. In the foundational skills for this level is the CCSS kindergarten foundational standard that students “follow words from left to right, top to bottom, and page by page” (Hileman & Zorzi Cline, 2017, p. 10). This skill is worth .01 points, directing teachers to check that students at 1Y understand the directional convention of print. 2Y increases this demand by ensuring that children understand that words are separated by spaces in print, and that they jump over the spaces between words (Mesmer & Williams, 2015).
Clearly calling out the spaces in words echoes Holden and MacGinitie’s (1972) finding that children need to be taught about spaces between words. 3Y requires that children recognize and name most upper- and lowercase letters of the alphabet. Also explicitly stated in CCSS, this skill takes concept of print to a smaller unit than the more broad concepts of books, pages, and words. Now, students are making the connection between the idea of the printed word and the speech sounds to the individual letters (Flanagan, 2007; Morris, 1993). The ability to connect words, speech sounds, and individual letters is mutually dependent upon the phonological and phonic skills demanded by 3Y.

In the IRLA, 1G is a level expected to begin in the sixth month of kindergarten. An entry requirement for 1G is the ability to make the primary or most frequent sound for each consonant, which is rooted in the phonological awareness (Ball & Blackman, 1991; Bradley & Bryant, 1983; Ehri & Wilch, 1987) and phonics (Henbest & Apel, 2017; Morris, 2014; Shapiro & Solity, 2016; Torgerson et al., 2006) literature. The denoted print concept at 1G is the first grade CCSS foundational skill “recognize and name all upper-and lower-case letters of the alphabet” (Hileman & Zorzi Cline, 2017, p. 35). Knowledge of and skill with the letters of the alphabet are separated into letter name (print concept) and letter sound (phonological awareness) in the IRLA, which is notable because this honors the research that there are different conceptual frameworks at play, both enacted by the symbols of the language. Also embedded in the print concept section of 1G is the first grade CCSS expecting recognition that spoken words are represented in written language by specific
sequences of letters (Ehri, 1998; Flanagan, 2007; Morris, 1993). Convergence of phonological awareness and phonics are placed in print concepts in the IRLA.

The level at which a typically developing student enters first grade is 2G. To enter this level, a student must be able to independently read and comprehend 2G text, which requires a foundational collection of known sight words, and the ability to navigate a book. Children read words and the spaces that bound them, they see initial consonants and can assign them sounds, they understand that text carries meaning, and can interact with it. The print concept presented in the IRLA at this level has to do with sentences: “Recognize and respond to punctuation (commas, periods, question marks) while reading” (Hileman & Zorzi Cline, 2017, p. 54). Instead of occurring within the context of fluency, this next element of print extends the convention of written language beyond the point of word and draws attention to other elements on the page. Invoking Justice et al.’s (2006) definition of print knowledge as extending to the rule-governed system of written language, the IRLA extends print knowledge beyond the simple orientation of a book as seen in earlier definitions of concepts of print (Clay, 1979) to include grammatical symbols denoting deeper text structures. Recognizing and responding to punctuation repeats as a strategic element in the 1B and 2B levels, rounding out the first-grade year. Print concepts are not presented in subsequent levels of the IRLA, which is in keeping with the research.

**Phonological awareness in the IRLA.** In the IRLA, phonological awareness makes implicit and explicit appearances. At the Read to Me (pre-kindergarten) level, “Active Listening to Read-Aloud,” and “Interactive Language” are given as foundational skills (Hileman & Zorzi Cline, 2017, p. 2). As seen in Bentin and
Leschem (1993), Clay (1991), and Iversen and Tunmer (1993), children who come from homes where reading and literacy is a fact of life come with more intact phonological awareness skills. By including active listening and interactive language as pre-reading foundational skills, the IRLA honors this knowledge.

At 1Y, students are expected to orally repeat a text pattern that they hear and see, and at 2Y students touch each word as they say them in the pattern. There is no explicit mention of the phonological processes taking place here, but the oral repetition of words makes a direct connection to the Foorman et al., (2016) call to bring students’ attention to words. Where students are touching the words as they say them in 2Y, they are assigning a word-sized phoneme to a word-sized symbol, and that process of mapping sound to sight is an element in many studies (Adams, 1990; Ball & Blachman, 1991; Clay, 1979; Henderson & Beers, 1980; Iversen & Tunmer, 1993, Ziegler & Goswami, 2005) used to show that sounds can easily be assigned to symbols. At 1Y and 2Y, the symbol being assigned a sound is a single-syllable sight word. Many of these single-syllable sight words are also known rimes, thereby taking advantage of Walton et al.’s (2001) study confirming rime analogy as a useful method for acquiring phonemic awareness at the same time as beginning to read.

The 3Y level requires acquisition of the sound-symbol relationship of the consonants. To enter this level, students must use the initial consonant of the unique word in the pattern (with the help of a picture on the same page) to cue the pronunciation of the word. Matching picture to initial consonant sound is a directly supported strategy from phonemic awareness research (Ball & Blachman, 1991; Bradley & Bryant, 1983; Ehri & Wilce, 1987). 3Y is where phonological awareness is
first explicitly denoted in the IRLA, and it is worth points. In the scoring system of the IRLA, a student can earn .02 points if they can “isolate and pronounce the initial sound of a spoken word” (IRLA, p. 10), a foundational skill pulled directly from the CCSS: F.K.2d. Within 3Y, students are to acquire the sounds of the consonants.

At 1G, the entry requirement is headed phonics, not phonological awareness, yet as the two are intertwined, this is where this phonemic skill resides. By mapping sounds to letters, students participate in the reciprocal relationship of the two skills. A student entering 1G is required to know at least 14 consonant sounds and to use them when trying to figure out an unfamiliar word in text. The phonological awareness component, while not an entry requirement, is explicit in the foundational skills section of 1G. Students can earn .01 point for orally producing groups of words that start with the same initial sound, a skill supported in the literature (Ball & Blachman, 1991; Hagans & Good, 2013; Hohn & Ehri, 1983; Lane, Pullen, Hudson, & Konold, 2009; Lewkowicz, 1980; Oudeans, 2003; Torgeson et al., 1992). Each with its corresponding standard from CCSS, five additional phonological elements are given here: recognize and produce rhyming words; count, pronounce, blend, and segment syllables in spoken words; blend and segment onsets and rimes of single-syllable spoken words; isolate and pronounce the initial, medial vowel, and final sounds (phonemes) in three-phoneme (CVC) words; and add or substitute individual sounds (phonemes) in simple, one-syllable words to make new words. None of these skills earns points within the 1G level. There is a guidance within the phonological awareness component providing direction and rationale: “The following Kindergarten standards are not essential for successful reading in either 1G or 2G, so they are not
scored or required here. However, they will be essential at the Blue levels, so they should be taught and practiced now in preparation for application to reading at 1B” (Hileman & Zorzi Cline, 2017, p. 36, emphasis in original text). This stance is directly related to the research showing that a granular knowledge of phonemic awareness is not essential for whole-word reading (Bentin & Leschem, 1993; Ehri, 1993; Ehri & Wilce, 1987; Glazzard, 2017; Goswami & Bryant, 2016). By advising the practice of rhyming words, blending, segmenting, isolating and substituting sounds, but not requiring mastery of them at this point of reading development, the IRLA brings teachers’ awareness to the phonemic competencies of students. Where children have pre-reading experience and rich exposure to language, phonological awareness will not require direct explicit teaching, but if it is found to be lacking, teaching it will provide the stability of knowledge required to break the code (Bentin & Leschem, 1993; Goswami & Bryant, 2016; Hohn & Ehri, 1983; Wise et al., 2000).

The phonological awareness entry requirement at 2G is also embedded in the phonics aspect of “self-prompt for initial blends and digraphs” (Hileman & Zorzi Cline, 2017, p. 46). As seen in the literature, the ability to blend two consonants to form a blended sound is an element of phonemic awareness (Lewkowicz, 1980). Students need to be able to say the sound for a minimum of 13 blends/digraphs upon entry, demonstrating they have acquired the concept of phonemic blending and its corresponding letter map. The skill of phonemic blending consolidates through 2G, and by the time the student is ready to move to the next level, they can “automatically say the sound of the blend or digraph while they look for meaning clues” (Hileman & Zorzi Cline, 2017, p. 54). It is important to note that while the IRLA requires the
acquisition of ever more granular phonological skills, meaning is explicitly emphasized.

Entry to 1B requires that students “combine initial sounds and Power Words to make new words” (Hileman & Zorzi Cline, 2017, p. 64), making use of rime analogy found to be successful for children with weak prereading skills by Walton et al. (2001). The rime analogy strategy employs a shared spelling sequence to predict a shared pronunciation, and is dependent on knowledge of onset-rime, a less granularly finite element than the phoneme level itself, but a phonological chunk of use to an early reader (Ziegler & Goswami, 2005). The assessment tools in the IRLA used to determine whether this skill is developing are nearly identical to tools described in the Walton et al. study. Students combine known words and letter sounds to make new words and to use words they know to figure out words they don’t know. Presented in zones, teachers can determine whether this skill is developed and to what degree of difficulty. The organization and presentation of zones enables teachers to determine what next element of rime analogy to develop. Within the 1B level, the discrete phonological awareness skills provided in the CCSS are given. They are not given high status with the inclusion of points to be earned, but they are indicated for teachers to check and teach if discovered to not yet be in-tact.

Phonological awareness is also checked in the IRLA at the 2B level; once again there is no phonological entry requirement, and within the level the foundational skills are a repeat of those from the 1B level. The 1R level includes “orally segment a multisyllable word into its syllables” as a final explicitly named phonological skill. As students move through the IRLA from pre-reading through kindergarten and first
grade, they actively turn sounds from the language they hear spoken and have mastered orally into the symbols of the same language written down, or represented in code (Clay, 1991; Foorman, et al., 2016).

**Word recognition in the IRLA.** Word Recognition is an expected element in all levels from 1G through 1R in the IRLA. Students cannot enter 1G, which begins at the sixth month of kindergarten, unless they are able to “read at least 25 high-frequency words by sight (out of context)” (Hileman & Zorzi Cline, 2017, p. 28). At kindergarten, children at 1Y and 2Y are pre-alphabetic, but growing toward understanding letters and the sounds they represent. At 3Y, children learn the most frequent sounds for letters, which we see in the phonological awareness and phonics sections of this paper, and they have gained print concepts to the level of understanding the boundaries of words. Supported by Ehri’s (2005) work, learning word units as wholes is entirely appropriate at this stage and will support further mapping of word parts phonologically to their appropriate graphemic representations.

At the 1G level, there is a list comprised of 60 common high-frequency words, called “Power Words”. Specifying that the words were selected using lists provided by Dolch (1936), Fry (2000), Johns (1997), Pinnell and Fountas (1998) and Zeno, Ivens, Millard, and Duvvuri (1995), the list of 60 power words are presented alphabetically. A comparison of this list to the Dolch list reveals three words in IRLA not on the Dolch list: can’t, lots, and love. To enter the level, students must know 25 at flash speed. Through the level, a student is expected to learn all 60 words. An additional 25 power words are needed to gain entry to 2G.
The 2G power word list is again made up of another 60 words. Before a student is finished with 2G, at the end of the third month of first grade, he or she will know at least 120 common high-frequency words at flash speed. Flash speed is not defined in the IRLA—it is not dependent upon naming a certain number of words in a given length of time. Rather, flash speed is an expectation of automaticity (LeBerge & Samuels, 1974; Lonigan & Shanahan, 2009). If a word is known and declared, then, as in the Perfetti and Hogaboam (1975) paper, the student is only engaging in a single process, indicating the word is known rather than needing to engage the code breaking process in addition.

1B and 2B take place over the second and third thirds of the first-grade year. These levels also have entry requirements demanding skill in automatic word recognition, but these lists call the words Tricky Words. Students must recognize them by sight (Clark, 2016; Watts & Gardner, 2013). Tricky words are those that are high utility, but whose decoding demands do not yet meet students’ known abilities. For instance, at 1B, students have not yet learned to decode two-syllable regular words, yet some of these words are high-use enough that students should be able to read them by sight: any, myself, never, something, always. Other words have letter pairings that, while phonetically regular, haven’t yet been learned by students: please, laugh, friend.

2B readers are working to learn the regular phonetic rules for double syllable words. They are getting more comfortable seeing longer strings of letters together as they practice compound words and words with inflectional endings, but there are words that may stop them in their tracks if they do not develop a knowledge of them
by sight. The 75 words given here are tricky for a wide variety of reasons but will appear in books at this level: thought, guess, beautiful, noise, sure, and often are among them.

At the 1R level, students are consolidating their phonics knowledge and are reading longer books. The words they must prove sight mastery of are generally regularly spelled but have infrequently occurring pairings. Included at this level are abbreviations such as Mr. and Mrs. Also included in this last list of sight words are tough, city, giant, quarter, and o’clock.

The research base does not provide a pure quantitative number of words that students should recognize by sight, nor is there evidence of a particular chronology of when certain words should be learned. What unifies the research is the idea of learning words by sight, particularly those of high utility and challenging orthography (Ehri, 2005; Ehri et al., 2001; Ehri & Saltmarsh, 1995). By including sight words from kindergarten, the IRLA leans into the research base, and eventually merges the research-grounded work about rapid naming, automaticity and sight words with an analytic phonics approach as students are coached to ‘find a word you know in a word you don’t’ to be able to read words automatically and smoothly in the pursuit of making meaning (Rasinski, 2013; Schwanenflugel et al., 2007).

**Phonics in the IRLA.** Based on Ziegler and Goswami’s (2005) psycholinguistic grain-size theory, the process of matching distinctive visual symbols to units of sound is at play in the yellow levels of the IRLA. The process of learning and applying mappings, otherwise known as phonological recoding, requires shared grain sizes in the orthography (symbol system) and phonology (sound system) of the
language. By uttering the single-syllable words of a pattern in a yellow level book at 1Y, students are activating their already-structured phonological system (Bloomfield, 1942; Gough & Hillinger, 1980; Reed, 2001). By distinguishing words separated by spaces one-by-one at 2Y, children are engaged in the reciprocal relationship between phonology and orthography and are building neural networks linking the two. The grain size at the yellow level is word sized. Once children have built a pathway for learning whole words, the acquisition of additional whole words gains speed. Rapid automatic naming of words is further considered in the word recognition section of this paper. Simultaneously, phonemes of smaller grain size are required to be orthographically mapped in to support meeting the consonant sound entry requirement for 3Y.

Skills to be gained in 3Y capitalize on the intersection of phonology and orthography, which is where phonics traditionally begins (Ehri et al., 2001). Students are to produce the primary or most frequent sound for most consonants and recognize and name most upper- and lowercase letters of the alphabet. These skills are often seen at the beginnings of systematic programs (Shapiro & Solity, 2016; Wright & Jacobs, 2003).

Distinguishing the phonic demand at the 3Y level from synthetic approaches is the expectation that meaning extracted from illustrations matching the text prompts students to produce the correct initial sound for the unique word in a pattern. The teacher is coached to watch “the mouths and lips of students to see if they are in the correct position for the first letter of the unknown word as they scan the picture for clues” (Hileman & Zorzi Cline, 2017, 2016 p. 19). Reliance on the convergence of
meaning, grapheme, motor skill, and phoneme to propel students through the initial mapping of sounds helps teachers recognize where students have skills upon which to build.

