A Mixed Methods Examination of Elementary Teachers’ Conceptions of Benchmark Reading Assessments

Marie N. Derby Ed. D.
West Chester University of Pennsylvania, md686656@wcupa.edu

Follow this and additional works at: https://digitalcommons.wcupa.edu/all_doctoral

Part of the Curriculum and Instruction Commons, Educational Assessment, Evaluation, and Research Commons, Educational Leadership Commons, Educational Methods Commons, Elementary Education Commons, Elementary Education and Teaching Commons, Language and Literacy Education Commons, Other Teacher Education and Professional Development Commons, and the Pre-Elementary, Early Childhood, Kindergarten Teacher Education Commons

Recommended Citation

This Dissertation is brought to you for free and open access by the Masters Theses and Doctoral Projects at Digital Commons @ West Chester University. It has been accepted for inclusion in West Chester University Doctoral Projects by an authorized administrator of Digital Commons @ West Chester University. For more information, please contact wcressler@wcupa.edu.
A Mixed Methods Examination of Elementary Teachers’ Conceptions of Benchmark Reading Assessments

A Dissertation

Presented to the Faculty of the

College of Education and Social Work

West Chester University

West Chester, Pennsylvania

In Partial Fulfillment of the Requirements for

the Degree of

Doctor of Education in Policy, Planning, and Administration

By

Marie Nanette Graziano Derby

May 2021

© Copyright 2021 Marie Graziano Derby
Dedication

This body of work is dedicated to my parents, Helen and Michael Graziano. Thank you for surrounding me with text.
Acknowledgements

I am acknowledging Dr. Heather Schugar for her tenacity, work ethic, and honesty. West Chester University is so fortunate to have Dr. Schugar on staff. Her ability to teach, with patience and a listening heart throughout the dissertation process was remarkable. Dr. Schugar inspired me to work harder than I ever have because she was there every step of the way, leading, pushing, pulling, and running alongside me. Our long conversations of literacy instruction and assessment revealed a shared passion which, I hope, leads to more research projects and several articles.

In addition to recognizing Dr. Schugar, it is important to me to thank the teachers of my past that shared Dr. Schugar’s passion for learning and working hard. Over the years when asked to write or speak about teachers who influenced me, I name Ms. Tuffner, and Mr. Gallagher, two high school teachers who never gave up on me, like Heather.

My professors at West Chester University in the Policy, Planning, & Administration Ed.D. program were top-notch, accessible, and passionate about sharing their expertise. I thank each one of them.

This process was hard and required a time commitment unlike anything else I have experienced. Thus, having support from my family, friends, and colleagues at work inspired me to continue through the toughest months. The cohort structure that West Chester uses was remarkably effective for me. My buddies in Cohort 3 became family and although we lost a year of in-person contact due to the Covid-19 pandemic, we cheered one another on via Zoom and created important bodies of work. Their compassion and deep desire to make the world a better place through, by, and for children is remarkable and inspirational.
Abstract

This study examined the case of 22 mid-Atlantic K-5 public school teachers’ conceptions of the Fountas and Pinnell Benchmark Reading Assessment Systems 1-3. The Namaste School District delivers ELA curriculum through reader’s and writer’s workshop. Triangulated data showed that participants held varying conceptions of the F&P. Specifically, they shared similar conceptions regarding the importance of one-on-one time with students, the amount of time it takes to administer the assessment, and the subjectivity of the F&P results. Results of this study provided similar results to other studies of teachers’ conceptions of assessment and adds to the literature debate regarding students’ instructional reading levels. The Complexity of Education Theory provided a dynamic theoretical framework through a sequential, explanatory quan → QUAL mixed methods design (Creswell & Plano Clark, 2018). In Phase I, teachers (N=22) completed Brown’s (2004) Teachers’ Conceptions of Assessment III-abridged version (TCOA-III). The 27-item, positively-packed, Likert-style scale had 6 indicator choices ranging from strongly disagree to strongly agree. The survey’s four constructs revealed teachers’ conceptions of the F&P as it improves teaching and learning, as it holds schools and students accountable, and the irrelevancy of it. Phase I data was analyzed and 6 participants, chosen to create a heterogenic group, were invited to continue to Phase II, semi-structured interviews.

Keywords: teacher conceptions, reading assessment, mixed methods, leveled reading, elementary literacy practices
# Table of Contents

List of Tables.................................................................................................................iv

List of Figures..................................................................................................................x

Chapter 1: Introduction ........................................................................................................1

  Purpose of the Study ........................................................................................................3
  Rationale for Study ...........................................................................................................4
  Problem Statement ...........................................................................................................5
  Research Questions .........................................................................................................6
  Significance of Methods ..................................................................................................7
  Significance of Study .......................................................................................................8
  Definition of Terms ...........................................................................................................8
  Limitations ......................................................................................................................10
  Summary ...........................................................................................................................11

Chapter 2: Review of Literature .........................................................................................12

  Benchmark Assessment Defined ..................................................................................13
  The Consequences of Labeling .....................................................................................15
  Benchmark Assessments & IRIs .......................................................................................16
  Fountas and Pinnell Benchmark Assessment Systems ..................................................17
  Reader’s Workshop as a Complex System .......................................................................17
  Teachers’ Conceptions Impact Reader’s Workshop Assessment Praxis .......................19
  Teachers’ Knowledge of Assessment .............................................................................19
Internal validity ........................................................................................................................................39
External Validity/Generalizability ........................................................................................................39
Reliability ...............................................................................................................................................40
Researcher’s Bias ...................................................................................................................................40
Procedures ............................................................................................................................................41
Quantitative data ...................................................................................................................................41
Qualitative data ......................................................................................................................................41
Analysis and Coding Procedures .........................................................................................................42
Data Integration .......................................................................................................................................44
Summary ...............................................................................................................................................44
Chapter 4: Results ................................................................................................................................46
TCoA III Three Item Reliability Scales .................................................................................................47
Participant Demographics ......................................................................................................................49
Case Study Participants ............................................................................................................................51
Assessment as Improvement .................................................................................................................55
Teachers’ Conceptions of Improvement Aggre. by Likert Scale Agreement Indicators. ................56
Teachers’ Conceptions of Improvement Realized Through Three-Item Scaling .......................57
Teachers’ Conceptions of Improvement as Realized Through Interview Data ........................58
Grouping for Bookclubs .........................................................................................................................59
Benefits of One-on-One Time with Students .....................................................................................60
Assessment as School Accountability ..................................................................................................62
Teachers’ Conceptions of School Acct. Aggre. by Likert Scale Agreement Indicators................62
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Teachers’ Conceptions of School Account. Realized Through Three-Item Scaling ........... 63</td>
</tr>
<tr>
<td></td>
<td>Teachers’ Conceptions of School Account. as Realized Through Interview Data ........... 64</td>
</tr>
<tr>
<td></td>
<td>Assessment as Student Accountability ...................................................... 65</td>
</tr>
<tr>
<td></td>
<td>Teachers’ Conceptions of Student Acct. Aggre. by Likert Scale Agreement Indicators .... 65</td>
</tr>
<tr>
<td></td>
<td>Teachers’ Conceptions of Student Acct. Realized Through Three-Item Scaling ........... 66</td>
</tr>
<tr>
<td></td>
<td>Teachers’ Conceptions of Student Account. as Realized Through Interview Data ........... 68</td>
</tr>
<tr>
<td></td>
<td>Assessment as Irrelevant ........................................................................... 69</td>
</tr>
<tr>
<td></td>
<td>Teachers’ Conceptions of Assessment as Irrelevant Aggre. by Likert Scale Agreement Indicators ........................................................................... 70</td>
</tr>
<tr>
<td></td>
<td>Teachers’ Conceptions of Assessment as Irrelevant Realized Through Three-Item Scaling ........................................................................... 71</td>
</tr>
<tr>
<td></td>
<td>Teachers’ Conceptions of Assessment as Irrelevant Realized Through Interview Data .... 71</td>
</tr>
<tr>
<td></td>
<td>Teachers’ In the Moment Decisions During Administration ................................. 73</td>
</tr>
<tr>
<td></td>
<td>Summary ....................................................................................................... 76</td>
</tr>
<tr>
<td></td>
<td>Chapter 5: Discussion ................................................................................... 77</td>
</tr>
<tr>
<td></td>
<td>Summary of the Study .................................................................................... 78</td>
</tr>
<tr>
<td></td>
<td>Application of Theoretical Framework to Findings .......................................... 79</td>
</tr>
<tr>
<td></td>
<td>Theoretical Framework Applied to the Researcher ............................................ 81</td>
</tr>
<tr>
<td></td>
<td>Theoretical Framework Applied to the F&amp;P .................................................... 82</td>
</tr>
<tr>
<td></td>
<td>Limitations of the Study ............................................................................... 83</td>
</tr>
<tr>
<td></td>
<td>Consequences of the Pandemic ....................................................................... 84</td>
</tr>
<tr>
<td></td>
<td>Emergent Themes ........................................................................................... 84</td>
</tr>
<tr>
<td></td>
<td>Overview of Study Results ............................................................................ 85</td>
</tr>
</tbody>
</table>
Accountability .................................................................................................................. 87
It Feels Like a Chore ........................................................................................................ 87
Time ................................................................................................................................ 87
Irrelevant .......................................................................................................................... 88
Limitations in Methodology ............................................................................................ 89
Limitations in Analysis ..................................................................................................... 89
Limitations in Generalizability ......................................................................................... 90
Implications for Future Research .................................................................................... 90
Dynamic Assessment to Increase Student Agency ......................................................... 93
Diversity is Everywhere ................................................................................................... 95
The Power of Relevant Text .............................................................................................. 95
Fountas and Pinnell's System Approach ........................................................................... 95
Dynamic Assessment to Increase Student Agency ......................................................... 93
Summary ........................................................................................................................... 96
References .......................................................................................................................... 99
Appendices ........................................................................................................................ 120
List of Tables

1. Participants’ Responses by Construct ......................................................... 49
2. Participants’ TCoA-III Four-Construct Scores ........................................... 50
3. Teachers’ Conceptions of Improvement Aggregated by Likert Scale Agreement
   Indicators ........................................................................................................ 56
4. Complementary Quantitative and Qualitative Data for Improvement ............. 57
5. Teachers’ Conceptions of School Accountability Aggregated by Likert Scale
   Agreement ........................................................................................................ 62
6. Complementary Quantitative and Qualitative Data for School Accountability .... 63
7. Teachers’ Conceptions of Student Accountability Aggregated by Likert Scale
   Agreement ........................................................................................................ 65
8. Complementary Quantitative and Qualitative Data for Student Accountability ... 66
9. Teachers’ Conceptions of the F&P Aggregated by Likert Scale Agreement ... 69
10. Complementary Quantitative and Qualitative Data for Irrelevancy ................. 71
List of Figures

1. Theoretical Framework ................................................................. 29

2. Theoretical Framework ................................................................. 77
Chapter I: Introduction

With input from nearly 1500 respondents across 63 countries and territories, the International Literacy Association’s (ILA, 2020) *What’s Hot in Literacy Report* identified that to improve literacy outcomes in the next decade, two of the most critical movements include increasing both equity and opportunity for all learners and improving professional development (p. 6). Increasing equity and opportunity for all learners is a timely goal as the National Academy of Sciences (2020) noted an “ongoing increase in the racial/ethnic diversity of students in U.S. schools” (p. 26). They project that by 2027, the White student population will decrease to less than 50% of the U.S. student body, while the Hispanic student body will increase. To concur, Garcia et al. (2009) found that U.S. schools’ demographic landscape is rapidly changing to include more children and youth from language and cultural minority groups. Teachers, on the other hand, are mostly White. In the 2015-2016 school year, 76.6% of teachers were White females, 56% of whom were between the ages of 30-49 (NASEM, 2020, p. 22). While there is research to support “that students of color benefit from having teachers of color” (NASEM, 2020, p. 25), in states such as Pennsylvania, the setting for this study, teachers of color are underrepresented and only 1.2% of male teachers in Pennsylvania are Black (p. 6). These statistics make Pennsylvania the most disparate state in the country in terms of hiring teachers of color (NASEM, 2020).

