An evaluation of a phonemic awareness program for middle school students and Its impact on academic achievement

Valdis Sigrun Rice
University at Albany, State University of New York, valdisrice@gmail.com

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd

Part of the Educational Psychology Commons

Recommended Citation

This Dissertation is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive. Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
An Evaluation of a Phonemic Awareness Program for Middle School Students and its Impact on Academic Achievement

by

Valdis S. Rice

A Dissertation
Submitted to the University at Albany, State University of New York
In Partial Fulfillment of the Requirements for the Degree of
Doctor of Psychology

School of Education
Department of Educational and Counseling Psychology

2021
ACKNOWLEDGMENTS

To begin to name all who have helped me during my time at the University at Albany over the past 10 years would be impossible, as I have had the privilege to work with and learn from amazing professors, supervisors, colleagues, students, and friends. I feel incredibly lucky to have so many individuals I can look to for wisdom, advice, and supervision.

To my chairperson, Dr. Deborah Kundert: a simple “thank you” does not begin to cover my appreciation of your immeasurable support and encouragement over the past few years. You have helped to make this process as enjoyable as possible. We laughed way more than I ever thought I would while working on a dissertation. I am so grateful for your intuitive ability to know when to give the extra nudge that I needed. Thank you for all the time, effort, and energy you have given to me and this project.

Thank you to my committee members, Dr. Dean Spaulding and Dr. Bruce Saddler, for your expertise, feedback, and thoughtful questions. I am so appreciative of your contributions.

My supervisors throughout the years, Dr. Britton Schnurr, Gwen Williams, Myles Goss, Dr. Callen Kostelnik, and Dr. Frank Salamone, have each uniquely demonstrated the perpetual influence a great mentor and teacher can have. Through my experiences working alongside them, I have developed more skills, confidence, and compassion than I thought possible.

Reflecting on my time as a student led me to realize that the people who surround me never doubted my abilities, even when I did. I have been cheered on throughout this process from start to finish by so many important people. Thank you to all my friends, my cohort, my Albany family, and family members who have helped me, especially my siblings, Margret and Johnny, and my parents, Helga and John, who each deserve at least their own chapter on all the amazing ways they have supported me.
An Evaluation of a Phonemic Awareness Program for Middle School Students and its Impact on Academic Achievement

Valdis S. Rice

University at Albany, State University of New York, 2021

Dissertation Chairperson: Deborah K. Kundert

ABSTRACT

With general instruction, most students’ reading skills develop overtime. For some, additional intervention in certain areas is necessary to acquire expected reading ability. One of five components of reading as identified by the National Reading Panel (NRP; National Institute of Child Health and Human Development [NICHD], 2000) included phonemic awareness (PA). Recognized as necessary to develop more complex skills, interventions for students’ PA abilities is thought by some to increase overall reading performance (Kilpatrick, 2013; 2015). One school district developed and implemented a new class program with the goal of increasing middle school students’ overall academic outcomes through targeted intervention with focus on PA within the class in response to the identification of students who were continuing to perform below grade level expectations in the area of reading as well as overall academically. This program uniquely addressed PA abilities in the middle school population, for which there is extremely limited research and recommendations related to best practice and implementation.

Archival data from the first year of the program were utilized to evaluate student outcomes one year after participation in the class. These data included report card grades and New York State (NYS) standardized test scores in English Language Arts (ELA) and Math. Analyses revealed no statistically significant differences in students’ overall academic performance across core class report card grades nor standardized test scores. Additionally, no
statistically significant differences in rates of passing the NYS ELA and math tests by demographic characteristics were found. Additional visual analyses were completed to offer the district additional information that may be of use in future program refinement despite non-significant findings utilizing data from the small sample. These analyses highlighted which academic variables students’ test scores or grades increased most, as well as which experienced minimal changes or decreased performance. This program evaluation included discussion of recommendations for the future of the current program and the district’s practices. Additionally, implications and limitations noted at the time of completion are presented.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title Page</td>
<td>i</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>ii</td>
</tr>
<tr>
<td>Abstract</td>
<td>iii</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>v</td>
</tr>
<tr>
<td>List of Tables</td>
<td>viii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>ix</td>
</tr>
<tr>
<td>List of Appendices</td>
<td>x</td>
</tr>
<tr>
<td>Chapter 1: Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Purpose of the Program Evaluation</td>
<td>3</td>
</tr>
<tr>
<td>Significance of the Program Evaluation</td>
<td>4</td>
</tr>
<tr>
<td>Chapter 2: Review of Literature</td>
<td>6</td>
</tr>
<tr>
<td>Program Evaluation</td>
<td>7</td>
</tr>
<tr>
<td>Program Evaluation Features and Approaches</td>
<td>8</td>
</tr>
<tr>
<td>Distinguishing Program Evaluation from Research</td>
<td>11</td>
</tr>
<tr>
<td>CIPP Evaluation Model</td>
<td>12</td>
</tr>
<tr>
<td>Improvement Sciences</td>
<td>14</td>
</tr>
<tr>
<td>Section Summary</td>
<td>17</td>
</tr>
<tr>
<td>Curriculum</td>
<td>17</td>
</tr>
<tr>
<td>Section Summary</td>
<td>20</td>
</tr>
<tr>
<td>Assessment</td>
<td>22</td>
</tr>
<tr>
<td>Collection and Utilization of Data</td>
<td>22</td>
</tr>
<tr>
<td>Current Status of Data Driven Decision Making</td>
<td>23</td>
</tr>
<tr>
<td>Accountability through Legislation</td>
<td>25</td>
</tr>
<tr>
<td>State Tests</td>
<td>26</td>
</tr>
<tr>
<td>Assessment within the Classroom</td>
<td>29</td>
</tr>
<tr>
<td>Section Summary</td>
<td>33</td>
</tr>
</tbody>
</table>
Reading ......................................................................................................................... 34
Reading Performance .................................................................................................... 35
History of Teaching Reading ...................................................................................... 36
Whole Language .......................................................................................................... 38
Phonics-Based Reading .............................................................................................. 39
Balanced Literacy ........................................................................................................ 40
Section Summary ........................................................................................................ 42
Phonemic Awareness .................................................................................................... 43
Differentiating Phonemic Awareness ........................................................................ 43
Identifying the Need for PA Instruction .................................................................... 44
Measuring PA ................................................................................................................ 46
Assets and Limitations of PA Instruction .................................................................. 51
PA Instruction with Middle School Students ............................................................. 52
Section Summary ........................................................................................................ 55
Chapter Summary ........................................................................................................ 55
Program Evaluation Questions .................................................................................. 57
Chapter 3: Methodology ............................................................................................... 58
Context .......................................................................................................................... 58
The District .................................................................................................................... 58
AIS Within the District ............................................................................................... 60
AIS Reading Class Development .............................................................................. 61
Participants ................................................................................................................... 63
Instrumentation ............................................................................................................ 63
Final Report Card Grades ......................................................................................... 63
New York State Tests ................................................................................................. 65
Procedure ....................................................................................................................... 68
Chapter 4: Results ......................................................................................................... 70
Data Review and Proposed Analyses ......................................................................... 70
Research Question One .......................................................... 70
Analyses .................................................................................. 73
Visual Inspection ................................................................. 73
Research Question Two .......................................................... 78
Analyses .................................................................................. 78
Visual Inspection ................................................................. 80
Chapter 5: Discussion .............................................................. 83
Program Evaluation Results ...................................................... 84
Implications ............................................................................. 86
Strengths ................................................................................ 87
Limitations and Recommendations ........................................... 87
Future Directions ................................................................. 96
Chapter Summary ................................................................. 97
References .............................................................................. 99
Appendices ............................................................................ 123
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Compilation of Plan-Do-Study-Act Best Practices</td>
<td>16</td>
</tr>
<tr>
<td>2. Traditional Features of Formative and Summative Assessment</td>
<td>30</td>
</tr>
<tr>
<td>3. Activities to Build Phonemic Awareness</td>
<td>47</td>
</tr>
<tr>
<td>4. Typical Schedule of Mastery Over Phonological Skills</td>
<td>54</td>
</tr>
<tr>
<td>5. District Demographic Characteristics</td>
<td>59</td>
</tr>
<tr>
<td>6. Participant Characteristics</td>
<td>64</td>
</tr>
<tr>
<td>7. 2018 and 2019 NYS State Test ELA Performance Levels and Scale Score Ranges</td>
<td>66</td>
</tr>
<tr>
<td>8. 2018 and 2019 NYS State Test Math Performance Levels and Scale Score Ranges</td>
<td>67</td>
</tr>
<tr>
<td>9. Descriptive Statistics for Academic Achievement Variables</td>
<td>71</td>
</tr>
<tr>
<td>10. Percentage of Missing Data by Variable</td>
<td>72</td>
</tr>
<tr>
<td>11. Results of Analyses Examining Academic Achievement</td>
<td>74</td>
</tr>
<tr>
<td>12. Results of Associations between NYS ELA Test Score and Demographics</td>
<td>81</td>
</tr>
<tr>
<td>13. Results of Association between NYS Math Test Score and Demographics</td>
<td>82</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Feedback Loop for Educational Program Evaluations</td>
<td>9</td>
</tr>
<tr>
<td>2. CIPP Evaluation Model Process</td>
<td>13</td>
</tr>
<tr>
<td>3. Instructional Hierarchy</td>
<td>21</td>
</tr>
<tr>
<td>4. Assessment Decisions</td>
<td>24</td>
</tr>
<tr>
<td>5. Phonological Awareness Activities Arranged by Increasing Complexity</td>
<td>45</td>
</tr>
<tr>
<td>6. Development of Phonological Skills</td>
<td>49</td>
</tr>
<tr>
<td>7. Changes in ELA Class Grades</td>
<td>75</td>
</tr>
<tr>
<td>8. Changes in Math Class Grades</td>
<td>75</td>
</tr>
<tr>
<td>9. Changes in Social Studies Class Grades</td>
<td>76</td>
</tr>
<tr>
<td>10. Changes in Science Class Grades</td>
<td>76</td>
</tr>
<tr>
<td>11. Changes in NYS ELA Test Scaled Scores</td>
<td>77</td>
</tr>
<tr>
<td>12. Changes in NYS Math Test Scaled Scores</td>
<td>77</td>
</tr>
</tbody>
</table>
## LIST OF APPENDICES

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Research Questions and Proposed Data Analyses</td>
<td>124</td>
</tr>
</tbody>
</table>
CHAPTER 1: INTRODUCTION

Within the scope of employment, school teachers and administrators are tasked with seemingly countless objectives. McIntyre et al. (2011) labeled teaching children to read as having the potential to be the most difficult task teachers face. Students who are labeled as struggling to read or at risk of not meeting expectations may have difficulty related to language ability or have less exposure to literacy outside the school setting (Carnine et al., 2006). Reading classes, particularly classes designed for students not meeting expectations, are created in an attempt to help students succeed. Determining whether a program is effective in positively impacting student performance can help teachers to meet goals as well as enable greater support of student needs.

Despite being responsible for educating an ever-changing culturally and linguistically diverse student-body (Merrell et al., 2012; National Center for Education Statistics, 2018), navigating achievement gaps between student groups (Miranda, 2014), and incorporating technology in meaningful ways to allow for appropriate instruction (McManis & Gunnewig, 2012), teachers largely rely on methods that were not developed to prioritize specialized instruction to students not meeting expectations (Detterman & Thompson, 1997). To understand what impact a specially designed class, or program, has had on students, school districts can closely examine success through program evaluation. Stufflebeam and Coryn (2014) identified evaluation as a key aspect to societal advances, allowing for the identification of favorable as well as disadvantageous outcomes through assessment of interests such as quality or efficiency. Program evaluation can be used to inform future planning and recommendations and may be of particular use when considering the design and implementation of a new reading class. A formal program evaluation allows for systematic examination of practices to assess quality (Joint
Inclusion of improvement science approaches in conjunction with traditional program evaluation techniques creates opportunities for stakeholders to receive additional information related to program creation and implementation through investigations of pre-planning and problem-solving techniques.

Successful reading curricula must consider research to support instruction as well as content design and delivery (McIntyre et al., 2011). The design of a curriculum for students at varying instructional levels with different background knowledge and supports can be challenging but is of high importance, as it allows students to access class material (National Association for the Education of Young Children, 2009). Schools may opt to provide intervention services through a Response to Intervention (RtI) approach in order to meet individuals’ needs regarding services, which should be determined by evidence-based assessment (Lichtenstein, 2014). Alternatively, as indicated by the NYS Education Department (NYSED, 2015) Part 100 Regulations, schools can provide academic intervention services (AIS). Students identified as qualified for AIS are in particular need of efficiently and effectively planned lessons, as the purpose of AIS is to propel students to meeting grade level expectations. While reading programs generally should be designed with the majority population in mind (Carnine et al., 2006), AIS class designs should focus on the specific needs of identified students. Proper utilization of varying assessment techniques can, in part, help teachers to adjust future instruction to allow for students to access content and benefit from lessons. Assessments provide school staff with information regarding day-to-day progress as well as learning that has occurred at the conclusion of instruction (Roskos & Neuman, 2012). To measure student growth, curricula should incorporate various assessment measures to accurately identify students’ skills.

Knowledge of competing reading theories, as well as current recommendations for the
most effective reading practices, feeds the content of curricula. The popularity and application of reading theory principles have experienced ups and downs, with more recent support for balanced literacy techniques within the classroom. Balanced literacy relies on a combination of explicit phonics instruction alongside focusing on texts’ meaning (Freppon & Dahl, 1998), marrying the phonics-based and whole language theories previously widely used. The National Reading Panel (NRP; National Institute of Child Health and Human Development [NICHD], 2000) highlighted teaching from a balanced literacy lens as the most effective, specifically identifying advantageous practices related to five specific areas of reading. One of these five facets included phonemic awareness (PA). PA involves the ability to distinguish and manipulate phonemes (Scanlon et al., 2010). It has been called a prerequisite skill for successful reading (Daly et al., 2015). Kilpatrick (2013; 2015) proposed that students, of any age, receive phonemic awareness (PA) intervention and instruction if they do not demonstrate appropriate skills, as shown through data from assessment measures. Yet, available literature related to PA instruction for students who are not in third grade or below is limited. Further, it is unclear whether increased PA abilities after intervention for older students impacts overall reading success and general academic performance.

**Purpose of the Program Evaluation**

This study evaluated a middle school’s introduction of a new program entitled Academic Intervention Services (AIS) Reading utilizing archival data from students in grades six through eight. The purpose of this evaluation was to identify the impact of a novel reading program on students’ overall academic achievement, as determined by performance on New York State (NYS) tests and academic report card grades across subjects. In addition, potential differences in outcomes based on student demographics were assessed. Information regarding the impact
curriculum and implementation has on students can be used to inform future class development, including recommendations for adapting new practices, with the goal of increasing AIS Reading’s efficiency and effectiveness.

**Significance of the Program Evaluation**

The proposed program evaluation provides the district specific and individualized feedback regarding a new class initiative aimed to increase students’ overall academic success through increasing phonemic awareness ability. Research regarding PA development and fluidity among middle school students is scarce. The current program evaluation is significant as it can offer contributions to the limited literature available. Utilization of explicit instruction methods has been recommended in order to effectively teach reading to elementary students, and specific inclusion of PA instruction is included across core reading programs commonly used (Reutzel et al., 2014). Yet, Reutzel and colleagues (2014) found reading programs did not advocate for continued progress monitoring of PA abilities. Further, the majority of research evaluating the effectiveness of PA interventions focuses on those that occurred for about 20 weeks and have found moderate effects (Hulme & Snowling, 2013), identifying a need for research on longer lasting interventions and potential impacts of implementation. The current district’s class design aimed to address these pitfalls. The program evaluation offers insight regarding what impact the program had on students’ academic achievement.

The proposed program evaluation is also important as it can provide the district with evidence for future planning based on data. The AIS Reading class proposed and implemented by the school district in focus was designed to span the school year, utilize explicit instruction techniques, and continually assess students’ progress, contrasting the participants and practices studied in previous research. The district will receive conclusions based on data centered on the
students and practices teachers and staff may wish to maintain or adapt, as evaluations can provide information related to all associated stakeholders (Joint Committee on Standards for Educational Evaluation, 2011). Information can be used to determine future characteristics of the class, such as the amount of instructional time or the length of the class offering (i.e., full year vs. half year). The results of this program evaluation can provide school administrators with insight to maintain practices and make changes to allow for more effective and efficient programming in future years.
CHAPTER 2: REVIEW OF LITERATURE

As one aspect of literacy, reading receives significant attention from varied audiences, including individuals such as students and parents, school staff members, researchers, and government officials. There has been much debate regarding how to best teach reading. To determine what instructional approaches are successful, program evaluation methods can be utilized. Program evaluation is increasingly used to examine educational initiatives and programs (Gall et al., 2015; Stufflebeam & Coryn, 2014). When coupled with improvement science techniques, evaluations can offer recommendations for design and refinement of programs (Castillo, 2020). Program evaluations combined with improvement science practices differ from traditional research, particularly in the ability to provide timely and relevant information to stakeholders about a specific initiative (Crow, 2019; Spaulding, 2014). Evaluations focused on educational programs must consider and understand the importance of curriculum, which drives what content is included and how instruction is shaped. Curricula guide all instruction, with a major component, assessment, to inform teachers’ decision-making processes regarding students’ progress and understanding of content.