The phonic requirement at the 1G level is a consolidation of initial consonant skills. By the time the student is ready to exit 1G, he or she has all initial consonants phonologically mapped by sight and has moved on to knowing the spelling-sound correspondences for common consonant digraphs and blends. There are 22 blends and four digraphs for students to read out of context to demonstrate this ability. Entrance to 2G is granted if a student can read 13 of these—the remainders are to be gained within this level. Students are expected to enter first grade having mastered the yellow levels and 1G. This is in line with some developmental constructs found in the literature (Chall, 1976; Ehri & McCormick, 1998; Morris et al., 2003; NRP, 2000) and slightly faster than some others (Bear et al., 2012; Dooley, 2010; Foorman et al., 2016).

As with all levels, entrance to one is equivalent to the exit of the one preceding it, indicating mastery. The start of first grade is in line with the start of 2G, and the phonic skill that must be at least developing by this time is blends and digraphs, as explained above in 1G. Throughout the 2G level, students master the use of “all initial consonant blends and digraphs as clues to unfamiliar words without prompting. When they come to an unfamiliar word, students should automatically say the sound of the blends or digraphs while they look for meaning clues” (Hileman & Zorzi Cline, 2017, 2016, p. 54). Teachers are directed to not have students sound out the rest of the word at this time. Phonics entry requirements at 1B, projected to be met by a typically
developing first grade student in the third month are: manipulate onsets and rimes and decode most one-syllable words. Previously discussed in the section in this paper on phonological awareness and word recognition, the entry checks for 1B require students to “use words you know to make new words,” and to “use words you know to figure out words you don’t” (Hileman & Zorzi Cline, 2017, p. 70-71). Models for these assessments appear in the Walton et al. (2001) study and are discussed in the Savage et al. (2003) study comparing synthetic and analytic approaches. Within 1B, teachers are to ensure that students can decode regular one-syllable words, words containing VCe and common vowel teams representing long vowel sounds, words beginning with silent consonant pairs (e.g. wr-, kn-, ph-, qu-, wh-), final blends and digraphs (e.g. -sh, -st, -ng, -lp), and r-controlled vowels. These elements are signatures of systematically teaching phonics (Ehri et al., 2001; Foorman et al, 2016; Torgerson et al., 2006). There are two additional pieces of guidance at the 1B level: “Cover parts of unfamiliar words with finger and look for familiar chunks inside” and “use familiar rhyming words to decode unfamiliar words” (Hileman & Zorzi Cline, 2017, p. 74). These directions emphasize a whole to part approach to decoding words, placing the phonics element of the IRLA in the analytic phonics camp (Clark, 2016; Ehri et al., 2001; Glazzard, 2017; Henbest & Apel, 2017; Torgerson et al., 2006).

The last third of the first-grade year is expected to be spent at the 2B level. To enter 2B, students must be able to decode most two-syllable words and words with inflectional endings. This continues reliance on using what you know to figure out what you don’t’ (Schwanenflugel et al., 2007). Students who have consolidated their ability to read regularly spelled single syllable words through analogy, or larger grain-
sized phonemes, are expected to be flexible in their word-solving skill to bridge syllables (Goswami & Bryant, 2016). The entry requirement requires a student be able to show enough skill to name 20 of the 50 words on the page, including compound words common to a child’s lexicon: teammate, raincoat, seesaw, daytime, etc., and inflectional endings -ed, -s, -ing and -y on commonly known words: eating, rainy, cars, played, etc. (Schwanenflugel et al., 2007). Once entered to 2B, children work to grow the scope of these skills. Finding known chunks in words to solve a whole two-syllable word, using syllable and consonant patterns to break long words into their syllables, decoding compound words and recognizing words with inflectional endings are all skills grown and mastered at this level. This continues the explicit, structured, developmental progress through the gradually more challenging elements of the phonic code (Henbest & Apel, 2017).

Second grade begins with 1R. The phonics component required for entry to this level is the ability to decode most regular three-syllable words. Students are checked on a phonics survey arranged by complex, though still regular, phonics rules: three-letter blends and trigraphs, endings/suffixes, vowel teams, two-syllable long vowels and three-syllable regular words. There are 60 words divided by skill, and a student need only read 20 to meet the phonics requirement for this level. The list of phonics rules above foretell the skills requiring consolidation in the level. Cover parts of unfamiliar words with finger and look for familiar chunks inside, the read by analogy skill, continues to be provided as a go-to prompt for the teacher. Students also apply some flexibility to words, decoding regularly spelled two-syllable words
with long vowels, and distinguishing long and short vowels when reading regularly spelled one-syllable words.

2R brings the expectation that the code is broken by the end of second grade. The phonic entry requirement is “decode almost any word familiar from everyday speech” (Hileman & Zorzi Cline, 2017, p. 126). The assessment teachers use is a tricky phonics check, and it requires that students “try various vowel/consonant sounds in decoding unfamiliar words until they recognize the word from everyday speech” (Hileman & Zorzi Cline, 2017p. 132). Fourteen phonic elements are laid out with five words each for students to read. There are 70 words on this page, but students need only 20 to enter this level. Where the teacher notes a pattern of errors (ex. y says long i), she knows to teach and provide practice with the discrete rule to help the student learn and consolidate this skill. Within the level, advanced phonics skills are checked and instructed where need arises. The tricky phonics skills emphasize flexibility with letter sounds. In the meta-study conducted by Ehri et al. (2001), there were indications that students who learned phonics from whole to part acquired a heightened ability to flexibly solve difficult words. The 2R level marinates these skills, enabling students to enter third grade without having to stumble over words.

**Vocabulary in the IRLA.** The IRLA is a pre-k through grade 12 tool. Vocabulary plays a significant role from grades 3-12, or the levels white and above. Noticing words and using context to learn them is a fundamental skill that enters at the white level, the ability to use morphemes, or word parts enters in the black level, grade 4, as does figurative language, word relationships, and nuance in word meaning
In addition to these discrete skills, the expected size of a students’ vocabulary increases by 1,500 words per year, grounded in Nagy and Anderson’s (1984) work. At the levels under investigation in this study, vocabulary makes only three explicit appearances, at 1Y, 2G, and 2R.

1Y states “students who do not know basic English vocabulary need extensive read-aloud, English language instruction, and/or real-life experiences” (Hileman & Zorzi Cline, 2017, p. 10). While not a list of words or morphemes for students to demonstrate mastery over, the language at this first level of kindergarten asks teachers to realize the spoken vocabulary capability of a student. For students just learning English, this is a way to help bridge the gap between what foundational skills may be in place but enables the teacher to understand a crucial gap in starting to make meaning from written English.

At 2G, students are heavily involved in learning sight words; by the end of 2G, students are expected to have an automatic sight word vocabulary of 120 words. Additionally, they are learning category words by sight: number words (one, two, three…); days of the week; colors; family words (mom, uncle, baby…); contractions; shapes; and directions (over, under, around…). The explicit vocabulary demand at 2G is “use and explain common antonyms frequently found in 2G-level books. Say one of the words below and ask the student to tell you the opposite. They don’t need to be able to read the word, just give its opposite orally” (IRLA, p. 55). Reflective of Clay (1991), texts that use the child’s own language support independence. The sight words chosen for the green levels, as discussed in the word recognition section of this paper, were chosen according to two metrics: frequency and familiarity to children’s
language or vocabulary. A native English speaking, late-kindergarten or early first
grade student can be expected to have many everyday words seen in the categories
presented above (Beck, McKeown & Kucan, 2002).

Including antonyms at this level and expecting them to be understood, not read
by the student, is an indication of a concept seen throughout the literature of checking
in to make sure that a skill is developing naturally (Apel & Henbest, 2016; Bentin &
Leschem, 1993; Flanagan, 2007; Stahl & Fairbanks, 1986). If the skill is not
developing, the literature advises that teachers notice it on time and intervene with
instruction that will bring the skill along. This is the spirit in which the IRLA has
been developed, and the inclusion of checking in on the language skill of thinking of
and naming antonyms is a clear example. At the start of 1st grade, a typically
developing native English-speaking child should have the opposites for big, over, old,
up, empty, near, happy, good, tall, and day in their personal vocabulary. A teacher
who learns that they do not is positioned well to intervene with instruction targeted to
the child and the skill.

From the end of the first third of first grade to the middle of second grade,
there are no explicit vocabulary demands in the IRLA. The skills in the blue levels,
1.3-1.99, focus on analytic phonics, solving big words based on known little words, or
word parts inside. Although emphasis throughout all levels in the IRLA is on making
meaning, attention at the blue levels is not on the meaning of individual words. If a
word doesn’t make sense, the student is expected to wrestle with it until it does. Using
initial sounds, analogies, word parts, and meaning, a student is engaged in a tussle
with meaning at both the vocabulary and comprehension levels. Using Beck et al.’s
framing of word complexity, the words students work with at grades K, 1, and most of 2 in the IRLA are tier 1 words, i.e. words of the most basic meaning that rarely require instructional attention to their meanings in school.

At 2R, another checkpoint for tier 1 vocabulary, students are expected to understand common abbreviations (e.g., Mr., St., Dec., Mon.), and have a working knowledge of prefixes and suffixes. At this point in the IRLA, word recognition and vocabulary are included under the same header, hinting that these are compatible terms, and the direction given for prefixes and suffixes is “decode words with common prefixes and suffixes.” (Hileman & Zorzi Cline, 2017, p. 135). The list contains three prefixes (un-, re-, mis-) and six suffixes (-ful, -able, -tion, -ly, -er/-ier, and -iest). While the focus is on decoding, this is a morphological task focused on both derivational and inflectional work, albeit at a beginning level.

Fluency in the IRLA. Fluency appears in the IRLA at 1G, the sixth month of kindergarten. Worth .03 points, students are expected to “read unfamiliar 1G-level text independently, sustaining concentration, monitoring comprehension, with 95-100% word accuracy, and when useful, re-reading” (Hileman & Zorzi Cline, 2017, p. 37). The research base claims that 1<sup>st</sup> or 2<sup>nd</sup> grade is ideal for fluency (Bear, 2012; Chall, 1976; Foorman et al., 2016; Jarmulowicz, 2007; Morris, 2003), yet the IRLA expects it from kindergarten.

Unpacking the expectation, there are nods to the research base. “Read unfamiliar 1G-level text independently…” (Hileman & Zorzi Cline, 2017, p. 37). In the repeated reading literature (Begeny et al., 2010; Martens et al., 2007; Samuels, 1979), students are confronted with text they have never seen before, but then read it
again and again with the explicit goal of increasing the words read per minute. The first expectation in the IRLA’s fluency direction—unfamiliar—does not attach itself directly to the repeated reading literature, but there is a twist. From the section in this paper on leveled texts, the reader will know that a 1G text is patterned and that the pattern is made up of sight words. If a 1G reader is properly placed, then the recurrence of the sight words can be argued to provide repeated reading opportunities. Additionally, the unique word in the pattern is directly matched to the illustration, heightening its imageability (Mesmer et al., 2012), making it comprehensible, and the initial sound in the unique word is known by a child operating at this level. In this way, fluency is linked to making the first phonics skill, initial sound, automatic.

Continuing through the IRLA’s expectations for fluency, “…sustaining concentration…” (Hileman & Zorzi Cline, 2017, p. 37) aligns with LeBerge and Samuels’ (1974) theory of processing capacity, which is repeated throughout the fluency literature. Time and again researchers declare that processes of decoding must be in place for fluency to develop (Clark, 2010; Ehri et al., 2001; Landerl, 2000; Savage et al., 2003). The IRLA does not wait for a single skill before introducing and growing the next skill. Instead, it directs teachers to automatize each skill as it grows. The demand of fluent reading follows this pattern throughout the IRLA. Each level introduces the next appropriate step for each skill (phonological awareness, phonics, word recognition, and word meaning), and then demands that it be consolidated through the practices of fluency and comprehension. “…monitoring comprehension…” and “…when useful, rereading” (Hileman & Zorzi Cline, 2017, p. 37) continue along this vein. The IRLA honors research asserting that text that is
meaningful to a child will be comprehended from the moment a child is aware there is meaning to be had (Fulmer & Frijters, 2011; Gambrell, 2015; Kuhn et al., 2006). Rather than withholding the meaning of books, or demanding they be re-read for increased number of words per minute, the IRLA provides students with just the next skill to just the next degree and then allows for consolidation while pushing forward with purposeful, attainable practice.

At each level, 95%-100% accuracy is required. O’Connor et al. (2010) tested their fluency intervention on students in two groups: one was assigned fluency work with text in which they were 80%-90% accurate and the other was assigned fluency work with text in which they were 91%-100% accurate. Though O’Connor et al.’s findings did not differentiate between the difficulty levels, students in both groups grew significantly. They called for further study but given that the intervention took place in a one-on-one environment, it is safe to say that students’ frustration levels were never reached because their tutor was instructed to provide support when the student got stuck. The IRLA is rooted in independent reading levels, and as such, it is expected that students choose books independently, based on interest, and be faced with challenges they know they are able to meet. An independent 1G reader is not reading Don Quixote but is absolutely able to read 1G text fluently and with comprehension.

**Formative assessment in the IRLA.** Black and Wiliam (2009) studied formative assessment by looking at formative interactions within more comprehensive theories of pedagogy. They sought to provide a unifying basis for formative practices and proposed a theoretical frame for studying classrooms using formative assessments.
Arriving at the conclusion that a formative interaction is one in which an interactive situation influences cognition, they proposed an initiation-response-evaluation structure for formative assessments, based on theoretical work by Mehan (1979).

Sadler (1989) enumerated three key points necessary in formative assessment: know what the standard (or goal) looks like, compare student performance against that standard, and engage in action in partnership with the learner to close the gap. Stiggins (2002) added that through this process assessment can build student confidence—adding to student motivation—that students can learn to self-assess where there is a clear goal, and that teachers can adjust instructional tactics in the moment, providing efficiencies in the teaching and learning process.

The IRLA relies on the formative assessment processes outlined by the CCSSO (2018): goal setting, eliciting evidence of knowledge or skill, self-assessment and actionable feedback to positively impact student learning. The following excerpt from a page in the introduction of the IRLA provides teachers with a clear vision for how formative assessment is realized in the IRLA:

The teacher watches the student at work. She analyzes his actions and his thinking about those actions and provides on-the-spot, targeted, and immediately useful feedback. The student continues working, applying the teacher’s feedback to his work. The teacher watches to assess the extent to which the student has improved, and the process repeats. Incremental, continuous student learning is supported by incremental, continuous teacher learning. (IRLA, 2017, p. vii)
One-to-one instruction in the IRLA. The IRLA is designed to support
teachers as they navigate teaching reading in grades Pre-K to 12. Consistent with Clay
(1991), who said that a teacher needs to have an overview of changes to watch for
along the way, the IRLA functions simultaneously as an assessment tool and an
instructional guide and is designed to be used in one-to-one instruction. The
introductory pages of the IRLA provide guidance about coaching, conferencing, and
building relationships. Student conferences are the mode of instructional delivery, and
conferences are intended to be “formative assessment where coaching and assessment
are simultaneous” (Hileman & Zorzi Cline, 2018, p. x).