In addition to and in support of rapid demographic changes, assessment standards in the U.S. were updated in 2015 while at the same time, President Obama’s *Testing Action Plan* promoted a restoration of balance “to America’s classrooms by ensuring fewer, better, and fairer tests” (ESSA Title 1 Part A & Title 1, Part B, p. 1). This restoration of balance refers to time and equitable assessments. Specifically, The Every Student Succeeds Act (2015) promoted a rebalancing of time in the classroom by directing school districts to
eliminate extraneous duplicative testing and replace it with innovative, dynamic techniques to promote learning and assessment for all students (ESSA Title 1 Part A & Title 1, Part B).

Christ et al. (2018) found that one way to promote a more equitable learning environment was to first promote a more equitable testing experience. Noticing a difference between students’ classroom performances and their Independent Reading Inventory (IRI) scores, they were prompted to examine this phenomenon. Their results revealed “students who rated the places and experiences in text as culturally relevant had higher IRI critical evaluation and connection scores. Moreover, students’ word recognition meaning-maintenance, rereading at miscues, and literal comprehension scores were higher” (p. xi). Had they not noticed this difference, those students’ original lower IRI scores may have placed them on a different, lower learning trajectory according to the teachers in the study (Christ et al., 2018).

The World Development Report (2018) noted that schools are failing learners as a result of “one or more of four key school-level ingredients for learning: prepared learners, effective teaching [assessment], learning focused inputs, and the skilled management and governance that pulls them all together” (p. 9). Teachers are the most important factor affecting learning in schools which is supported by the fact that “U.S. students with great teachers advance 1.5 grade levels or more over a single school year, compared with just 0.5 grade levels for those with an ineffective teacher” (World Development Report, 2018). Thus, it is imperative to examine teachers’ beliefs and understandings of the context in which they work. Developing an understanding of local practice builds a foundation for improving assessment techniques such as discourse with colleagues and participating in thoughtful research (Darling-Hammond et al., 2005). Thus, as Barnes et al. (2014) noted, “The ethics of assessment practices and teachers’ beliefs about those practices is an area ripe for investigation” (p. 298).
Cultural, ethnic, assessment literacy, and belief system differences, if left unchecked, may lead teachers to unfair conclusions regarding students’ abilities (Kuhl et al. 2013). Teachers’ conclusions regarding students’ abilities and how their home lives have either prepared them well for school or have not, are known to alter educators’ decision-making processes and students’ educational trajectories (Bandura, 1978; Barnes et al., 2015, 2017; Christ et al., 2018; Darling-Hammond et al., 2005; Harris & Brown, 2009; Kontovourki, 2012; Pellegrino & Chudowsky, 2003; Robinson-Zanartu, 2013; Sanchez, 2017; Yamtim & Wongwanich, 2014).

**Purpose of the Study**

Thus, in light of the aforementioned research regarding the roles of teachers’ conceptions, and the fact that the Namaste School District uses the F&P as one of the assessment data points when considering students for RTII support services, this study examined what teachers’ conceptions of the F&P were, how teachers’ operationalized administration of the F&P, and how teachers’ conceptions of the F&P impacted their decision-making processes. The study of this case utilized a sequential, explanatory quan → QUAL mixed methods design (Creswell, 2018; Yin, 2018) to build upon current research to better understand the why behind teachers’ assessment decisions and their approach to administering benchmark reading assessments (Brown, 2004).

Ortiz & Flanagan (2013) recognized the complexities of evaluating learners, systems of education, and teachers and reminded educators, researchers, and policymakers that the results of a research study, mirror the theories utilized in its design. Therefore, the complexity of education theory guided this study and was utilized as the theoretical framework (Davis & Sumara, 2006; Mason, 2018). The complexity of education theory “concerns itself with environments, organizations, or systems that are complex in the sense that very large numbers of constituent elements or agents are connected to and interacting with each other in
many different ways” (p. 36). These interactions create emergence; thus, the complexity of education theory is a theory of emergence.

**Rationale for Study**

Organizations successfully transition to new policies, new curriculums, or new ways to assess students because they already had capacity-building mechanisms in place (Carnoy, 1978; Kahl et al. 2013). Thus, to incorporate the 2015 updated standards of assessment, a school district with built in capacity is better designed to incorporate the changes (Carnoy, 1978). Effective acquisitions of change in organizations, such as school districts, occur because efficient systems have the capacity to absorb and deal with new expectations placed upon them from a source which is outside of the system (Carnoy (1978, 1985, 2017).

In addition, the National Academy of Science (2001) reported that teachers use intuitive or qualitative methods to interpret assessment results more often than formal statistical models. Bias, beliefs, worldviews, and contextuality influence teachers’ intuitive and inferential decisions; therefore, teachers’ knowledge of all assessments (assessment literacy), and their conceptions, requires constant updating (Calkins, 2018; Fountas & Pinnell, 2018; Gaskins & Labbo, 2007; Kontovourki, 2012; Thompson, 1992).

Assessment literacy (AL) is now a highly-researched phenomenon in education (Barnes & Dacey, 2014; Brown, 2004, 2006, 2008, 2009; Campbell, 2013; Fives & Buehl, 2012; Kahl et al. (2013); Remesal, 2011) and this understanding of how one’s conceptions of particular assessments and students’ home lives is casually related to one’s own decision-making processes is a component of assessment literacy. In addition, Bonner et al. (2018) reported the need for researchers in the U.S. to study the relationship between teachers’ perceptions and practices within assessment. Moreover, the autonomous expression of teachers’ professional knowledge is necessary for “sustained improved instruction” (p. 90) as is the knowledge and nimbleness of administrators (Schneider & Andrade, 2013). Thus, this
study was nested within current research and extended the body of knowledge regarding the impact of teachers’ conceptions of the F&P, housed within the complexity of education theory which allows for the emergence of new ideas within a bounded system.

**Problem Statement**

The quality of information obtained through assessment can impact the quality of educational decisions (Cronbach, 1970), and the quality of data obtained is evaluated through the collection of validation and reliability evidence. The quality of assessment information includes awareness of (a) any limitations of assessment information, including the degree of inaccuracy in any measure, (b) dangers in over-reliance on any single measure, and (c) any unfair consequences for students (Cronbach, 1970; Hall, 2000; Linn, 2000; Popham, 2000; Thorndike, 1997). Unfortunately, many teacher-made or classroom assessments and intuitive judgments lack such quality indicators (McMillan, 2001b, p. 27).

Burns et al. (2015) concluded from their study that the instructional level estimates based on accuracy while reading three books rated to be at the same difficulty level by an IRI had a study kappa coefficient of .50 (.70 is considered strong). Their second research question, to what extent do students accurately read books that are rated at their instructional level using IRI data showed that 28% of the median scores fell within the instructional level range (93%-97%). Fifteen percent of the participants were reading books that were too difficult and 56% of participants were reading books deemed to be too easy. Overall, they reported that the F&P overestimated the reading levels of students at or below the 25th percentile and underestimated reading levels for students above the 25th percentile (p. 3).

Relatedly, Howard (2009) cautioned school districts to take care when designing or purchasing assessment materials used to screen students for Response to Intervention (RtI) groups. One consideration is that “packaged assessment has become a financial gold mine for publishers” (p. 100), and Howard (2009) warned that choosing a tool from a popular list may
lead to alarming consequences. Howard advised school districts to develop teacher ownership of test results through professional learning communities, but her article did not acknowledge the impact or relationship that teachers’ conceptions of assessment have on the experience and the outcome.

Assessment education's effectiveness, offset by teachers’ conceptions, emotions, needs, and prior experiences within assessment, is a ripe field of study (Brown, 2006a, 2008b; DeLuca & Bellara, 2013; Smith et al., 2014; Xu & Brown, 2014). In concurrence, Hill (2017) explained that teachers benefit when and if they learn to understand themselves as individuals and professionals. This endeavor is complicated and achieving the expansion of teachers’ conceptions of assessment involves many interactive and simultaneous factors. The first step in learning is the awareness and wherewithal to seek answers from a more knowledgeable other. Hill (2017) reported that one explanation for teachers’ inability to express their learning needs is the lack of built-in reflective muscles. In addition to understanding how political, social, cultural-historical, and institutional factors influence teacher praxis, Hill (2017) explained teachers also need to evaluate how their assessment practices impact students, society, and institutions (p. 4).

**Research Questions**

This study’s primary research question was What are teachers’ conceptions of the F&P? Two sub-questions supported the data collection of the primary research question:

1. In what ways do teachers operationalize the administration of *The Fountas and Pinnell Benchmark Assessment Systems 1-3*?
2. How do teachers’ conceptions of assessment inform their administration of benchmark reading assessments?

To answer these research questions, I used a two-phase mixed-methods study, which emphasized the lived experiences of K-5 teachers who administer the F&P. Teachers’
conceptions about assessment have ethical implications for their students. Analyzing the quantitative survey data led me to purposely choose participants who, through semi-structured interviews, provided a voice to the studied phenomenon. Triangulation of data added to the research regarding teachers’ conceptions of assessment specifically related to the administration of the F&P (Jicks, 1979).

**Significance of Methods**

Assessment is a complicated endeavor, and researchers are moving towards using different theoretical frameworks and methods to enable the gathering of all available data (Brown 2007; Carpenter, 2008; Creswell, 2018; Darling-Hammond et al. 2005; Denzin, 2011; Flores et al., 1991; Garcia, 2002; Greene, 2010; Harraway, 1998; Ortiz & Flanagan, 2013; Remesal, 2011; Robinson-Zanartu, 2013; Sanchez, 2017; Schwartz, 2002; Xu & Brown, 2016; Zemack et al., 2007). Researchers use mixed methods when one technique is not enough to capture the data they seek. Mixed-method designs allow the researcher flexibility to use two or more methods such as quantitative and qualitative data collection techniques. One reason to use mixed methods is to show how the results from one data collection source validate the results from the other source of data (Creswell & Plano Clark, 2011, 2018; Guetterman et al., 2015; Maxwell, 2012). Another reason to use mixed methods is if the context of the research typically relies on or considers one method better. The F&P assessment is a combined quantitative and qualitative reading inventory. Thus, a mixed methods study mirrors the F&P.