Initially, program evaluation and improvement science will be discussed as they relate to educational programs, providing the context on which this proposed study was based. A closer examination of theoretical approaches to teaching reading across time is then presented, illuminating the reasons behind widespread use of balanced literacy approaches in current schooling. Throughout history, approaches to teaching reading have fluctuated between competing theories of how students learn to read resulting in what Chall (1967; 1983) referred to as the “Reading Wars” between differing approaches. The NRP (NICHD, 2000) highlighted balanced literacy approaches as the most effective and noted five aspects of reading instruction
that are considered vital for students’ learning. Stemming from the research supporting balanced literacy as the most effective method for teaching reading, PA is described in detail. PA was labeled as one of these important skills, however, limited information regarding instruction across age ranges is available, especially for secondary students. PA is distinguished from other aspects of phonological awareness and reasons supporting the teaching of PA are explored. Research demonstrates the utility of explicit PA teaching for younger students (Daly et al., 2015; Kilpatrick, 2013, 2015; McEwan, 2009; Moats, 2010; Scanlon et al., 2010). While it has been argued that anyone with PA deficits should receive instruction, regardless of age (Gillon, 2018; Kilpatrick, 2013), clearly defined strategies and suggested implementation for older students who are in middle school remains unclear. The level of students’ PA deficits at the state and national level, including what PA deficits may be present among older students, are currently unknown. Specific information regarding the percentage of students believed to have PA deficits within the identified district in focus is also currently unavailable, as PA is not typically specifically assessed in all students unless concern regarding a deficit exists. Further, researchers’ assertions regarding the impact of PA instruction for older students on overall academic performance conflict. Proposed questions for the current program evaluation are offered at the conclusion of the chapter.

Program Evaluation

Successful program evaluations in the school setting provide opportunity to allow districts to utilize data to offer higher quality services and, in turn, benefit the school community (Salvia et al., 2017). Program evaluations include the collection and examination of data in order to inform individuals how closely the implementation of a specific program aligns with its intended purpose (Spaulding, 2014). The Joint Committee on Standards for Educational
Evaluation (2011) asserted program evaluations can and should be utilized to further develop and refine the needs of a program. Program needs are viewed as required for the desired consequences to be possible (Stufflebeam & Shinkfield, 2007). Use of program evaluation within the field of education specifically has increased, in part due to funding demands by the federal government as well as grants (Gall et al., 2015; Stufflebeam & Coryn, 2014). Renewed interest and development of program evaluations occurred in the late 1950s, stemming largely from a desire to strengthen government initiatives (Stufflebeam & Coryn, 2014). Successful evaluations can help to reduce delayed or uninformative feedback to programs, further increasing the potential quality of services and programs offered (Linfield & Posavac, 2019).

**Program Evaluation Features and Approaches**

Either an internal or external evaluator(s) conducts an evaluation. External evaluators are not employed by the program’s institution, unlike internal evaluators (Spaulding, 2014). Both types of evaluators can bring their own strengths to the process. While internal evaluators may possess increased familiarity and knowledge of the program, as well as a working relationship with stakeholders, external evaluators may offer a greater degree of impartiality (Spaulding, 2014). Depending on the objectives, program evaluations may produce formative and summative data to align with the goals of evaluation. Linfield and Posavac (2019) labeled feedback as essential for successful organizations. Displayed in Figure 1 is the feedback loop system explained by Linfield and Posavac (2019) as adapted to fit educational programs. Scriven (1967) asserted formative evaluations are undertaken to establish decisions to help to design and form the program. As such, formative evaluations occur more often while programs are being developed and refined (Spaulding, 2014; Stufflebeam & Coryn, 2014). Alternatively, summative evaluations produce data that are presented at the end of a specific project or program endeavor.
Figure 1

*Feedback Loop for Educational Program Evaluations*

Program Evaluation

---

Plans and Financial Commitment (Inputs)

Education Programs (Outputs)

Program Results (Outcomes)

School Building
School District
State Education Department
Professional Groups
Local Education Agencies

and are used to examine outcomes related to how successful the program was (Scriven, 1967). Summative findings can be used to make informed decisions, such as whether stakeholders decide if a program should continue to be implemented, be discontinued, or be altered in some way (Linfield & Posavac, 2019).

Stufflebeam and Coryn (2014) further distinguished program evaluation through the identification and delineation of 23 major approaches that fit within five categorizations based on the review of seminal evaluation literature. These categories include: pseudoevaluations, quasi-evaluation studies, improvement- and accountability-oriented evaluation approaches, social agenda and advocacy evaluation approaches, and eclectic evaluation approaches (Stufflebeam & Coryn, 2014). Of these approaches, 17 were noted as legitimate methods for producing informative evaluations assuming the approach is correctly applied (Stufflebeam & Coryn, 2014). While commonly utilized, the six approaches categorized as pseudoevaluations should not be considered as sound approaches to evaluation (Stufflebeam & Coryn, 2014). Quasi-evaluation models yield stakeholders with timely data in relation to specific questions posed, but may be unable to capture a program’s success in whole as a result of this narrow focus (Stufflebeam & Coryn, 2014). Improvement- and accountability-oriented evaluations engage in the study of multiple facets of a program (e.g., cost, feasibility, equity) and typically have a long-term focus on determining a program’s value (Stufflebeam & Coryn, 2014). Through social agenda and advocacy approaches, evaluations are completed to target enhancing social justice, largely through equity (Stufflebeam & Coryn, 2014). Lastly, eclectic approaches are varied methods that are based in pragmatic practices as opposed to a single theoretical or methodological approach, while still adhering to evaluation standards and proper technical applications (Stufflebeam & Coryn, 2014).
Distinguishing Program Evaluation from Research

Program evaluations are intended to aid in data-based decision making within a specific application, while research is generally conducted to discern information that is relevant across various settings (Spaulding, 2014). Within research, objectives are largely set by the principal investigator(s) while program evaluations’ foci are typically determined by stakeholders and the client in conjunction with one another (Fitzpatrick et al., 2011). Evaluations should illuminate needs and future directions for programs’ stakeholders (Fitzpatrick et al., 2011; Joint Committee on Standards for Educational Evaluation, 2011). They are completed to clarify areas in need of change, resulting in greater significance upon immediate execution of recommendations than historically found in research (Spaulding, 2014). Some program evaluation models contend that evaluations that produce recommendations and findings that are not applied are to be a poor use of evaluators’ time and, overall, pointless (Patton, 1997; Spaulding, 2014). These distinguishing features set evaluations apart from research. Results of program evaluations are typically shared directly with the stakeholders, as opposed to published, further discerning research and program evaluations (Fitzpatrick et al., 2011; Spaulding, 2014).

The Joint Committee on Standards for Educational Evaluation (2011) put forth 30 standards for conducting evaluations with the aim of increasing the quality of evaluations performed with regard to how well stakeholders’ needs are addressed. These standards were divided amongst five categorizations: utility, feasibility, propriety, accuracy, and accountability (Joint Committee on Standards for Educational Evaluation, 2011). Stufflebeam and Coryn (2014) labeled these standards as a vital tool to reduce the occurrence of poorly executed, deceitful, and unnecessary evaluations, as they hold evaluators to an expectation of producing useful and informative summaries (i.e., utility), engaging in non-invasive and viable
investigations (i.e., feasibility), considering legal and ethical principles (i.e., propriety), sharing factual and adequate information (i.e., accuracy), and engaging in assessment of how closely the evaluation meets these standards (i.e., accountability). Alternatively, Mathison (2008) stated research is customarily evaluated through examination of validity. Validity measures include whether the research can demonstrate findings are due to causality as well as how widely results can be applied (i.e., internal and external validity, respectively). To demonstrate validity, researchers must utilize tests and procedures that are found to be reliable and valid. A valid test measures what it was intended to measure (Goodwin & Goodwin, 2017; Sattler, 2008). Reliability is required in order to attribute meaning to scores yielded by measures (Goodwin & Goodwin, 2017). Valid tests must be found to be reliable (Reynolds & Livingston, 2014). While both research and program evaluation use guiding principles to systematically examine specified questions or objectives (Morrison & Harms, 2018), the two must be clearly outlined.

**CIPP Evaluation Model**

The current program evaluation approach, CIPP evaluation model, falls within the improvement and accountability-oriented evaluation categorization, as its objectives include providing stakeholders with information while the evaluator(s) maintains a close relationship with stakeholders who utilize findings to make decisions (Stufflebeam & Shinkfield, 2007). CIPP includes examination of context, input, process, and products (Stufflebeam & Shinkfield, 2007). The current program evaluation looked most closely at the process and products. The steps of the CIPP model are seen below in Figure 2. Advantages of the CIPP model include its inherent link to ongoing refinement by working to create a system that is providing feedback constantly to promote future changes and adaptations for more favorable products (Stufflebeam & Shinkfield, 2007).
Figure 2

CIPP Evaluation Model Process

Program

Context is Evaluated

Change needed?

No

Yes

Define and Identify Goals and Barriers

Solution in Place

Solution Found

Yes

No

Input is Evaluated

Solution Found

Yes

No

Need to Test/Pilot?

Yes

No

Develop and Test Solution

Process is Evaluated

Still Worthy of Investment?

Yes

No

Still Worthy of Investment?

Yes

No

Objective Met?

Yes

No

Abandon Program

**Improvement Sciences**

The field of improvement science offers novel approaches to enhancing educational programs. These frameworks can be applied individually, or as a complement to program evaluation techniques. Improvement science methods can also be employed as a precursor or reaction to completed program evaluations. Similar to program evaluation, improvement science is rooted in enhancing procedures in place and is increasingly noted by educational leaders as an area of interest (Crow, 2019), and distinguishes itself from traditional research (Crow, 2019; Spaulding & Hinnant-Crawford, 2019).

The methodology of improvement science provides unique frameworks for all aspects of improvement (Spaulding & Hinnant-Crawford, 2019). There is emphasis placed upon understanding the relationships between various levels of stakeholders within an organization or practice being studied with improvement sciences methods, as there is recognition that, while top level individuals may determine how or what programs are implemented, others within organizations are responsible for the daily application (Castillo, 2020). It is this emphasis, in part, that allows for meaningful change at a systemic level (Castillo, 2020). Bryk and colleagues (2015) highlighted the importance of specifying what accomplishments are desired, what changes must occur for improvement to take place and why are these changes effective, and how to measure the impact changes have on improvement in order to address all rudimentary principles of the improvement science model.

Castillo (2020) identified improvement science methods, also referred to as implementation science, as being uniquely appropriate for the application and refinement of practices as the methods incorporate great focus on collaboration and the process of practices. Castillo (2020) asserted program evaluation approaches are more useful for improving the
implementation of evidence-based practices on larger scales, as evaluation methods can be utilized to also dictate future literature. Implementation stages are utilized to determine how evidence-based practices should be planned, implemented, and judged (Bertram et al., 2015); however, debate exists regarding how to evaluate these stages. Saldana (2014) proposed use of a tool that spans varied phases of practice implementation, citing a need for increased standardized evaluative measures. Christie et al. (2017) labeled improvement sciences as more functional than program evaluation in some regard, due to the clearly expressed directions provided for stakeholders and organizers to use moving forward. Bryk et al. (2015) highlighted the importance of using improvement science techniques to enable organizations, particularly school systems, to carry out evaluations with high utility. To do this, emphasis on small scale changes and continuous assessment were made, as opposed to a tendency to use program evaluation as a way to correct overarching problematic design and implementation after application (Bryk et al., 2015). These small-scale foci help to prevent what Bryk et al. (2015) termed solutionitis, which occurs when stakeholders act too quickly before true consideration and comprehension of problems faced.

Langley et al. (2009, p. 5) explicitly outlined the following three questions as necessary to facilitate effective improvement efforts: What are we trying to accomplish?; How will we know that a change is an improvement?; What changes can we make that will result in improvement? In conjunction with a Plan-Do-Study-Act (PDSA) model, these questions guide stakeholders toward answering outlined objectives (Langley et al., 2009). The PDSA model involves four phases of program implementation, beginning with planning and ending with ongoing sustainability practices (Langley et al., 2009). Best practices for the PDSA are found in Table 1.
Table 1

Compilation of Plan-Do-Study-Act Best Practices

<table>
<thead>
<tr>
<th>Stage</th>
<th>Tasks</th>
</tr>
</thead>
</table>
| Exploration and Planning   | - Develop an implementation team  
- Assess elements of implementation including (a) feasibility, (b) potential barriers, and (c) overall goodness of fit between program and needs  
- Conduct a needs assessment and evaluate staff and organization readiness  
- Assess match between community needs, EBP needs, and resources  
- Facilitate focus groups with staff (and youth, if applicable) to discuss the changes that will be coming  
- Plan for data collection                                                                                      |
| Initial Installation       | - Implementation team(s) begin to work directly with program developers, external consultants, and/or other important stakeholders  
- Assure the availability of resources necessary to install implementation infrastructure (e.g., staffing, space, new policies)  
- Develop practitioner/staff competence through trainings  
- Observe, describe, and document initial installation efforts  
- Test implementation on a small scale; acknowledge that implementation may need to be adjusted or discontinued based on incoming data |
| Full Implementation        | - Address barriers and develop solutions quickly as issues arise  
- Ensure funding streams are reliable and adequate  
- Provide ongoing trainings  
- Have system in place for continued training, coaching, performance assessment protocols, measuring fidelity, and using data to make decisions for continued improvement  
- Achieved when 50% of providers are routinely using the new practice with high fidelity  
- Study the results of implementation                                                                                   |
| Improvement and Sustainability | - As necessary, make cultural adaptations to the intervention and/or implementation process  
- Provide ongoing trainings  
- Identify and problem-solve program drift and/or threats to fidelity  
- Provide ongoing technical assistance in the realm of data collection  
- Facilitate reflective group practices to collect qualitative data from staff/practitioners |

Note: Table from Sweet, M. (2021). Initial implementation of the ARC model in a residential treatment center: A program evaluation [Unpublished doctoral dissertation]. University at Albany; Information for this table was gathered from Barnett et al. (2018), Chance et al. (2010), Fixsen, et al. (2005), Langley et al. (2009), Metz and Albers (2014), and National Implementation Research Network.
At the time of this evaluation, the program being examined could be classified as within the ‘Study’ stage of the PDSA model. According to reports from the district, the planning phase involved a teacher training from an individual with expertise in PA, time to create and organize the curriculum to be utilized with the class, and identifying and securing materials to be used throughout the year. The second phase is represented by the school year in which teachers worked directly with students to implement the class designed. Continued collaboration between teachers, small scale implementation (limited participants), and the creation of new class sections throughout the year align with some best practice recommendations during this phase of the PDSA model.

Section Summary

Program evaluations are one source utilized by stakeholders to make decisions regarding programs. As such, Linfield and Posavac (2019) underscored that findings and recommendations are often not directly applied due to reliance on other sources by programs. Program evaluation models, however, offer organizations a route to improved programs based on collection and analysis of program specific data, allowing for recommendations directly aligned with organizational resources. Program evaluation differs from research in this way, and complements improvement science methodologies.

Curriculum

The delivery of a new educational program within a school involves the creation and design of a new curriculum, which is defined as what will be instructed and how instruction will occur (Apple & Au, 2018; Behar-Horenstein, 2018; Kelly, 2004). Theory drives understanding of concepts, but curriculum planning allows for these concepts to be taught. Ikeda et al. (2010) posited that before intervention, or the introduction of a supplemental class, districts should
ensure that the general curriculum is effective (i.e., the majority of students should be making sufficient progress). In recent years, curricula have received scrutiny regarding equity and inclusion (Luke, 2013), further emphasizing the significance of well-designed curricula allowing access for all students. As noted by Kamps et al. (2008), successfully remedying reading deficits in particular requires the knowledge and resources to target these issues through the implementation of interventions, or curriculum in the case of this reading class, in addition to accurately identifying the problem.

Emphasis on evidence-based interventions and curricula that utilize them has grown significantly over time, particularly in response to the push for accountability. Yet, intervention use and assessment with high fidelity is difficult for many teachers (Benner et al., 2011). At the middle school level, varied intervention fidelity has been demonstrated to negatively impact student gains from implemented practices (Benner et al., 2011). These challenges may be due, in part, to a lack of resources or support for teachers (Kamps et al., 2008), or unintentional error in implementation and assessment that go without correction (Reed et al., 2014). Teachers who are responsible for content not typically targeted for certain age groups or abilities may be faced with creating curricula by utilizing various methods, such as combining evidence-based manualized treatment plans or previously implemented reading curricula with increased attention in the content area of focus. The complexity of organizing and disseminating support for new initiatives or programs is particularly notable in secondary level schools as middle school students have many teachers across different subjects. Boyle and Charles (2016) underscored the importance of support, particularly for teachers, when considering the creation of a curriculum to be maintained over time. In upper elementary years, curriculum is commonly divided by class subject (Apple & Au, 2018). Deng (2013) differentiated disciplines from subjects, asserting that
select subjects may represent only portions of the discipline they belong to and indicate a need for specific content focus. For example, deficits in reading across grade levels can negatively impact students’ performance across class subjects and overarching disciplines, leaving them unable to access curriculum in areas other than their reading or English Language Arts classes (Cromley, 2009; Espin & Deno, 1993; Fiester, 2013; McArthur & Castles, 2017). School staff may engage in curriculum mapping to identify content and selected skills expected to be mastered by students within certain grade levels (Breymier, 2018). Engaging in curriculum mapping processes requires explicit definition of expected outcomes and the identification of the content that will result in these objectives, creating known links between action and expectations (Breymier, 2018).