Summary

Learning how to read is not a simple process. Teaching beginning readers is
likewise not a simple process. The ingredients for reading: print concepts,
phonological awareness, word recognition, phonics, fluency, and vocabulary are
marshalled by the executive functioning system of the learner in service of making
meaning or comprehending text. Instructors must understand—for each child—which
of the ingredients are firmly in place and which need further development according to
known developmental sequences. What is known and presented through the research
reviewed in Chapter 2 are the elements named in the IRLA required for readers to
decipher and comprehend text of ever-increasing complexity, and the most appropriate
developmental sequencing for each. Using a theoretical framework that combines the
taxonomical structures of Hambrick (1984) and Travers (1980), the wide variety of
elements and timing required for learning to read can be organized into a clear,
comprehensive frame that directs use of formative assessment in one-to-one
conferences organized by instructional ingredient and timing anchored in research. The realization and employment of such a tool could have a profound impact of the teaching of reading to beginners. The two-fold purpose of this study is to fill the gap in the research on the IRLA by determining whether the inclusion and timing of each element presented in the foundational levels of the IRLA is appropriate and justified and to discover whether learning informed by the IRLA yields increased results on third grade summative assessments.
Chapter 3: Methodology

Chapter 3 states the purpose of this research and describes the Independent Reading Level Assessment (IRLA). Methodology used to examine the elements for learning to read within the IRLA and to discover whether learning informed by the IRLA yields increased results on third grade reading achievement will be presented and rationalized. Chapter 3 also includes information on the sampling strategy and participation, information about the instrumentation, disclosure statement, and description of the data analysis.

Research Questions

The purpose of this research was two-fold: to determine whether the inclusion and timing of each element presented in the foundational levels of the IRLA was appropriate and justified, and to discover whether learning informed by the IRLA yielded increased results on third grade summative assessments. Chapter 2 presented a combined taxonomic structure as the theoretical framework for organizing the elements required for learning to read, including reading elements and developmental timelines.

Independent Reading Level Assessment

The Independent Reading Level Assessment, or IRLA (Hileman & Zorzi Cline, 2017), operates as a criterion-based assessment when fully employed by a teacher. It relies on the reading research base and includes every reading CCSS as well as those language standards key to reading success. According to the authors, teachers work to identify what skills a student possesses, and what skills need to be gained next. The level, or classification representing a collection of skills, denotes the
most complex text that a student can independently read and comprehend without any support from a teacher. Each level has entry requirements which must be passed by the student to gain entry to the level. Within each level are skills appropriate to the developmental sequence that can be practiced with text identified based on its complexity as being at that level, and it is the work of the teacher—guided by the IRLA—to identify which skills need practice. When a teacher can no longer identify skills within the level that the student needs to gain mastery of, the teacher checks the student for the entry requirements at the next level. This check is triggered by teacher observation of the reading behaviors and skills displayed by the student, not by a quantitative sum. The structure of this tool, especially employment of the entry requirements, ensures that skills develop in concert with one another rather than allowing one skill, say phonics, to outpace another, like vocabulary, resulting in students who can name any word in print but are hardly able to extract meaning from connected text.

The authors go on to say that the use of the IRLA is characterized by a teacher working with a student through regular formative assessment conferencing. In a conference, the student brings a book from the level he is working within and talks to the teacher about what he has been reading. Prompted or questioned by the teacher, the student reads a short passage from his book aloud. The teacher listens, surveying the skills listed for that level in the IRLA and homes in on errors, hesitations, or misunderstandings. Through conversation with the student, a single skill needing work is singled out and the teacher provides coaching and modeling about that skill. The student practices with the teacher, they set a goal for further practice, and the
student resumes reading independently. To facilitate this work in the classroom, at least 30 minutes each day are set aside for independent reading in class, during which students practice their goal from self-selected authentic books at their color level. A single level, 1 Blue, is presented here in service of representing all levels.

**Exploration of 1B.** Every level follows the same basic structure: entry requirement overview, entry requirement assessments (cold read and comprehension check and foundational skill or skills check), and tables corresponding to the elements of reading at the developmental stage of that level. These tables are followed by the CCSS for the grade containing the level with sample question stems for teachers to reference as they ensure the standards are met.

To identify the highest level at which a student can basically and independently read and comprehend text, a student must demonstrate mastery of the entry requirements for that level. Using a phonics infrastructure as an initial indicator, the teacher selects a ‘Cold Read’, a short text within the indicated level of the IRLA for the student to read aloud and discuss. When a student can read and comprehend unfamiliar text at that level, the remainder of the entry requirements are assessed and must be passed with a stated degree of surety for a student to enter a level. At 1B, for example, the entry requirements fall into three categories, displayed in Table 6. Word lists unique to the level are provided in the IRLA for each element to be assessed.
The taxonomic structure of the IRLA as discussed in Chapters 1 and 2 presented, according to Hambrick’s (1984) structure, a grouping of environmental and strategic aspects enabling each skill to strand out according to the known developmental sequence of each skill. Additionally, Travers’ (1980) classification scheme was given as present in the IRLA, demanding isolated categories. Travers’ structure is realized in the IRLA as color levels, each containing a slice of the developmental sequence of each skill. As seen above, entry to a level is assessed to ensure a student has all necessary skills to independently and basically comprehend text at that level. Once a student has entered, the work to be done within that level is presented according to the elements of reading. A single slice, the level 1B, is presented in Table 7 below to illustrate this concept.
### Table 7

**1B Foundational Skills**

<table>
<thead>
<tr>
<th>Environment/ Uncontrollable Attribute</th>
<th>Comprehension of Complex Text</th>
<th>Executive Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 Blue</strong></td>
<td><strong>Engagement and Independence</strong>: Read regularly and independently, sustaining engagement, in 1B-level materials for at least 30 minutes every day in the classroom.</td>
<td><strong>Home Reading</strong>: Have established a home reading habit and read for at least 30 minutes every night without prompting.</td>
</tr>
<tr>
<td><strong>Genres</strong>: Choose to read both informational and literary text at the 1B level.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strategic Content or Fixed Variables</th>
<th>Print Concepts</th>
<th>Phonological Awareness</th>
<th>Word Recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sentences</strong>: Recognize and respond to punctuation (commas, periods, question marks) while reading.</td>
<td>* Distinguish long from short vowel sounds in spoken single-syllable words.</td>
<td>* Orally produce single-syllable words by blending sounds (phonemes) including consonant blends.</td>
<td><strong>Power Words</strong>: Know 100-300 high-frequency words and use them as &quot;islands of certainty&quot; while reading in 1B-level materials.</td>
</tr>
<tr>
<td>* Isolate and pronounce initial, medial vowel, and final sounds (phonemes) in spoken single-syllable words.</td>
<td>* Segment spoken single-syllable words into their complete sequence of individual sounds (phonemes).</td>
<td><strong>1-Syllable Key Words</strong>: Read (25) common 1-Syllable Key Words by sight. Once memorized, each of these Key Words can be used as a chunk to decode words in 25 of the most useful word families.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phonics</th>
<th><strong>Cover parts of unfamiliar words with finger and look for familiar chunks inside.</strong></th>
<th>* Use familiar rhyming words to decode unfamiliar words.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>* Decode almost any one-syllable word that follows a regular vowel pattern (<em>hid/hide</em>) or is built from a familiar chunk (<em>lift</em>).</td>
<td>* Decode words containing -e and common vowel team conventions for representing long vowel sounds.</td>
</tr>
<tr>
<td></td>
<td>* Decode words beginning with these sounds: wr-, kn-, ph-, qu-, wh-.</td>
<td>* Use final blends and digraphs to figure out unfamiliar words (<em>desk, wish</em>).</td>
</tr>
<tr>
<td></td>
<td>* Use final blends and digraphs to figure out unfamiliar words (<em>desk, wish</em>).</td>
<td>* Decode words containing “r”-controlled vowels.</td>
</tr>
</tbody>
</table>

| Fluency | * Read 1B text comfortably, with confidence, purpose, and understanding. | * Read 1B text orally with 95-100% accuracy, appropriate rate, and expression on initial readings. | * Read in whisper voice. |
The sequence of skill instruction within a level is not prescribed, instead, consideration is given for the connected interaction of skills as seen in the literature review. Consistent with Clay (1991), who said that a teacher needs to have an overview of changes to watch for along the way, the IRLA functions simultaneously as an assessment tool and an instructional guide and is designed to be used in one-to-one instruction. The Formative Assessment Protocol leads teachers through a series of questions intended to be internalized and used in natural conversation with students as skills are probed in service of determining what the next skill requiring instruction and/or development may be. The protocol is shown in Table 8 (Hileman & Zorzi Cline, 2018, p. xi).
### Table 8

**Formative Assessment Protocol**

<table>
<thead>
<tr>
<th>Question for the Teacher</th>
<th>Rationale/Clarity</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Is this level easy enough for this student?</strong></td>
<td>Can the student read the words and ideas fluently and problem-solve 99% of challenges without teacher help of any kind?</td>
<td>No. Stop and re-focus the conference on identifying the student’s correct level. Yes. Continue to work on identifying the student’s Power Goal.</td>
</tr>
<tr>
<td><strong>Is this student an engaged reader?</strong></td>
<td>Does this student read at home? Regularly finish books? Read for his/her own reasons, not just because school assigns it?</td>
<td>No. Stop and make an action plan. Yes. Continue to work on identifying the student’s Power Goal.</td>
</tr>
<tr>
<td><strong>Where should I coach this student?</strong></td>
<td>Does this student need more coaching in this level or is s/he ready for coaching towards the next level?</td>
<td>This level. Next level.</td>
</tr>
</tbody>
</table>

**Look at the IRLA**

a) Entry Requirements

b) High-Point Values

c) Transition/Exit Requirements

Ask the student what s/he thinks s/he should work on.

**Student Power Goal:**

Make sure the student can say what s/he will learn/do, why, and how s/he will know when it is accomplished.

**What next?**

- Teach now
- Assign to Small Group with others who need this same PG.
- Identify a way the student can work on PG without me and set a date when I will check his/her progress

**Action Plan:**

Guidance given to teachers through the *Formative Assessment Protocol* functions within the taxonomic structure. As process, one-to-one instruction is

employed to marshal the strategic content in service of the environment. The cues
given to the teacher adhere to this structure as well, harnessing the power of the
organizational device. By asking *Is this level easy enough for the student?* and
clarifying with a check for fluency, meaning making, and problem-solving, the
protocol directs the teacher to consider the environment of comprehension as laid out
by Snow (2002), complete with consideration of the skills the reader brings to the text,
the text and its demands, and the activity done by the reader with the text. Next, the
question about reader engagement checks on the second environmental construct—the
reader. Engagement plays a significant role in the IRLA. Governed by executive
functioning (working memory, inhibitory control, and cognitive flexibility), the active
focus of engagement must be attainable by a student before skills can be acquired or
practiced. Where the protocol guides teachers to learn about the degree of
engagement of a student, it is gauging the environment of the individual to ensure it
has the capacity to proceed. If not, stopping to make an action plan focusing on
gaining engagement is required.

The protocol refers to color level, where it prompts the teacher to consider
where to coach the student. Given the structure of the IRLA, it is possible that a
student has gained all of the skills in one level, but not to the degree of mastery
required to gain entry to the next level. The teacher must be facile enough with the
collection of skills at each level to determine where to coach the student. Notable in
the protocol is the direction to ask the student. Independence is an overall goal of the
IRLA, denoted in the title. To be independent is to make decisions in one’s own best
interest. Occasionally a teacher will listen to a student read and two or more next
steps will be evident. The IRLA coaches teachers to choose the single highest leverage next step, arranged in the tool itself as entry requirements and skills with high-point values. When multiple skills meet those cues, presenting them to the student yields purposeful energy that can propel students forward. Lastly on the protocol is the question, “What next?”. A goal will have been identified, and depending on the nature of the goal and the student’s understanding of it, the teacher must decide whether to teach it in the moment, defer its instruction until a small group meeting where other students with the same goal can come together (moving the one-to-one nature to an individualized one as defined by Frey (2006)), or to find a way for the student to practice the goal independently with a timeline for accountability.

**Rationale for Methodology**

The purpose of this research was two-fold: to determine whether the inclusion and timing of each element presented in the foundational levels of the IRLA was appropriate and justified, and to discover whether learning informed by the IRLA yielded increased results on third grade summative assessments. Answering the first purpose required a careful look at the content validity of the IRLA. Warner (2008) discussed content validity as the degree to which the content of a tool matches the domain of material. To determine whether the inclusion and timing of each element presented in the foundational levels of the IRLA was appropriate and justified, the content of the IRLA was analyzed and compared to the theoretical dimensions and content areas found throughout the research base. To determine whether learning informed by the IRLA yielded increased results, the study examined a cluster of schools that had implemented the IRLA for three years. This was determined through
SchoolPace, the electronic database accompanying the IRLA. Approximately 2,100 schools used SchoolPace for at least one year between the 2011-2012 and 2017-2018 school years, representing states having adopted the Common Core State Standards (CCSS) and having in-common summative assessments provided by either the Smarter Balanced Assessment Consortium (SBAC), the Partnership for Assessment of Readiness for College and Careers (PARCC), or state-based summative assessments, as well as states not having adopted the CCSS. To ensure both a breadth of districts to study as well as in-common assessment tools, this study established criteria for schools to be included in the cluster as: must have used the Smarter Balanced Assessment (SBA) to measure performance on CCSS, must have begun using the IRLA and SchoolPace for any grade K-3 in 2015-2016, and must have had continuous use through the 2017-2018 school year as indicated in SchoolPace. Application of the above criteria identified 53 schools for inclusion.

SBA was first used during spring of 2015. The establishment of 2015-2016 as year one for IRLA implementation enabled a baseline score to be procured from SBA (2014-2015 school year scores) prior to installation of the IRLA. Narrowing the schools for study to these 53 enabled a close look at progress over time against the same measurement tool. Additionally, feasibility for expanding this type of study as use of the IRLA and SchoolPace increased over time was determined without an undue expenditure of time and resources.

PARCC, the second widely used summative assessment measuring CCSS provided another potential cluster. Applying the same criteria as stated above but switching PARCC for SBA revealed 23 schools. PARCC and SBA are different
assessments, so although the two groups could not be studied in one cluster, trends might have presented fair comparisons. However, PARCC was not included in this study. Additionally, many of the states using IRLA as discovered through SchoolPace used state-specific summative assessments. Five schools in Arizona; one each in Kentucky, Minnesota and Ohio; two in New Jersey; 13 each in Florida, New York, and Texas; and 20 in Pennsylvania have 2015-2016 as their first use of IRLA and SchoolPace with continuous use since that time. It was impractical to study each of these nine clusters separately as part of this study.

Sampling Strategy

Fifty-three schools were identified for study according to the criteria listed above. Those schools are located in Delaware (four schools), Oregon (38 schools), and Washington (11 schools). All states were required by law to provide outcome data to the public (§200.2(5)(ii) ESSA, 2017), though the responsibility for producing and distributing the data fell to states and districts. Each state listed above had a different timeline for reporting and provided results in a different way. Because they were unified by the same assessment, though, the data were collected from each state and compiled in a unified Excel spreadsheet. Demographic data were included in each state’s data set; due to privacy laws, the data were anonymous and small group sizes were suppressed.

Once data from each of the three states (DE, OR, WA) was procured for each of the four years under study (2014-2015 – baseline, 2015-2016 – year one, 2016-2017 – year two, 2017-2018 – year three) and consolidated, ANOVAs were run to discover
any statistically significant differences in gain scores from the baseline year to year one, year two and year three. The groups were analyzed by the whole data set.