**Significance of Study**

Fives and Buehl (2012) synthesized 627 peer-reviewed empirical studies written in English and published prior to 2009 in order to better understand the existing research on teachers’ beliefs. Fives and Buehl (2012) were confounded to find that although teachers’ beliefs occur within systems, “few empirical investigations have examined beliefs as complex
systems” (p. 477). Fives and Buehl (2012) ultimately defended their stance regarding teachers’ beliefs in their literature review. Critiques of studying teachers’ conceptions have referred to beliefs as a moving target, situated within contexts, and seemingly ineffectual because of policy and leadership constraints (Fives and Buehl, 2012). Steadfastly, Fives and Buehl (2012) recommended that teachers slough off the hindrances that their beliefs cause by becoming aware of their own beliefs and by considering the basis for them (p. 488).

To concur, Klingbeil and colleagues (2015) noted that additional research regarding teachers’ assessment literacy is warranted, because teachers in their study made ‘judgment-based’ decisions and were found to be lacking in diagnostic accuracy.

Examination of teachers’ conceptions of assessment within a system’s approach aligns with Fives and Buehl’s (2012) and Gerhard and Mayer-Smith’s (2008) work. Gerhard and Mayer-Smith (2008) explained that complexivists “look at life as structured in networks and webs of relationships” (p. 6). Based on this premise, the complexity of education is a good fit as a theoretical model for this study.

**Definition of Terms**

The following terminology is relevant to this study:

**Assessment Literacy**

Assessment literacy consists of an individual’s understanding of the fundamental assessment concepts and procedures deemed likely to influence educational decisions (Popham, 2018).

**Benchmark Assessment**

Benchmark assessments are given periodically throughout a school year to establish baseline achievement data and measure progress towards a standard or set of academic standards and goals (NWEA, 2013).

**Beliefs**
Beliefs are the meanings connected to psychological objects or phenomena and are an environmentally contingent and culturally defined lens through which sense is made of events, people, and interactions (Brown, 2004).

**Cognitive footprint**

Cognitive footprints are the evidential markers of teachers’ decision-making practices as they relate to latent conceptions of assessments revealed.

**Formative Assessment**

Formative assessment is a planned, multistep process, relying on assessment-elicited evidence of students’ status, used by teachers to adjust their ongoing instructional procedures or by students to adjust their current learning tactics (Popham, 2008).

**Performance Assessment**

**Praxis**

Praxis is our belief system and content knowledge in action (Pinar, 1996, p. 400).

**Social Justice**

Engagement in the active transformation of individual actions, communities, and institutions to be more equitable for groups marginalized due to their disadvantaged status in society to create a more unified system (Fietzer & Ponterotto, 2015).

**Teachers’ Conceptions of Assessment**

Teachers’ conceptions of assessment can be understood in terms of their agreement or disagreement with four purposes to which assessment may be put, specifically: (a) improvement of teaching and learning, (b) school accountability, (c) student accountability, or (d) reading assessment as irrelevant (Brown, 2004, 2006, 2008, 2009; 2016).

**Limitations**

Creswell and Plano Clark (2018) recommended that researchers use a subset of participants from the initial quantitative phase of the study as this subset, in their opinion, is
the best suited to provide more enriching details during the qualitative phase. For me, this was a limitation because teachers were free to abandon their participation in the study at any time. To delimit this factor, I applied a flexible interview schedule to meet my participants’ needs best and addressed any concerns they had regarding their data protection.

Another limitation of the study is that teachers appear to hold multiple and, at times, contradictory conceptions without being disturbed by such contradiction (Brown & Harris, 2009, p. 70). To delimit this factor, I focused on discussing teachers’ conceptions of assessment to the target: the F&P. Five and Buehl (2012) guided researchers to assess teachers’ conceptions within a specific context, which is the approach I took with this study.

Because of the COVID-19 pandemic, I interviewed teachers via Zoom, and this may have inhibited their participation as typically researchers record interviews with an audio device such as a recorder or mobile. To delimit participant angst, I reiterated my use of a password-protected laptop to house encrypted Zoom interviews.

My positionality within the setting and my choice of the TCoA-III survey tool, and semi-structured interview protocol may also be seen as limitations. To delimit those factors, reliability and validity measures were utilized such as member checking (Kornbluh, 2015), descriptive statistical analysis of quantitative data, and triangulation of all three data points. My biases and worldviews were potential limitations to this study (Padilla-Diaz, 2015; Lichtman, 2013; Marshall & Rossman, 2016; Patton, 2002; Richards, 2005). For example, one of my biases, from my reading specialist certification, my ESL certification work, and the work I have completed thus far towards the National Board Teacher certification have taught me the importance of using an assessment tool for the sole purpose of its design. Another bias that I have, grown from more than 30 years as an educator is the lack of authority classroom teachers have to make decisions regarding their craft (Morse, 1991).
Finally, another limitation of my study might stem from the theoretical framework I chose which was guided by my worldview. I see fit to use the complexity of education theory as my theoretical framework in order to see the context of the “interconnectedness and interrelatedness” of how teachers’ conceptions of the F&P are operationalized during the administration of said assessment (Gerhard & Mayer-Smith, 2008, p. 10).

More recognizable theoretical frameworks in education include: constructivism, socio-cultural theory, transformative frameworks, or the feminist framework. While those theories offer important lenses and frames, the complexity of education theory is better suited to this study. Although the concept of complexity is accepted and readily used in other scientific research, it has not infiltrated education research and therefore this study may not be widely accepted (Davis, 2006).

Summary

In this chapter, I introduced the increasing racial diversity of students in U.S. schools, recognized Pennsylvania as the most disparate state in their hiring practices of teachers of color and gave a historical context for the study (NASEM, 2020). I further discussed that although evidence shows teachers’ conceptions of assessment impact their decision-making processes and despite the call by researchers to enhance and modify pre- and in-service professional development of teachers’ assessment literacy little had been done in the field to recognize the fact that students’ educational trajectories rely on the expertise of their classroom teachers. In the next chapter, I reviewed the literature related to components of the F&P benchmark reading inventory, the components of the reader’s workshop model, and the factors that comprise teachers’ conceptions of assessment. In addition, the theoretical framework that I used to guide this study is explained in detail.
Chapter II: Review of Literature

The following section will address the significant bodies of literature relevant to this study: (a) benchmark & reading assessment research, (b) the reader’s workshop model, (c) teachers’ conceptions of assessment research, and (d) the theoretical framework that I propose to guide this study. Teachers’ conceptions of assessment impact educators’ decision-making processes (Brown, 2016; Harris, 2009; Remesal, 2011), and teachers’ conceptions are essential to recognize and research. Teachers are the most crucial factor for student success; therefore, studying their ‘wholeness’ is vital (Brown, 2016; Campbell, 2013; Darling-Hammond et al., 2000; Remesal, 2013; Wiliam, 2006).

The literature review builds foundational knowledge to support my research proposal to study this case using a sequential explanatory quan → QUAL mixed methods design to answer the following questions: What are teachers’ conceptions of the F&P?

1. In what ways do teachers operationalize the administration of The Fountas and Pinnell Benchmark Assessment Systems 1-3?

2. How do teachers’ conceptions of assessment inform their administration of The Fountas and Pinnell Benchmark Assessment Systems 1-3?

Shanahan (2020) noted how Freud impacted the world of education and assessment by encouraging parents and teachers to avoid frustrating activities to keep children calm. Combining the field of psychology and reading assessment, Betts (1940) designed a system based on Freud’s proclamations. Recent studies suggest no relationship between difficult text and motivation, or frustration, exists (Shanahan, 2020). Students who are interested in reading a book will do so regardless of the difficulty level of the text (Shanahan, 2020). Debunking otherwise recognized scientific data, such as data collected during administration of an IRI or a BAS, is supported by the work of Hyslop-Marginson (2010), Kuhn (1970), Nilsson (2013), Popper (1963), Weber (1993), and now Shanahan (2020).
The value of administering an IRI or a BAS is in the observed reading behaviors, not the instrument itself or the subsequent percolation of the independent and frustration level (Powell, 1970). Powell noted the child does not have an instructional level; she has a performance level, promoting learnability. The instructional level is the teacher’s task; it is the indicated place that teachers use to plan instruction (Fountas & Pinnell, 2008; Powell, 1970, Shanahan, 2020). However, it is common to hear teachers, principals, parents, and students say, “I am a level K” for example. To this point, publicly indicating an instructional reading level comes with consequences (Black & Wiliam, 1998; Benke & Erekson, 2018; Brown, 2008; Delpit, 2012; Hoffman, 2017; Ladson-Billings, 1995). Furthermore, Shanahan noted that “restricting students to easier materials usually means preventing them from dealing with content at their age and may serve to isolate these children from their social peers” (p. 22). Shanahan reiterated that the instructional level is the level that the teachers, not the students, use.

**Benchmark Assessment Defined**

Recent articles discuss the differing definitions of formative and benchmark assessments. Fountas and Pinnell (2018) themselves recently changed how they identify the F&P. In previous printings, they designated the F&P as a standardized test, while in the latest printing (third edition), they define it as a criterion-referenced test. According to the F&P teacher’s manual, a benchmark is a standard against which to measure something. Specifically, they noted for the F&P, the standard is a student’s read aloud and how they talk about the text during the comprehension conversation. Teachers are directed on page 16 to have students read until they read with less than 95% accuracy or when comprehension falls below the limited threshold. Next, the administration directions note that teachers, to maintain standardization, need to take care not to embellish the introduction or enter into any additional conversation with the student about the text (Fountas & Pinnell, 2018, p. 20). To
continue, if a student reads just below the 95% accuracy threshold with repeated errors, the teacher may want to engage in the comprehension conversation with the student (Fountas & Pinnell, 2018, p. 24).

Directions for assessing comprehension require teachers to have a conversation with students about the text. Checking for comprehension should not include teachers asking the questions that are on the teacher’s guiding and scoring paperwork. Teachers are looking for behavioral evidence of a student’s understanding. Prompting should not result in a lower score and avoid leading (Fountas & Pinnell, 2018).

When testing multilingual students, the manual recommends that teachers make an extra effort to draw them into a conversation as they may know more than they can explain in English—make a judgment but be sure you have clear evidence of understanding. Noting differences in dialects and accents of multi-lingual readers while adhering to the scoring recommendations (which recommends that teachers do not mark errors for ELs based on accents or other particulars related to the child’s native language) may delimit false positives for reading support (researcher’s knowledge).

The comprehension conversation is a flow of talk, which assists teachers’ discernment of the purpose of the assessment tool which is to identify students’ instructional reading level or stated in a different way, the materials and teaching points that the teacher will use based on the results of the F&P. Identification of both the hard level and the independent level guide teachers’ decision-making processes (Fountas & Pinnell, 2018). The third edition of the F&P, designed to reduce the subjectivity of the comprehension conversation, according to Fountas and Pinnell, includes more direct questions for teachers to use as a guide in their discernment of students’ comprehension levels.