Behar-Horenstein (2018) differentiated between curriculum as a plan, which focuses on what a teacher aims to instruct, from curriculum as a system, which denotes the organization and processes driving content selection and delivery. Howell and Nolet (2000) extended these distinctions between three categorizations: intended curriculum (i.e., curriculum as a system), taught curriculum (i.e., curriculum as a plan), and the learned curriculum, which captures the knowledge and skills students actually acquire. All aspects must be considered by educators for success.

Across all curricula and developmental levels, the Instructional Hierarchy (IH; Haring & Eaton, 1978) can be utilized to guide general instruction. Evidence based instruction of basic skills including reading, writing, and math abilities often employs IH approaches (Ardoin et al., 2018; Martens et al., 2007). The IH is applicable to general curriculum as well as targeted intervention instruction, while providing teachers with an inherent connection between the structure and delivery of content and assessment. Utilization of an IH can aid teachers in
accurately introducing material that matches students’ current performance levels (Haring & Eaton, 1978). The IH includes four levels within its model, designed to represent the process that takes place as learning occurs (Haring & Eaton, 1978). This model is visualized in Figure 3. According to Haring and Eaton (1978), the IH begins with acquisition, which entails attaining a new skill with accuracy. Accurate responding during the acquisition phase is vital, as students should not be pushed to advance until materials are understood correctly. Following acquisition in the IH, students develop fluency. Fluency involves accurate responding in an automatic fashion. When students have demonstrated acquisition of knowledge and fluency, generalization follows. The ability to apply skills in novel situations demonstrates successful generalization. Fourth, students should engage in adaptation. Adaptation involves utilizing skills in modified ways in specific novel experiences. Emphasis on the process involved in reading skill development highlights the significance of proper instruction (Scanlon et al., 2010). Use of the IH within curricula design enables teachers to systematically expose students to foundational skills before more complex applications. Teachers must be attuned to developmentally appropriate practices for individual students and direct focus on skills logically expected to follow.

Section Summary

Curriculum determines what is taught and how instruction is provided. It provides a necessary map of content and instruction to guide teachers within the classroom. In addition to general classroom teaching, curriculum should also inform interventions and allow for the targeting of specific skills. The match between curricula and interventions is well established at the elementary level, but curricula design can be particularly challenging at the middle school level.
Assessment

Student assessment should be routinely used to gather data regarding pupil progress, benefiting both individuals as well as larger school systems (Schaughency et al., 2010). Results inform teacher decision making regarding class content, a need for differentiation, or need for increased exposure to materials. The data gathered through assessment feed data-based problem solving (Pluymer, 2014), allowing for the identification of problems, explanation behind its occurrence, and implementation of intervention and progress monitoring to decrease or eliminate the issue. The collection of assessment data can be done using various approaches, and the decisions that are made as a result of these data are often dependent upon educators’ objectives and goals. The current program evaluation focuses specifically on tests as the main method of collecting assessment data. Mandinach (2012) identified data-driven decision making (DDDM) as non-negotiable within the school setting, calling attention to the need for methodical collection and analysis of assessment data to motivate decisions. The need for DDDM was strengthened by the introduction of federal legislation explicitly requiring monitoring and displays of student progress.

Collection and Utilization of Data

Salvia and colleagues (2017) asserted ways in which data are collected as part of the assessment process, and include record reviews, interviews, observations, and testing. Hosp et al. (1987) first termed these four methods RIOT. Though often immediately thought of when considering assessment, testing is only one approach to data collection. Record reviews should be completed when working with students to further elucidate information that may inform future instruction, such as when difficulties may have begun or what interventions a student has previously received (Salvia et al., 2017). Interviews, which can include casual conversation or
standardized rating scales, can be used to collect background data from various sources (Salvia et al., 2017). According to Salvia et al. (2017), both systematic and non-systematic observations can help to supply additional data about a student, especially when considered within the context of a particular setting. Lastly, tests serve to collect data that offers more precision than other methods (Salvia et al., 2017). It is advised that tests are included when making decisions for students that may impact instruction (Salvia et al., 2017).

The data compiled from these different assessment methods can be utilized to direct future decisions depending upon the objectives of stakeholders, as displayed in Figure 4. Decisions based on data range from progress monitoring to determining when instruction should be modified and beyond (Salvia et al., 2017). Different assessment methods are the tools by which educators are able to address many of the questions schools face with regard to demonstrating student progress.

**Current Status of Data Driven Decision Making**

Legislative actions, discussed in more detail in the following section, have called for implementation of DDDM to increase student outcomes, however, no specified applications have been created or mandated. Hargreaves and Braun (2013) have asserted that utilization of DDDM within the schools is more closely related to complying with mandates than as a way to impact outcomes. Dunn et al. (2013) emphasized teachers’ low self efficacy in effectively utilizing DDDM within the classroom, further hindering its application. While there have been calls to revitalize DDDM to more closely align with improvement (Mandinach, 2012), uniform processes do not exist. Adoption of DDDM at the national and state levels vary. Manidinach and Schildkamp (2021) highlighted the importance of using more than just test scores to understand student performance, including student demographics, health, and environmental factors as key.
Figure 4.

*Types of Assessment Decisions and their Purposes*

- **Screening**
  - Identify unrecognized issues
- **Progress Monitoring**
  - Track progression toward individual goals/general expectations
- **Accountability**
  - Determine progress toward objectives
- **Program Evaluation**
  - Determine effectiveness of programs
- **Plan/Modify Instruction**
  - Increase competence and capacity
- **Special Education Eligibility**
  - Conclude eligibility for special education/related services
- **Resource Allocation**
  - Determine what resources are necessary

Within the current district, there were multiple pathways to access AIS through the district’s defined DDDM process, including examination of NYS assessments of all students and district assessments. The AIS identification process is further discussed in chapter three. Within New York, NYSED (2010) has provided guidance to districts with regard to what minimum RTI requirements and data should look like, but individual districts may have individualized plans. Day to day, decisions by school personnel may be made by different people using different techniques and emphasizing different areas of data. As is commonly noted in education, districts nationwide are lagging behind best practice recommendations, whether it be due to misuse or misunderstanding of data (Manidinach & Schildkamp, 2021).

**Accountability through Legislation**

Utilization of assessment to elucidate student performance has sharply increased in recent educational history. The use of assessment to determine eligibility for special education services increased dramatically with the passing of laws such as Public Law 94-142, originally the Education for All Handicapped Children Act (EAHCA) in 1975, which granted students with disabilities the right to a free, appropriate education (Merrell et al., 2012). Prior to this legislation, students with disabilities were often denied participation in education, as it was thought that these children were unable to make educational gains (Merrell et al., 2012). Continued educational reforms through the Individuals with Disabilities Education Act (IDEA; 1990, 1997, 2004) have created further need for assessment, particularly regarding the need for Response to Intervention (RtI) services and interventions more broadly as greater focus on inclusion occurred within education (Merrell et al., 2012).

In response to these laws and other related educational legislation, a need to evaluate the
effectiveness of schools was identified. Shifts from the use of assessment to assign grades within the classroom to the use of assessment as the basis for education improvement and reform occurred (Glaser & Silver, 1994; National Research Council, 2002), and increased focus on accountability resulted. The push for schools to demonstrate accountability challenged previous notions that student success can be measured by the amount of money spent per student within a district (Ornstein, 1988). Salvia et al. (2017) identified the push for accountability as a way to determine whether current practices in schools are yielding desired results. Due to this, schools experienced increased demands to critically evaluate the programs offered and how students respond (Ornstein, 1988).

The National Research Council (2002) labeled accountability as receiving attention from legislators and assorted policy makers because of the ease and low costs associated with creating assessments when compared with other potential approaches to reform, such as increased teacher training or professional development. To demonstrate progress in accordance with requirements set forth by Every Student Succeeds Act ([ESSA]; 2015), individual states were tasked with disseminating plans of accountability that outlined how student progress was to be measured. ESSA (2015) required several varied measures of growth, preventing schools from relying on one or few markers to demonstrate student success and called for individual Local Education Agencies (LEAs) to develop plans that aligned with those put forth by their respective states.

**State Tests**

Stemming from this push for accountability, state tests, a form of summative assessment, were introduced as one way to measure student progress. Nationwide, it became mandated that public school students be administered standardized tests in English Language Arts (ELA), as well as mathematics, yearly from third to eighth grade (ESSA, 2015; No Child Left Behind
These high-stakes tests lead to direct consequences for students, parents, and school staff, including impacts on federal funding received or grade promotion and graduation (Thomas, 2005). NCLB (2001) called for schools unable to demonstrate student progress repeatedly through testing results to be penalized, potentially resulting in major restructuring and school closures.

Strengths noted in regard to widespread implementation of these high stakes tests include the increased capacity to compare student performance across states due to increased systematization of content and the explicit call for educational standards (Fulmer, 2011). Prior to widespread state tests, differences between states with respect to requirements for grade promotion/retention and grade-related content and expectations muddied many comparisons drawn (Fulmer, 2011). These tests can be used to demonstrate a portion of individual student progress on a year-to-year basis and provide data to evaluate how well the school as a whole performed on state measures, demonstrating one way of assessing school effectiveness (Nelson, 2013). State tests also offer a uniform and interpretable way for communities and parents to be informed of student performance (Thomas, 2005).

Preceding the required implementation of state tests by legislation, some LEAs utilized group achievement tests to assess and disseminate data on student progress as well as inform high stakes decisions such as special education referrals (Salvia et al., 2017; Stone et al., 1988). Due to the demand for states to report on student progress as defined by state standards, these high stakes tests should represent content taught. Group achievement tests were not designed to align with standards (Salvia et al., 2017). Standardized state tests offered refinement in comparison with group achievement tests, presenting another strength of state test utilization, as group achievement tests did not lead to clear planning suggestions for instruction (Salvia et al.,
High stakes tests, however, are not without limitations. Lack of systematic criteria for evaluating the alignment of curriculum content and state tests may obscure the ability to proclaim students are or are not making educational gains (Fulmer, 2011). NCLB (2001) required content standards that align with state tests, however, Fulmer (2011) asserted that current research is lacking reliable and valid ways to assess this alignment. Another potential limitation of the data stemming from state assessments is tied to the requirement of LEAs to test at least 95% of their student population, including those identified as having disabilities (ESSA, 2015; NCLB, 2001).

Students with disabilities can be granted testing accommodations or participate in Alternate Assessments (AA), if the curriculum of state tests is deemed inappropriate (IDEA, 2004). These accommodations, however, have been questioned in regard to the accuracy with which test data from students with disabilities demonstrate growth as required by the accountability movements. Pressures to retain federal funding, dependent upon student performance, have been identified as a potential incentive to inflate test scores through accommodation provisions (Popham, 2001). Increased resources and time devoted to “teaching to the test” have been tied to decreased time allotted for school activities including significant time for additional instruction (Nelson, 2013). Afflerbach (2008) identified the intense focus on state tests as causing decreased attention to daily assessment of high utility that informs instruction to aid readers who struggle specifically, though it could be hypothesized this would impact students across other subjects as well. Even if data from state tests are informative, Afflerbach (2008) asserted that the decreased attention on daily performance of students is harmful with regard to actual student growth.
**Assessment within the Classroom**

Within the classroom, teachers commonly use various types of tests to monitor student progress, including formative and summative assessments. Merrell et al. (2012) stated that well-designed curricula include procedures for monitoring student performance, allowing for the identification of discrepancies between taught and learned curriculum. It is vital, though, that testing does not become the primary activity of the classroom teacher. Increased time engaging in testing and data analysis reduces classroom instruction (Gunning, 2012; Hosp et al., 2016). To avoid devoting large amounts of classroom time to assessment, efficiency should be considered in the selection of tests. When curricula are arranged so that information builds on previous instruction, assessment can help to pinpoint unlearned concepts (Howell & Nolet, 2000).

This discussion of types of assessment, formative and summative, recognizes the traditional definitions and characteristics (See Table 2), as well as acknowledges that these types are not mutually exclusive categories (Dixson & Worrell, 2016). As Dixson and Worrell (2016) highlighted, formative and summative assessments are complementary tools because they inform inter-related purposes. Furthermore, a traditional formative assessment tool may be used summatively, and likewise a traditionally summative assessment tool may be use formatively (National Research Council, 2001). The determination of the type of assessment measure is based on the purpose for which the assessment results are to be used (as previously discussed and presented in Figure 4). For example, Curriculum-Based Measures (CBMs) may be identified as formative assessment for the purpose of screening, planning instruction (e.g., baseline determination), and monitoring student progress. In addition, CBMs may be classified as summative assessment for the purposes of special education accountability and program evaluation.
### Table 2

*Traditional Features of Formative and Summative Assessment*

<table>
<thead>
<tr>
<th>Feature</th>
<th>Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose</strong></td>
<td></td>
</tr>
<tr>
<td>Formative</td>
<td>Identify areas in which students are not achieving and improve instruction</td>
</tr>
<tr>
<td>Summative</td>
<td>Assess instruction through student performance and utilize data to evaluate appropriate student programming</td>
</tr>
<tr>
<td><strong>Objectives</strong></td>
<td></td>
</tr>
<tr>
<td>Formative</td>
<td>Determine areas of instruction that are effective and that may need improvement</td>
</tr>
<tr>
<td>Summative</td>
<td>Identify students’ comprehension of material and readiness to progress</td>
</tr>
<tr>
<td><strong>Utilization</strong></td>
<td></td>
</tr>
<tr>
<td>Formative</td>
<td>Low-stakes measures that can be informally administered at any point before or while instruction is occurring</td>
</tr>
<tr>
<td>Summative</td>
<td>High-stakes measures that are typically formally administered when instruction has concluded</td>
</tr>
<tr>
<td><strong>Precision</strong></td>
<td></td>
</tr>
<tr>
<td>Formative</td>
<td>Typically ranges from low to high levels of rigor/precision of measurement</td>
</tr>
<tr>
<td>Summative</td>
<td>Typically ranges from moderate to high levels of rigor/precision of measurement</td>
</tr>
</tbody>
</table>

Typically, formative assessment and evaluation practices occur as instruction is being delivered (Gunning, 2012; Roskos & Neuman, 2012; Schaugency et al., 2010; Shepard, 2019), offering teachers flexibility in teaching focus (Stanley & Alig, 2014). Formative assessment allows for frequent evaluation of student progress and informs decisions regarding instruction to be made in response to data obtained (Roskos & Neuman, 2012; Schaugency et al., 2010; Shepard, 2019). Formative assessment does not require formal testing, however, as frequent checks for understanding, questioning, or group response work can offer insight as to whether students are making progress (Stanley & Alig, 2014).

CBMs are one example of a type of formative assessment that can be used. CBMs are defined by Salvia et al. (2017) as “a standardized set of procedures that allow one to directly measure important skills in a relatively short amount of time” (p. 347). CBMs are typically characterized by their standardized directions, brief administration, and alignment with the curriculum (Hosp et al., 2016). The short amount of time required to administer CBMs is a major strength, as it is significantly shorter than that needed to administer and score tests such as the group achievement tests previously discussed (Salvia et al., 2017). As a result, students are able to spend more classroom time engaged in instruction as opposed to being assessed (Salvia et al., 2017). Additionally, CBMs have been found to be valid and reliable in assessing student performance (Chung et al., 2018; Deno, 2003; Hosp et al., 2016), and are directly correlated with targeted intervention focus (Floyd, 2010). The prioritization of reliability and validity has been identified as an advantage of CBMs (Tindal, 2013). The ability to use CBMs as a progress monitoring tool allows for increased ease of screening students and repeated measures of performance over time (Salvia et al., 2017). CBMs are also used as benchmarking assessments and can help to identify students not meeting progress, thus in need of intervention (Ardoin &
Christ, 2008). Another asset of these measures is their ability to be developed to measure academics across skills, including reading, writing, math, and vocabulary knowledge (Salvia et al., 2017).

CBMs pose some limitations as identified within the literature. While they can assess a wide variety of skills, CBMs are limited to assessing basic skills, as opposed to higher-level thinking skills (Salvia et al., 2017). The creation of CBMs by school officials can increase alignment with curricula; however, it is a very time-consuming process and can counteract the efficiency of the measures previously described (Salvia et al., 2017). Despite attempts to increase clarity and decision-making rules, teachers have demonstrated difficulty utilizing analysis of CBM results to correctly inform instructional decisions (den Bosh et al., 2017; Fuchs & Fuchs, 2002). Further, researchers continue to contest evidence-based support for CBMs regarding exactly how CBMs should be used with respect to the number and frequency of repeated administrations and how to monitor progress (Ardoin et al., 2013). Espin et al. (2017) noted that many of the analysis methods with CBMs were developed for elementary-aged students, suggesting more research is needed with older students.

Unlike formative assessments such as CBMs, summative assessments are used to determine what students have learned at the end of formal instruction (National Research Council, 2002; Roskos & Neuman, 2012). Schaughency et al. (2010) defined summative assessment as a process to collect data needed for accountability purposes. These summative assessments range from teacher designed tests at the end of a specific unit to an exam at the mid-point of the semester to the aforementioned state mandated standardized tests that arose from the accountability movement. Summative assessment results that come from classroom administered tests are often used in the calculation of student report card grades.
Report card grades may offer a more accurate portrayal of students’ command of the curriculum, when compared to state tests, as classroom-based tests that inform grades typically are based on the content taught (Duckworth et al., 2012; Willingham et al., 2002). Stanley and Alig (2014) asserted that when summative assessments are administered, teachers should already have data to predict student performance from prior formative measures. As such, teachers can differentiate instruction or engage in reteaching prior to summative assessments in an attempt to increase student knowledge.