**Participants and Setting**

The application of criteria (SBA state, initial IRLA use in 2015-2016, continued use since first implementation) to the full collection of schools in SchoolPace yielded 53 schools from five districts in three states. These districts are represented in Table 9 below with the number of schools. For this study, the collection of schools was aggregated to yield third grade data and demographic information.

Table 9

*Cluster for Analysis Including Baseline Year*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>Beaverton (33)</td>
<td>Cape Henlopen (4)</td>
<td>Federal Way (11)</td>
<td>Molalla River (4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Instrumentation

The IRLA is a developmental reading taxonomy that spans grades PreK-12 and “includes every Common Core State Standard for reading, as well as those language standards key to reading success” (Hileman & Zorzi Cline, 2017, p. i). Developed by reading teachers with practical knowledge informed by the research base of the process of learning to read, the IRLA was an instrument subjected to analysis in this study. Additionally, this study used state reported public data on the SBA. A computer adaptive assessment, SBA was in use for the first time in the spring of 2015 (SBA, 2018). Administered online, each test was unique to the student taking it. When a student answered an item correctly, the next item became a little more difficult, and when a student answered an item incorrectly, the next item became a little less difficult. This technique was intended to keep students engaged by mitigating frustration or boredom and provided a more accurate score (SBA, 2018). Third grade students were presented questions organized by claims which focused on specific categories within the overall score. In English Language Arts, the claims were reading, writing, speaking/listening, and research. The claims were further divided into content categories; literary and information within reading; organization/purpose; evidence/elaboration and conventions within writing; listening within speaking/listening; and research within research). Due to the adaptive nature of SBA, although more than 500 questions were possible, the testing experience for a 3rd grader included about 40 computer adaptive tasks which included machine-scored items and short-text items of varying difficulty. Specified in the SBA ELA/Literacy Summative Assessment Blueprint from 11/10/16, one or two short-text items in
reading and one short-text item in writing were designed for hand-scoring. SBA was estimated to take about three-and-a-half hours for a typical 3rd grader.

SBA was developed by a consortium whose initial membership included 30 states in 2010. With a federal grant of $178 million, the consortium was tasked with developing an assessment system to provide valid, reliable, and fair information about student achievement in math and ELA in respect to the CCSS at grades 3-8 and high school (SBA Technical Report, 2016). SBA developers used an evidence-centered design as the approach to the development and validation of the summative assessment, and the technical reports released annually after results are collected and analyzed by the consortium speak to the sources of validity evidence based on test content. Cut scores were determined following work done in three phases: online panel, in-person workshop, and cross-grade review prior to administration of the first test in spring of 2015 (SBA Achievement Level Setting Final Report, 2015). Cut scores have not been adjusted, nor has the administration of the assessment been changed over the intervening years.

Each year following administration and scoring, SBA has produced a technical report describing essential validity elements required for evidence of SBA as a valid measure of achievement towards the CCSS. Although internal reliability coefficients have not been available due to the nature of this adaptive assessment, items and their responses are analyzed for reliability annually and findings are provided in the technical report. Reliability has been reported as moderate to high for each year of the third grade ELA measure (CRESST, 2017).
Smarter Balanced provides a sample items website (http://sampleitems.smarterbalanced.org/) where the public can interact with items in a similar way to a student taking a test. A practice test is also available.

**Procedures and Timeline**

SBA data were obtained through state education department websites through September and October of 2018 for the 2014-2015, 2015-2016, 2016-2017, and 2017-2018 school years. Third-grade scores were extracted for all schools included in the study. Microsoft Excel 2018 software was used to organize the data into a single spreadsheet. Data were verified in a number of ways: randomly double-checking cells against the master downloads during creation of the master spreadsheet; manually verifying the sum of male and female numbers for all schools for all years to ensure 100% total population; and lastly, by carefully double checking back to the master at least 10 cells each time the spreadsheet was accessed.

**Data Analysis**

To determine whether the inclusion and timing of each element presented in the foundational levels of the IRLA was appropriate and justified, the first research question, a content validity study was conducted. A matrix was developed to show related evidence for each element and was substantiated though consideration of the empirical evidence as being the best evidence provided by experts in the field. This procedure examined the strength of the content validity of the specific individual elements included in the IRLA, and in the IRLA as a whole (through the foundational skills levels) (Warner, 2008; Wynd, Schmidt, & Schaefer, 2003). To determine whether there was an impact on third grade reading achievement due to
implementation of the IRLA over time, the second research question, the total third grade population was isolated from the total state SBA download in a separate sheet and the EZAnalyse plug-in for Excel 2018 was used to generate a histogram and calculate ANOVAs comparing the baseline year, 2014-2015, to each subsequent year. ANOVAs are a statistical method used to test differences between two or more means and analyze their variance. A p-value of <.05 was used to determine whether post-hoc analysis would be required. The Tukey post-hoc test was identified for use to determine specifically where differences in means lie, and the Bonferroni post hoc was identified for application to analyze whether there is a probability of having made a Type I error, the determination that the results are significant when they are not.

**Disclosure Statement**

As K-12 director of curriculum and instruction in a mid-sized low-SES urban district, I worked with a team for over a year to investigate and operationalize the best possible K-12 literacy program research could point us to, and in the course of this work learned about the IRLA.

At this point in my career, I work as a coach for American Reading Company (ARC), the company behind the IRLA. My job is to train teachers and administrators in the effective implementation of the ARC Core literacy program and the IRLA, which is a component of ARC Core.

**Summary**

The purpose of this research was two-fold: to determine whether the inclusion and timing of each element presented in the foundational levels of the IRLA was appropriate and justified, and to discover whether learning informed by the IRLA
yielded increased results on third grade summative assessments. After selecting schools using the IRLA over time based on participation in SchoolPace, groups were analyzed for variance over time. The hypothesis for this research was that if the IRLA is aligned with the strongest possible research, results will show that the reading outcomes for students in districts using the IRLA increased with each year of instruction informed by this tool.
Chapter 4: Results

Chapter 4 restates the purpose of the research and provides the hypothesis of the study. A narrative description of each element of reading as it strands through the IRLA is given as analysis of the taxonomy to establish whether each element is appropriate and justified for inclusion. A statistical analysis to discover change in student achievement is presented to provide answer to the question of impact due to the use of the IRLA, including descriptive statistics to provide an overview of the participating districts, schools, and students. Analysis of whole group third grade Smarter Balanced Assessment (SBA) scores are presented to provide an answer to the research question.

Research Question and Hypothesis

The hypothesis for this study is that if the taxonomically organized elements within the IRLA are aligned with the research base, results will show that implementation of the IRLA yields progressively more positive reading outcomes for students in schools using the IRLA over time. The purpose of this research is two-fold: to determine whether the inclusion and timing of each element presented in the foundational levels of the IRLA is appropriate and justified, and to discover whether learning informed by the IRLA yields increased results on third grade summative assessments.

Findings: Inclusion and Timing of Elements

Each element present in the IRLA was studied and presented in the literature review of Chapter 2, which focused on defining and understanding what empirical evidence said about it. Further, a description of each element’s appearance throughout
the levels from 1Y to 2R was presented in Chapter 2, and Table 5 in Chapter 2 denoted where each skill was explicitly placed within the IRLA. Chapter 3 described the use of a matrix to determine the content validity of the elements in the IRLA, the results of which are presented in Chapter 4, which discusses findings for each element to understand the extent to which the research base supports or challenges their inclusion and timing.

Comprehension of complex text. Research on comprehension says the ability of the reader, including foundational skills and the knowledge of the reader, are essential elements for a child learning to read (Adams, 1990; Afflerbach et al., 2008; Elwér et al., 2015; Fielding & Pearson, 1994; Gough & Tunmer, 1986; Liben & Liben, 2017; Pearson & Liben, 2018). The IRLA is designed to help a teacher know the full scope of foundational skills and to tune into them when listening to a reader. The IRLA expects a grade-level-appropriate depth of knowledge to support a child as he or she makes meaning while reading. When teachers match books of the IRLA color level with the skills demonstrated by a student, comprehension as discussed in the literature review is supported by the IRLA and is informed by the comprehension research.

The leveling system of the IRLA helps teachers match readers to books that will grow student skills within a level, providing practice with skills that will ensure access to the next level. Identification of a level for a student is entirely skills based, which supports seamless differentiation up and down the spectrum of skills. Schools using the IRLA are encouraged to purchase books by color level. Typical for a classroom is a purchase of eight baskets within a range of color levels common for
that grade; each basket contains 30 unique titles. At the yellow and green levels, many of the books are written and published by ARC Press, a division of ARC. This is because there are not enough rich, high-quality, non-fiction trade books on the market to support development of the skills required by the IRLA at these levels. Overall, though, 99% of the books included in the baskets are not produced by ARC Press but are sourced from other publishers (M. Lynch & M. Weil, personal communication, July 31, 2018). ARC builds baskets with the intent of providing a high volume of high-interest, authentic text to classrooms so that students’ interests are piqued, and motivation is activated (Fulmer & Frijters, 2011; Gambrell, 2015; Guthrie et al., 2012).

**Executive functioning.** Throughout the foundational skills available to be assigned as power goals are evidence of the underlying skill of executive function. At 2Y, students are tracking, self-correcting, and gaining engagement and independence, which are reading behaviors that require significant inhibition control and that build working memory (Blair & Raza, 2007; Foy & Mann, 2013). However, description of executive function in the IRLA is implicit, and therefore incomplete, creating potential misunderstandings or gaps in teacher attention to the development of this suite of cognitive activities (Jones et al., 2016; Willoughby et al., 2017).

**Print concepts.** The print concept strand is a strategic content that appropriately belongs in the taxonomic structure of the IRLA from the start of the tool through the end of first grade (Bear, 2012; Chall, 1976; Dooley, 2010; Jarmulowicz, 2008). By directing teachers to explicitly attend to directionality of print, spaces between words, the phonological matching of words to sounds, initial sound cues, and
grammatical elements represented by punctuation, the IRLA is directly aligned with the research on this component and does not leave the realization of how text works to chance (Clay, 1979; Flanagan, 2007; Justice et al., 2006; Meisels & Piker, 2001; Morris, 1993).

**Phonological awareness.** Treatment of phonological awareness in the IRLA is aligned with research along the developmental path of whole to part (Glazzard, 2017; Goswami & Bryant, 2016; Walton et al., 2001; Zeigler & Goswami, 2005). Rather than beginning with the smallest sound parts of words and blending to make ever greater wholes, the IRLA begins with a child’s listening to whole stories, moving to patterned sentences, words as wholes, then initial sounds, blends, and finally onset-rime. The research supports a whole to part progression, and when coupled with the growth of phonics, word recognition, and comprehension, the IRLA provides a powerful path for readers to gain from the variety of connections among the elements of early learning skills.

**Word recognition.** The research base does not provide a pure quantitative number of words that students should recognize by sight, nor is there evidence of a chronology of when certain words should be learned. What unifies the research is the idea of learning words by sight, particularly those of high utility and challenging orthography (Adams, 2011; Clark, 2016; Ehri, 2005; Ehri et al., 2001; Ehri & Saltmarsh, 1995; Schwanenflugel et al., 2007; Watts & Gardner, 2013). By including sight words from kindergarten, the IRLA leans into the research base and eventually merges the research-grounded work about rapid naming, automaticity, and sight words with an analytic phonics approach as students are coached to find a word you know in
a word you don’t to be able to read words automatically and smoothly in the pursuit of making meaning.

**Phonics.** The progression of phonics skills through the IRLA unfolds systematically and explicitly and meets the research-based recommendations of What Works Clearinghouse as given in the Foorman et al. (2016) teaching guide for foundational skills to support reading for understanding. While both synthetic and analytic phonics approaches are validated in the research base (Foorman, 2016; NRP, 2000; Savage et al., 2003), the phonics approach in the IRLA is analytic, teaching students to process text from whole to part, rather than part to whole (Glazzard, 2017; Goswami & Bryant, 2016; Schwanenflugel et al., 2007; Watts & Gardner, 2013). There is a growing research base supporting an analytic approach as providing a heightened ability to flexibly solve difficult words (Ehri et al., 2001; Henbest & Apel, 2017), to accurately spell words (Landerl, 2000; Zeigler & Goswami, 2005), and to strengthen phonological blending (Glazzard, 2017; Savage et al., 2003).

**Vocabulary.** Teaching vocabulary word by word has justifiably fallen out of favor as the vastness of the English language makes this a crippling task (Nagy & Anderson, 1984). Instead, focusing on the meanings of word parts, or morphemes, as the building blocks for meaningful words and phrases has been elevated throughout the research (Apel & Henbest, 2016). The IRLA’s foundational levels under investigation in this paper do not exploit morphological awareness to the extent that its later levels do, leading to questions about whether more explicit morphological work should enter the foundational levels. In a study by Apel, et al. (2013) kindergarten and first-grade students were working on prefixes and suffixes. The morphemes given in
their study enter the IRLA at 2R, the second half of 2nd grade. While within a year to a year and a half of the practice explored in this promising research, it may be suggested that moving prefix and suffix awareness into the first-grade year would provide a foundational scaffolding for students’ eventual manipulation of morphemes to discover a wide variety of word meanings.

By not including explicit demands for expansion of vocabulary at the yellow, 1G, 1B, 2B, and 1R levels, and by explicitly demanding a robust sight word vocabulary and frequent sustained work with authentic texts, the IRLA and the books leveled to it do not include demands on a child’s vocabulary outside of the child’s world.

**Fluency.** While the research base claims that 1st or 2nd grade is ideal for fluency (Bear, 2012; Chall, 1976; Foorman et al., 2016; Jarmulowicz, 2007; Morris, 2003), the IRLA expects it from kindergarten. However, throughout the IRLA levels, fluency is an element that directs the teacher to notice, instruct where needed, and indicate growth where it occurs. Independent reading as expected by the IRLA demands a match between reader ability and complexity of text. When a student is reading at his or her independent level, the skill of fluency is appropriate as soon as a student attains independence with appropriately complex connected text and continues throughout the grades. The language of fluency changes slightly through the IRLA levels, but the essence remains the same: 95%-100% accurate with the leveled text matching the highest level a student can read without support, with rate and expression appropriate to the text, and with confidence, purpose, and understanding. There are some notable additions: at 1B and 2B (the second and third thirds of 1st grade),
students are asked to whisper read. This heralds a transition from reading fluently aloud to reading fluently silently. At 1R, the start of second grade, students are asked to use punctuation as a guide to expression and to read silently while maintaining comprehension.

Formative assessment as the vehicle for making strategic choices and one-to-one assessment as the process through which the strategic content is strengthened were also analyzed against the research base.

**Formative assessment.** The IRLA is in alignment with the most recent definition provided by CCSSO (2018) and formative assessment theories. It follows the McManus (2008) assertion that formative assessment is a process, not an instrument: “Formative assessment is not a test or a tool. Formative assessment is a continuous improvement process employed by expert teachers and coaches” (Hileman & Zorzi Cline, 2017, p. vii). Reliance on a teacher’s ability to know the target is fundamental to the goal setting component of formative assessment and could be sourced as the impetus for the development of the taxonomy itself. When a teacher meets with a student for a one-to-one conference, the first consideration upon listening to a student read a leveled text is finding evidence regarding whether the level is easy enough. Unless the skills, standards, or goals are known by the teacher, this question is unanswerable. Using the collections of skills within a level, the teacher can evidence which skills are strong, which skills are adequate, and which skills need further instruction and practice. The taxonomic structure of the IRLA supports the formative assessment process as it is presented in theory (Bennett, 2011; Black & Wiliam, 1998, 2008; CCSSO, 2018; Sadler, 1989; Shepard, 2000; Taras, 2005).
The IRLA epitomizes the spirit of formative assessment as seen theoretically (CCSSO, 2018), and as stated by McGill-Franzen (2006), “To know where to start instruction, you must know what the child can do” (p. 7), and by Davis (2012) “At the heart of teaching is the fundamental insight that learners can only acquire new knowledge on the basis of what they already know and understand, and so a teacher must constantly monitor and diagnose learners’ existing cognitive and motivational states” (p. 569).