Currently in the field, there are diverse suggestions for when and how to assess students’ growth in reading. Abrams, McMillan, & Wetzel (2014) postulated that benchmark
assessments and formative assessments have conflicting definitions. Kamhi & Catts (2017) claimed that “the only compelling reason to administer a standardized test of reading comprehension is when these tests are necessary to qualify students for special education services” (p. 2). Researchers are also comparing differences in diagnosing reading comprehension deficits by using different tools. For example, Keenan and Meenan (2012) found only a 43% overlap between tests in diagnosing comprehension difficulties” (p. X).

**The Consequences of Labeling**

Powell (1970) noted that the reading inventory results are fruitful for instruction if the child’s performance level, the teacher’s instructional level, and the difficulty level of the text match. On exploring whether IRIs had become more dependable than in the past, Spector (2005) and more recently Nilsson (2013) examined nine revised IRI manuals and showed that IRI data is often unreliable. Spector’s study revealed no gains in reliability in over two decades. Moreover, Gaskins and Labbo (2007) reported this could lead to false positives, which potentially keep students from the instruction that is best suited to their needs. So rather than relying on it, Spector (2005) and Gaskins and Labbo (2007) reported that educators should consider looking for what does work for the child and implement that instruction.

Klingbeil et al. (2015) reported the F&P resulted in high false-negative rates, and they deemed it insufficient in isolation to make decisions regarding multi-tiered support systems. To concur, high-stakes decisions based on instructional levels increase the risk of harm for misplacement (Ball & Christ, 2012; Spector, 2005).

Another controversy lies within students’ book choices. While choosing book from the library, The American Library Association in their Library Bill of Rights (2014) admonished labeling books with leveling systems which guide readers to specific texts. Students have the freedom to choose any kind of book(s) regardless of readability. However,
within the classroom, students’ identified independent and instructional reading levels guide teachers decision-making processes for book clubs, guided reading groups, and independent reading choices.

**Benchmark Assessments & IRIs**

Controversy exists regarding whether: (a) students’ instructional reading levels are identifiable and transfer to equally-leveled text, (b) classroom teachers have adequate professional development to make leveling decisions, and (c) consequences exist when labeling a student’s instructional reading level (Burns et al., 2015; Delpit, 2012; Hoffman, 2017; Ladson-Billings & Tate, 1995; Nilsson, 2013; Popham, 2011; Powell, 1970). This literature review assumes a level of reader awareness of the history of components of, and similarities between independent reading inventories (IRI) and benchmark assessment systems. Both inventories used to identify students’ instructional reading levels require a skilled, trained administrator to document word-level identification, fluency rates, reader accuracy levels, and text comprehension. Historically presumed, literate individuals have an identifiable independent, instructional, and hard level (Applegate & Applegate, 2008; Betts, 1940; Fountas & Pinnell, 2010; Powell, 1970; Walter, 1974; Pearson, 1979). Interestingly enough these beliefs are being debunked by current research (Nilsson, 2013, Shanahan, 2020). Shanahan reported how Killgallon’s (1942) study which was recently replicated, concluded that “students could often comprehend text well despite evidencing many more oral reading errors than Bett’s criteria prescribed” (p. 18). Nilsson (2013) confirmed that IRIs, QRIs, & CRIs exhibit different thresholds for accuracy measures. Furthermore, Bett’s instructional reading level information was garnered from the work of Killgallon, who reported that students’ comprehension may be below the 75-89% comprehension outcome, an unknown threshold (Shanahan, 2020)
Fountas and Pinnell Benchmark Assessment Systems

Fountas and Pinnell (2011) identified the instructional level to be the point at which students read with 95-97% accuracy with excellent or satisfactory comprehension or 98-100% accuracy with limited comprehension. Fountas and Pinnell report their system can precisely and reliably identify the instructional and independent reading levels of students. Their 26-point-text gradient delineates text characteristics at each point or level. In the text gradient levels, A-Z defined by Fountas and Pinnell, level A text is for emerging readers and level Z is for seventh grade readers and beyond (Fountas & Pinnell, 2012, p. 2). While teachers use the benchmark books to identify each student’s reading levels, booksellers and textbook companies often use the Fountas and Pinnell text gradient system to identify and label their books’ readability levels. Identifying and labeling the difficulty of text level in books has significantly impacted trade book purchasing, professional development manuals, and workshops as assessment is big business (Carnoy, 2018). In contrast, the Library Bill of Rights developed by the American Library Association does not encourage, use, or advocate for leveling systems.

However, using empirical data is aligned with a positivist epistemology. Comprehension is a sociocultural activity that takes place in situated conditions. Factors interfering with a reader’s ability to comprehend text, according to Snow (2002), include self-perception, motivation, cultural recognition, and purpose of the task. She directed teachers to understand that learning and literacy are bound within each student, and therefore, assessing students’ reading comprehension needs to account for these factors.

Reader's Workshop as a Complex System

Reader’s workshop by its very nature is a framework that supports differentiation and instantiates the possibilities provided through heterogeneously grouping students (Calkins, 2015). Calkins (2018) reported how complex reading assessment is — saying, “any effort to
assess reading will inevitably conflate an assessment of reading with an assessment of writing or talking, of understanding and answering questions” (p. 5). To counter these challenges, Calkins recommended having the understanding that all assessment data is fallible and therefore needs to be conducted fluidly so that teachers can provide learners with crystal clear goals. According to Calkins, assessing and teaching within bands of text level difficulty makes the most sense.

In fact, assessing within bands is exactly how the Fountas & Pinnell Benchmark Assessment System texts “were demonstrated to be both reliable and valid measures for assessing students’ reading levels” (Field Study of Reliability and Validity of the F&P, 2009, Heinemann). Resnick and Hampton (2009) called the assessment system complex and comprehensive. They reported the system’s purpose is to capture students’ progress in the fundamental subskills of reading which, consequently, is the purpose of the F&P. One reason it is difficult to identify students’ reading level is as a result of the variability in their content knowledge:

Mastery of content knowledge plays an increasingly larger and complex role in reading comprehension. In other words, it is difficult to predict, given a classroom’s instructional focus and students’ background knowledge, what a student might “typically” know. (Resnick & Hampton, 2009, p. 10)

The consideration that reading comprehension differences between students is causally related to their content knowledge and cognitive maneuvers is changing how reading comprehension is assessed. Another factor which is changing how reading comprehension is assessed is the research regarding how teachers’ conceptions of assessment impact their decision-making processes and the complexities that reading presents (Kim, 2017).
Teachers’ Conceptions Impact Reader’s Workshop Assessment Praxis

Research shows that teachers’ conceptions of assessment impact their openness to emerging or changing protocols. A documented report shows that teachers’ agency development is related to their autonomy, latent conceptions, identity as assessor, context, and critical inquiry of their assessment practices (Cowie et al., 2014; Scarino, 2013; Willis et al., 2013). Although research of teachers’ conceptions of assessment regarding the F&P was not found during the literature review for this study, there is evidence of a relationship between pre- and in-service educator belief systems and their response to assessment education, their decision-making practices, and the complexity of evaluating students’ growth and learning (Carpenter, 2008; DeLuca et al., 2019; Fives et al., 2015; Greene, 2010; Sanchez, 2017; Shepard, 2000; Stiggins, 1995; Xu & Brown, 2016; Zemack et al., 2007). Young and Kim (2010) reported, “Teacher’s beliefs act as a filter through which new ideas, are perceived, interpreted, and executed” (p. 13). Teachers’ conceptions inhibit teachers from learning new things such as new assessment practices, or the adoption of a new curriculum. Thus, an examination of teachers’ conceptions, which is currently a call by researchers, is timely and important work.

Teachers’ Knowledge of Assessment. Teacher education, specifically learning how to include students themselves in the process of assessment, is currently driving changes in education (Xu & Brown, 2016; Greene, 2010; DeLuca et al., 2019a; Zemack et al., 2007; Carpenter, 2008; Sanchez, 2017; Remesal, 2011). The consensus of current research is that “teacher assessment knowledge [is] generally inadequate relative to standards and expectations” (Campbell, 2013; Deluca, 2012; Pearson et al., 2014; Popham, 2009; Remesal, 2011; Stiggins, 2010; Young & Kim, 2010; Xu & Brown, 2016). Teachers use large-scale assessments such as the PSSA [state exam] and in-class assessments to determine differences and educational trajectories among readers (National Research Council, 2002).
For the most part, teachers use observation and interpretation—two of the three assessment pillars—to intuit, infer, and make decisions regarding formative, benchmark, and summative evaluations (National Research Council, 2002). The National Research Council’s (2001) report explained that assessment is, to some degree, imprecise and only partially reflects what a person knows and can do (p. 2). Therefore, drawing the most accurate inferences requires using all three corners of the assessment triangle: cognition, observation, and interpretation. This study takes the position that teachers’ conceptions are considered latent attributes that once attained, become part of teachers’ working cognition.

**Teacher Beliefs.** Teacher assessment literacy is dependent on a combination of cognitive traits, affective and belief systems, and socio-cultural and institutional influences, all of which are central to teacher education. They pinpoint the complexity of and entail a need for reconceptualizing teacher with insights from prior studies and new perspectives from teacher education. One component of belief systems is self-efficacy. Bandura (2006) proclaimed, “People choose which challenges to undertake, how much effort to invest in the pursuits and how long to persevere in the face of difficulties” (p. 28). He described the interactions of environment, biology, cognition personal agency, motivation, self-efficacy, and other affective behaviors and beliefs as they can affect human functioning. Peoples’ beliefs in their ability to affect and have control over events in their lives is a critical component of agency and a factor of belief systems (Fulmer et al., 2015).

**Teacher Professional Development**

Furthermore, Datnow and Park (2014) and Mandinach (2012) warned that teachers’ education does not reflect the demands placed upon them in schools. As a result of insufficient professional development, teachers often do not recognize the factors that alter their decision-making processes, thus reducing the reading experience's authenticity and lowering the reliability and validity of the assessment results (Sanchez, 2017; Turcios-Cotto
& Milan, 2013). Teachers can improve these skills by reflecting upon and analyzing their cognitive footprints (Hoffman-Kipp et al., 2003; Remesal, 2011), leading to deeper assessment literacy. Remesal’s (2011) study determined that “there were belief systems about assessment, related with the monitoring of both teaching and learning processes. On the other hand, there were belief systems about the certification of learning and the teachers’ accountability” (p. 475). The teachers held beliefs that assessment can lead to positive changes or they saw it as a negative activity “not leading to change or as a disruptive measure” (Remesal, 2011, p. 475). A documented report shows that teachers’ agency development is related to their autonomy, identity as assessor, and critical inquiry of their assessment practices (Cowie et al., 2014; Scarino, 2013; Willis et al., 2013). Furthermore, teachers find it helpful to be involved in learning communities where they have a common language to share, negotiate, and make decisions about their assessment practices (Lukin et al., 2004; Wyatt-Smith et al., 2010). Other researchers also found that educating teachers to deconstruct traditional beliefs and practices improves their assessment literacy (Chen et al., 2010; Li, 2017; Popham, 2011; Xu & Brown, 2016; Yamtim & Wongwanich, 2014). Despite agreement, researchers have only just begun to explore teachers’ acquisition of this reflective process.