Report card and classroom grades, however, face limitations when considered as a way to assess student progress. Duckworth and colleagues (2012) underscored the potential impact homework and behavior in the classroom may have on the calculation of classroom grades. Variations between homework assignments, types and frequencies of exams, emphasis placed on attendance or participation can all alter students’ classroom grades, but may not accurately represent their knowledge and progress. Additionally, report card grades as well as school engagement have been demonstrated to typically decrease for students as they begin middle school (Barber & Olsen, 2004; Maulana et al., 2012). Contested purposes and attributed meanings of report card grades add to the limitations when considering their utility in assessing and demonstrating student progress. For example, Munk and Bursuck (2001) found that the purpose parents attributed to report cards varied between individuals, ranging from categorizing report cards as a means of communication between home and school to a measure of students’ effort in class. Salvia et al. (2017) identified the importance of not relying on one single summative assessment as an indicator of achievement.

Section Summary

Systematic assessment can be implemented to determine which students are at risk of not
meeting academic achievement expectations. Generally, assessment is a process that may include all or some aspects of RIOT. The way in which assessment is carried out may differ depending upon the decisions that will be drawn from data and the purpose of the data collection (Salvia et al., 2017). Testing is often a main focus of the RIOT process, particularly due to increased focus on accountability and demonstration of progress within schools. Testing, compared to other assessment methods, can offer educators the ability to collect data that is more specific and tailored to a specific question or objective (Salvia et al., 2017). Different types of testing assessment approaches offer unique assets and limitations, underscoring the importance of utilizing varied assessment measures within DDDM. Federal legislation mandated all children receive an education and outlined the duty of schools to prove the implementation of effective instruction through progress captured by assessment. Typically, tests utilized include summative and formative assessments, which can range from state tests to CBMs and report card grades. Each of these tests offer unique assets and disadvantages.

**Reading**

Studying reading theory and historical approaches to teaching is vital in ensuring schools are providing students with education that supports skill development. While consensus exists to some extent regarding certain aspects of literacy and reading, specific approaches and instructional practices are still contested. For example, features of conditions for learning, such as clear expectations, demonstration, and engagement, are identified (Cambourne, 1995), yet the application of these conditions continues to vary. Debate between whole language and phonics-based reading instruction has led to competing theories driving and directing the reading educations received by students. Current statistics within the United States demonstrating the number of students performing below reading expectations call attention to the importance of
effective and efficient teaching methods.

**Reading Performance**

Students within the United States continue to exhibit skills deemed to be below performance expectations. Data from the 2019 Nation’s Report Card, put forth by the National Assessment of Educational Progress (NAEP), indicated that twelfth, eighth, and fourth grade students’ reading performance decreased from 2015 data collection (U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, NAEP, 2019). The NAEP assesses student performance nationwide across subject areas. This observed decline in reading has been noted over time by researchers. Significant increases in reading rates have not been demonstrated between 2009 to 2019 (U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, NAEP, 2019). As these students continue to grow older, engagement with reading does not appear to increase. The National Endowment for the Arts (2004) has called attention to the finding that less than half of the American adult population reads literature.

It is estimated that over 65% of students do not demonstrate competency in reading comprehension by the time they complete middle school (Reardon et al., 2012). Research notes varied terms to label these students, including poor readers and students with low reading ability, with reading difficulties, with reading problems, and students with dyslexia (Al Dahhan et al., 2016). Robb (2010) described four categories to capture middle school readers: struggling, reluctant, grade-level, and proficient. The *struggling reader* group is identified as consisting of students who read below grade-level and have difficulty comprehending any information they may read (Robb, 2010). *Reluctant readers* can read grade-level material, but only engage in reading when required (Robb, 2010). Robb (2010) classified students who meet expectations as
grade-level readers and those students who exceed expectations and engage in reading willfully as proficient readers.

Students are also not always captured by early screening procedures designed to identify students at risk for not meeting expectations, representing a portion of the population who have possible reading disabilities that emerge after the third grade (Catts et al., 2012; Leach et al., 2003). Seidenberg (2013) attributed poor reading skills among students, in part, to a disconnect between research and education, preventing swift application of research within schools. America’s students’ current performance rates call attention to a deeper need for effective teaching of reading.

**History of Teaching Reading**

It is argued that, regardless of the instructional method used, about 75% of students will learn to read sufficiently (Liberman & Liberman, 1990). The history of reading instruction helps to explain the varied approaches used with students over decades. Study of these practices calls attention to how these have changed, become more effective, and demonstrate how other factors, such as learning principles, have been considered over time (Chartier, 2004). A search for the “correct” way to teach reading has enabled what Chall (1967; 1983) entitled the “Reading Wars” between meaning-based and code-based reading approaches. Weaver (1994) explained that these competing methods have become more commonly referred to as whole language and phonics, respectively. The importance of knowledge related to the history of instruction is stressed by the International Literacy Association (2018), which asserted that an understanding of historical changes that have occurred is needed to meet the requirements to be a professional within the field.

America’s tie to England and the Protestant church can be cited as the initial major push
for Americans to begin to learn to read and to instruct reading, as the church called for all individuals to be able to read and interpret the Bible (Banton-Smith, 2002; Barry, 2008). Historical reading tools used in the United States included various hornbooks, psalters, primers, and spellers, which were used in church and also to instruct reading (Banton-Smith, 2002; Barry, 2008). Much greater focus was put on the content, as opposed to teaching methods and developmentally appropriate material during the 17th and first half of the 18th centuries (Banton-Smith, 2002; Barry, 2008). Some evidence exists, however, to support emphasis on memorization of the alphabet and further introduction of reading principles afterward (Banton-Smith, 2002).

Secularization of the schools shifted focus from church-related readings and content to nationalism as the country gained independence (Banton-Smith, 2002). In the early 19th century, educators began to emphasize the importance of meaning within instruction, critiquing former practices of rote memorization with limited context for students (Barry, 2008; Monaghan, 2005). Educators advocated for students to be taught whole words, as opposed to in parts like the alphabet method previously used, and stressed ability to recognize words by sight (Monaghan, 2005). During the Civil War, the United States saw some efforts to alter the alphabet to be phonetic, meaning each letter would correspond to one sound (Barry, 2008). This was later dismissed; however, phonics-based approaches continued to impact education. The popularity of phonics-based approaches spanned from the late 1800s through the 1920s, when reading instruction increased focus on sight words (Weaver, 1994). The popularity of phonics-based instruction returned throughout the 1960s (Weaver, 1994). Debates regarding whole-word versus phonics-based instruction continued over the course of the 19th and 20th centuries (Chartier, 2004). These variations in teaching continue to impact students today, with researchers still
debating the best approaches (Bowers, 2020).

*Whole Language*

The whole language method of teaching reading was, at a time, viewed as “conventional wisdom” (Pearson, 2004, p. 219). Manning et al. (1990) stated whole language proponents believe students’ experience with reading should occur genuinely, identifying the isolation of teaching skills from significant and engaging content as one reason for poor literacy skills amongst students. There is increased emphasis on reading for meaning (Manning et al., 1990). Similarities between learning to read and learning to speak are attended to by whole language principles (Arrow & Tunmer, 2012). A whole language approach to reading stresses the connection of the material to real life situations (Goodman, 1986). These approaches are also referred to within the literature as a meaning-emphasis approach (Chall, 1967), or a psycholinguistic model (Anderson, 1984; Goodman, 1967). This method is also referenced as a “top-down” approach to reading (Gunning, 2003).

Goodman (1967) asserted efficient reading involves making use of the ability to identify cues that lead individuals to correctly guess words. This assertion contrasts phonics-based approaches, which stress individual letter and sound identification (Anderson, 1984). Within the whole language perspective, individualized reading skills are not explicitly focused on as they are believed to complicate the process and confuse readers (Gunning, 2003). Phonics-based strategies, such as sounding out an unknown word, are included within whole language, but are not systematically taught as they are in other approaches (Bowers, 2020). Instead, these strategies are employed when other whole language strategies, like guessing the words from context, fail (Bowers, 2020).

Pearson (2004) noted that the whole language approach to teaching reading was not
solely implemented nationwide during the height of its popularity, and that implementation was not always purely aligned with the approach’s core concepts. Specifically, whole language reading was, at times, confused and improperly applied as “whole-class” learning which calls for all students getting the same instruction regardless of performance (Pearson, 2004). Students may face limitations when taught through a whole-language perspective, as the experiences of readers are heavily relied upon within this model (Suraprajit, 2019). Students from varying diverse cultural or socioeconomic backgrounds may lack some of the experiences called upon by reading material. Sole use of whole language principles has largely fallen out of favor, particularly with the publication of the NRP (NICHD, 2000), which heavily emphasized the importance of phonics when teaching students to read. Castles et al. (2018), however, suggested the remnants of whole language approaches in schools may be due to a large gap between research and implementation.

*Phonics-Based Reading*

Unlike whole language approaches, teaching reading through phonics involves prompting students to learn how to identify and create associations between letters and the sounds that correspond (Scanlon et al., 2010). As previously stated, phonics-based reading instruction began to gain traction over whole language reading (Weaver, 1994). Pearson (2004) attributed this shift to various factors, but particularly tied an increase in phonics-based instruction to the passing of NCLB (2002) and increased accountability requirements of LEAs. Pearson (2004) also indicated developments within reading research further supported increased phonics-based instruction. Phonics-based approaches align with the goal of early reading instruction, which has been identified as mastering word recognition and decoding in order to enable later fluency and reading comprehension (Arrow & Turnner, 2012). Students are able to read independently more
quickly when provided with the knowledge to decode words through phonics-based approaches (Weaver, 1994). Research has demonstrated that children who received phonics-based instruction, compared to those who did not, read at a slower pace but with greater accuracy (Chall, 1967; Connelly et al., 2001). Further, explicit teaching of phonics, including for students who are not performing at expected reading levels, leads to more effective reading than does implicit phonics (Jager-Adams, 1997). Arrow and Tunmer (2012) asserted that phonics instruction is still advantageous despite the irregularity of the English language.

A major limitation of phonics-based instruction lies within its implementation, as opposed to its efficacy. Prior to the publication of the NRP (NICHD, 2000), it was found that most explicit instruction based in phonics was not being utilized in classrooms systematically (Jager-Adams, 1997). This has continued over time despite support for these methods. Pearson (2004) asserted that, in order to implement strict phonics-based instruction, significant and complex changes would be needed with regard to teacher training, class designs and materials, and LEAs’ curricula and standards. Additionally, phonics-based approaches are seen by some as rigid and unadaptable (Pearson, 2004). While these requirements are not necessarily reasons to dismiss phonics-based reading, they do highlight the low feasibility of this occurring effectively and efficiently. Castles et al. (2018) also stated that the gaps between research and implementation cause limitations to phonics-based instruction, specifically identifying limited resources that accurately explain the theory behind the efficacy of phonics-based reading. Over time, strictly phonics-based approaches, however, were disregarded.

**Balanced Literacy**

The history of teaching reading demonstrates the implications research and paradigm shifts have over instruction. In response to considerations of whole language and phonics-based
instruction, support for a balanced literacy approach, which consists of a combination of both meaning- and code-based instruction, arose (Arrow & Turnner, 2012; Castles et al., 2018; Ehri et al., 2001). Balanced literacy should incorporate the explicit teaching of phonics while still maintaining emphasis on the meaning of text (Freppon & Dahl, 1998). Arrow and Turnner (2012) asserted that the “one-size-fits all” approach brought about independently by both whole language and phonics-based instruction is not effective for all readers, necessitating this balanced approach. Ehri et al. (2001) cited the need for systematic phonics instruction as part of a balanced literacy approach due to its greater effectiveness when compared with non-systematic phonics instruction.

Major support for balanced literacy came from publication of the NRP (NICHD, 2000). The NRP (NICHD, 2000) was a federally funded meta-analysis to clarify the most effective reading instruction practices as well as make recommendations regarding implementation. Shanahan (2005) stressed the NRP’s (NICHD, 2000) findings represented legitimate assessment of the research at the time and accounted for numerous potential biases within its design. The NRP focused specifically on alphabets (comprised of PA and phonics), fluency, comprehension, which included discussion of vocabulary and comprehension strategies, teacher education, and technology (NICHD, 2000). The NRP asserted that a balanced literacy approach to teaching reading was the most effective method (NICHD, 2000). Further, it specified that phonics alone should not be utilized as a stand-alone reading program (NICHD, 2000). The NRP underscored a need for meaning-based instruction as well in order to be considered a complete reading program (NICHD, 2000). While some critiques of the NRP’s (NICHD, 2000) methodology and summaries have been made (Camilli et al., 2003, 2006; Hammill & Swanson, 2006), further investigation of these analyses led to conclusions that these critiques do not
undermine the NRP’s findings (Stuebing et al., 2008). Additional phonics-based instruction beyond what is needed to master skills is not considered disadvantageous for students; however, too little instruction can be detrimental to the development of expected reading skills (Arrow & Tumner, 2012).

Despite the support for balanced reading teaching, these combined methods have not been met with complete acceptance. Seidenberg (2013) described the introduction of balanced literacy methods as a “Treaty of Versailles solution” (p. 10) that helped to lessen debate between the whole language and phonics camps without addressing the theoretical differences between the two. At times, balanced literacy is thought to simply be a renaming of whole language methods, as the balanced literacy approaches are characterized as not explicitly focusing on phonics-based aspects (Bowers, 2020). Sweet (2004) declared that without sufficient training of teachers in conjunction with complete support received from both from local and state governments, publishing companies responsible for text books and teaching materials, teacher training programs and continued professional development opportunities, proper balanced literacy instruction could not occur.

Section Summary

A need for effective reading instruction is demonstrated by the number of students within the United States of America who are identified as not meeting reading achievement expectations. Historical influences as well as theories regarding how children best learn to read have shaped the implementation of reading instruction strategies over time. As a result, the Reading Wars (Chall, 1967, 1983) ensued. Sole implementation of either whole language instruction or phonics-based instruction have fallen out of favor, particularly with the publication of the NRP (NICHD, 2000), which called for systematic phonics-based instruction as part of a
comprehensive reading program. Balanced literacy approaches combine both meaning- and code-based strategies (Castles et al., 2018). The NICHD (2000) compiled the results of thousands of reading studies, featuring five areas of reading instruction within the NRP report: PA, phonics, fluency, vocabulary, and text comprehension. Balanced literacy continues to face criticism, specifically related to a propensity to often largely mirror whole-language based approaches (Bowers, 2020).

Phonemic Awareness

As previously discussed, PA was one skill evaluated by the NRP, and was identified as a prerequisite for acquiring the remaining reading skills discussed (NICHD, 2000). Daly et al. (2015) defined prerequisite skills as anything required before more complex learning can occur. The NRP concluded that numerous groups of students, including those meeting expectations, at risk for not meeting expectations, and students with reading disabilities, benefited from PA instruction (NICHD, 2000). To read successfully, individuals must grasp the alphabetic principle, which is an understanding that the letters correspond to unique sounds (Scanlon et al., 2010). It is one aspect of the alphabetic principle, underscoring the importance of PA in students’ journey to proficient reading (Moats, 2010). The ability to decode, which is necessary for fluent reading, is supported by PA acquisition (Anthony & Francis, 2005; Daly et al., 2015).

Differentiating Phonemic Awareness

PA differs from phonics, as well as phonological awareness, despite some interchangeable term use within literature and by reading instructors (Kilpatrick, 2015; Scanlon et al., 2010). PA is distinguished from phonics by its exclusive emphasis on the ability to identify and manipulate phonemes in spoken language (Scanlon et al., 2010). A phoneme is the smallest unit of spoken language (Gillon, 2018; Kilpatrick, 2015). Estimates vary due to differences in
dialects; however, Moats (2010) asserted the English language contains approximately 40 to 52 phonemes. Kilpatrick (2015) contended that the majority of phonemes are represented by single letters. PA does not involve written words, only sounds (Kilpatrick, 2015). Daly et al. (2015) included the blending of sounds, segmenting of sounds, substitution, and deletion of sounds as examples of PA. The task of isolating an initial phoneme is considered the most basic, followed by the ability to blend phonemes, and then engage in segmentation (Daly et al., 2015). Phoneme segmentation skills, of all phonological awareness components, have been found to be the best predictor of first grade students’ ability to read and spell (Nation & Hulme, 2001).

PA is one portion of phonological awareness, which involves a broader variety of tasks such as phonology, the study of speech sounds and rules of production (Scanlon et al., 2010), and phonological processing (Carnine et al., 2006; Gillon, 2018). Chard and Dickson (1999) detailed, from simple to most complex, activities needed for proficient reading within phonological awareness. This information can be found in Figure 5. Activities that involve PA are identified as more complex in Figure 5, including the blending and segmenting of individual phonemes (Chard & Dickson, 1999). Differing from PA, phonological awareness involves one’s ability to manipulate varied parts of spoken words, such as syllables, rimes, onsets, and phonemes (Scanlon et al., 2010). McEwan (2009) described phonological awareness as an “umbrella” and “the various levels of the speech system as its spokes” (p.18). PA is also not interchangeable with the term phonics, which is defined as “a way of teaching reading and spelling that stressed symbol-sound relationships” (Yopp & Yopp, 2000, p. 131). When considering the phonics, PA is one of many areas of instruction.