**One-to-one instruction.** Consistent with the research, the IRLA does not specify how long a teacher should spend in one-to-one instruction with a student. Although not explicitly stated in the IRLA, professional development provided at the onset of using the IRLA teaches that conferences should be about 5-7 minutes long, and that each student should have at least one conference every 14 days (two school weeks). There is not direction in the literature that speaks to the duration or frequency of instructional conferences of this style. Equitable conferencing schedules are likewise coached by professional developers teaching how to use the IRLA, meaning that students who are further behind should meet on a more frequent basis. This is a direct connection to the research on one-to-one instruction from special education applications (Frey, 2006).

Inconsistent with the research on one-to-one instruction is the idea of focusing on a single element of reading at a time. The aim of the IRLA is for teachers to find a *power goal*, “the ONE thing a student most needs to practice/learn between now and the next conference” (Hileman & Zorzi Cline, 2017, p. x). Unlike the Reading Recovery model, where a session is intended to take 30 minutes and formally progress
through a series of activities (“rereading two or more familiar books, taking a running record, letter identification, writing a story, cut-up story to be rearranged, new book introduced, new book attempted” (Clay, 1993, p. 14)), a conference guided by the IRLA is flexible, fluid, targeted, and brief. “The ability to listen to a student read for 2 or 3 minutes and identify the most important ONE thing he needs to practice/learn next in order to improve (the Power Goal) is the fundamental skill of teaching reading” (Hileman & Zorzi Cline, 2017, p. xi).

Repeated throughout the literature on teaching early reading is the sentiment that whenever possible, know what each individual student can do and teach the next thing. Hiebert (1981) advised that by knowing what children can do, teachers can be better instructors. Ehri and McCormick (1998) advocated an instructional style that is individualized, differentiated, and grounded in the knowledge of what a student can do. Findings that individualizing instruction is the most effective course of action are undeniable (Adams, 1990; Bear et al., 2012; Compton-Lily, 2009; Connor et al., 2014; Ehri, 1998; Flynn, 2016; Foorman et al., 2016; Goswami, 2009; Hiebert, 1981; Konold, Juel, McKinnon, & Deffes, 2003; McGill-Franzen, 2006; Meijer, Veenman, & van Hout-Wolters, 2005; Morris et al., 2003; Mandel Morrow, Tracey & Del Nero, 2011; Shanahan et al., 2010).

**Impact of the IRLA on Results of Third Grade Summative Assessments**

To discover whether learning informed by the IRLA yielded increased results on third grade summative assessments, schools that began using the IRLA in the 2015-2016 school year were identified. In this way, a baseline of summative assessment data would be set for comparison by using SBA scores from the 2014-2015 school
year, the first year SBA was available. Schools were identified for study according to criteria set forth in Chapter 3: schools that had used SchoolPace, the optional electronic database that supports use of the IRLA, beginning in the school year 2015-2016 continuously through the 2017-2018 school year, and were located in states using SBA. Initial identification of schools meeting these criteria yielded 53 schools from three states: Delaware (four schools), Oregon (38 schools), and Washington (11 schools).

**Participants.** State education departments provide publicly accessible databases, and those from Delaware, Oregon, and Washington from the 2014-2015, 2015-2016, 2016-2017, and 2017-2018 school years were downloaded and arranged to create a master spreadsheet from which analysis could be done. The construction of the master spreadsheet revealed the necessity of eliminating five schools, four from Oregon and two from Washington. This discovery is described below.

Beaverton School District, in Oregon, was originally counted as having 34 elementary schools. While this was true for the 2017-2018 school year, in 2014-2015, there were only 33 schools from this district actively using SchoolPace. Molalla River School District, also in Oregon, has four elementary schools, all of which were included in the original count. Further investigation revealed that only one of the four schools used IRLA in the 2015-2016 school year. Therefore, only one school in Molalla River was included in this study. Portland Public Schools has one school meeting the criteria for inclusion, bringing the overall count for the Oregon schools to 35 from the original 38.
Federal Way School District in Washington State has 23 schools. Twelve of those schools were involved in a pilot with the IRLA in the 2015-2016 school year. Those 12 schools piloted in a variety of grades and were the schools originally selected for inclusion in this study. Of the 12 schools included in the pilot, two schools had fewer than 20 students participate in the use of IRLA (Camelot and Olympic View). Due to such small involvement at those schools, they were eliminated from the sample, bringing participation from 12 to 10 in this district.

One district in Delaware met the requirements for entry to this study. That district, Cape Henlopen, had four elementary schools from 2014-2015 to 2016-2017, but in 2017-2018 one elementary school was added. The original four schools were included in the study, but there is a noticeable decline in the number of students included in the study in 2017-2018 due to this shift. Table 10 reflects the total number of districts and schools by state and the number of students whose data contributed to this study over the four years derived from the SBA data downloaded and compiled in the master spreadsheet used for the calculations to examine change.

Table 10

Total Districts, Schools, and Number of 3rd Grade Students Included in the Study

<table>
<thead>
<tr>
<th></th>
<th>Number of Districts</th>
<th>Number of Schools</th>
<th>Number of 3rd Graders 2014-2015</th>
<th>Number of 3rd Graders 2015-2016</th>
<th>Number of 3rd Graders 2016-2017</th>
<th>Number of 3rd Graders 2017-2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delaware</td>
<td>1</td>
<td>4</td>
<td>417</td>
<td>431</td>
<td>433</td>
<td>310</td>
</tr>
<tr>
<td>Oregon</td>
<td>3</td>
<td>35</td>
<td>3,093</td>
<td>3,160</td>
<td>3,018</td>
<td>2,946</td>
</tr>
<tr>
<td>Washington</td>
<td>1</td>
<td>10</td>
<td>692</td>
<td>743</td>
<td>791</td>
<td>703</td>
</tr>
<tr>
<td>Total</td>
<td>5</td>
<td>49</td>
<td>4,202</td>
<td>4,334</td>
<td>4,320</td>
<td>3,959</td>
</tr>
</tbody>
</table>
To gain an understanding of the demographic representation of students included in this study, a breakdown is presented in Table 11. Of note, the largest racial groups represented were Whites, at either 45% or 46% across all four years, Hispanic/Latino, at 23% or 24%, and Asian, ranging from 8% to 13% across the study years. Multi-Racial students represented were between 5% and 7%. African American students in this sample range from 3% to 5%. Representation of Native Hawaiian represented less than 1% in 3 of the 4 years. Asian/Pacific Islander were 2% of the sample in the first year and were not represented in the subsequent years. Indian/Alaska Native students ranged from zero to four total students across the study years.

Identification by program is also presented in Table 11. Between 15% and 18% of students in this study were English learners, 9% to 11% received Special Education services, and 43% to 45% qualified for free or reduced lunch, indicating a low socio-economic status as identified by federal guidelines.
Table 11

Demographic Breakdown of Students in the Study

|---------------------------|-----------|-----------|-----------|-----------|
| Total                     | 100%  
  ($n = 4,202$) | 100%  
  ($n = 4,334$) | 100%  
  ($n = 4,320$) | 100%  
  ($n = 3,959$) |
| Female                    | 51%  
  ($n = 2,145$) | 49%  
  ($n = 2,111$) | 48%  
  ($n = 2,101$) | 47%  
  ($n = 1,872$) |
| Male                      | 49%  
  ($n = 2,057$) | 51%  
  ($n = 2,223$) | 51%  
  ($n = 2,219$) | 53%  
  ($n = 2,087$) |
| African American          | 3%  
  ($n = 128$) | 5%  
  ($n = 225$) | 5%  
  ($n = 198$) | 4%  
  ($n = 178$) |
| Asian                     | 8%  
  ($n = 325$) | 12%  
  ($n = 511$) | 13%  
  ($n = 543$) | 10%  
  ($n = 399$) |
| Asian/Pac Islander        | 2%  
  ($n = 83$) | 0%  
  ($n = 0$) | 0%  
  ($n = 0$) | 0%  
  ($n = 0$) |
| Indian/Ak Native          | 0%  
  ($n = 0$) | 0%  
  ($n = 3$) | 0%  
  ($n = 4$) | 0%  
  ($n = 0$) |
| Hispanic/Latino           | 23%  
  ($n = 971$) | 24%  
  ($n = 1,027$) | 24%  
  ($n = 1,043$) | 24%  
  ($n = 966$) |
| Multi-Racial              | 5%  
  ($n = 223$) | 7%  
  ($n = 321$) | 7%  
  ($n = 315$) | 6%  
  ($n = 256$) |
| Native Hawaiian           | 0%  
  ($n = 0$) | 1%  
  ($n = 35$) | 1%  
  ($n = 37$) | 1%  
  ($n = 33$) |
| White                     | 45%  
  ($n = 1,911$) | 46%  
  ($n = 2,000$) | 45%  
  ($n = 1,949$) | 45%  
  ($n = 1,782$) |
| English Learners          | 18%  
  ($n = 761$) | 19%  
  ($n = 828$) | 19%  
  ($n = 804$) | 15%  
  ($n = 592$) |
| Special Education         | 9%  
  ($n = 387$) | 11%  
  ($n = 473$) | 10%  
  ($n = 443$) | 11%  
  ($n = 429$) |
| Low SES                   | 45%  
  ($n = 1,885$) | 45%  
  ($n = 1,964$) | 44%  
  ($n = 1,892$) | 43%  
  ($n = 1,715$) |

Change in performance of third grade students. To determine whether there was an impact on third-grade reading achievement due to implementation of the IRLA over time, the mean scores of the total population of third-grade students from the sample for each of the four years were studied. Across all schools in the study, there was an average positive change in proficient scores, a combined measure of Level 3, Meets, and Level 4, Exceeds, from 2014-2015 to 2017-2018 of 1.38. Of the 49 schools included in the study, 27, or 55%, realized increases in their SBA scores over the four years. Conversely, 21, or 45% of schools saw reductions in their SBA scores.
scores. One school saw no difference. Analysis revealed a range of 77% difference in scores from -33% to +44%. The median change for students scoring a Level 4 dropped by .02%, and the median change for students scoring a Level 1 increased slightly, by .30%. Level 2 median scores dropped 1.76%, while Level 3 realized a gain of 1.39%. Table 12 presents the difference in each level of SBA scores from 2014-2015 to 2017-2018.

Table 12

*Difference in SBA Scores from 2014-2015 to 2018-2019 by Level*

<table>
<thead>
<tr>
<th>School</th>
<th>Proficient (Level 3 or 4)</th>
<th>Level 4</th>
<th>Level 3</th>
<th>Level 2</th>
<th>Level 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>43.50</td>
<td>35.30</td>
<td>8.10</td>
<td>-18.30</td>
<td>-25.20</td>
</tr>
<tr>
<td>2</td>
<td>25.80</td>
<td>3.20</td>
<td>22.60</td>
<td>-15.60</td>
<td>-10.20</td>
</tr>
<tr>
<td>3</td>
<td>19.70</td>
<td>11.70</td>
<td>8.00</td>
<td>-8.00</td>
<td>-11.70</td>
</tr>
<tr>
<td>4</td>
<td>15.50</td>
<td>7.90</td>
<td>7.50</td>
<td>-5.90</td>
<td>-9.50</td>
</tr>
<tr>
<td>5</td>
<td>12.70</td>
<td>11.60</td>
<td>1.10</td>
<td>-6.60</td>
<td>-6.00</td>
</tr>
<tr>
<td>6</td>
<td>12.70</td>
<td>0.80</td>
<td>11.90</td>
<td>-1.50</td>
<td>-11.30</td>
</tr>
<tr>
<td>7</td>
<td>12.40</td>
<td>12.20</td>
<td>0.10</td>
<td>-0.10</td>
<td>-12.30</td>
</tr>
<tr>
<td>8</td>
<td>11.50</td>
<td>9.70</td>
<td>1.80</td>
<td>-7.40</td>
<td>-4.00</td>
</tr>
<tr>
<td>9</td>
<td>11.50</td>
<td>8.40</td>
<td>2.90</td>
<td>-13.00</td>
<td>1.50</td>
</tr>
<tr>
<td>10</td>
<td>9.90</td>
<td>14.00</td>
<td>-4.20</td>
<td>-3.50</td>
<td>-6.30</td>
</tr>
<tr>
<td>11</td>
<td>9.66</td>
<td>2.14</td>
<td>7.52</td>
<td>-2.39</td>
<td>-7.27</td>
</tr>
<tr>
<td>12</td>
<td>9.20</td>
<td>6.50</td>
<td>2.70</td>
<td>-10.70</td>
<td>1.50</td>
</tr>
<tr>
<td>13</td>
<td>7.90</td>
<td>5.60</td>
<td>2.30</td>
<td>-1.10</td>
<td>-6.70</td>
</tr>
<tr>
<td>14</td>
<td>7.60</td>
<td>1.70</td>
<td>5.90</td>
<td>-2.60</td>
<td>-5.10</td>
</tr>
<tr>
<td>15</td>
<td>7.40</td>
<td>4.10</td>
<td>3.30</td>
<td>-4.90</td>
<td>-2.60</td>
</tr>
<tr>
<td>16</td>
<td>7.40</td>
<td>2.10</td>
<td>5.20</td>
<td>-5.50</td>
<td>-1.90</td>
</tr>
<tr>
<td>17</td>
<td>6.30</td>
<td>9.60</td>
<td>-3.40</td>
<td>4.10</td>
<td>-10.30</td>
</tr>
<tr>
<td>18</td>
<td>6.10</td>
<td>-13.20</td>
<td>19.30</td>
<td>-3.70</td>
<td>-2.50</td>
</tr>
<tr>
<td>19</td>
<td>4.32</td>
<td>-12.00</td>
<td>16.31</td>
<td>-3.54</td>
<td>-0.32</td>
</tr>
<tr>
<td>20</td>
<td>3.90</td>
<td>3.30</td>
<td>0.60</td>
<td>7.80</td>
<td>-10.10</td>
</tr>
<tr>
<td>21</td>
<td>3.80</td>
<td>8.80</td>
<td>-5.10</td>
<td>1.20</td>
<td>-5.10</td>
</tr>
</tbody>
</table>
Analysis of variance in performance of third grade students. The average scores for the total population of third graders for each year were calculated and compared using an ANOVA. Year 4 proficient SBA scores average 1.38 points higher than Year 1 SBA scores, but data fluctuated across the years and the differences were
not statistically significant. The ANOVA results indicated that none of the groups differed significantly over time in achievement at SBA Levels 1, 2, 3, 4, or proficient (a combined total of 3 and 4). At the proficient level, 3 and 4 combined, there was not a significant effect for treatment, $F(3, 192) = .183, p = .908$. SBA Level 4 indicates exceeding the standards. At this level there was not a significant effect for treatment, $F(3, 192) = .219, p = .883$. Level 3, indicating meeting standards, did not have a significant effect for treatment, $F(3, 192) = .456, p = .714$. Level 2, indicating close to meeting standards, also did not have a significant effect for treatment, $F(3, 192) = .732, p = .534$. Level 1 indicates a not meeting standard score, and it did not indicate a significant effect, $F(3, 192) = .531, p = .661$. The total group means and standard deviations are presented in Table 13.