### A Systems Approach to High Literacy Outcomes for All Children

Much like Carnoy (1975), Fountas and Pinnell (2018) recognized that teachers experiences are value-added and offer a vital perspective of the system in schools. They described the different types of reading that occur throughout a typical day, and the only time that students are engaged with text at their instructional level, per their recommendation, is during guided reading. Within the systems approach, Fountas and Pinnell reported that “By selecting and reading books of their own choosing, students build mileage and stamina as they freely exercise full control of the reading process” (p. 16). Additionally, Fountas and
Pinnell reported that talking proficiency is a vital component within the system of students’ continuum of literacy learning. As reading comprehension is different from speaking, clarity of the purpose of assessment tools within a system is vital (p.14). Specifically, Fountas and Pinnell (2018) reported:

Many great ideas flounder and fail because of weak implementation. We need to get away from terms such as school reform and to stop expecting that the adoption of a new set of standards or the purchase of a new core program will fix everything overnight. It is not a process of fixing; it is a process of studying, improving, and becoming a learning organization. The work of Bryk et al. (2015) on learning to improve has presented a learning process in which schools get better at getting better. The school becomes a self-extending system because of how the system operates. As expertise increases and success occurs, learning accelerates. (p. 18-19)

Similar to Carnoy’s work, Fountas and Pinnell (2018) urged districts to value what teachers bring to the table; reporting that if the teacher is valued as the single most important factor related to student outcomes and the system in the school fosters collaborative conversation, communication, and interdependence, “We all grow a culture of reflective practice which “fosters teacher agency and encourages acts of leadership by individuals and promotes the kind of generous teamwork that benefits everyone” (p. 16).

The Role of Background Knowledge

Prior knowledge, often called background knowledge, affords readers an advantage while reading (Garcia, 2002; Garcia & Traugh, 2002; Hattan, 2019; Johnston, 1984; Pearson et al., 1979). Readers’ (helpful) background knowledge, such as topic knowledge, domain knowledge, personal knowledge, or world knowledge, plays a significant role in aiding comprehension (Hattan, 2019). O’Reilly, Wang, & Sabatini (2019) studied 3,534 high school students who took a background knowledge test before working on a reading comprehension
test. Results indicated that a determined knowledge threshold line is indicative of comprehension. Results showed 87% of students in the below-threshold group (n=835) scored below the grand-mean comprehension score of 15, which was equivalent to the grand-mean comprehension score across all students in this sample. On the other hand, 91% of students (n=1,356) whose comprehension score was above the mean also scored above the knowledge threshold. Understanding the role of background knowledge before, during, and after assessment administration may increase testing fairness. The F&P instructs evaluators to follow a standardized test protocol. One of the first tasks of that protocol is for evaluators to read the introductory sentence to the examinee, which may for most students elicit helpful background knowledge. Hattan (2019) noted that inaccurate background knowledge decreases comprehension for some students, “Students tend to rely on their previous understandings rather than integrating the new information with their old knowledge” (p. 452).

In terms of identifying a student’s instructional reading level, another decision that teachers make is evaluating how their students’ prior knowledge influences benchmark assessment results. The Fountas and Pinnell benchmarking books Saving Up by Kitty Colton, City Hawks by Maryann Dobeck, Snake Myths by Katherine Herenger, Animal Instinct by D.M. Longo and Could Be Worse by Sharon Fear contain experiences that may be unfamiliar to students [researcher’s observation]. Yet, exhibited misunderstanding of text is not definitively related to vocabulary acquisition or background knowledge. Researchers found evidence to show that a student’s “communicative competence” should replace “language proficiency” (Duran, 1989). Researchers measuring students’ communicative competence with peers and teachers revealed that students’ communicative competence “cannot be understood as a characteristic of the individual student per se. It may be affected by the communicative strategies used by other students and the teacher and differences in the social
relations among these interlocutors” (Duran, 1989, p. 155). Thus, when teachers assess students’ instructional reading levels, they should consider that students’ abilities to communicate their understanding of text is at times limited by their communicative competence (Brown & Hirschfeld, 2008).

Culturally Relevant Text

In assigning meaning to text, assisted by what readers already know, background knowledge bridges the space between the text and the student. Thus, research has well documented that assessments are inherently geared toward success for White students and children from the middle class. Herman et al. (2010) used the term ‘fair’ in saying “benchmark assessments are ‘fair’ if no advantage exists between different subgroups with the same level of knowledge and skills” (p. 6). Without determining the background knowledge that a student has prior to a teacher administering a benchmark assessment, it is rather difficult to determine the fairness of the assessment. Moreover, some aspects of IRIs have been contested in the literature to include teacher’s awareness of students’ accents on word pronunciations, exposure to content, and the unreliability and transferability of IRI data (Calkins, 2018, Fountas & Pinnell 2020, Nilsson, 2013). Relatedly, Gandy (2013) suggested that teachers should use extra caution when assessing students for whom English is not their native language.

Fairness

Tierney’s (2013) work brings forth the researchable concept of fairness in assessment. He concluded that transparency in classroom assessment, differences in student opportunities, and teachers’ perceptions regarding equal or equitable concepts surfaced to the top of issues that require more research. Specifically, Tierney noted that teachers need to actively engage in critical reflection regarding the purposes of assessment and how their biases affect their assessment praxis. The next study in the literature review shows how a classroom teachers’
praxis determined which books her students could read. Kontovourki’s (2012) study is useful in considering the concept of fairness in the classroom as it relates to assessments, teachers’ decision-making, and students’ opportunities.

Kontovourki (2012) conducted a case study to examine how reading leveled books and assessing students’ reading levels shaped and defined their lived experiences. Kontovourki reported some students showed diminished enthusiasm for reading because they were guided to read books within their identified reading level. The reason for students’ diminished enthusiasm for reading was causally related to how the teacher used the benchmark reading results to define students’ book-choice parameters. According to Kontovourki (2012), the test results “defined what students could do and whom they were allowed to be as readers” (p. 154). Recognizing the predicament, they were now in, the students began to negotiate with the teacher. Specifically, they negotiated to have access to the books that the teacher said were off limits.

**Socioeconomic Status**

The nation’s K-12 schools enroll 25 million students from low-socioeconomic households—nearly half of all public-school students (Jensen, 2015). Teachers’ awareness of their students’ needs includes “the effects of contextual factors” such as (a) economic resources, (b) class membership, (c) ethnicity, (d) neighborhood, and (e) school culture (Snow, 2002, p. 7). Assessment-literate teachers understand the complexities of applying strategies and structures that mitigate the negative effects of poverty on student learning. Assessment-literate teachers are aware that factorial characteristics of students from impoverished homes are revealed in “students’ oral language practices, in students’ self-concepts, in the types of literacy activities in which individuals engage, in instructional history, and in the likelihood of successful outcomes” (Snow, 2002, p. 7). Assessment literate teachers recognize how dialectical thinking due to lived experiences may affect the results of
conversation or test score (Black & Wiliam, 2017; Hamamura et al., 2008) In the context of this study, two of the four elementary schools typically qualify for Title 1 funding.

**Transferability of Benchmark Data**

Burns and colleagues (2015) noted the difficulty in designing an assessment system to accurately determine if the instructional level found through the administration of a benchmark reading inventory is transferable to original books, because existing knowledge plays a significant role in reading successfully. Burns and colleagues (2015) questioned how well the F&P identifies students’ instructional reading levels, because their findings suggested an unequal match of identified instructional reading levels with equally-leveled text. Goldberg (2019) found the F&P kit not worth the time it takes to administer, the inconsistencies in Lexile levels as readers move through the kit, and Goldberg found fault with the fact that the F&P kit was field tested on grade level readers. Fountas and Pinnell did not use Lexile parameters to design the books in the F&P; therefore, according to Goldberg (2019), they cannot place students in like-level trade books. Goldberg’s article, to me, is a perfect example of why school districts need to deeply analyze assessment kits, guided reading programs, and other published materials for use in schools.

Fountas and Pinnell (2010) adhered to their initial goal for creating the F&P, which provides teachers a mechanism to observe students’ reading behaviors while reading increasingly difficult text so that instruction is more targeted for those students. According to Fountas and Pinnell, “A gradient is not and was never intended to be a way to categorize or label students, whose background experiences and rate of progress will vary widely. We have never written about leveling students” (Goldberg, 2019). They posited that having students choose books that may be hard to read at first, but are of interest, stretches the student for a time. Common Core Standards are in full implementation, so stretching students to read complex text is warranted. The standards are rigorous in their goals, and students who are
reading books below their reach are being held back from experiencing the thrill of the challenge (Fountas and Pinnell, 2010; Shanahan, 2020).

**Proponents of Benchmark Assessment for Improved Learning**

Snow (2002) sought to help teachers understand the “reading comprehension phenomenon” by better understanding how to (a) more precisely identify reader capabilities and limitations and (b) discern the impact of different text features on readers with varying capabilities (p. xv). To measure students’ growth in comprehension, he recommended using research-based benchmarks. In support of teachers’ evaluations based on assessment results, Snow (2002) noted that without such benchmarks, “we as a society risk aiming far too low in our expectations for student learning” (p. xix). Herman et al. (2010) were also proponents of benchmark assessment for improved learning interpreted benchmark assessments as *an in-between* for classroom assessment and annual assessments noting, “Benchmark data flows into the annual assessment, which in turn transfers into subsequent years of teaching, learning, and assessment” (p. 2). In turn, Herman et al. (2010) suggested that local policymakers ask these three questions prior to implementation:

1. What purposes do you expect benchmark assessments to serve?
2. What criteria should you use to select or create benchmark assessments?
3. What organizational capacity is needed to successfully support a benchmark assessment program? (p. 2)

Herman et al. (2010) also noted four related, but different, purposes for benchmark assessments. They are (a) communicate expectations for learning, (b) plan curriculum and instruction, (c) monitor and evaluate instructional and program effectiveness, and (d) predict future performance.
Teachers’ Conceptions of Assessment

As early as 1974, educators realized the need to examine the impact of teachers’ conceptions in the classroom on student performance. Other more familiar terms for teachers’ conceptions include but are not limited to untested assumptions, perceptions, and reasonable beliefs (Brown, 2002). They were charged with planning and dispersing the large, newly allocated federal funds for education research (1965 Education Act). They planned and commissioned education researchers to use Herbert Simon’s theoretical and methodological frameworks to examine how teachers’ conceptions impact the classroom environment, decision-making, and student achievement. Just as researchers and educators today choose how and what to research based on policy goals and funding, the Education Act of 1965 historically changed education’s trajectory.