**Identifying the Need for PA Instruction**

McCormick et al. (2002) highlighted children who have learned to read have developed
Figure 5

**Phonological Awareness Activities Arranged by Increasing Complexity**


https://doi.org/10.1177/105345129903400502
PA skills, however, these skills are not often explicitly recognized. The majority of children will develop PA with seemingly little explicit instruction (Linan-Thompson & Vaughn, 2010; McEwan, 2009; Scanlon et al., 2010). McEwan (2009) outlined that other students have difficulty acquiring PA, regardless of their cognitive ability, due to some combination of genetic and environmental factors. Environments rich in caregiver support and exposure to reading and word games were linked with PA skills (McEwan, 2009). When these factors are not present, McEwan (2009) established that there is a need for explicit, systematic PA instruction by teachers utilizing evidence-based practices. Carnine et al. (2006) outlined eight activities students can engage in to build PA. These activities are found in Table 3.

Children who are introduced to literacy from an early age and receive proper support may still face challenges associated with PA (Scanlon et al., 2010). According to Scanlon et al. (2010), difficulty with PA can be caused by varying factors including children’s tendency to attend to the meaning of language instead of sounds, trouble distinguishing sounds due to the speed at which sounds are produced, the presence of coarticulation of sounds (blending), and confusion caused by the similarity of phonemes. Daly et al. (2015) further discussed additional individual risk factors including memory deficits, developmental delays and cognitive deficiencies, hearing and speech disabilities, and the presence of other diagnoses that may make skill acquisition more difficult. PA skills are not developed along with natural maturation (Liberman & Shankweiler, 1985), so there must be targeted intervention when students are not acquiring these skills as expected.

**Measuring PA**

The importance of PA acquisition in the reading process coupled with the information some children will not develop the ability to discriminate and manipulate phonemes calls for a
### Table 3

**Activities to Build Phonemic Awareness**

<table>
<thead>
<tr>
<th>Activity</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phoneme Isolation</td>
<td>Children recognize individual sounds in words</td>
<td>The first sound in <em>fit</em> is /f/</td>
</tr>
<tr>
<td>Phoneme Identity</td>
<td>Children recognize the same sound in different words</td>
<td>The first sound /r/ is the same in <em>rat, rim,</em> and <em>rice</em></td>
</tr>
<tr>
<td>Phoneme Categorization</td>
<td>Children recognize the “odd” word in a set of words</td>
<td>In the set of words, <em>top,</em> <em>tin,</em> and <em>hen,</em> the word <em>hen</em> does not belong since it doesn’t begin with /t/</td>
</tr>
<tr>
<td>Phoneme Blending</td>
<td>Children listen to a sequence of phonemes and combine them to form a word</td>
<td>The sounds /s/ /u/ /n/ make the word <em>sun</em></td>
</tr>
<tr>
<td>Phoneme Segmentation</td>
<td>Children break a word into its separate sounds, saying each individual phoneme</td>
<td>The word <em>grab</em> is composed of the sounds /g/ /r/ /a/ /b/</td>
</tr>
<tr>
<td>Phoneme Deletion</td>
<td>Children recognize the word that remains when a phoneme is removed from a word</td>
<td>The word <em>spark</em> without the /s/ is <em>park</em></td>
</tr>
<tr>
<td>Phoneme Addition</td>
<td>Children make a new word by adding a phoneme to an existing word</td>
<td>If you add the phoneme /s/ to the word <em>pot,</em> you have the word <em>spot</em></td>
</tr>
<tr>
<td>Phoneme Substitution</td>
<td>Children replace one phoneme for another to make a new word.</td>
<td>Start with the word <em>rug.</em> Change the /g/ to /n/. The new word is <em>run</em></td>
</tr>
</tbody>
</table>

need to be able to assess PA. Measuring PA provides a way to identify students at risk for not meeting expectations, track growth among students receiving general systematic instruction, and to monitor growth for those receiving more intense intervention. Without mastering phonological awareness, Daly et al. (2015) noted that older students engage in less reading time compared with others. While proper instruction and intervention can increase PA and overall phonological awareness, these students may continue to find reading to be a difficult task (Daly et al., 2015). Gillon (2018) identified the assessment and measuring of older students’ PA skills as necessary when students have been previously identified as having a reading disability or have clear reading difficulties and more information to understand the root cause is required.

Understanding expected development of PA skills helps to dictate measures teachers may use to determine student performance. The greatest growth in increased phonological awareness as a whole appears to occur between kindergarten and first grade for most students, with PA proficiency typically occurring by second grade (Daly et al., 2015). Typically, children can successfully segment beginning and ending phonemes by mid-kindergarten (Schuele & Bourdreau, 2008). By the end of kindergarten, most children are expected to be able to segment and blend words with two or three phonemes (Schuele & Bordreau, 2008). Targeted instruction for young students should start with focus on syllables, which are easier to distinguish than phonemes, and then move to phonemes as understanding of syllables occurs (McCormick et al., 2002). Kilpatrick’s (2015) progression of typical phonological skill development, which includes PA, can be found in Figure 6.

Engagement of students in instruction of PA as well as integration into daily classroom activities is stressed as vital by McCormick et al. (2002). McCormick et al (2002) declared that assessment of PA can also be integrated easily into curricula, and can consist of standardized
Figure 6

*Development of Phonological Skills*

- **Early PA**
  - Rhyming,
  - Alliteration,
  - First Sounds

- **Basic PA**
  - Blending
  - and Segmenting

- **Advanced PA**
  - Phonemic Proficiency

Figure adapted from Kilpatrick, D.A. (2015). *Essentials of assessing, preventing, and overcoming reading difficulties.* John Wiley & Sons.
assessments with norm comparisons or more informal measures. Kilpatrick (2015) echoed this emphasis on distributed practice throughout the school day. Torgesen and Mathes (2002) cautioned against assessing students’ overall phonological awareness for the purpose of identifying students in need of intervention before mid-kindergarten, as students’ text exposure prior to starting school may account for low performance at the beginning of the school year. Some standardized assessments that can be utilized to measure PA include the subtests that make up the Phonological Awareness Composite on the Comprehensive Test of Phonological Processing, Second Edition (CTOPP-2; Wagner et al., 2012), which assesses students’ performance on elision, blending, and phoneme isolation tasks and the Woodcock Reading Mastery, Third Edition (WRMT-III; Woodcock, 2011) which also includes focus on phonological awareness. Others with more specific focus on aspects phonological awareness include the Test of Phonological Awareness Second Edition: PLUS (TOPA-2+; Torgesen & Bryant, 2004) and the Phonological Awareness Test, Second Edition (PAT-2: NU; Robertson & Salter, 2007). The latter measures, however, have age limits of eight years old and nine years, 11 months old respectively (Robertson & Salter, 2007; Torgesen & Bryant, 2004). Gillon (2018) labeled the TOPA-2+ (Torgesen & Bryant, 2004) as a tool to elucidate more specific skills or deficits than broader measures, such as the CTOPP-2 (Wagner et al., 2012).

Gillon (2018) suggested seven more informal ways to measure students’ PA as a way to monitor performance. These activities to assess PA closely align with those suggested by Carnine et al. (2006) as ways to improve students’ PA, but are instead structured as questions a teacher, or other educational professional, may ask students to better understand their grasp on the different aspects of PA. These tasks are area specific tasks, meaning that each assesses one particular area of PA and include: alliteration awareness (also called phoneme categorization),
matching, isolation, blending, deletion or elision, segmentation, and spoonerisms (Gillon, 2018). Hosp et al. (2016) detailed the use of CBMs to assess PA, such as a phoneme segmenting CBM and a word blending CBM. Advantages of using low-inference CBMs include brief administration time, allowing large numbers of students to be assessed quickly, low or no costs for measures, and teachers’ ability to align CBMs with current classroom focuses (Hosp et al., 2016).

**Assets and Limitations of PA Instruction**

As previously noted, the need for PA instruction is well established in the research literature. During the first two years of schooling, PA has been identified as one of two best predictors of students’ ability to learn to read (NICHD, 2000). Further, the NRP reported that PA can be explicitly taught and helps children to improve reading and spelling abilities (NICHD, 2000). Suggate (2016) conducted a meta-analysis analyzing long-lasting impacts of reading interventions with varying targets and concluded that students who received interventions with a PA focus demonstrated generalization to additional skills, stressing the impact PA intervention can have on students’ academics overall. As students engage in reading and writing, PA is typically increased (McGee & Dail, 2010). As such, McGee and Dail (2010) argued that PA instruction in pre-school enables students to more successfully complete and grasp kindergarten instruction. Program evaluations have been previously utilized to assess the success of PA instruction with pre-school age students and have found increased reading comprehension and decoding abilities amongst the students with higher PA (Byrne & Fielding-Barnsley, 1993; 1995).

Yet, research does not recommend focus on PA alone as being considered sufficient to teach children to read (NICHD, 2000). Shanahan (2003) cautioned viewing PA as a “magic
bullet” to achieve reading success, particularly due to PA rarely being implemented alone within the literature examining reading interventions with students not meeting expectations. When specifically considering the inclusion of PA instruction in classrooms for younger students, Yopp and Yopp (2000) cautioned against replacing additional areas of reading instruction to only focus on PA. Kilpatrick (2015) further stated that PA intervention alone is unlikely to result in definite reading improvement. O’Connor (2011) maintained that older students may not benefit from PA instruction as younger students do, though Kilpatrick (2015) has challenged this assertion. The NRP found that programs with PA instruction that lasted approximately 30 minutes over repeated sessions, with total instructional time ranging from 5 to 18 hours per instructional year were most effective in increasing student performance for younger students (NICHD, 2000). McGee and Dail (2010) underscored another issue related to PA and its teaching: misunderstanding by teachers. Teachers, or other school staff members, may not accurately understand what is PA instruction and instead, unknowingly, be engaging students solely in phonics-based teaching, leaving some students to receive less actual PA instruction than is assumed to be occurring (McGee & Dail, 2010).

**PA Instruction with Middle School Students**

A major limitation regarding research available on PA instruction is related to the age of most participant populations. The majority of improvement demonstrated across studies analyzed by the NRP (2000) was amongst young, typically reading students, and so effects on older students who are not meeting reading standards remain unclear due to limited studies (Shanahan, 2003). Limited peer reviewed research regarding intense PA instruction with students past the third grade is available, as most children demonstrate proficiency as younger students. Teaching phonological awareness, which includes PA, as part of a more comprehensive reading instruction
approach is recommended for older students who are not meeting expectations (Gillon, 2018). Torgesen (2000) introduced the term *treatment resistors* for older students who have not responded to instruction as expected, which consists of many students with low PA. Gillon (2018) declared that phonological awareness instruction with older students should be delivered in small groups or individually. Gillon (2018) further stated that these interventions should occur either daily or over approximately two hours during the week. It is important to note, however, that these recommendations pertain to phonological awareness instruction overall, not specifically PA. Spires et al. (2018) called attention to the relationship between poor reading and overall academic performance, particularly for middle school students who are expected to engage in reading for different disciplinary subjects to further their learning. Kilpatrick (2013) extended this claim to highlight the ties between poor academic performance and impacts on students’ behavior, stating the impacts of behavior on education continue to grow as students continue to age. Kilpatrick (2013) asserted that a student who has not developed PA proficiency is never “too old” to receive intervention. Specifically, Kilpatrick (2015) defined PA as the ability engage in segmentation, identification, and the manipulation of phonemes. McGuinness et al. (1996) demonstrated increased overall word recognition and nonsense word decoding among children ranging in age from 6-16 after completion of a, then, novel reading program with explicit inclusion of PA instruction. In contrast, Bhat et al. (2003) found no significant change in word recognition after middle school students received phonological awareness training. Bhat et al. (2003) did note improved phonological skills, however, and predicted that a lack of integrated phonics instruction may explain stated results. Expectations for when command over PA skills usually occurs for typically performing readers and readers not meeting expectations, as outlined by Kilpatrick (2015), are shown in Table 4.
Table 4

Typical Schedule of Mastery Over Phonological Skills

<table>
<thead>
<tr>
<th>Level of Phonological Awareness</th>
<th>Typical Readers</th>
<th>Readers Not Meeting Expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syllable Level</td>
<td>Pre-K to Kindergarten</td>
<td>Pre-K to Second Grade</td>
</tr>
<tr>
<td>Onset-Rime Level</td>
<td>Pre-K to Early First Grade</td>
<td>First Grade to Third/Fourth Grade</td>
</tr>
<tr>
<td>Phoneme Level</td>
<td>Late First to Early Third Grades</td>
<td>Often Never</td>
</tr>
</tbody>
</table>

**Section Summary**

PA has been shown to be an important part of developing overall reading, however, it remains misunderstood in regard to its definition as well as its utility in intervening with students of all ages. Differentiation of PA from other facets of phonological awareness is vital for correct incorporation to classroom instruction. PA instruction cannot solely improve a student’s reading (NICHD, 2000; Shanahan, 2003); yet its thoughtful integration within the classroom and targeted intervention can improve students’ reading (Kilpatrick, 2013; 2015, McGee & Dail, 1996; McGuinness et al., 1996; Suggate, 2016). Various assessments and measures of PA can help teachers to identify students in need of further instruction as well as monitor progress amongst those receiving additional support through comparison with local norms as well as those presented by standardized assessments.

It has been demonstrated that PA may help to improve the skills of older readers who are not meeting expectations (Kilpatrick, 2013; McGuinness et al., 1996). Still, neither a detailed understanding of what effective implementation at the middle school level should look like nor a clear understanding of how these impacts occur is currently available to school professionals seeking to intervene with students.

**Chapter Summary**

The current lack of available research regarding PA instruction at the middle school level creates an opportunity for a program evaluation to provide meaningful feedback and recommendations for future implementation. Program evaluations provide stakeholders with data to inform future decision making by comparing a program’s purpose and how closely actual implementation align (Salvia et al., 2017; Spaulding, 2014). Effective program evaluations help to shape the quality of practices implemented (Linfield & Posavac, 2019). When considering
how to teach students to read using effective programs, the importance of curriculum design and utilization of proper assessment cannot be understated. With pressures from federal legislation, as well as LEAs and community members, school staff must ensure effective instruction is delivered and demonstrate such through assessment data. Effective instruction is inherently tied to incorporation of evidence-based practices and proper curricula design, which must be developmentally appropriate as guided by the IH (Haring & Eaton, 1978). Broadband assessment as well as PA specific measures can help to inform curricula decisions, but specific investigations of programs can provide further information regarding advantageous curriculum, assessment, and implementation decisions.

The ways of teaching reading in schools have shifted over time, due to societal influences, politics, and theory development and implementation. Most recently, the “Reading Wars” (Chall, 1967; 1983) between proponents of whole-language and phonics-based reading approaches have largely subsided, leading to a widespread push for and acceptance of balanced literacy implementation. The majority of support for balanced literacy stemmed from the NRP (NICHD, 2000), which labeled a combination of phonics and whole language approaches as the most effective way to teach students how to read. Yet, Reardon et al. (2012) estimated that a majority of students are not meeting performance expectations on measures of reading by middle school.

Bowers (2020) noted that implementation of balanced literacy is often incorrect, employing more emphasis on whole-language approaches than integration of systematic phonics instruction. In response to the need to elucidate explanations for current reading performance and make recommendations for instruction, the NRP was published (NICHD, 2000). PA was identified as one of five areas of instruction necessary for successful reading by the NRP
(NICHD, 2000), yet students cannot solely receive PA instruction to become proficient readers (Kilpatrick, 2015; NICHD, 2000; Shanahan, 2003; Yopp & Yopp, 2000). PA instruction as part of a larger reading curriculum, however, has been shown to increase students’ reading and overall academic abilities (McGuinness et al., 1996; Kilpatrick, 2013; Suggate, 2016). Kilpatrick (2013; 2015) argued for PA instruction to be delivered to all readers who demonstrate weak PA abilities, stating that there is no maximum age to improve PA awareness.

**Program Evaluation Questions**

The proposed questions that were addressed in this program evaluation included:

1. Does participation in the PA class result in increased student achievement from pre to post-test as assessed by:
   a. End-of-year final report card grades by subject
   b. NYS ELA test scores
   c. NYS math test scores

2. After participation in the PA class, is there an increased rate of passing on NYS ELA and NYS math tests overall and by demographic characteristics, including:
   a. Gender
   b. Race/ethnicity
   c. Classification status
   d. Eligibility for free or reduced-price lunch
   e. Grade level
CHAPTER 3: METHODOLOGY

This chapter includes the methods and procedures necessary to respond to the planned questions for the program evaluation. First, an overview of the school district and description of the PA program are provided. This is followed by a discussion of participants, instruments, and procedural components.

Context

A general understanding of the district, the criteria and procedures for AIS class participation, as well as the rationale behind the development of the AIS Reading class in focus is necessary for program evaluation.