Table 13

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Proficient (SBA 4 &amp; 3)</td>
<td>55.21</td>
<td>17.50</td>
<td>56.96</td>
<td>18.68</td>
</tr>
<tr>
<td>Exceeds (SBA 4)</td>
<td>32.37</td>
<td>16.50</td>
<td>33.47</td>
<td>16.46</td>
</tr>
<tr>
<td>Meets (SBA 3)</td>
<td>22.84</td>
<td>5.63</td>
<td>23.35</td>
<td>6.09</td>
</tr>
<tr>
<td>Nearly Meets (SBA 2)</td>
<td>23.69</td>
<td>7.19</td>
<td>22.38</td>
<td>7.43</td>
</tr>
<tr>
<td>Does Not Meet (SBA 1)</td>
<td>20.61</td>
<td>12.44</td>
<td>20.59</td>
<td>12.84</td>
</tr>
</tbody>
</table>

$N = 49$

From these data, there was no evidence to demonstrate that learning was influenced by use of the IRLA as measured by summative state assessment scores for third-grade students.
Summary

Chapter 4 presented the results of the two-fold study conducted to determine whether the inclusion and timing of each element presented in the foundational levels of the IRLA was appropriate and justified, and to discover whether learning informed by the IRLA yielded increased results on third-grade summative assessments. Each element of reading as presented in the taxonomic structure presented in Chapter 2 was followed throughout the IRLA levels with attention to where research supported or contradicted inclusion and timing. Analysis of SBA data from the schools and districts selected for participation in the study was presented. Findings from the analyses done in Chapter 4 will be discussed in Chapter 5.
Chapter 5: Discussion

The purpose of this research was two-fold: to determine whether the inclusion and timing of each element presented in the foundational levels of the IRLA was appropriate and justified, and to discover whether learning informed by the IRLA yielded increased results on third grade summative assessments. To determine whether the inclusion and timing of each element presented in the foundational levels of the IRLA was appropriate and justified, Chapter 2 considered the research base for each of the elements included in the IRLA, and Chapter 4 analyzed the ways in which each element is realized within the IRLA. Comprehension as defined by Snow (2002), was unpacked in three parts: the reader and the skills the reader brings to the text; the text and its demands; and the activity, made up of purpose, processing, and outcome, done by the reader with the text. Comprehension was presented in the taxonomically based theoretical framework as the uncontrollable environment within which reading work takes place. Also presented as an environmental aspect is that of the child’s executive functioning. Informed by Jones et al. (2016), executive functioning links working memory, inhibitory control, and cognitive flexibility as the control center through which learning takes place.

Strategic elements, supporting the comprehension claim of the skills the reader brings to the text were presented in turn. Print concepts, or the knowledge about orthographic rules of written language (Justice et al., 2006), phonological awareness, the awareness of phonemes, syllables and words (Ziegler & Goswami, 2005), word recognition, referring to the conglomerate skills of sight words, automaticity and rapid naming (Chall, 1967; Dolch, 1936; LaBerge & Samuels, 1974; Perfetti & Hogaboam,
1975), phonics, the sound-symbol system of language (Ehri et al., 2001; Goswami & Bryant, 2016; Gough & Hillinger, 1980; Henbest & Apel, 2017; Morris, 2014), vocabulary, the body of words that make up a language (Apel et al., 2013; Foorman et al., 2016; Nagy & Anderson, 1984), and fluency, the accurate, automatic, and prosodic reading of text to support comprehension (Allington, 1983; Hudson et al, 2005; Kuhn & Stahl, 2003), compose the strategic elements or skills a student requires to make meaning of increasingly complex text. Formative assessment, a continuous feedback loop between teaching and learning (Afflerbach et al., 2008; Black & Wiliam, 1998; Sadler, 1989; Stiggins, 2002), was equated to the strategic choices required in Hambrick’s (1984) taxonomic structure, and one-to-one instruction, the provision of instruction in the most individualized setting possible (Al Otaiba et al., 2011; Bloom, 1984; Connor et al., 2009; Connor et al., 2013; Frey, 2006; Iversen & Tunmer, 1993) was discussed as the process through which learning occurs as guided by the IRLA.

To discover whether learning informed by the IRLA yielded increased results on third grade summative assessments, Chapter 3 detailed the methodology used for making this discovery: schools from three states using Smarter Balanced Assessment (SBA) were identified, and third-grade summative scores were analyzed for each of the four years SBA had been used. To qualify for inclusion in the study, schools must have begun use of the IRLA in 2015-2016, one year after the first administration of SBA, ensuring a consistent baseline. ANOVAs were conducted to compare the difference in average score from year to year. Chapter 4 presented the statistical analysis in search of the answer. Chapter 5 discusses the findings of the analyses done
in Chapter 4 and makes suggestions for potential revision to the IRLA to strengthen its connection to the research.

**Findings: Inclusion and Timing of Elements in the IRLA**

Elements in the IRLA are inclusive of the elements named in NCLB’s identified *big five*—comprehension, fluency, vocabulary, phonics, phonemic awareness—and further include word recognition, print concepts, and executive functioning. Additionally, IRLA uses a one-on-one instructional delivery method informed by formative assessment. This study examined the literature base, compared findings to the outlay of skills in the IRLA and found that the elements present are appropriately included and are presented in research-based developmental progressions. However, there are some considerations for potential change to the IRLA discovered through the literature review and subsequent analysis of content validity comparing the research to each level in the IRLA.

To understand the foundational levels (RTM-2R) of the IRLA as a developmental reading taxonomy, this paper began with a survey of the literature to fully understand the organizational support that taxonomies provide. Combining Hambrick’s (1984) classification theory and Travers’ (1980) separation of classes, a gridded structure that lists the elements of learning to read as strategic content, further classified by environment, strategic choices, and process are displayed horizontally, and the timing, or developmental sequencing for each, provides separation by class, or color level. An inventory of each color level presented in a graphic representation (Table 5 in Chapter 2) of the combined taxonomic structure revealed missing elements: print concepts do not extend to 1R and 2R, phonological awareness skips
1Y, 2Y, 2G, and ceases after 1R. Word recognition and phonics do not begin until 3Y, and fluency does not begin until 1G. Vocabulary is present in only three non-sequential levels: 1Y, 2G, and 2R. Examination of the skills and their progressions provides rationale and suggestions for most of the vacancies within levels, particularly where they cease due to mastery as in print concepts, but a recommendation of this paper for enhancement to the IRLA is to fully embrace a multi-dimensional taxonomic approach in the introduction and training of the IRLA.

The inside front cover of the IRLA describes that the taxonomy “outlines the big ‘jumps’ in reading skills, strategies, and concepts that distinguish one IRLA reading level from the next” (Hileman & Zorzi Cline, 2017, front cover). While this is true, it is one dimensional, and essentially provides an analytic progression of phonic skills: “initial consonants, high-frequency words, initial blends & digraphs, onset + sight word/rime, two-syllable words, multisyllabic words, irregularly spelled words” (Hileman & Zorzi Cline, 2017, front cover). Precision in representing the taxonomy as slices through a synergistic collection of skills interdependent with one another and calling out a signature or key skill may provide a greater clarity for users.

In the taxonomic slot designated for environmental factors, comprehension and executive function are named as the uncontrollable attributes. Comprehension is a fully realized component of the IRLA, and is further supported by leveled text, providing a simultaneously controlled and authentic environment through which comprehension can be grounded and grown. Executive function, which considers the environment of the learner’s mind, could be better developed. A user of the IRLA should know more than simply that executive function is foundational for learning.
Rather, an understanding of the three elements that make up the executive functioning system (working memory, inhibition control, and cognitive flexibility), and how they can be developed and supported in a beginning and advancing reader should be presented in the introductory pages of the IRLA to give teachers a better foundation for the work of the read-to-me and yellow levels. Executive functioning is a key factor through all levels to the stamina and engagement demands made on students to read independently for 30 minutes in school and an additional 30 minutes at home each day. Jones et al. (2016) emphasize that executive functioning is a term used across the lifespan, and it evolves through development. Acknowledging working memory throughout the IRLA would support strategic elements such as word recognition and fluency. Learning about inhibitory control would directly support developing the focus required to sustain reading text of growing difficulty. Understanding cognitive flexibility would be a crucial skill to help students who stop to figure out an unknown word or get interrupted for a different instructional task. Presenting executive functioning with greater precision and explicitly connecting it beyond the yellow levels would provide teachers with deeper understanding of the cognitive environment of the students in their classrooms.

The strategic elements of the IRLA (print concepts, phonological awareness, word recognition, phonics, vocabulary, and fluency), are all presented in research-based progressions and are all appropriately included. Only two elements invite consideration for modification based on this researcher’s findings: phonological awareness and vocabulary.
Phonological awareness is not explicitly named in the collection of skills at 1Y or 2Y, rather, “pattern/picture integration” (Hileman & Zorzi Cline, 2017, p. 10) is included. The ability to repeat a pattern is a reinforcement of phonological awareness and should be explicitly called out as such in 1Y. Similarly, at the transition point between 2Y and 3Y, initial consonants are required with the description “produce the sound of (or get mouth into the ready position for) the initial consonant” (Hileman & Zorzi Cline, 2017, p. 16). At 2G students must produce the “sound of a blend or digraph while they look for meaning clues” (Hileman & Zorzi Cline, 2017, p. 54). Picture cues used in conjunction with phonic cues are required for these two phonological tasks. While not a potential enhancement to the IRLA, an explicit link to this skill (get the mouth ready to produce the sound of…) as a phonological one may deepen teacher understanding of phonological awareness in this earliest stage of reading.

Guided by the growing body of research prompted by Apel et al. (2013), vocabulary instruction at the foundational levels of the IRLA could be expanded to include morphemic awareness, particularly of affixes, at the first grade, or 2G, 1B, and 2B levels. Children are aware of the influence of affixes on words and could leverage the emphasis on meaning throughout the word recognition strategy to grow awareness and use of morphemic units. Although promising, more research would need to be considered to deepen the vocabulary demands of the IRLA at this time.

Each element studied for this paper cross-referenced others, creating reinforcing pathways between themselves. However, one element was referenced in the research on multiple others that is not included in the IRLA: spelling. Reading and
writing are reciprocal processes and given that the IRLA is a developmental reading taxonomy, it makes sense that spelling, a convention used in writing, would not feature. Weiser and Mathes (2011) discuss how decoding and encoding are mutually beneficial processes in the pursuit of gaining literacy skills, though, and found that students who received encoding instruction to learn phoneme-grapheme relationships outperformed groups not receiving encoding instruction in phonemic awareness, spelling, decoding, fluency, comprehension, and writing. Flanigan (2007) found phonemic spelling ability increased with instruction in the concept of a word. Ehri and Wilch (1987) and Bradley and Bryant (1983) found that knowing the sounds in words made it easier for children to spell words with accuracy. Ehri (2005)—building a link between phonological awareness, phonics, word reading, and spelling—asserted that spellings are phonemic maps, but until children have knowledge of letters and their sounds, they must learn words by sight, which in turn, bolsters their ability to spell by analogy. Foorman et al. (2016) directs teachers to have students decode, analyze, write, and recognize words based on the IES meta-analysis of early reading skills.

The IRLA is not devoid of mention of spelling. Spelling patterns as discussed in Cunningham (2012) are companion to the phonics developmental sequence page in the introduction (Hileman & Zorzi Cline, 2017, p. v), and spelling-sound correspondences as provided throughout the CCSS are referenced in 1G, 2G, 1R, and 2R. In each of the preceding levels, descriptors recognize, identify, or know are given. Sound-spellings in the IRLA are for decoding, but the research shows that encoding is
equally powerful. An enhancement to the IRLA worth consideration is the encoding of sound spellings.

**Independent reading.** Every level of the IRLA is led by a category called *Range of Reading and Level of Text Complexity*. Akin to the taxonomically defined environment of comprehension of complex text, independent reading is an element arguably more important than all the rest. Embedded in the range of reading category are engagement, independence, and home reading, calling for students to independently read 30 minutes in school and 30 minutes at home every day. A bedrock concept for American Reading Company, independent reading is more important than any given element in the presented taxonomic arrangement. The research base for independent reading and its benefit on growing reading ability is vast and beyond dispute. Anderson, Wilson, and Fielding (1988) conducted a survey of 155 fifth-grade students to catalogue time spent out-of-school and found that reading independently was the best predictor of reading achievement. Paul (1996) studied reading performance data on 659,214 students K-12 and found that students who read more grow faster. Predicted growth for low-achieving students was 1.66 grade levels more than typical growth for every 60 minutes per day of independent reading. Allington (2014) praised the work of Kuhn et al. (2006) and Schwanenflugel et al., (2006) for studies comparing wide reading interventions where students read what they chose and wanted to read to repeated reading interventions and finding greater fluency gains.

The IRLA expects that students read independently from self-selected books that are high interest, keyed to the skills a student possesses, and that propel them to
want more. When students are engaged with self-selected books and are reading independently, the taxonomy enables teachers to identify and precisely instruct on-time individualized reading skills. Independent reading, for all its import, though, is not explicitly called out in the introduction of the IRLA. New users of the IRLA may fall into a false sense of the IRLA as test or assessment tool rather than as a robust everyday propellent helping students find themselves while growing their abilities as readers. One last enhancement this researcher would suggest is to elevate independent reading to a place of prominence in the introduction of the IRLA.

**Findings: Results on Third Grade Summative Assessments**

**SBA proficient.** A proficient score on the SBA is indicated with an achievement Level 3 or Level 4 score. Each level indicates likely success in future coursework in English language arts/literacy, with a difference being that at Level 3, the student has ‘met’ the achievement standard and ‘demonstrates progress’ and at a Level 4 the student has ‘exceeded’ the achievement standard and *demonstrates advanced progress* (Smarter Balanced, 2018).

This study looked at proficient scores and their change across the four years of SBA testing. Year 4 proficient scores averaged 1.38 points higher than Year 1 proficient SBA scores, but data fluctuated across the years and the differences were not statistically significant. Standard deviations were high, and there was not a significant effect for treatment. While the data indicates that there is not an impact on overall proficiency due to use of the IRLA, the sample size was small, at 49 resulting in low statistical power. While the results show no impact, this may be due to a number of limitations described later in this chapter.
**Variability.** Chapter 4 presented the total differences between Year 1 (2014-2015), Year 2 (2015-2016), Year 3 (2016-2017), and Year 4 (2017-2018) at each of the SBA levels: Proficient (3 and 4 combined), Level 4, Level 3, Level 2 and Level 1. ANOVA results describing the variance in means were small and yielded no statistical differences. However, at the school level, there was a 77% variance in proficient scores between years one and four across the 49 schools. Figure 1 shows the array of gains and losses between years one and four, notably the two instances of schools gaining more than 20 points (Schools 1 and 2) and the three schools losing more than 20 points between the first and fourth years (Schools 47, 48, and 49). Given the small sample size, the variance in differences from Year 1 to Year 2, and low effect sizes, it is not possible to confirm an impact on SBA scores as a result of the use of the IRLA.

*Figure 1.* Difference in proficient scores between 2014-2015 and 2017-2018 for the 49 schools included in the study.