Teachers’ conceptions and their impact continue to be an area of interest for researchers, as well as if the movement in education moves away from standardized tests. Hence, a more teacher- and student-controlled evaluation model will be vital (researcher’s opinion). According to Brown (2002), who developed the Teachers’ Conceptions of Assessment (TCoA) survey tool, these conceptions, if realized, may increase student achievement in response to teachers’ awareness of the impact that their conceptions have in planning, teaching, and assessing students. The focus on teacher education and efficacy to make important decisions about student progress and to include the students themselves in the process of assessment is driving changes in education (Carpenter, 2008; DeLuca et al., 2019a; Greene, 2010; Remesal, 2011; Sanchez, 2017, & Tillema; 2011; Xu & Brown, 2011; Zemack et al., 2007).

Preexisting Beliefs

A later study from Brown and Harris (2009) implored stakeholders to realize that teachers’ conceptions are a strong factor in their decision-making processes regarding
acceptance or rejection of change: change of policy, change of practice, and change of any kind. They reported, “These data suggest that before the reform, teachers must first be persuaded of its utility and allowed to ‘buy-in’ to the program; without it, reforms or resources are unlikely to maximally achieve their intended goal” (Harris, 2009, p. 87). The identification of pre-existing beliefs is vital for a school to be its best. Brown and Harris’s 2009 study supports the need for the study I am proposing. If teachers’ beliefs affect their decision-making processes during the F&P administration, transparency regarding those conceptions will illuminate the results to include the teacher (Brown, 2016; Campbell, 2013; Remesal, 2013).

**Theoretical Framework**

The complexity of education theory suggests that it is the tangential interactions among elements or agents, essential or not, “that are responsible for the phenomena, the patterns, properties, and behaviors that characterize or constitute a particular environment or system” (Mason, 2008, p. 38-39).

The complexity of education theory provides a firm framework for my study (see Figure 1). Historically, education research in assessment follows an empirical protocol and mirrors the standardized procedures and product of said assessments; however, there is little to no research regarding the standardized F&P. Stemming from physics and mathematics in the 1950s and 1960s, complexity theory, according to Davis (2006), is becoming the umbrella for other disciplines, namely studies of anthropology and sociology.
**Figure 1**

*Theoretical Framework*

*Note:* The dark rectangle represents the Namaste School District, the boundaries of the study. The spheres inside the rectangle, contain the complexity of each system. The connecting links represent the interactions that occur within and between the complex systems. The complexity theory is in the center as the best place to capture the simultaneity of the phenomena under study.

Assessment and research in education historically mimic how scientists and medical practitioners gather data and use the results to make sense of the phenomenon being studied or prescribe a cure for an ailment. Wolf and colleagues (2017) called for a reconceptualization of assessment, noting, “Our conceptions of learning and what is worth being learned evolve at the high speed of culture rather than the gradual speed of biology” (p. 67). Thus, in my opinion, transferring the current F&P practices to dynamic assessment practices provides the one-on-one time that teachers find valuable, and students’, having the opportunity of guidance from their teacher, may have the chance to learn about themselves as readers and practice with the teacher, in situ (Afferbach, 2016). In support of the aforementioned suggestion, Calkins (2015) promoted the idea that teachers acting as
researchers, more frequently than two or three times a year, are situated to observe and teach what needs to be taught when it needs to be taught (Calkins, 2015).

**Complexity of Education Theory**

According to Haggis (2008), the complexity of education theory examines phenomena from an inside perspective to view the mini-systems’ iterative movements while also examining the complete system. This theory stems from Ludwig von Bertalanffy’s (1937) work combined with the law-building work of Lorenz’s (1963) “butterfly effect” theory. Lorenz (1963) proclaimed that when a butterfly flutters its wings, it moves molecules, which in turn move more molecules, eventually changing the weather on the far side of the world (p. #). Ludwig von Bertalanffy, a biologist, is considered the initiator of general systems theory (Vornberg, 2013).

Whereas systems thinking examines contexts and relationships by pulling apart the whole to see how each element works independently and then together, complexity thinking seeks to observe the wholeness of the system and regards each element as a “mini-system” (Conchran-Smith, 2014; Haggis, 2008; Hilpert & Marchand, 2018; Morrison, 2006). Complexity thinking respects and recognizes each element's complexity, person, aspect, and condition as mini-systems that iteratively change from within and subsequently interact and change the other mini-systems in the larger system.

In the context of this dissertation, the complexity theory frames the study and proposes that relationships occur within nested contexts. For example, interactions among students, teachers, schools, districts, policies, practices, and tools. These interactions are involved, which is why this theory works well as the study’s framework. This theory does not trap itself in absolutes; it compels the researcher to consider how their involvement implicates the phenomena they study. (Davis & Sumara, 2006, p. 16).
The complexity of education theory supports the ability to look at assessment from within a system to see the moving parts and their interactions. According to Cochran-Smith et al. (2014), this theory has been critiqued and challenged for the following reasons:

(a) a proclivity for retrospective description
(b) its rejection of linear causality
(c) does not provide causal explanations with implications for practice
(d) does not deal with the values and power inequalities inherent in the normative enterprise of education. (p. 2)

To mitigate challenges to the theory, Hilpert & Marchand (2018) recommended that researchers delineate the study’s context, use multiple analysis levels, and recognize and incorporate change over time. Incorporating their recommendations increases the chances that the research design, and subsequently the study’s product, will encapsulate the complexity theory's abilities. Otherwise, design and product become a linear representation of the linear phenomenon.

Summary

This chapter reviewed the literature and the complexities for the following topics: the reader’s workshop model, reading assessments [IRIs, F&P, formative, dynamic], teachers’ assessment literacy, the contextual components of the setting, and the theoretical framework that was used to guide this study. I reported on the literature regarding the impact of assessment results on students’ learning trajectories as well as the benefits and challenges of IRI use. The next section of this paper describes the methods that were used to examine teachers’ conceptions of the F&P and how those conceptions impacted their decision-making processes before, during, and after administration.
Chapter III: Methodology

Assessment is a complicated endeavor, and researchers are moving towards using different theoretical frameworks and methods to gather data (Brown, 2007; Carpenter, 2008; Creswell, 2018; Darling-Hammond et al., 2005; Flores et al., 1991; Garcia, 2002; Greene, 2010; Harraway, 1998; Ortiz & Flanagan, 2013; Remesal, 2011; Robinson-Zanartu, 2013; Sanchez, 2017; Schwartz, 2002; Xu & Brown, 2016; Zemack et al., 2007). When researchers are compelled to derive an in-depth understanding of the phenomena or case to be studied, Yin (2018) explained that a case study design is fitting. Thus, “among other features, case study research assumes that examining the context and other complex conditions related to the case(s) being studied are integral to understanding the case(s)” (Yin, 2018, p. 2).

Therefore, I have designed this study to use multiple methods of data collection. Moreover, Yin (2018) explained that inquiry questions that begin with the words ‘what’ or ‘how’ are the types of questions that best match case study design (p. 5). As a new researcher, Yin’s structured framework which promotes the researcher to use construct validity, internal validity, external validity, and reliability provided a firm framework from which to work (Yazan, 2015).

Researchers use mixed methods when one technique is not enough to capture the data they seek. Mixed-method designs allow the researcher flexibility to use two or more methods such as quantitative and qualitative data collection techniques. One reason to use mixed methods is to show how the results from one data collection source validate the results from the other source of data (Creswell & Plano Clark, 2011, 2018; Guetterman et al., 2015; Maxwell, 2012). Another reason to use mixed methods is if the context of the research typically relies on or considers one method better. Brown & Harris, (2011) reported that
mixed methods have the ability to provide a mechanism for collected data to not only show complementary results but that it also has the ability to allow distinct results to emerge. Similarly, Yin (2018) proposed the benefits of studying a case using mixed methods is a choice that some researchers utilize to gather the data they seek to answer their research questions.

The case of K-5 public school teachers’ conceptions of assessment was guided by the complexity of education theory (Davis & Sumara, 2008), and examined data collected in a sequential explanatory mixed-methods design with an emphasis on the qualitative data. This design provided, in my opinion, the best devices for me to capture data (Bazely, 2016; Ivankova & Stick, 2007; Johnson & Turner, 2003; Moseholm & Fetters, 2017). While using an explanatory design, Creswell recommends a larger quantitative sample and a smaller qualitative sample be used. He posits that this helps the researcher "obtain a rigorous, in-depth qualitative exploration and a rigorous high-power quantitative examination of the topic" (Creswell, 2018, p. 188). While Yin (2018) stated that studying a case is not limited to style or type of methods, only that the data collected is relevant. He encouraged researchers to use a collection of methods to triangulate evidence on the same research questions, resulting in less criticism of the data collected.

I used the case-selection variant design to understand how the purposely chosen interview participants operationalized their conceptions of the F&P. Creswell and Plano Clark (2018) defined the case-selection variant design as an appropriate choice for researchers who emphasize the qualitative data. The quantitative data, in this design, was used to select participants for Phase II. In Phase I collected quantitative data using Brown’s Conception of Assessment survey tool. In Phase II, I used semi-structured interviews to enrich and expand the quantitative data.
The study of the effects of teachers’ conceptions in the classroom and specifically in assessment is not a new line of research but reports show that although deeply studied, teachers’ professional development is not commensurate with researchers’ suggestions. In their 2016 study, Xu and Brown reported that teachers revealed significant professional assessment development needs which is deeply supported in the literature. Xu and Brown (2016) found in their review of literature that most studies were qualitative, the one proposed here will incorporate quantitative information along with the qualitative data.

**Overview**

The purpose of this study is to examine how K-5 public school teachers operationalize the administration of the Fountas and Pinnell Benchmark Assessment System (F & P [BAS]) and how their conceptions of assessment impacted the administration and subsequent results of said inventory. Developing an understanding of local practice adds to the body of education research and builds a foundation for improving assessment techniques at the local level. As reflective practitioners, teachers examine their praxis through useful techniques such as discourse with colleagues, journaling about their thoughts on daily experiences with students, and participating in thoughtful research. The main question for this study was, *What are teachers’ conceptions of the F&P?* The sub-questions that guided this study were:

1. In what ways do teachers operationalize the administration of benchmark assessments?
2. How do teachers' conceptions of assessment inform their administration of benchmark reading assessments?

**Research Design**

According to Brown (2019), “Research originating in New Zealand with the TCoA self-report inventory has been replicated in multiple locations and languages (i.e., English in New Zealand, Hong Kong, and India; Greek in Cyprus; Arabic in Egypt; Spanish in Spain,
Ecuador) and at different levels of instruction” (p. 1). Specifically, data showed that teachers’ belief and value systems play an integral role in their decision-making process. This mixed methods study was designed to uplift the voices and experiences of elementary school teachers in the mid-Atlantic region of the U.S. This study takes an emic look at the phenomenon of teachers’ conceptions of the F&P and how those conceptions are operationalized during the administration of said tool. Results will enhance and deepen our understandings of the implications of teachers’ AL and conceptions of reading evaluations.

This study used a mixed-method explanatory sequential approach building upon the work of DeLuca et al. (2019) and Brown (2004, 2006, 2009, 2011, 2018). Mixed-methods studies allow researchers the forum to integrate and consolidate data (Creswell, 2018; Vogl, 2019). Teachers’ stories need to be heard. Creswell (2018) stated that an interview is a well-used technique which is a method that best captures the experiences of participants.