The District

The district is classified by the National Center for Education Statistics as a large, suburban local district. Based on 2018-2019 enrollment data according to the NYSED (2019a), the majority of students in the district are White. Less than 25% of students identify as Black or African American, Hispanic or Latino, Asian or Native Hawaiian/Other Pacific Islander, multiracial, and American Indian or Alaskan Native (NYSED, 2019a). Across all grade levels, approximately 11% of students are classified as having a disability and receive special education services (NYSED, 2019a). Nearly one in five students is considered economically disadvantaged (NYSED, 2019a). The student population is evenly split between individuals who identify as male and female (50% each). District wide, about 220 students, or 2%, are English Language Learners (NYSED, 2019a). Approximately 21 students are identified as individuals without housing (NYSED, 2019a). The overall student to teacher ratio is reported as about 15 to one by the National Center for Education Statistics. Demographic characteristics data for all students in the district are presented in Table 5.
### Table 5

**District Demographic Characteristics**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>50</td>
</tr>
<tr>
<td>Female</td>
<td>50</td>
</tr>
<tr>
<td><strong>Race/Ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>0</td>
</tr>
<tr>
<td>Black or African American</td>
<td>3</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>5</td>
</tr>
<tr>
<td>Asian or Native Hawaiian/Other Pacific</td>
<td>10</td>
</tr>
<tr>
<td>Islander</td>
<td>79</td>
</tr>
<tr>
<td>White</td>
<td>3</td>
</tr>
<tr>
<td>Multiracial</td>
<td></td>
</tr>
</tbody>
</table>

*Note. Information for this table was gathered from New York State Education Department.*

AIS Within the District

According to a publicly available plan for supporting students, the district offers various services for students in order to support educational gains. AIS is provided by the district as one response to increase student performance across disciplines (e.g., reading and mathematics). The district identifies their AIS classes as a tier one support, meaning that it is available to all students (i.e., general education students, special education students). The purpose of AIS is to provide increased and repeated exposure to instruction already provided in the general education classroom. Additionally, it is intended that AIS reduce obstacles or barriers that prevent students from succeeding academically. Intensity of AIS is determined in accordance with students’ needs. The current district offers three intensity levels of AIS, categorized as low, moderate, and high. Low intensity AIS is described as the general classroom instruction. Moderate intensity AIS can occur in the classroom or a separate location from three to five times a week for up to 45 minutes, providing students with targeted intervention in addition to general instruction. Progress monitoring in moderate intensity AIS should occur biweekly. High intensity AIS occurs in a location outside the classroom five times a week for up to 90 minutes. Students in high intensity AIS are provided differentiated course materials and should be progress monitored every 10 instructional days.

For students in grades three through eight, qualifications for AIS, according to NYSED (2015), include scoring below pre-determined proficiency levels on State assessments, having limited English proficiency, or through universal district-specific procedures for identifying students at risk of not meeting NYS standards in core academic subjects. Core academic subjects include English Language Arts (ELA), mathematics, social studies, and science (NYSED, 2015). AIS is intended to provide additional curriculum exposure as well as address potential barriers
prohibiting students from meeting standards (NYSED, 2015). Students identified as in need of AIS are in particular need of efficiently and effectively planned lessons, as the purpose of AIS is to propel students to meeting grade level expectations. At the middle school level within the current district of interest, criteria for selecting students for AIS are evaluated by the Academic Administrator and Instructional Support Teams (ISTs). Data reviewed should include assessments (including NYS exams, final exams, and grade-wide exams routinely administered within the district), progress reports, and report cards. Teacher recommendations and input are also considered. Throughout the year, a principal may also designate a student eligible for AIS, if asked by the IST.

Special education supports should not be replaced by AIS; however, students with disabilities are able to receive AIS instruction. Within the district’s plan for Student Support Services, it is also stated that English Language Learners (ELLs) are eligible to participate in AIS. Similarly, AIS cannot replace English as a New Language (ENL) instruction. Particular attention is called for regarding the appropriateness of AIS for students receiving ELL and special education services, specifically related to the need, appropriateness, and fit for individual students.

**AIS Reading Class Development**

According to the district, the current AIS Reading class was designed to support students as performing below skill levels expected of middle school students in the areas of PA as well as general decoding and fluency. School employees were responsible for selecting individual students identified as in need of targeted instruction and supplemental teaching, which they were intended to receive through this new course offering. Students in grades six through eight were eligible to participate in the class, including students with Individualized Education Programs.
(IEPs) who receive special education and those in general education. The AIS Reading class was offered five times a week for a 40-minute period and class sizes ranged from 8 to 12 students. If this class was added to students’ schedules, they missed other areas of instruction dependent upon their grade level (sixth graders missed a general reading class; seventh and eighth graders missed ‘specials,’ such as art, technology, etc.).

Specific objectives for the class, in addition to increasing overall academic performance, included increasing students’ ability to read at least 130 words correctly per minute and for students to demonstrate proficiency on specific measures of PA in order to achieve goals of having students decode at grade level. School staff indicated that planning for the class began approximately six months before implementation during the 2018-2019 school year. No formal training was received by teachers responsible for daily instruction. A one-day conference was used as an introduction to PA focused teaching for the instructors. Teachers noted that they read two books on the subject of PA instruction, both written by Kilpatrick (2013; 2015), and engaged in planning activities throughout the summer.

Teachers integrated PA drills presented by Kilpatrick (2013) within reading class curricula already used within the district. Throughout the year, teachers used various progress monitoring assessments, however, the assessments and schedules of administration that were selected differed by each teacher. Daily practices within the classroom also differed by class section. Both used variations of small group and whole group instruction, but miscellaneous activities were implemented including one minute PA drills from Kilpatrick (2013) as well as other PA-focused activities such as phonics instruction from System 44, a reading program with a specific focus on phonics-based instruction principles (Houghton Mifflin Harcourt, 2015), and teacher created PA activities such as irregular word bingo games, splitting words into phonemes,
and assigning students to identify syllables within news articles

**Participants**

Participants’ demographic information including gender and racial identity, special educational classification status, eligibility for free or reduced lunch, and grade level was acquired through the district. Participants consisted of 34 students enrolled in grades sixth to eighth were enrolled in the AIS Reading class. General demographic information across students was provided by the district and displayed in Table 6. Teachers responsible for the class sections offered were also asked to provide any additional relevant information or provide clarifications related to the class creation and implementation.

**Instrumentation**

To complete the proposed evaluation, various forms of archival data were retrieved including students’ core class final report card grades and NYS test scores from before participation in the AIS Reading class (i.e., 2018) and at the end of the academic year (i.e., 2019).

**Final Report Card Grades**

Final report card grades examined included students’ core academic classes. According to the district’s publicly available handbook, core classes include: ELA, math, science, and social studies. ‘Reading’ is also noted as a core class, however, only sixth grade students take this class, and students who participate in the AIS Reading class are not enrolled in Reading. The district’s academic year contains four quarters, each lasting 10 weeks. Report card grades are calculated according to various factors, and might include classroom assignments, homework assignments, tests or quizzes, and participation. As students spanned multiple middle schools and various grade levels, specific information on how report card grades were calculated was not available.
### Table 6

**Participant Characteristics**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>$N$</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>21</td>
<td>61.8</td>
</tr>
<tr>
<td>Female</td>
<td>13</td>
<td>38.2</td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Black or African American</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>5</td>
<td>14.7</td>
</tr>
<tr>
<td>Asian or Native Hawaiian/Other Pacific Islander</td>
<td>1</td>
<td>2.9</td>
</tr>
<tr>
<td>White</td>
<td>26</td>
<td>76.5</td>
</tr>
<tr>
<td>Multiracial</td>
<td>2</td>
<td>5.9</td>
</tr>
<tr>
<td>Special Education Classification</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>23</td>
<td>67.6</td>
</tr>
<tr>
<td>No</td>
<td>11</td>
<td>32.4</td>
</tr>
<tr>
<td>Autism Spectrum Disorder</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Deaf-Blindness</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Deafness</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Emotional Disturbance</td>
<td>1</td>
<td>4.3</td>
</tr>
<tr>
<td>Hearing Impairment</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Intellectual Disability</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Multiple Disabilities</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Other Health Impairment</td>
<td>3</td>
<td>13.0</td>
</tr>
<tr>
<td>Orthopedic Impairment</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Specific Learning Disability</td>
<td>19</td>
<td>82.7</td>
</tr>
<tr>
<td>Speech or Language Impairment</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Traumatic Brain Injury</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Visual Impairment</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Free Lunch Eligibility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>19</td>
<td>33.9</td>
</tr>
<tr>
<td>No</td>
<td>15</td>
<td>44.1</td>
</tr>
<tr>
<td>Reduced Lunch Eligibility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>2</td>
<td>5.9</td>
</tr>
<tr>
<td>No</td>
<td>32</td>
<td>94.1</td>
</tr>
<tr>
<td>Grade Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6th</td>
<td>9</td>
<td>26.5</td>
</tr>
<tr>
<td>7th</td>
<td>22</td>
<td>64.7</td>
</tr>
<tr>
<td>8th</td>
<td>3</td>
<td>8.8</td>
</tr>
</tbody>
</table>
District-wide, however, final exams were administered at the end of each year in each core class and made up a pre-determined percentage of students’ final report card grades in the class. For sixth graders, final exams made up 5% of a final grade. For seventh graders, this increased to 10%, and for eighth graders, 15% of final grades were dependent upon students’ performance on the final exam in each class.

**New York State Tests**

In accordance with NCLB (2001) and its replacement, ESSA (2015), NYS assessments in ELA and mathematics are administered yearly to students beginning in third grade. The NYSED (2019b, p.1) identifies the purpose and intention of the assessments as follows:

“…The tests are designed to measure how well students are mastering the learning standards that guide classroom instruction and help to ensure that students are on track to graduate from high school with the critical thinking, problem solving, and reasoning skills needed for success in college and the modern workplace. The tests also show how schools and districts are progressing with the learning standards and can support professional development for teachers.”

Scores are presented as performance levels ranging from 1 through 4 (NYSED, 2018a; 2018b; 2019c; 2019d). The NYSED (2018a; 2018b; 2019c; 2019d) presented the following definitions for each level: A level 1 indicates students are performing well below expected according to standards for their grade; students who receive a level 2 are performing “partially” below expectations; level 3 scores are indicative of student proficient performance; and students who obtain a level 4 are deemed to “excel” beyond the expectations for their grade. Each level is associated with a scale score, which are shown in Tables 7 and 8. The reliability for the 2018 NYS exams was demonstrated to be acceptable across third through eighth grade exams for ELA (2018 = .87-.89; Questar Assessment, Inc., 2018) and math (2018= .91-.94; Questar Assessment, Inc., 2018). Similar levels of reliability were observed in 2019 exams (ELA= .87-.91; math= .92-.95; Questar Assessment, Inc., 2019). Questar Assessment, Inc. (2018; 2019) asserted
### Table 7

2018 and 2019 NYS State Test ELA Performance Levels and Scale Score Ranges

<table>
<thead>
<tr>
<th>Grade</th>
<th>Year</th>
<th>NYS Level 1</th>
<th>NYS Level 2</th>
<th>NYS Level 3</th>
<th>NYS Level 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2018</td>
<td>509-593</td>
<td>593-608</td>
<td>609-621</td>
<td>622-661</td>
</tr>
<tr>
<td>6</td>
<td>2018</td>
<td>514-589</td>
<td>590-601</td>
<td>602-613</td>
<td>614-657</td>
</tr>
<tr>
<td></td>
<td>2019</td>
<td>502-589</td>
<td>590-601</td>
<td>602-613</td>
<td>614-656</td>
</tr>
<tr>
<td>7</td>
<td>2018</td>
<td>511-590</td>
<td>591-606</td>
<td>607-622</td>
<td>623-654</td>
</tr>
<tr>
<td></td>
<td>2019</td>
<td>510-590</td>
<td>591-606</td>
<td>607-622</td>
<td>623-657</td>
</tr>
<tr>
<td>8</td>
<td>2018</td>
<td>507-583</td>
<td>584-602</td>
<td>603-616</td>
<td>617-651</td>
</tr>
<tr>
<td></td>
<td>2019</td>
<td>507-583</td>
<td>584-602</td>
<td>603-616</td>
<td>617-651</td>
</tr>
</tbody>
</table>

*Note.* Table adapted from New York State Education Department (2018a) *Definitions of performance levels for the 2018 grades 3-8 English language arts tests.*

### Table 8

2018 and 2019 NYS State Test Math Performance Levels and Scale Score Ranges

<table>
<thead>
<tr>
<th>Grade</th>
<th>Year</th>
<th>NYS Level 1</th>
<th>NYS Level 2</th>
<th>NYS Level 3</th>
<th>NYS Level 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2018</td>
<td>527-591</td>
<td>592-603</td>
<td>604-615</td>
<td>616-654</td>
</tr>
<tr>
<td>6</td>
<td>2018</td>
<td>528-591</td>
<td>592-603</td>
<td>604-615</td>
<td>616-656</td>
</tr>
<tr>
<td></td>
<td>2019</td>
<td>529-591</td>
<td>592-603</td>
<td>604-615</td>
<td>616-649</td>
</tr>
<tr>
<td>7</td>
<td>2018</td>
<td>524-592</td>
<td>593-605</td>
<td>606-617</td>
<td>618-644</td>
</tr>
<tr>
<td></td>
<td>2019</td>
<td>520-592</td>
<td>593-605</td>
<td>606-617</td>
<td>618-643</td>
</tr>
<tr>
<td>8</td>
<td>2018</td>
<td>527-595</td>
<td>596-609</td>
<td>610-621</td>
<td>622-651</td>
</tr>
<tr>
<td></td>
<td>2019</td>
<td>520-595</td>
<td>596-609</td>
<td>610-621</td>
<td>622-653</td>
</tr>
</tbody>
</table>

*Note.* Table adapted from New York State Education Department (2018b). *Definitions of performance levels for the 2018 grades 3-8 mathematics tests.*


assessments of content and construct validity were completed across tests and supported.

All students were expected to participate, with the exception of those identified as having “severe cognitive disabilities,” who may be found eligible to take the NYS Alternate Assessment (NYSED, 2019e). Guardians may opt to have their child not participate in state testing with no penalty to the child (NYSED, 2019f). Students who did not complete state tests remain eligible for special education and additional academic support if need is determined (NYSED, 2019f). NYSED (2019f); however, the district encourages parents to allow participation, asserting that low participation levels increase the difficulty associated with identifying with performance gaps.

As of spring 2016, the assessments are untimed and beginning in the 2018-2019 school year, the ELA and math tests occurred over two days instead of three days as in the past (NYSED, 2019b). In spring 2017, computer-based testing (CBT) became available to schools who opted to administer the state tests electronically (NYSED, n.d.). In the spring of 2019, NYSED (2019g) stated that about 19% of students who took ELA tests state wide used CBT. A comparison between students who participated through CBT and students who took paper-based tests (PBT) showed students demonstrated similar competencies regardless of test format (NYSED, 2019g).

**Procedure**

A file of de-identified archival data for students who participated in the 2018-2019 AIS Reading class was released by the school district. These data included each student’s gender, race/ethnicity, grade level, special education classification category (if applicable), eligibility for free or reduced-price lunch and various test scores as well as final report card grades. These scores included the 2018 and 2019 Performance Levels and Scale Scores for the NYS ELA and
math tests. Grades released included final report card grades from the Spring of 2018 and Spring of 2019 for core classes. Data indicated whether the student was placed in the AIS Reading class during the 2019-2020 school year.
CHAPTER 4: RESULTS

This program evaluation was performed to examine to what extent student achievement was impacted by participation in a district’s new AIS Reading class implemented at the middle school level for selected students in grades six through eight. This investigation also sought to determine if academic outcomes differed by certain student demographic categories. Students’ academic achievement was measured by final report card grades, NYS ELA test scores, and NYS math test scores. This chapter contains summaries of the analyses performed to address the proposed research questions, as described in Appendix A

Data Review and Proposed Analyses

Descriptive statistics for each academic achievement variable are presented in Table 9. Not all academic variables for every student were provided. Displayed in Table 10 are the percentages of data missing for each of the variables. With regard to missing 2018 final report card grades, 10.29% of grades across all four core classes were missing. In 2019, 2.94% of final report card grades across all four core classes were missing. It was observed that at least one year of NYS test score data for approximately one half of participants (47.06% of NYS ELA scores and 51.47% of NYS math test scores) were missing.

Proposed data analyses are found in Appendix A. It was proposed that the first research question be addressed using one-tailed repeated measure t-tests with Bonferroni corrections for significance. Chi Square tests of independence were proposed to address research question two.

Research Question One

Does participation in the PA class result in increased student achievement from pre to post-test as assessed by: (a) End-of-year final report card grades by subject (b) NYS ELA test scores (c) NYS math test scores
Table 9

Descriptive Statistics for Academic Achievement Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>M</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final Report Card Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td>32</td>
<td>69.59</td>
<td>11.88</td>
<td>29-96</td>
</tr>
<tr>
<td>2019</td>
<td>33</td>
<td>71.73</td>
<td>11.14</td>
<td>50-92</td>
</tr>
<tr>
<td>Math</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td>32</td>
<td>71.38</td>
<td>12.24</td>
<td>50-93</td>
</tr>
<tr>
<td>2019</td>
<td>33</td>
<td>70.12</td>
<td>12.75</td>
<td>44-91</td>
</tr>
<tr>
<td>Social Studies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td>31</td>
<td>73.52</td>
<td>10.24</td>
<td>53-91</td>
</tr>
<tr>
<td>2019</td>
<td>33</td>
<td>71.52</td>
<td>12.01</td>
<td>52-94</td>
</tr>
<tr>
<td>Science</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td>27</td>
<td>72.26</td>
<td>13.18</td>
<td>50-96</td>
</tr>
<tr>
<td>2019</td>
<td>33</td>
<td>71.39</td>
<td>13.04</td>
<td>48-94</td>
</tr>
<tr>
<td>NYS Test Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td>18</td>
<td>584.94</td>
<td>14.77</td>
<td>560-612</td>
</tr>
<tr>
<td>2019</td>
<td>18</td>
<td>583.67</td>
<td>16.92</td>
<td>560-620</td>
</tr>
<tr>
<td>Math</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td>16</td>
<td>589.31</td>
<td>14.38</td>
<td>555-617</td>
</tr>
<tr>
<td>2019</td>
<td>17</td>
<td>592.35</td>
<td>15.99</td>
<td>559-626</td>
</tr>
</tbody>
</table>

*Note. n = number of cases, SD = Standard Deviation*
### Table 10

**Percentage of Missing Data by Variable**

<table>
<thead>
<tr>
<th>Variable</th>
<th>2018</th>
<th>2019</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Final Report Card Grades</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELA</td>
<td>2</td>
<td>1</td>
<td>5.88</td>
</tr>
<tr>
<td>Math</td>
<td>2</td>
<td>1</td>
<td>2.94</td>
</tr>
<tr>
<td>Social Studies</td>
<td>3</td>
<td>1</td>
<td>8.82</td>
</tr>
<tr>
<td>Science</td>
<td>7</td>
<td>1</td>
<td>20.59</td>
</tr>
<tr>
<td><strong>NYS Test Score</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELA</td>
<td>16</td>
<td>16</td>
<td>47.06</td>
</tr>
</tbody>
</table>
| Math                         | 18   | 17   | 52.94%

*Note. n = number of cases*
Analyses

Repeated measures $t$-tests, also known as paired-samples $t$-tests, were utilized to examine whether there were any differences in academic achievement within subjects over time. Repeated measures were used due to the within subject nature of the program (Field, 2013). As a test, assumptions of the repeated measures $t$-test include normality of sampling distribution of means that participants are independent from one another (Field, 2013). The Shapiro–Wilk test was used due to its ability to test for normality with small sample sizes (Field, 2013), and all data across academic achievement variables were found to be normally distributed. To address the increased chance of a Type I error with multiple tests, the Bonferroni correction was used to modify acceptable probability values (Field, 2013). Results indicated non-statistically significant differences in final report card grades across all four core classes, NYS ELA test scores, and NYS math test scores between the 2018 and 2019 school year. These results are presented in Table 1.