At the proficient level, scoring a 3 or 4 on SBA, School 1 gained 44% and School 2 gained 26% between Years 1 and 4. They are from the same district in
Oregon. An analysis of their demographic data across the four years reveals no discernable change in total enrollment, program size, or demographics. School 1 had a principal change in 2015-2016, the first year IRLA was implemented. Table 14 presents the demographic data from Schools 1 and 2.

Table 14

Demographic Descriptors of the Two Schools Gaining More Than 20% from Year 1 to Year 4

<table>
<thead>
<tr>
<th>School Year</th>
<th>School 1: Gained 44%</th>
<th>School 2: Gained 26%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14-15</td>
<td>15-16</td>
</tr>
<tr>
<td>Total Enrollment</td>
<td>580</td>
<td>582</td>
</tr>
<tr>
<td>Low Socio-Economic Status</td>
<td>44%</td>
<td>45%</td>
</tr>
<tr>
<td>Special Education</td>
<td>11%</td>
<td>11%</td>
</tr>
<tr>
<td>English Learners</td>
<td>25%</td>
<td>27%</td>
</tr>
<tr>
<td>Am Indian/Ak Native</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Asian</td>
<td>3%</td>
<td>3%</td>
</tr>
<tr>
<td>Black/African American</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>Hispanic/Latino</td>
<td>26%</td>
<td>27%</td>
</tr>
<tr>
<td>Multi-Racial</td>
<td>5%</td>
<td>6%</td>
</tr>
<tr>
<td>Native Hawaiian/Pac Is</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>White</td>
<td>65%</td>
<td>63%</td>
</tr>
<tr>
<td>New Principal</td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>

Three schools lost more than 20% at the proficient level between the first and fourth years. Similar to the schools with large gains, there is demographic stability in most areas. It is notable that the overall low socio-economic (SES) populations are
higher than those from the schools with large gains, although School 47, ranging from 51% to 58% in their low SES percentage is only about 10% higher than School 1, ranging from 43% to 45%. Also notable is that School 1 and School 47 are from the same district. The school with the greatest loss, School 49, saw the greatest shift in demographics, shifting from 13% Black/African American in the first year to 23% in the fourth year with an inverse shift in the White population, changing from 39% in the first year to 24% in the fourth. Accompanying this shift is an increase in total enrollment of 148 students. Table 15 presents the demographic data from schools 47, 48, and 49.
Table 15

Demographic Descriptors of the Three Schools Losing More Than 20% from Year 1 to Year 4

<table>
<thead>
<tr>
<th>School Year</th>
<th>School 47: Lost 21%</th>
<th>School 48: Lost 30%</th>
<th>School 49: Lost 33%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Enrollment</td>
<td>450</td>
<td>524</td>
<td>512</td>
</tr>
<tr>
<td>Low SES</td>
<td>58%</td>
<td>58%</td>
<td>52%</td>
</tr>
<tr>
<td>Special Education</td>
<td>14%</td>
<td>15%</td>
<td>16%</td>
</tr>
<tr>
<td>English Learners</td>
<td>34%</td>
<td>34%</td>
<td>36%</td>
</tr>
<tr>
<td>Am Indian/AK Native</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>Asian</td>
<td>11%</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>Black/Af Am</td>
<td>1%</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td>Hispanic/ Latino</td>
<td>32%</td>
<td>31%</td>
<td>31%</td>
</tr>
<tr>
<td>Multi- Racial</td>
<td>7%</td>
<td>7%</td>
<td>7%</td>
</tr>
<tr>
<td>Native Hawaiian/ Pac Is</td>
<td>1%</td>
<td>2%</td>
<td>1%</td>
</tr>
<tr>
<td>White</td>
<td>48%</td>
<td>47%</td>
<td>49%</td>
</tr>
<tr>
<td>New Principal</td>
<td>x</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Implications Embedded in the Statistical Analysis

This study revealed no statistically significant difference in the percentages of students who were reaching proficient levels in the schools meeting conditions for inclusion in the study. Consideration of the variance in performance across the schools realizing the greatest and lowest gains reveals no patterns that could explain the variance. In common across the schools is the implementation of the IRLA in
2015-2016, though there must be a cause for the variability in results. The IRLA is a tool to help teachers ensure the development of skills in concert with one another and is to be employed in one-to-one conferences during 30-minute independent reading embedded in a daily literacy block. The development of skills in concert with one another encourages a deeper look at teacher understanding, training, and application of the tool in accordance with its design. One-to-one conferences must happen with a frequency that is beneficial to students; 30-minute independent reading times must not be hindered by disengaged students (resulting in potential disciplinary actions); and school-level disruptions to a daily literacy block cannot be tolerated. Significant growth is not likely to be realized at third grade summative levels until the IRLA is mastered by teachers and supported by administrations. A purely quantitative look at IRLA results is inadequate for telling a more complete story of the success or failure of this tool in teachers’ practice.

Limitations of the Study

The remainder of this chapter will consider the limitations of the quantitative aspect of the study, will discuss limitations of IRLA implementations, and will explore the literature on implementation science as an effective guidance system that may improve the chances for this important innovation to realize its potential on a greater scale. Finally, recommendations for further study will be presented.

Unit of analysis. To best understand the impact of the IRLA on summative test results, a student-by-student study would be most transparent, and therefore most informative. To protect student privacy, states do not report to this level of granularity. Instead, this study looked at the smallest possible publicly available grain size, which
is by school, rendering a more robust look at variations in classrooms and individuals impossible. A future study using the same analysis design should be conducted in partnership with districts providing student-by-student SBA, PARRC or state-specific scores. Specificity of growth over time could be studied more closely and potentially more conclusively, rather than sacrificing the opportunity to study variations within groups.

**Pre-test information is not available.** Another limitation of this study is the absence of pre-test data. The 3rd grade SBA is the first time students are subjected to a large-scale summative assessment. Determining whether the conclusions in the data are the result of an instructional intervention (application of the IRLA to instruction) or the type of test or testing situation is not possible to tease out, providing no points for comparison. While compelling to compare 3rd grade SBA data to 3rd grade IRLA data, there are limitations inherent in the IRLA data that makes this a less-than-optimal study. These limitations will be explored in the section on limitations of IRLA implementations.

**Quantity of data.** Ensuring that data would be truly like across states and districts limited the study to only three states yielding five districts and ultimately 49 schools. This represents 34% of the schools beginning use of the IRLA in the 2015-2016 school year, and while a more robust sample is desired, identifying parameters and staying within them was necessary. Compounding the challenge of a small sample size is the danger of making Type II errors in the analysis of data. The statistical power in a study done with a small sample size is weak and makes rejecting the null hypothesis difficult (Cohen, 1962).
Duration of data. IRLA has been in existence for 20 years, and there are schools in every state who have used it to inform reading instruction over that time (American Reading Company, 2018). However, as with any innovation, many schools abandoned their practice, others did not implement with consistency across grade levels, and still others did not maintain a relationship with American Reading Company, rendering them invisible to the database that informed the selection of schools for the study. Additionally, the completion and adoption of CCSS in most states in 2010 required revision to the IRLA, and the change of state standardized assessments in response to the CCSS required the short time frame for this study of the impact of the IRLA on instruction as determined by a state summative assessment. This study used the first year, 2014-2015, of the standardized assessment SBA as a baseline year, which allowed for the use of only three years’ data.

Another duration challenge within this study is revealed when cohorts are considered. Those third-grade students with SBA data in the 2017-2018 school year had the greatest amount of their primary education informed by the IRLA, but not all; their kindergarten year did not include IRLA-based instruction as they were 1st graders in the 2015-2016 year, the first year the IRLA was used in their schools.

Implementation variations. Two of the five districts included in this study technically met the requirements for entry to the study; they were in SBA states and began use of the IRLA in the 2015-2016 school year. However, those districts were piloting the IRLA in select schools at that time, and the total number of students included in the study in 2015-2016 were noticeably smaller. During a pilot situation,
schools often have limited training and limited materials. A complete breakdown of the total number of students by grade across the five districts is included as Table 16.

Table 16

<table>
<thead>
<tr>
<th></th>
<th>2015-2016</th>
<th>2016-2017</th>
<th>2017-2018</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>K</strong></td>
<td>3003</td>
<td>4978</td>
<td>6003</td>
</tr>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt;</td>
<td>3136</td>
<td>5259</td>
<td>4887</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt;</td>
<td>3217</td>
<td>5298</td>
<td>5122</td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt;</td>
<td>3257</td>
<td>5545</td>
<td>5251</td>
</tr>
</tbody>
</table>

Limitations of IRLA Implementations

Each school choosing to include the IRLA in its practice does so with guidance from American Reading Company, but ultimately each is on its own to put this evidence-based practice in motion, which inevitably presents limitations to the study of its effectiveness. This section will consider some challenges the implementation of the IRLA in schools and districts presents.

**Educator knowledge and readiness.** Chapter 1 presented the chaotic landscape of reading instruction informing the taxonomic basis of the IRLA, and Chapter 2 explored the many components necessary to consider when teaching reading. To effectively employ this tool in a primary-grade classroom requires teacher readiness to change instructional tack and a depth of knowledge and facility of management that is not common for most elementary teachers (Zimmerman, 2006). Of the schools in the study, the variable of teacher knowledge and readiness is both unknowable and uncontrollable, presenting itself as a limitation.
Teachers are not solely responsible for the implementation of the IRLA. Site and district administrators manage decision making and funding, and without their knowledge of the complexities of the IRLA, the creation of conditions for its successful implementation may be overlooked. In each district and school included in this study, the leadership conditions vary, making the administrative layer of implementation unknowable and adding to the limitations of the study.

**Accuracy of IRLA data.** Each level of the IRLA is assigned a point value commensurate with a grade level equivalent. For instance, the level 1 Blue, used as an illustrative example in Chapter 3, begins with a point value of 1.3, indicating the reading level of a typical first grader starting the third month of learning literacy skills. When a student meets the entry requirements for 1 Blue, the teacher indicates this by clicking the “Ready For” button in SchoolPace, and the student is recorded at 1.3. Throughout the level, the skills prioritized for learning carry points, and although teachers are encouraged to use and record points, the SchoolPace system does not require their use. Rather, by meeting the entry requirements at the next level, students receive points in another batch. The level 2 Blue Ready For button indicates a student is a 1.6, or a first grader with six months’ skill mastered.

Between kindergarten and second grade, a typically developing student has opportunities for leveling up four times in kindergarten (1Y = .01, 2Y = .10, 3Y = .25, and 1G = .6), three times in first grade (2G = 1.0, 1B = 1.3, 2B = 1.6), and two times in second grade (1R = 2.0 and 2R = 2.5) (Hileman & Zorzi Cline, 2017). The progression is shown in Table 17.
The use of the IRLA requires decision making on the part of teachers. By relying on entry requirements, the decisions are standardized. Where entry requirements are more frequent, as in kindergarten, the standardized checkpoints support a more stable view of accurate levels and the quantification of progress through grade-level skills progression. Where the entry requirements spread across a wider range, skills continue to be acquired, but their quantification becomes less rigidly standardized and invites educators to question the objectiveness of student progress. Teachers who struggle with an absolute relationship between a single skill and the points it earns in this tool may inhibit their use of it, resulting in challenges with implementation and accuracy of student level within schools and across systems.

**Teacher training.** A further limitation of the implementation of the IRLA is the variation in teacher training and ongoing professional development. While professional development was not investigated and so no findings were presented in Chapter 4, consideration of the training opportunities is appropriate in a discussion of limitations. The decision making that is required with the use of the IRLA is complex and teachers gain competence and confidence with the IRLA by using it frequently and with expert guidance. Professional development literature tells us that without a

### Table 17

*To-Scale Progression of Points from Kindergarten Through Second Grade*

<table>
<thead>
<tr>
<th>Kindergarten</th>
<th>First Grade</th>
<th>Second Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1Y</td>
<td>2Y</td>
<td>3Y</td>
</tr>
<tr>
<td>.1-.24</td>
<td>.25-.59</td>
<td>.60-.99</td>
</tr>
</tbody>
</table>
guide to coach users about how to prioritize skills or provide feedback or consultation about behaviors observed in readers, teachers may develop habits counter to the intent of the tool (Darling-Hammond, Wei, Andree, Richardson & Orphanos, 2009; Milburn, Girolametto, Weitzman & Greenberg, 2014; Rezzonico, et al., 2015). While the skills, their progressions, and complementary skills are clearly linked to research-based practices and standards, there remains a depth to each level that needs to be developed and cultivated in partnership with a practiced expert. ARC provides a variety of training opportunities for districts: initial introductory workshops, ongoing coaching in one-to-one settings, small group meetings, fishbowl demonstrations, and leadership coaching are among the most common. To understand the quantity of coaching provided for each of the five districts included in this study, professional development records (dates and levels only) from ARC were gathered and analyzed for the four years under consideration (T. Fields, personal communication, Dec. 5, 2018 and January 4, 2019). While the information gathered does not enable analysis of individual teachers or grade levels receiving the professional development, the general commitment of districts and schools can be ascertained by a simple tally of the quantity and type of professional development deployed. Table 18 denotes the amount of training each of the five districts received during the years under consideration in this study.
### Table 18

**Professional Development Received by the Five Districts Over the Four Years in the Study**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Train</td>
<td>Coach</td>
<td>Train</td>
<td>Coach</td>
</tr>
<tr>
<td><strong>BSD (33)</strong></td>
<td>D</td>
<td>2 days</td>
<td>31 days</td>
<td>14 days</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MRSD (4)</strong></td>
<td>D</td>
<td></td>
<td></td>
<td>2 days</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PPS (1)</strong></td>
<td>D</td>
<td></td>
<td>1 day</td>
<td>2 days</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td></td>
<td>20 days</td>
<td>27 days</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>FWSD (23)</strong></td>
<td>D</td>
<td>1 day</td>
<td></td>
<td>20 days</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CHSD (4)</strong></td>
<td>D</td>
<td></td>
<td></td>
<td>1 day</td>
</tr>
<tr>
<td></td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


### Recommendations for Implementing the IRLA

Learning to read is a complex activity, and the IRLA is a complex tool. The current study considered whether there would be an impact on third-grade summative test scores after three years of implementation of the IRLA and found no discernable change. By looking solely at SBA results, the vast multiplicity of conditions in each
of the 49 schools whose data were examined was not able to be taken into consideration.

Selection and implementation of the IRLA in conjunction with supporting materials is something that districts are advised to do in partnership with ARC leadership support and coaching. From the literature on implementation, a recommendation of this paper is that ARC use the drivers described in implementation science to help districts ensure the full complement of drivers are aligned to support implementation of this important innovation.

**Conditions for using the IRLA.** As discussed, the IRLA is a tool to help teachers ensure the development of skills in concert with one another and is to be employed in one-to-one conferences during 30-minute independent reading times embedded in a daily literacy block. In accordance with research on 1:1 instruction (Al Otaiba et al., 2011; Clay, 1993; Compton-Lily, 2009; Connor et al., 2006; Connor et al., 2009; Connor et al., 2013; Iversen & Tunmer, 1993; Juel & Minden-Cupp, 2000; Pinnell et al., 1988; Pinnell et al., 1994; Schwartz et al., 2012; Slavin et al., 1991) and the amount of independent reading students must do to make meaningful gains (Allington, 2014; Anderson, Wilson & Fielding, 1988; Beers & Probst, 2017; Paul, 1996), at least 30 minutes each school day are to be designated for students to read independently and for teachers to conference 1:1 with students to assess and instruct guided by the IRLA. Books at the designated IRLA levels of text complexity are optimal for students to both read and practice reading. Lastly, leadership that not only provides time and materials, but also that supports the inherent complexities of the IRLA and maintains a healthy educational environment for implementation is optimal.
This collection of conditions is neither guaranteed nor known in the schools being analyzed in this study. Inability to control for time spent on independent reading, instructional delivery, or leadership conditions presents further limitations to the study.