Specifically, this study employed non-equivalent groups in two phases to investigate how teachers’ mindsets are related to their approach to assessment. In Phase I, 22 participants completed the Teacher’s Conceptions of Assessment (TCoA) survey tool (Brown, 2006), followed by open-ended questions. Phase II included semi-structured interviews of six cases. Qualitative methods give voice to populations whose experiences may not be well-represented in the research literature, such as teachers (Creswell, 2018). DeLuca (2019) noted that most of the research about teachers’ approaches to assessment has been conducted with preservice, undergraduate students.

**Participants**

The target population for this study was K-5 elementary school teachers who had worked a full year within the chosen school district and have previously administered the F&P. The elementary teacher population for the Namaste School District was roughly 135 teachers, at the time of the study, which included special education teachers. Twenty-two
teachers, or 17% of the total eligible teachers completed the TCoA-III survey tool in November 2020. Based on the survey data, six participants were invited to participate in Phase II, semi-structured interviews which were conducted in December 2020. These specific participants were chosen to participate in the interviews because they, as a group, held a wide range of beliefs about the F&P assessment.

**Protection of Human Subjects**

Data collection preserved the privacy and identity of participants. Participants’ names were redacted and replaced with pseudonyms. All related research documents were housed at the researcher’s home or on the researcher’s password protected laptop and will remain there for three years from the completion of this study. Zoom interviews were encrypted and housed in the university’s iCloud account.

**Description of the Setting**

The setting of this study was purposeful and convenient. As my place of employment, this mid-Atlantic public-school district was chosen for their use of reader’s workshop and the F&P reading assessment. This district, located outside of a major city in Pennsylvania, is on the smaller side with four elementary schools, one middle and one high school. Three of the elementary school houses approximately 600 students or less and one elementary school houses upwards of 300 students. Teams of teachers typically have between 3-5 classrooms.

The population of the community is 30,000 people in roughly 40 square miles. Sixty-one percent of the population is between 18 and 64 years of age with 80% of the population White, 5% Black, 10% Asian, and 5% Latino. The median household income is $115,000/year with approximately 5% of children and seniors at the poverty level. This school district has a high school graduation rate of 96%, with most constituents holding a bachelor’s degree and roughly 80% are homeowners (censusreporter.org).
**Instrumentation**

This study utilized 2 instruments to collect data, a survey, and a semi-structured interview protocol.

**Survey**

With permission, I adapted Brown’s (2002) *Teachers’ Conceptions of Assessment Survey III-Abridged version* (TCoA-III) instrument for use in this study. I adapted this tool to focus less on assessment as a whole and more specifically on the F&P benchmark assessment (see Appendix D). Found to be reliable in other settings, the TCoA-III survey took approximately 15 minutes for participants to complete. The TCOA-III survey tool collected quantitative data for this study using 27 positively-packed, Likert-style questions that addressed how the F&P could be used to address the following constructs: (a) F&P for Improvement, (b) F&P for School Accountability, (c) F&P for Student Accountability, and (d) F&P for Irrelevance. In addition, two qualitative, open-ended questions were added at the end of the survey that asked participants to identify the benefits and challenges of the F&P as well as to share how their knowledge of their students’ background knowledge impacts their operationalization of the administration of the F&P.

**Semi-Structured Interview Protocol**

A semi-structured interview protocol allowed me to stay focused during interviews. The interview protocol was designed to create a safe, open environment to ensure participants' comfort. Opening interview questions asked about how their day went. We exchanged small chit chat. From there, I asked a broad question: *How do you prepare to administer the F&P?* Working within the setting, in this instance, as a reading specialist and now as a researcher, seemingly provided comfort and participants seemed eager to begin, curious, and excited to be part of education research. Finally, I explained how I was
operationalizing the word *conceptions* in this study and we moved forward into the discussion (see Appendix E).

At the end of the interview process, I asked participants if they had anything to add, and I informed them of my next steps. My next steps included transcribing the interview, journaling about the interview experience, and taking the data back to the participants to have them verify or member check the data. I also informed participants that their information would be kept secured in a password locked computer, and if they feel any sense of discomfort which was brought on by the interview, they could call Dr. Heather Schugar, my advisor, for guidance.

**Internal Validity**

As Creswell (2018) noted, it is an internal threat to data analysis to not explain surprising or contradictory quantitative results with qualitative data. Having an unequal quantitative and qualitative sample size had the potential to pose problems. In addition, keeping results from the different databases separate and failing to resolve disconfirming results were also threats to validity (Creswell, 2018, p. 251). To increase internal validity, I explained contradictory results which at times revealed complementary viewpoints instead (Green, 2006). Using a graphic organizer to show both sets of data at complementary or disconfirming results increased and maintained internal validity. Creswell and Plano Clark (2018) recommended asking questions in the qualitative phase to clear up contradictory quantitative information. In addition, member checks of qualitative data were conducted so that participants could validate their interview data.

**External Validity/Generalizability**

The data gathered in this study was limited to the perspectives of the participants in this mid-Atlantic public-school district during the time of the study. Although researchers aspire to have their data generalizable to other settings (Fraenkel et al., 2012), I understand
that this data is bounded in the setting and can be used by the teachers, the Teaching & Learning Department, and other stakeholders to develop professional programs, alter policies, and increase discourse about the F&P particular to how the Namaste School District utilizes this assessment.

**Reliability**

To increase the reliability of the quantitative survey results, I embedded a clause in the directions reminding participants to keep the F&P in mind for each and every question. The 27 Likert style questions were also altered to start with the statement, “The F&P assessment.” To increase the reliability of the qualitative data, I used member-checking which, according to Fraenkel et al. (2012), increases the validity of the data thereby increasing its reliability. Member checking ensured that participant interview data and qualitative written responses to questions were accurate and revealed the intentions that the participant had set forth. Triangulation is another way to increase validity and reliability; it is a process to show how participants experience the administration of the F&P BAS through the quantitative, qualitative written responses, and the semi-structured interviews.

**Researcher’s Bias**

My position in this setting was emic. The setting was my place of employment, and the participants are my colleagues. I do not hold a supervisory role here and upon my committee’s request, semi-structured interviews were conducted with participants from the other three elementary schools. Participants for Phase II were chosen with names removed from the data set. Two initially chosen participants were replaced because they worked in the same building as I do. By announcing my position, readers of this study can form their own opinions about the results.
Procedures

The procedure timeline began in June 2020, with satisfactory completion of comprehension exams, followed by study approval from the West Chester University IRB. Defense of the study’s proposal was conducted on November 12th. Due to the pandemic, several changes were made to the original design study and timelines were delayed. The study’s original data collection start date was late September early October, but again, as a result of the pandemic, was delayed until November 2020. Quantitative data analysis took three weeks and then semi-structured interviews were conducted in December. Data analysis continued and the results were recorded and defended on April 28, 2021.

Quantitative Data

According to Brown (2002), the TCoA-III makes visible teachers’ latent conceptions of assessment whereby enhancing the development of teacher education programs, the design of assessment policy, and guides further research into educational assessment practices. His self-administered, abridged version aggregates teacher beliefs about assessment around four significant ideas. Brown's (2006) final design of the TCoA survey tool summarized assessment as a mechanism to improve teaching and learning, as a way to evaluate the success of and hold schools and students accountable, and as an irrelevant component of education. His design, of 27 positively packed Likert scale items, contain three item scales that load onto nine first order factors. As this is a multidimensional survey tool, results do not surmise a single overall score.

Qualitative Data

For the qualitative portion of the study, six participants were chosen (identities redacted) however two of the subjects were excluded and replaced because they work in the same building as I did. Creswell (2012) argued that qualitative data allows the researcher to learn the viewpoints of the participants, develop theories, and assess a process over time (p.
Thus, I conducted six semi-structured interviews with participants who exhibited targeted scores in the quantitative phase of the study and did not meet the exclusion criteria above. To collect qualitative data, semi-structured interviews were conducted in person, via Zoom, and lasted approximately 60 minutes. Interviewees decided which location and time was best for them. The COVID-19 pandemic changed the way researchers conducted research this year, so an in-person interview, which was my preferred method, was not possible per IRB guidelines.

Richards (2009) offered five standards to consider using while collecting qualitative data: accurate, contextual, thick description, useful, and reflexive. Creswell (2018) concurred with Richard; what you observe and collect needs to be recorded with as much detail as possible. Following data collection, to capture as many details as possible, I journaled/memo in my researcher’s logbook. Researchers need to consider how they are part of the study and consider the situations that they have created. To ensure accuracy, I reviewed transcriptions to identify inaccuracies and misinterpretations. Then I used member checking, a technique that researchers use to verify their interpretation of collected information (Kornbluh, 2015). In this case, member checking was conducted by sharing transcribed interviews with participants so they could 'check' and verify the data.

**Analysis and Coding Procedures**

Quantitative data was analyzed using parametric and non-parametric methods. Currently, there are tensions in the field regarding how small sample size data, surveyed with a Likert-style tool, can be analyzed. For example, Barry (2017) claimed, “The use of averages cannot account for the importance of capturing and understanding variability” (Chapter 1, n.p.). He said, “A good way to remember not to use means for Likert scale data is to think the average of Agree and Strongly Agree is not Agree and a Half” (Chapter 1, n.p.). Thus, nonparametric tests such as the Mann Whitney Wilcoxon is better suited for analyzing Likert-
style scales. However nonparametric tests have less power, although some “nonparametric tests can manage ordinal data, ranked data, and not be seriously affected by outliers” (Minitab, 2015, n.p.). Therefore, after consultation with my advisor and Dr. Brown [creator of the TCoA], I used item-scale analysis to ensure that each of the three item-scales of which there are nine, met the minimum reliability coefficient markers, which they did. Once confirmed that each of the three item-scales met those requirements, I analyzed the mean, mode, and median for each question separately, and as a three item-scale. I reported both the mean and the median and indicated where the two diverged. Quantitative data was analyzed using descriptive statistics (e.g., item-specific means), Data analysis was conducted to preserve the privacy and identity of participants, and participants' names were replaced with pseudonyms to maintain privacy.

Qualitative data was uploaded and sorted in Dedoose. The four constructs of the TCoA survey tool were the foundation for the initial coding of interview data: (a) improvement (b) school accountability (c) student accountability, and (d) irrelevance. I checked for consistency of categorization three times while processing the data. Data processing took approximately four weeks. Once the qualitative data was coded for the four constructs of the teachers’ conceptions of assessment survey tool, an iterative process ensued (Yin, 2018) where participants words were analyzed for the recognition of emergent themes. Several emergent themes were identified and validated by my committee chair for whom I shared the qualitative data with during a guided work session. Emergent themes consisted of: time, use of the F&P, one on one with students, and test equity. I journaled to capture details that I may have otherwise forgotten; I listened to the interviews one month and then two months after data collection; I also reread the transcripts. This process allowed me to increase my consistency of understanding participants interview data as researchers need to consider
how they are part of the study and reflect upon the situations that they have created which is what I was able to do through the reflective journaling process (Richards, 2009).