Visual Inspection

While emphasizing the non-significance of the preceding results, it was determined that further visual inspection of the data may help to elucidate any patterns or clarify trends that may be present and inform future investigations. As previously stated, evaluations of the group level data demonstrated no significant increases in academic achievement outcomes from the 2018 to 2019 school years across all variables. At the individual level, some changes were observed. These differences in class grades are displayed visually in Figures 7 through 10 below, and differences in NYS test scores are displayed in Figures 11 and 12.

Four distinct observations were made based on the review of the visual analyses results. First, it was noted that, across academic outcome variables, the highest percentage of students
Table 11

Results of Analyses Examining Academic Achievement

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
<th>t</th>
<th>df</th>
<th>p*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final Report Grade</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELA</td>
<td>1.806</td>
<td>10.281</td>
<td>.978</td>
<td>30</td>
<td>.168</td>
</tr>
<tr>
<td>Math</td>
<td>-1.258</td>
<td>10.614</td>
<td>-.660</td>
<td>30</td>
<td>.257</td>
</tr>
<tr>
<td>Social Studies</td>
<td>-.933</td>
<td>9.652</td>
<td>-.530</td>
<td>29</td>
<td>.300</td>
</tr>
<tr>
<td>Science</td>
<td>-.500</td>
<td>8.228</td>
<td>-.310</td>
<td>25</td>
<td>.380</td>
</tr>
<tr>
<td>NYS Test Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELA</td>
<td>-.267</td>
<td>13.672</td>
<td>-.076</td>
<td>14</td>
<td>.471</td>
</tr>
<tr>
<td>Math</td>
<td>4.667</td>
<td>8.478</td>
<td>1.907</td>
<td>11</td>
<td>.042</td>
</tr>
</tbody>
</table>

*p < .008 with Bonferroni Correction
Figure 7

Changes in ELA Class Grades

Figure 8

Changes in Math Class Grades
Figure 9

Changes in Social Studies Class Grades

![Changes in Social Studies Class Grades](image)

Figure 10

Changes in Science Class Grades

![Change in Science Class Grades](image)
Figure 11

Changes in NYS ELA Test Scale Scores

![NYS ELA Test Scaled Score Changes](image)

Figure 12

Changes in NYS Math Test Scaled Scores

![NYS Math Test Scaled Scores Changes](image)
were found to increase their scores on the NYS math test by more than five points. The percentage of students with a decrease of more than five points on the NYS math test scores was small (16%). Second, overall, minimal changes (+/- five points) were noted in students’ NYS ELA test scores, math class final report card grades, and science class final report card grades from 2018 to 2019. Additionally, it was noted that the lowest percentage of students with minimal grade changes (+/- five points) were observed in ELA class and Social Studies class. Lastly, with the exception of the NYS math test, across all other achievement variables, approximately one-third of students demonstrated a decrease of more than five points on their NYS ELA test score or in their final report card grades.

Variation in individual performance between 2018 and 2019 does not equate to statistically significant change at the group level. These further inspections, however, may be of interest to the district. These visual inspections also further highlight the missing data across variables.

Research Question Two

After participation in the PA class, was there an increased rate of passing on NYS ELA and NYS math tests overall and by demographic characteristics, including (a) gender (b) race/ethnicity (c) classification status (d) eligibility for free or reduced-price lunch (e) grade level?

Analyses

To address the second research question, which specifically focused on NYS test scores and student demographics, chi-square tests were proposed as the method of data analysis. Assumptions of the chi-square tests of independence were first examined. These assumptions include independence of variables and expected frequencies (Field, 2013). It is assumed that no
expected value is smaller than five (Field, 2013). Of the five demographic categories of interest, none met the assumption of expected values.

Due to this, Fisher’s exact test was utilized. Field (2013) identified Fisher’s exact test as having the capability to compute chi-square statistics when samples are small. The term “passing” in the research question is defined as achieving a state test score of either a level 3 or level 4, as these are qualified as students who demonstrated “proficient performance” at level 3 and who “excelled beyond expectations for their grades” at level 4 (NYSED 2018a; 2018b; 2019c; 2019d). Students who achieved a performance level 1 or 2 were labeled as performing “well below” grade level expectations and “partially below” expectations respectively (NYSED 2018a; 2018b; 2019c; 2019d). Data from students who demonstrated scores of 1 or 2 were not analyzed for the purpose of this question, as the evaluation aimed to focus on increased performance.

To complete the analyses, NYS performance levels from the 2018 and 2019 ELA and Math tests were split into dichotomous categories (i.e., “Not Meeting Expectations” or “Not Passing” for scores of 1 or 2 and “Meeting or Exceeding Expectations” or “Passing” for scores of 3 or 4). As previously stated, only scores of 3 or 4 were analyzed. For the Gender demographic, students were grouped as Males or Females. For the Race/Ethnicity category, students were collapsed into a “Non-White” and “White” grouping due to the samples’ characteristics. Students were split into two groups when considering Classification status (i.e., “Classified for Special Education” and “Not Classified for Special Education”) as well as Eligibility for Free or Reduced Lunch (i.e., “Eligible” and “Not Eligible”). As no NYS test scores were available for the eighth-grade participants in 2019, comparisons between only sixth and seventh-grade students were made for the Grade Level category. No statistically significant associations across
all investigations were found, meaning improved performance on state tests does not appear to significantly differ by demographic characteristics. These non-significant results are displayed in Tables 12 and 13.

**Visual Inspection**

Similar to Research Question One, no significant results were found during analyses. Additional investigation, via visual inspection, was completed, while stressing the non-significance of findings at the group level. A high rate of missing NYS test scores resulted in extremely small samples (less than five per cell), therefore conclusions based on demographic data may be incomplete. Visual analyses highlighted that, regardless of demographic group, the majority of participants from these small samples demonstrated NYS test scores that were either well below or partially grade level expectations after participation in the PA class. No individuals from diverse backgrounds met expectations on the NYS ELA or NYS Math test in 2018 or 2019. These visual analyses of the small sample may offer some patterns of interest to the district, despite non-significance at the group level.
### Table 12.

**Results of Association between NYS ELA Test Score and Demographics**

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Met or Exceeded Expectations on NYS ELA Test</th>
<th>2018 n (%)</th>
<th>2019 n (%)</th>
<th>Fisher’s Exact Test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td>1 (16.7)</td>
<td>0 (0)</td>
<td>1.00</td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td>2 (33.3)</td>
<td>3 (50)</td>
<td></td>
</tr>
<tr>
<td><strong>Race/Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-White</td>
<td></td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1.00</td>
</tr>
<tr>
<td>White</td>
<td></td>
<td>3 (50)</td>
<td>3 (50)</td>
<td></td>
</tr>
<tr>
<td><strong>Classification</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classified</td>
<td></td>
<td>2 (33.3)</td>
<td>1 (16.7)</td>
<td>1.00</td>
</tr>
<tr>
<td>Not-Classified</td>
<td></td>
<td>1 (16.7)</td>
<td>2 (33.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Eligibility for Free or Reduced Lunch</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eligible</td>
<td></td>
<td>1 (16.7)</td>
<td>1 (16.7)</td>
<td>1.00</td>
</tr>
<tr>
<td>Not Eligible</td>
<td></td>
<td>2 (33.3)</td>
<td>2 (33.3)</td>
<td></td>
</tr>
<tr>
<td><strong>Grade Level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6th Grade</td>
<td></td>
<td>0 (0)</td>
<td>1 (16.7)</td>
<td>1.00</td>
</tr>
<tr>
<td>7th Grade</td>
<td></td>
<td>3 (50)</td>
<td>2 (33.3)</td>
<td></td>
</tr>
</tbody>
</table>
Table 13.

Results of Association between NYS Math Test Score and Demographics

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Met or Exceeded Expectations on NYS Math Test</th>
<th>Fisher’s Exact Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2018 n (%)</td>
<td>2019 n (%)</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1 (16.7)</td>
<td>1 (16.7)</td>
</tr>
<tr>
<td>Female</td>
<td>2 (33.3)</td>
<td>2 (33.3)</td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-White</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>White</td>
<td>3 (50)</td>
<td>3 (50)</td>
</tr>
<tr>
<td>Classification</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classified</td>
<td>1 (16.7)</td>
<td>1 (16.7)</td>
</tr>
<tr>
<td>Not-Classified</td>
<td>2 (33.3)</td>
<td>2 (33.3)</td>
</tr>
<tr>
<td>Eligibility for Free or Reduced Lunch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eligible</td>
<td>1 (16.7)</td>
<td>2 (33.3)</td>
</tr>
<tr>
<td>Not Eligible</td>
<td>2 (33.3)</td>
<td>1 (16.7)</td>
</tr>
<tr>
<td>Grade Level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6th Grade</td>
<td>1 (16.7)</td>
<td>1 (16.7)</td>
</tr>
<tr>
<td>7th Grade</td>
<td>2 (33.3)</td>
<td>2 (33.3)</td>
</tr>
</tbody>
</table>
CHAPTER 5: DISCUSSION

Focus on supporting students not meeting expectations across school subjects has increased over time due to rising accountability standards (NCLB 2001) and heightened attention placed upon individualized instruction, inclusion, and intervention implementation (Merrell et al., 2012; Miranda, 2014). This increased reliance on and integration of DDDM within schools has resulted in expanded efforts to collect and utilize data to inform decisions shaping instructional practices and responding to areas in need of improvement (Manidinach & Schildkamp, 2021). Yet, districts continue to face difficulties in the application and evaluation of efforts to increase student performance. Educators are faced with selecting and implementing evidence-informed practices that are not often specifically designed for their setting. The effort and time necessary to accurately determine impact of an intervention poses additional challenges. Despite varied approaches to reading instruction, students nationwide continue not to demonstrate proficiency (U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, NAEP, 2019). Program evaluation can be utilized to provide districts with specific investigations of current practices and analyses of data to discern impacts (Salvia et al., 2017).

The current program evaluation was designed to address questions about a novel reading program, AIS Reading, on students’ overall academic achievement. Academic achievement variables assessed included NYS test scores and final report card grades for core classes. Further, this program evaluation aimed to identify any noted variations in performance based on student demographics. This chapter provides discussion of the results presented in chapter four in the context of the available literature. Additionally, the district’s strengths and areas in need of future focus and improvement are offered. Limitations of the current evaluation and future implications
for the district and the AIS Reading class are included, along with a summary of the program evaluation to conclude the chapter.

**Program Evaluation Results**

This program evaluation was conducted to determine whether students’ academic achievement improved after participating in a newly implemented AIS Reading class. The CIPP model utilized focuses specifically on a program’s context, input, process, and products and (Stufflebeam & Shinkfield, 2007). Considering the stage of the class at the time of this program evaluation, the main foci of the current project concerned the *process* and *products* phases, including the assessment of program implementation and success. This CIPP model is complementary to Langley and colleagues’ (2009) PDSA model based in improvement science. Similar to the CIPP model, the PDSA approach challenges stakeholders to consistently assess design, implementation, and progress of initiatives to ensure resources are being used to achieve proposed goals and outcomes. From this lens, the current district falls in the *Study* phase of PDSA, which includes the examination of implementation results as well as the identification of potential problems and solutions (Barnett et al., 2018; Chance et al., 2010; Fixsen, et al., 2005; Langley et al., 2009; Metz and Albers, 2014; and National Implementation Research Network).

As previously discussed, research regarding the impact of PA instruction at the middle school level is extremely limited. Kilpatrick (2013) and Gillon (2018) asserted that PA instruction should be provided to anyone who does not demonstrate proficiency, yet the intensity, frequency, and methods of delivery have not been identified in the available literature. Evaluating students’ academic outcomes was meant to provide the district with data to determine if objectives were met, and how to proceed if not. One observation noted included the limited NYS test score data for all students across both years. This may be due to the option for students
and families to “opt-out” of the tests, meaning that participation is not required. Further, the small sample size prevented use of statistical tests with greater power. The missing data render results that must be considered tentative as a consequence. The conclusions that can be drawn by the current evaluation are also limited by the small sample of participants, particularly as many students who were part of the AIS Reading class did not take the NYS tests in 2018 and/or 2019, which decreased available data to analyze. These issues impacted the analyses that could be performed.

Examinations of student achievement outcomes at the group and individual level were completed. Results for all of the research questions at the group level were found to be non-significant; student achievement did not improve across final report card grades nor state test scores. More specifically, at the group level, no significant improvement in student achievement was observed across either class final grades or state test scores. Investigations at the individual level were also conducted. These visual analyses of participants were completed to provide the district with patterns within the data and potential areas of future focus despite the aforementioned lack of increased student achievement at the group level. Visual analyses related to research question one highlighted performance on the NYS math test to increase by more than five points the largest percentage of students. Minimal changes were noted across other academic outcome variables, and about one-third of students’ performance decreased by more than five points across all variables except the NYS math test; however, there is no established relationship of PA and math achievement in the literature. While no significant differences on achievement were found to exist between demographic groups, the district may find certain visual patterns of interest. Specifically, no non-white students who participated in the AIS Reading class and took the NYS ELA and math tests in 2018 and 2019 achieved a score of three
or four. Overall, regardless of demographic category, more students from the current sample did not meet expectations on the state tests than those who did.

At the time of the current evaluation, it is not clear that the district’s objectives for the 2018-2019 AIS Reading class were met based on the available data. This lack of clarity is due, in part, to missing data from the one-year span being evaluated as well as other limitations discussed below. Despite this, however, it would seem that the investment in curricula designed to aid students continuing to struggle despite previous intervention is a worthwhile allotment of resources. These additional efforts taken by the district align with increased attention to accountability and ensuring students are learning through effective instruction and practice (Ornstein, 1988; Salvia et al., 2017). While current data cannot offer evidence of improved achievement, continued refinement and evaluation may facilitate increased academic achievement for these students over time. Though the current evaluation was unable to assert whether district met its goals, continued intervention and examination of progress is key to future success (Salvia et al., 2017).

Implications

Implications for the AIS Reading class based on these preliminary findings were informed by the CIPP model and improvement science approaches identified, which are designed to enhance and increase the successes of programs through consistent and constant examination of decisions and corresponding effects (i.e., DDM). The focus on program improvement over time considers various associated factors, including costs and feasibility, in determining recommended next steps for stakeholders (Stufflebeam & Coryn, 2014). Difficulties associated with the implementation of novel programs are not rare, and consideration of challenges and specific plans to address these issues can lead to more success and target future
implementation, as opposed to the tendency to target issues after discovery in response to problems (Bryk et al., 2017). Moving forward, it is recommended that the district return primarily to the planning phase described within the PDSA cycle for various reasons provided.

**Strengths**

While unable to conclude the impact the AIS Reading class had on student achievement during the 2018-2019 year, strengths were identified. The AIS Reading class was designed and executed in response to the staff identification of students with continued PA difficulties despite prior reading intervention at various levels. The current district hypothesized that targeting these noted deficits would help students to increase academic achievement. The district is lauded for employing efforts to address reading difficulties impacting students across grade levels. The teachers and staff members who worked to study specific PA teaching strategies and design a class curriculum employed undeniable efforts, and their seeking out an expert in the field for professional development to introduce the concepts to be focused on throughout the year by the teachers in class is noteworthy. Further, the district’s continued interest in investigating the impact of the class and seeking out recommendations for future practice are commended. While incredibly important, continued dedication to refining practices to benefit students can be challenging and time-consuming.