Clearly, there are many factors rendering a quantitative look at the impact of IRLA on summative state test results an ineffective way to study the impact of this tool. Reflecting on the limitations of the study illuminates issues having to do with the systematic and controlled implementation of the IRLA across school systems.

**Implementation Science as a Tool for Effective Implementation**

The findings and associated limitations presented in this paper are grounded directly in implementation science. Implementation science, as presented by the National Implementation Research Network (NIRN), considers the science to service gap of evidence-based innovations. Essentially, it asserts that although service organizations (in this case schools and school systems) know much about effective, evidence-based innovations such as the IRLA, gaps in the realization of the full potential of these innovations persist. There is no self-executing innovation, and NIRN presents a framework for the implementation of evidence-based innovations where intentional measures must be in place for an innovation to deliver its promised outcome. The limitations listed above for studying the implementation of IRLA in practice are all elements discussed in the implementation science literature, the “study of factors that influence the full and effective use of innovations in practice” (NIRN, 2018, para. 3). The factors that influence the full and effective use of the educational innovation that is the IRLA warrant consideration.
**Implementation drivers.** This research paper has determined that the inclusion and timing of the elements in the IRLA are appropriate and justified but failed to find statistically significant impact on learning as a result of the IRLA’s presence across 49 schools. Presence is not enough; rather, the evidence-based practices of the IRLA must be implemented with informed deliberation. Implementation is driven by the collective will of an organization. There are three categories of drivers at play in implementation science: competency, organization, and leadership (Fixsen, Blase, Metz, & Van Dyke, 2013; Fixsen, Blase, Naoom, & Duda, 2015). Competency drivers include selection, training, coaching, and performance assessment of staff. Organization drivers include: decision support data systems; facilitative administration and systems interventions; and leadership drivers focused on ensuring the right technical or adaptive leadership strategies for supporting the change management process.

Because of their complexity, school systems often fracture focus, asserting implementation of a variety of initiatives simultaneously, such as effective behavior supports, trauma informed practices, attendance, mathematics, tiered systems of support, and literacy (Schmoker, 2011). Implementation science takes the stance that in order to do something well, that thing must be of singular focus and the entire system should ensure adherence to that focus. Therefore, each driver is considered on an accountability scale, or implementation lens that is finely focused on the singular innovation being implemented (Fixsen et al., 2013). The implementation lens ranks each driver on a scale from pre-exploration, through exploration, installation, initial implementation, and full implementation. Systems considering implementing an
innovation are encouraged to rate each driver on this scale and ensure the will and ability to support it prior to implementing. Systems finding that an innovation is not taking root as expected are encouraged to use the scale to troubleshoot and make systems adjustments to provide a more fertile ground for the innovation to grow.

**Competency: Selection.** The selection driver refers to the recruitment and selection of staff. For example, if the IRLA is to be implemented in a school, the staff hired for that school should have the knowledge, skills, and abilities to teach reading (Mason & Schroeder, 2010). This should go without saying, but a system deeply dedicated to impacting reading outcomes must be able to discuss and enact hiring and retention practices that assure any new members to the teaching team are aligned with the vision for teaching reading inherent in the IRLA.

**Competency: Training.** For schools installing the IRLA as an innovation, training is crucial. The training driver acknowledges that an innovation requires that practitioners receive program knowledge from trainers with knowledge of the history, theory, philosophy and values of the program and can introduce the elements and practices required. With professional development, teachers can learn and try these elements in a supportive environment and trainers can gather formative and summative information on the initial performance of the newly trained teachers to inform future coaching interactions across the system and with individuals (Darling-Hammond & McLaughlin, 2011).

**Competency: Coaching.** To ensure that the elements of the innovation brought to teachers in its training are carried out with adherence to the program, such as the IRLA, ongoing and effective coaching is required. Implementation of an
innovation requires behavioral changes, and these do not happen quickly, easily, or accurately without advice, encouragement, feedback, and guided opportunities for practice (Darling-Hammond et al., 2009; Milburn et al., 2014; Rezzonico et al., 2015). Coaching provides continued formative data for the whole system to continue to improve coaching methods, teacher learning, and student outcomes.

**Competency: Performance assessment.** Evidence-based innovations have built-in criteria. Within the IRLA, entry requirements, one to one conferencing, engaged reading, and designation of power goals are criteria that teachers are expected to skillfully leverage, and use of these criteria signifies fidelity to the innovation. Training and coaching must be in place to ensure teachers know how to responsively measure entry requirements, finesse conferencing conversations and adapt to varying levels of engagement. But without a measure of fidelity in the form of a performance assessment, use can adapt in ineffective directions (O’Donnell, 2008). When systems use performance assessment to measure effectiveness of implementation, the data can inform necessary adjustments and identify common strengths. Support can be dispatched with a greater focus and chance for impact when the system knows what is working and what is not.

**Organization: Decision support data systems.** A school system must be informed by data for all staff to make good decisions. When an innovation provides data on a variety of levels, it can support the system in assessing its overall performance, can support decision making among grade levels and within classrooms, and can alert the system when evidence shows that something is succeeding or is amiss so corrections can be made or celebrations can be had (Bryk, Gomez, Grunow,
& LeMahieu, 2015; Wayman, 2005). The SchoolPace data system provides informative and actionable data for individual students, classrooms, grade levels, schools, and school systems in accordance with this driver. Where data systems exist and are immediately accessible and useful to all members of a school system, there is a strong indication that the organizational structure of the system is healthy enough to benefit from the implementation of an innovation (Metz, Naoom, Halle & Bartley, 2015).

**Organization: Facilitative administration.** Leadership of a school and district is a vitally important factor in the implementation of any initiative or innovation (Eagle, Dowd-Eagle, Snyder, & Holtzman, 2015). This driver explicitly describes the necessity of leadership using a variety of data inputs to inform, support, and organize staff around the focus of the innovation. Leadership can be either an individual or team that aligns policies, procedures, structures, culture, and climate with the needs of the practitioners to effectively implement the innovation (Nauman, 2017).

**Organization: Systems interventions.** “System interventions take on issues that impact the ability to provide effective services within organizations” (Fixsen et al., 2015, p. 21). This driver instructs that leadership within a system ensure availability of financial, organizational, and human resources to support implementation. Within this driver lay the elements that enable most of the other drivers. Without financial resources, the materials required for implementation of evidence-based innovations may be inadequately purchased, staffing may be shorted, and training opportunities may be limited. Any of the above variables can erode the evidence-based nature of the innovation. This is true also of organizational resources.
Weak warehouse structures can hinder accurate materials distribution, inadequate classroom shelving can interrupt students’ access to materials, and lack of storage solutions can contribute to damage or loss of materials over time. Supportive system interventions are peripheral but crucial to effective implementations.

**Leadership.** This driver represents a continuum of leadership from adaptive to technical. Systems are encouraged to ensure that leadership (teams or individuals) is adaptive in that they are willing to champion change, invest in innovation, and practice flexibility when necessary. Systems are also encouraged to ensure that leadership is technical in how they interview candidates, conduct performance assessments, and rally system interventions. Both types of leadership are required for the complex work of implementing an evidence-based innovation (Heifetz & Laurie, 1997).

**Recommendations for Further Study**

The review of reading literature supports the design, content, and delivery built into the IRLA, validating it as a powerful tool for the instruction of early readers. The lack of discernable results in the present study is due to a wide variety of unknown variables, inviting a host of further research opportunities. ARC is a company whose mission is to ensure all students are reading at grade level, and it produces and sells materials and professional development in service of that mission. A program analyst or research team may be worth consideration by ARC to actively explore areas exposed in this study so that school districts committed to impacting the learning of their students can rest assured that the IRLA and other ARC materials do what they set out to do. Regardless of who carries out further study, recommendations for continued investigation of the IRLA follow.
**Annual replication of the current study.** As noted in the limitations, the third graders whose test scores were used for this study had only been exposed to instruction guided by the IRLA for three of their four years of school. Additionally, the pilot situation limiting the number of students in the first year could have impacted the results. An annual update of the spreadsheet used for this study and continued ANOVA information would provide the opportunity to mark growth over time, potentially providing ARC ways to impact change for new and returning systems.

**Student-by-student growth within systems.** There are approximately 4,000 students represented in the data on the 49 schools included in this study whose data could be analyzed to yield a wealth of information. Important for ARC and schools (currently implementing and prospective) are answers regarding subgroups of students: is there even or disproportionate growth within programs such as special education or English learners? Are students of poverty growing at the same rates as higher SES populations? Are some classrooms seeing greater growth over time? Analysis of data to this level of granularity would require obtaining data through partnership between ARC, schools, and districts and could inform improvements of practice at all levels of the school/ARC relationship.

**SBA correlations to IRLA.** Predicated on an if/then relationship (if the IRLA and summative state test scores are correlated, then the focus on instruction guided by IRLA to increase IRLA scores will result in increased state test scores), correlational studies between IRLA scores and state summative assessment scores would be important to discover and relay. Finding this out will help us make the case to schools to invest time and effort in the skills and practices embedded in the IRLA rather than
falling prey to the distractions of test preparation or other assessment practices that do not provide teachers with the immediately actionable information that the IRLA directs. Discovery of correlational cutoffs will also bring confidence and measurable responses to the question that lingers from districts about whether color levels alone assure meeting proficient on state summative assessments, or whether more precision in the assignation of points is required.

**Interrater reliability study.** Data is entered to SchoolPace by a classroom teacher during a one-to-one conference. Data-focused districts approach scores obtained in this manner skeptically. Ongoing interrater reliability checks performed in accordance with the principles of the performance assessment driver outlined in the implementation science literature will enable understanding of whether the points and levels designated in SchoolPace are accurate to student ability and are not skewed by teacher relationship with student, teacher relationship with administration, teacher relationship with grade level teams, parent pressures, or other factors. A recommendation for future study of the IRLA’s impact on third grade summative assessment results is to design and conduct an interrater reliability study across a variety of schools and compare accuracy to outcome measures.

**Foundational skills toolkits.** The IRLA names skills across time. Deft teachers understand the skills and can effectively use conferences to identify and name the deficit skill, present strategies for practicing it, and set the student to work on the skill with text at the appropriate level of complexity. Not all teachers are equipped to do this without further support. To provide support, ARC released Foundational Skills Toolkits, one per color level at the yellow through red levels, with lessons for each of
the foundational skills named within the IRLA. The lessons are intended to support teachers as they gain knowledge of the skills for teaching early reading; they are not intended to be taught sequentially, nor are they intended to be used with all students. They are intended to support teachers and students at a point of need.

The toolkits have become very popular and fit a curricular desire. It would be worth doing a comparison study between school systems using the IRLA but not the toolkits, using the IRLA and the toolkits as intended, and using the IRLA and the toolkits as a sequence of lessons for all students. Building an understanding of the effectiveness of the toolkits will aid in the further development of this type of curricular tool as continued intervention for students not making expected progress, as teacher development, and as support for the reading taxonomy itself.

**Professional development.** Implementation science is clear about the need for initial professional development when installing an innovation and about the need for ongoing coaching. A thorough review of the coaching literature and analysis of levels accuracy and student gains on outcome measures would provide a reference point to share with districts in initial partnership stages to help design an implementation pathway. It is necessary that districts become self-sustaining in their implementation of any innovation, but models of gradual release vary. Understanding what optimal professional development plans are for the implementation of this innovation, the IRLA, would help districts plan for efficient and effective installation.

**Levels beyond 2R.** This study considered the foundational levels of the IRLA—grades kindergarten through two. The IRLA continues all the way through the 12th grade, emphasizing vocabulary acquisition, the use of genre structures and
author’s craft to support comprehension, and the expansion of historical and global context to enhance comprehension and invite critical interaction with the greater world. When students are in the levels above 2R, is there interaction or impact with summative state test results? This is an advisable next step in the research on the full impact of the IRLA.

**Implementation variables.** In light of the connection made between the innovation, the IRLA, and implementation methodology, a study assuring the conditions of an innovation as defined by Fixsen et al. (2013) and NIRN (2018) should be undertaken and findings used to potentially make revisions to the IRLA itself. Michie, Fixsen, Grimshaw, and Eccles (2009) claim the gap between the promise of an innovation and the result of an innovation persists due to inadequately described detail of the innovation itself. Upon assurance of adequate definition of the IRLA as an innovation, performance assessments must be designed and tested for a balance of rigid fidelity to some elements—such as 30 minutes each day—and flexibility of others—such as determining a power goal within point of observed need.

**Conclusion**

The ability to read supersedes every other academic skill, but in the United States, achievement data for fourth-grade students has remained virtually unchanged over the past 15 years. Countless studies and metastudies conducted over the past 50 years have yielded clear and actionable information about teaching and learning processes that teach students to read. However, achievement data remains stagnant, with unacceptable numbers of students still unable to read at grade level as they progress through the intermediate grades and into high school.
This dissertation argues that the landscape of knowledge for teaching reading is cluttered and unwieldy, overwhelming teachers and resulting in low levels of achievement. Using organizational schemes from the sciences, the Independent Reading Level Assessment (IRLA), tames the chaotic landscape of evidence-based practices by assembling them in a taxonomically based structure (Hambrick, 1984; Körner, 1970; Mosenthal, 1987; Travers, 1980) that capitalizes on the synergistic nature of the elements required for learning to read and the developmental sequences of each to provide teachers with a tool that enables them to simultaneously discover student strengths and uncover weaknesses in service of becoming readers.

The IRLA is a well-crafted tool. This study shows it organizes the best-known skills and sequences into columns of synergistic actions, enabling teachers to bring a world of literacy to students as they learn the necessary skills to turn print into meaning. The variables for realizing the potential of this tool in classroom, school, and district settings are diverse, and study of the IRLA cannot be relegated to a pure quantitative investigation.

School systems approach the IRLA with a variety of concerns, one of which is whether it will impact state test scores. While this question can easily be used to editorialize on the state of education in the U.S., it is a current unavoidable practicality that cannot be discounted by a publisher, and ARC should be able to conclusively respond with *of course it does!* However, at this time, ARC cannot demonstrate a direct line between the implementation of the IRLA and increased test scores with certainty, as the vastness of variables interfere with measurable outcomes.
While messy, ARC should engage in continued research of this tool and the conditions for its appropriate implementation. By entering in partnerships with districts who agree to a research-based rigor, ARC has fertile ground for studying the true impact of the IRLA. And, by using data from districts who do not agree to the level of rigor but purchase IRLA anyway as counterpoint, comparisons can be made to support the conditions necessary for this tool to help students realize their full potential as readers.


Dooley, C. M. (2010). Young children's approaches to books: The emergence of comprehension. *The Reading Teacher, 64*(2), 120-130. doi:10.1598/RT.64.2.4


O’Connor, R. E., Swanson, H. L., & Geraghty, C. (2010). Improvement in reading rate under independent and difficult text levels: influences on word and
comprehension skills. *Journal of Educational Psychology, 102*(1), 1-19. doi: 10.1037/a0017488