**Limitations and Recommendations**

In completing the current program evaluation, various limitations were noted pertaining to the AIS Reading class at the focus of the investigation, as well as the field more broadly. The recommendations that come, in part, due to the limitations discussed can be used to inform future decision making and possible areas of future research or evaluation focus. As previously discussed, the district’s status at the conclusion of the 2018-2019 school year and the beginning
of this evaluation can be described as falling within the products and processes phases, specifically involving the testing/piloting of a new program and subsequent examination of the initial implementation efforts. In response to the testing and piloting phase, the process is evaluated and data are evaluated to determine if objectives were met (Stufflebeam & Shinkfield, 2007). If it cannot be determined if objectives were met, the district is asked to decide if the purpose and objectives are still worthy of investment (Stufflebeam & Shinkfield, 2007). The following recommendations are made with the assertion that the district’s mission in aiding students who are struggling to achieve grade-level standards and expectations in reading is worthy of continued efforts. To do so, it is recommended that district staff return largely to the exploration and planning phase captured within the PDSA best practice literature. Exploration and planning will provide the district with the ability to more specifically identify needs, resources, implementation barriers and aids, and the explicit plan for data collection and utilization (Sweet, 2021).

The drive for the creation of the current AIS Reading class was rooted in the conclusion that students had PA deficits that had not been addressed through previous intervention. The current evaluation was unable to determine if current academic difficulties exist as a result of a lack of PA skills as the current assessment data were not specific enough to capture PA abilities of students. Moving forward, it is suggested that the district more specifically identify the PA deficits, or other reading skills in need of improvement, of students who are selected to participate in the current class. As a result of the need for stricter entry criteria for the AIS Reading class and confirmation regarding PA ability, the district would likely benefit from returning to the planning stage to more precisely identify students’ needs and implement curricula in response through screening tools and establishing students’ baseline skill levels.
Langley and colleagues’ (2009) questions of “What are we trying to accomplish?” and “How will we know that change is an improvement?” are particularly important for the district to address as they proceed. By more concretely identifying through assessment the reading issues that these students present, and creating a more specific objective definition of the program’s purpose as well as identifying how improvement will be measured, the district can more clearly define who would likely benefit from the class. This will also require an explicitly pre-defined identification process for students in addition to the creation of criteria to determine when students have benefitted from the class and are deemed as no longer needing to participate in AIS Reading. It is recommended that staff further re-evaluate Salvia et al.’s (2017) RIOT assessment process. It appears as though some aspects of record reviews and interviews were used in selecting students for the 2018-2019 class year, but creating a standardized format in which these techniques are utilized can help, in part, to create a consistent identification process. The additional input that comes from the assessment process, apart from testing, aids in planning and implementation. Determining whether PA skills remain the main focus of the AIS Reading class will impact all future decisions, including student participation, curricula, and progress monitoring. When district staff is able to assert that deficits in PA continue to be the area of focus for the AIS Reading class, or identifies another skill more pertinent to students who struggle in reading or general achievement, educators can begin to re-shape classroom practices and processes, such as identification of students and progress monitoring.

During the first year of the AIS Reading class, uniform formative assessments were not used across all sections of the class offered. While there was some overlap in tools used, these assessments were not necessarily administered at common times or across uniform time intervals, nor recorded in a standard manner. At the time of the evaluation, the district’s actions
related to use of assessment fell within the study phase. Some formative assessments were used, but the data collected as a result were not available for all students who participated. In response to this identified limitation, it is recommended the district return to the planning stage of the PDSA cycle. The program may benefit from the identification and utilization of selected assessment tools that can be uniformly administered to students across sections to collect data on growth. More than one measure should be used, ideally, to create multiple forms of data. For example, continued use of the Phonological Awareness Screening Test (PAST) or DIBELS may be considered. The types of tests as one part of assessment presented by Salvia et al. (2017) might be explored. It will be important that district staff select tools to use thoughtfully and ensure they capture data of interest. When specifically looking at progress monitoring assessments, it is recommended that consistently utilized progress monitoring tools include narrowband measures specifically focused on students’ PA abilities to assess growth of this skill. These tools need to be uniformly given to all students, across classes, so that data over time are comparable. In addition to these narrowband measures, the district may also consider general measures of reading fluency as well, as PA is a foundational skill for fluent reading (Kilpatrick, 2013). Consideration of progress monitoring tools already used in the school should be made, as the district can then compare reading fluency rates and PA skill demonstration of students who participate in the AIS Reading class with students who receive other curricula, including those who do not receive AIS services. This can help to determine whether there is a gap between student performance, and may direct future changes to curricula within AIS or the general education setting. Once assessment tools are selected for progress monitoring, the calculation of expected rates of improvement (ROIs) as well as examination of students’ actual ROIs can offer staff insight as to how students are progressing during implementation as well as the average
ROI, and what ROI students who demonstrate proficiency levels to graduate from the AIS Reading class may have. The introduction of a formal schedule for assessment will also help to support consistent data across students and classes.

Reliance on only final report card grades as a measure of academic achievement also brings about some limitations. The incorporation of report card grades as a way to monitor growth is useful, however, these grades are not specific enough to capture increased PA skills. Within education, report card grades continue to be predominately used to communicate student performance (Guskey & Link, 2019), making their inclusion in the current evaluation a logical decision. Guskey and Link (2019) further asserted that the reliability of grades as measures may have been positively impacted by recent increased teacher training on general assessment techniques. Additionally, grades appear to be more accurate predictors of future performance, such as college graduation, than standardized tests (Galla et al., 2019). The current evaluation also utilized NYS test scores, however, many students did not participate in these annual assessments, as seen in the current data provided. In the district, slightly less than 20% of all students refused the 2019 NYS tests (19.2% for the NYS ELA test and 18.8% for the NYS math test) according to NYSED (2019h). Of these students, however, 37.1% of students with disabilities refused the 2019 NYS ELA test and 38.1% of students with disabilities refused the 2019 NYS math test (NYSED, 2019h). Further, about 30% of students who refused both 2019 NYS tests were also identified as students who were economically disadvantaged (NYSED, 2019h). As previously indicated, there is a need for more specific tools that capture PA skills. Despite the advantages that evaluating report card grades may offer, Guskey (2013) labeled percentage grades as unreliable for multiple reasons, including subjectivity associated with grading student materials as well as the common practice to establish minimum grade passing
percentages for students on report cards. Further, the weight of various class assignments and expectations across teachers and grades differs, leaving final report card grades that may not accurately portray a student’s performance in comparison to himself or peers with different teachers or in different grades. Following the previous recommendation, additional tools may be added for assessment by the district. To respond to the current limitation of utilizing report card grades and small samples of NYS test scores, it is suggested that, assuming staff incorporate additional measures, the district can return to the initial installation stage described by the PDSA model. This involves assuring that new resources are secured and implementors are made aware of them, aiding teachers responsible for collecting data through assessment by providing training or consultation, and continued checking in with educators regarding implementation efforts and challenges. Examination of report card grades should not cease, as they offer information displaying student performance in a way that is easily accessible and communicated with students and caregivers, but rather occur in conjunction with additional measures that can provide further information for future evaluators and stakeholders.

During this novel stage of implementation, it does not appear that fidelity was measured or addressed. Fidelity refers to “the degree of accuracy of a measuring instrument or model” (American Psychological Association, n.d.). To understand the effectiveness of a program utilizing evidence-based practices, the degree to which implementation occurred as prescribed, or fidelity, should be assessed (Ricardo et al., 2020). At the time of the current evaluation, the district aimed to study current practices. It is common in educational research related to intervention implementation not to report data related to fidelity (O’Donnell, 2008). Data measuring fidelity are labeled as particularly important when results are inconclusive or unclear (Swanson et al., 2011). Additional information pertaining to the accuracy and consistency of
implementation can clarify results and impacts. As such, it is recommended the district include measures of fidelity in future classes. To do this, the staff must return to the planning phase and determine what aspects of fidelity are necessary to assess in order to inform program objectives. Swanson et al. (2011) also asserted that fidelity is a crucial aspect to consider prior to scaling up an intervention. As previously stated, explicitly and objectively defining the purpose of the program as well as the components involved within the curricula will be necessary to also collect fidelity data. Once this is accomplished, it is suggested that the district outline steps in implementation and collect data related to how closely and accurately these steps are implemented. O’Donnell (2008) further recommended educators engage in continued fidelity data collection to allow for examining reliability and validity. As part of measuring fidelity, it is recommended that students’ attendance is also monitored. At the time of the current program evaluation, attendance information was not made available. Attendance rates will provide stakeholders with data related to how many days of instruction students are exposed to (i.e., dosage).

In addition to the addition of fidelity measures, it is proposed that a formal process for gathering feedback from stakeholders be included. These stakeholders may include administrators, teachers, students, and parents. Stufflebeam and Shinkfield (2007) labeled any individual who will be impacted by or contribute to an evaluation as a stakeholder. Stufflebeam and Shinkfield (2007) called for a process of stakeholder involvement as part of the CIPP model. Stakeholders can be presented with the opportunity to define the values and objectives of the program (Stufflebeam & Shinkfield, 2007). Teachers responsible for the AIS Reading class during the 2018-2019 year did have this opportunity, however, increased and formalized collections of stakeholder input can help to further empower and increase buy in of all involved
(Stufflebeam & Shinkfield, 2007). To accomplish this, district staff may consider a schedule of surveys or planned discussion to address multiple stakeholder groups’ concerns and celebrations. In considering the current evaluation, it is important to highlight that the data analyzed were collected prior to the COVID-19 pandemic, which resulted in drastic changes to schooling overall as well as the class beginning in March of 2020. The purpose of this program evaluation was to examine data collected during the first year of this novel reading class, and so while the current evaluation’s data were not impacted by COVID-19, future decisions made using recommendations that stem from this evaluation may need to consider the pandemic for the purposes of planning. This program evaluation did not address for any of the changes made or implemented during 2019-2020 nor 2020-2021 academic years, and since March 2020 the pandemic has been the major issue impacting children and schools in various ways (Benner & Mistry, 2020; Leech et al., 2020; Pressley, 2020). The evaluator was aware that the district utilized remote school options, and delivery of instruction significantly changed. The impact of COVID-19 cannot be understated, nor underestimated, in future consideration of class design and impact on students. Relying on the CIPP model and PDSA approach, the district may require serious reconsideration of the context by which the AIS Reading class may be offered in future years as a result of the pandemic. Once information is known regarding how the school year and day may be structured in the wake of COVID-19, focus on the planning stage can help the district to assess if continued implementation is feasible in the new environment and what adaptations might be necessary. Plans for data collection and assessment may also require redesign as a result of a context shift. Previously identified implementation barriers and aides may no longer be present, which may result in the continued need for teachers and staff
responsible for implementation to revisit initial program goals and the steps previously outlined to achieve these objectives.

The current evaluation was also limited by information unavailable within the field more broadly, including a lack of data regarding what PA performance levels and deficits exist nationwide at the middle school level specifically. Without this information, staff cannot compare their students’ abilities to determine whether students struggling with PA difficulties is a relative weakness within the district, or something common nationwide. There are also no best practice recommendations regarding optimal instruction frequency and intensity for PA at the middle school level. As previously stated, the NRP (NICHD, 2000) made recommendations for younger students, but this guidance has not been adapted for students who are older. In future implementations of the AIS Reading class, assuming the district continues to move forward with focus on PA skills, increased record keeping and data on the district population’s PA proficiency may help to fill some gaps the literature does not address. Lastly, the district must determine if continued focus on PA skills with students in sixth through eighth grade is necessary and aligns with long term objectives for the students. The literature contradicts impacts of intervention with older students; continued intervention is recommended by some, as research demonstrates that older students can respond to appropriately targeted intervention and demonstrate growth (Edmonds et al., 2009; Kilpatrick, 2013; Olson & Wise, 2006; Solis et al., 2012). Yet, others have asserted that extensive intervening with students above grade three often does not result in meaningful impact on overall reading outcomes (Wazneck et al., 2013). Upon the inclusion of additional screening of students for PA deficits as previously recommended, the district may be able to more easily determine the focus of future interventions. Similarities, or differences, in students’ needs may drive future foci of AIS Reading classes.
Future Directions

Assuming the district continues to target intervention for students who are not meeting expectations in reading, continued evaluation and examination of data will benefit the program and its efforts. Along with continued efforts, the district may consider examining the data of students who are enrolled in the AIS Reading class for more than one year, which was unavailable at the time of this evaluation due to focus on the first year of implementation. While not currently visible, the district may find latent effects not detected by original visual and statistical analyses of the current evaluation.

In response to the return to planning phase, the district may wish to evaluate success before continued scaling of implementation. As previously stated, the current evaluation looked at process and products. In the future, the district may consider formative evaluation of context and input. Stufflebeam and Shinkfield (2007) labeled context evaluation as the judging the needs, utilizing opportunities, and responding to problems to ensure adequate decision making and quality implementation. Input evaluation provides stakeholders with direction related to the program by supplying assessment of implementation and suggestions for improvement to better align with program objectives (Stufflebeam & Shinkfield, 2007). Increased attention to the context and input related to the current AIS Reading class likely would allow for the district to receive feedback and guidance prior to continued implementation, as the current evaluation was unable to conclude impact of the class on student performance due to a multitude of factors (i.e., small sample size, limited student test score data, and absence of data specifically related to PA skills).

In the future, if the district felt it successfully addressed limitations identified related to the current program’s context and input phases, utilization of implementation on a smaller scale
may be beneficial for further refining the launch of the program. Langley et al. (2009) described using the PDSA cycle repeatedly as one way to ensure the development of a program that can endure change over time. By using this approach to implementation, the district can respond to issues identified through formative results more quickly, allowing changes to be made moving forward before the program is too intertwined or too wide-reaching to make sweeping change easily. Langley and colleagues (2009) asserted that to do this effectively, support must be provided for proposed changes both during the program and after.

When considering the cyclical process of improvement science, it is not unexpected that the current program received recommendations to consider engaging in return to planning phases (Langley et al., 2009). It is expected that programs undergoing evaluation will experience redesign and refinement of practices to better support objectives. The current program has varied strengths as well as areas in need of further development to support student and teacher success.

**Chapter Summary**

The impact of reading across the lifespan cannot be understated. Specific to school performance, Kilpatrick (2013) highlighted reading as vital for success in all classes and areas, including students’ behavior and interactions with peers. Over time, competing theories regarding how to best teach students to read, and when and how to intervene with students not acquiring skills as expected, have challenged educators to engage in curricula and intervention development that facilitates student achievement. Reliance on formal program evaluations can help organizations, including school districts, to assess the effectiveness and quality of a program, including what action may be necessary to fulfill program objectives (Stufflebeam & Coryn, 2014).

The NICHD (2000)’s NRP identified five components of reading necessary to read
successfully, including PA. PA involves an individual’s ability to both identify and manipulate phonemes in spoken language (Scanlon et al., 2010). The current program, AIS Reading, aimed to target PA skills of students ranging from sixth to eighth grade in order to increase reading ability as well as overall academic outcomes. The district took initiative to further support students who continued to not to meet grade level expectations after receiving varied previous intervention supports through the creation of the current program.

Archival data from the first year of the AIS Reading class’s implementation were used to evaluate students’ overall academic achievement after one year of participation in the program. Additionally, potential differences in increased levels of passing on NYS tests were examined based on students’ demographics. Data analyses revealed no statistically significant increases in students’ report card grades nor their NYS test scores. Students’ rate of passing the NYS tests also were not found to be statistically different based on demographic characteristics. Visual analyses were performed for both research questions, offering some additional information that may be of interest to the district despite non-significance at the group level. These visual analyses may offer areas of future further inspection for the district. Through an improvement science lens and the CIPP model for program evaluation, recommendations for future practice were made, which largely consisted of advisement to return to the planning phases of the program to refine objectives, plans, and implementation. Limitations of the novel program at the time of implementation were noted, and limitations for the field more broadly were considered. While the implementation of the AIS Reading class was not found to result in increased academic performance of students, emphasis on the need for continued evaluation and changes to programs for success was presented.
References


Cromley, J. (2009). Reading achievement and science proficiency: International comparisons
from the programme on international student assessment. *Reading Psychology, 30*, 89-118. https://doi.org/10.1080/02702710802274903


https://www.congress.gov/114/plaws/publ95/PLAW-114publ95.pdf


https://doi.org/10.1086/509524


https://www.hmhco.com/programs/system-44/research-results

Hulme, C., & Snowling, M.J. (2013). Learning to read: What we know and what we need to understand better. *Child Development Perspectives, 7*(1), 1-5.

https://doi.org/10.1111/edep.12005


https://doi.org/10.1177/074193250102200503


https://doi.org/10.1598/RRQ.32.2.2


https://nces.ed.gov/ccd/districtsearch/


https://nces.ed.gov/FastFacts/display.asp?id=96#


115


Robb, L. (2010). *Teaching Reading in Middle School*. Scholastic.


Wazneck, J., Vaughn, S., Scammacca, N.K., Metz, K., Murray, C.S., Roberts, G., & Danielson,


**APPENDICES**  
Appendix A: Data Analyses Plan

*Research Questions and Proposed Data Analyses*

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Data</th>
<th>Proposed Analyses</th>
</tr>
</thead>
</table>
| 1. Does participation in the PA class result in increased student achievement from pre to post-test as assessed by:  
  a. Final report card grades by subject  
  b. NYS ELA test scores  
  c. NYS math test scores | Variables: 2018 and 2019 final report card grades, 2018 and 2019 ELA and math test scores | One-Tailed Repeated-measures t-tests |
| 2. After participation in the PA class, is there an increased rate of passing on NYS ELA and NYS math tests overall and by demographic characteristics, including:  
  a. Gender  
  b. Race/ethnicity  
  c. Classification status  
  d. Eligibility for free or reduced-price lunch  
  e. Grade level | Variables: student demographics, 2018 and 2019 NYS ELA and math test scores | Chi Square Tests of Independence |