Data Integration

Guetterman, Fetter, and Creswell (2015) recommended the following best practices when integrating quantitative and qualitative data sets: “(a) label quantitative and qualitative results, (b) be consistent with the design, (c) be consistent with the integration approach, and (d) identify inferences or insights generated” (p. 560). According to Creswell (2018), the integration phase between the quantitative and qualitative data is the reason for using mixed methods. When personal experiences help to explain statistical results, value is added to the research. I looked for common concepts and determined how the qualitative data confirmed or disconfirmed teachers’ conceptions of the F&P as revealed through the quantitative survey. Data was reported using words and visual display/tables which integrate collected results in a joint display. The intent is to match the results by "comparing them and illustrating them with a side-by-side joint display to identify points of convergence or divergence" (Creswell, 2018, p. 298). I made inferences based on the integrated data which is explained in the discussions section of the report. I interpreted results and provided, in Chapter V, implications and recommendations for future studies.

Summary

In this chapter I explained the methodology for the study, the research design, and procedures, followed by participant selection. I described the instrumentation for the study and specifically how the quantitative and qualitative data collection proceeded. Coding technique and software utilization were also outlined within this chapter. I explained how the work of Yin (2018) and Creswell & Plano Clark (2018) along with the complexity of education theory (Davis & Sumara, 2008) influenced the research design and provided structures to increase the validity, reliability, generalizability, and the limitations of the study.
Chapter IV will explain the results of the study followed by Chapter V which will include my conclusions and recommendations for future research.
Chapter IV: Results

This chapter examines the case study data collected in November and December of 2020 from the participants in the Namaste School District. Teachers’ collective conceptions of the Fountas and Pinnell Benchmark Assessment Systems 1-3 (F&P), revealed complexities of assessment that are ripe for revisiting. To capture the complexities of this phenomenon, I used an explanatory mixed methods approach to answer the following two research sub-questions: (1) In what ways do teachers operationalize the administration of *The Fountas and Pinnell Benchmark Reading Systems 1-3*? and (2) How do teachers’ conceptions of assessment inform their administration of benchmark reading assessments? The explanatory mixed methods design combined survey and case study designs in order to demonstrate what and why surrounding teachers’ conceptions of the F&P assessment.

Teachers’ conceptions, which are formed through lived experiences, education, belief systems, and context (e.g., place of employment, local policies, grade level, and team) have an impact on their decision-making processes (Brown et al., 2019; Brown & Remesal, 2012; Davis & Sumara, 2006; Dewey, 1933; Freire, 1972; Hoffman-Kipp et al., 2003; Ladson-Billings, 1999; Piaget, 1976; Pina, 2006; Van Manen, 1977; Vygotsky, 1978). To ascertain teachers’ conceptions of the F&P, I utilized the Teachers’ Conceptions of Assessment-III abridged version (Brown, 2004, 2008) with the 22 K-5 elementary school participants. A positively-packed polytomous Likert scale collected teachers’ reactions to statements, such as: *the F&P improves teaching and learning, the F&P holds students and schools accountable, and the F&P is a valid and relevant assessment.* Brown’s (2008) TCoA-III survey constructs were designed to measure teachers’ latent conceptions. Cronbach’s and McDonald’s reliability statistics for this scale were acceptable (Scale $\alpha = 0.910$ and $\omega = 0.918$). Each of the four constructs in the TCoA-III have nested items
that aggregate to bring forth a visible representation of teachers’ latent conceptions of assessment. Survey items are set to nest and load according to Brown’s original design but have been known to load differently, reflecting each setting’s uniqueness. My methodological plan included maximum variation sampling which aimed to capture and explain ‘cut-across’ themes. Maximum variation sampling recognizes the importance of capturing conceptions, in this instance, from a variety of lived experiences, educational variety, and context differences (Patton, 2006). Finding ‘cut-across’ themes between collective case study participants who have established heterogeneity determined through the quantitative survey, builds a stronger case for the phenomena under study (Davis, 2006; Patton, 2006). Quantitative data was analyzed using the Statistical Package for the Social Sciences (SPSS) software and Jamovi, a more user-friendly platform.

After completing the survey, six participants were chosen for their varied teaching experiences and quantitative data results, and all agreed to continue to Phase II of the study. Phase II consisted of 30-minute semi-structured individual Zoom interviews. Interviews were recorded, transcribed, cleaned, and then uploaded to Dedoose, a software coding program. Transcripts were sent back to the participants, for member checking (Candela, 2019; Creswell, 2005; Stake, 1995; Yin, 2014). Then, applying Yin’s (2014) five-step structured coding technique and the complexity of education theory, I continued to code and recode. To keep participants anonymous, identifiable information was redacted from the transcripts and teachers were given pseudonyms. To ensure the validity of the collected quantitative data, I employed several techniques recommended to me by Dr. Brown.

**TCoA-III Three-Item Reliability Scales**

According to Zijlmans et al. (2018), “Reliability is usually estimated for a total score, but it can also be estimated for item scores. Item-score reliability can be useful to assess the
repeatability of an individual item score in a group” (p. 998). Although, even one item, can represent a scale if it is found to be reliable, the TCoA-III uses three related statements. Looking at the data item by item and as they nest with other similarly worded items, created a detailed analysis format for me to follow.

Dr. Brown advised that I conduct scale reliability estimation for each of the three-item scales (see Appendix A), then repeat using the aggregated improvement and irrelevant scales (see Appendix A). He noted, “If these approached defensible (preferably using McDonald’s omega instead of Cronbach’s alpha) then calculate scores” (Brown, December 5, 2021, personal communication). Dr. Brown suggested that I use Jamovi instead of SPSS to secure McDonald’s omega for the item scales.

I used Jamovi software to determine the reliability of each of the three-item scales in the TCoA-III. The three-item scales include relational questions such as: (a) The F&P assessment interferes with my teaching, (b) The F&P assessment makes me teach in a way that is against my beliefs, and (c) The F&P assessment has little impact on my teaching. Item reliability was analyzed for McDonald’s omega, as Dr. Brown advised (see Appendix A). Three-item scale reliability and single item reliability data is reported within each of the four construct sections in this chapter.

Although coefficient alpha is used for reporting throughout this chapter, Cronbach (2004) himself stated, “I no longer regard the alpha formula as the most appropriate way to examine most data” (p. 403), therefore I also used McDonald’s omega. To report the descriptive statistics, I used the median as well as the mean. Ideally, categorical, polychromous data does not lend itself to being averaged in the same way as integer data. For example, the differences between strongly disagree and moderately disagree are not equidistant, as with integers (Sullivan & Artino, 2013). Thus, the median and mode offer a more exacting report of data. Nonetheless, most of the mean and median scores
were comparable and they were used to analyze respondents’ choices. Respondents’ answers to survey statements which are similar, allow the researcher to cross-check for consistency in Likert scale indicators. The process of analyzing Likert scale survey items in this manner increases the reliability and validity of the results.

**Participant Demographics**

Participants included 22 K-5 public school teachers. Three of these teachers had between 4-8 years of teaching experience (14%); 5 teachers had between 9-15 years (23%), 7 teachers had over 15 years of teaching experience (32%), and the remaining 7 teachers had over 20 years of teaching experience (32%). Many of these teachers (60%) taught kindergarten through third grade, while 30% taught fourth or fifth grade, one teacher taught all levels of elementary school, and one teacher abstained from sharing their grade level information. Of these teachers, 23% held a bachelor’s degree, 70% participants held master’s degrees, and 0.09% held doctoral degrees in education. In addition, 32% held a reading specialist certification, 30% held a special education certification, and .5% participant held an English as a Developing Language certification. Ethnicity and gender identification/preference data was not available as per the policy of the Namaste School District.

The four survey constructs of Assessment as Improvement, Assessment for School Accountability, Assessment for Student Accountability, and Assessment as Irrelevant, along with the guiding principles of the complexity of education theory, will serve as the organizational structure for this chapter. Table 1 provides an overview of participants’ responses to the items grouped into the constructs of improvement, irrelevance, school accountability, and student accountability.

**Table 4.1 Participant Responses by Construct**

<table>
<thead>
<tr>
<th>Constructs</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
</table>
The mean score for the improvement variable (12 items) was 42.19 (SD = 8.81). The range for this construct was 26-57 and showed that participants’ conceptions regarding whether the F&P improves teaching and learning varied widely. The mean score for the irrelevant construct (9 items) was 22.81 (SD = 5.23). This reveals a tighter collective conception which includes some level of disagreement to the F&P survey indicators. School accountability (3 items) had a mean score of 7.50 (SD = 2.43) and a range of 3-12, with most teachers indicating that the F&P is not a good way to evaluate and hold schools accountable. Student accountability (3 items) had a mean score of 11.09 (SD = 1.77) with a range of 7-14. This data shows that teachers slightly agreed that the F&P results can be used to hold students’ accountable. Table 2 provides an overview of all survey participants’ scores on the four constructs of the TCoA-III.

**Table 2 Participants’ TCoA-III Four-Construct Scores**

<table>
<thead>
<tr>
<th>Name</th>
<th>School Accountability</th>
<th>Student Accountability</th>
<th>Improvement</th>
<th>Irrelevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dolores</td>
<td>7</td>
<td>11</td>
<td>49</td>
<td>15</td>
</tr>
<tr>
<td>Mason</td>
<td>12</td>
<td>13</td>
<td>54</td>
<td>23</td>
</tr>
<tr>
<td>Sam *</td>
<td>7</td>
<td>13</td>
<td>43</td>
<td>22</td>
</tr>
<tr>
<td>Pat</td>
<td>7</td>
<td>11</td>
<td>51</td>
<td>13</td>
</tr>
<tr>
<td>Annush</td>
<td>5</td>
<td>13</td>
<td>42</td>
<td>21</td>
</tr>
<tr>
<td>Evie *</td>
<td>3</td>
<td>10</td>
<td>39</td>
<td>28</td>
</tr>
</tbody>
</table>
Ryan 10 10 43 21
Anhil 11 13 49 23
Ky 6 13 46 19
Lilly * 12 14 57 20
Mick * 10 10 50 17
Yolanda * 8 12 33 28
Cruz 6 11 35 30
Kathy 8 10 42 21
Mica 5 8 26 27
Ansika 6 9 31 23
Reg * 6 7 37 17
Dee 8 11 36 28
Mohammed 6 12 47 27
Carry 7 12 47 15
Nan 10 11 50 26
Anna 5 10 26 32

Note. * = Participants invited to Phase II.

In the next section, I will introduce the Phase II participants, all of whom partook in Phase I and were chosen for their particular experiences and answers to the survey questions. Interview data brought to life how teachers’ conceptions of the F&P are operationalized before, during, and after administration.

**Case Study Participants**

Six purposively-chosen teachers were invited to continue to Phase II of the study. I worked within the context of this study, so to mitigate favoritism or researcher bias, I decided to choose participants based on the data alone, names removed. I analyzed the survey data to identify teachers that scored high, low, or in the middle for each of the four constructs. For ethical reasons, my committee recommended that I interview teachers from buildings other than my hub school. Thus, two teachers from the initial list were replaced. Next, I briefly describe why each participate was invited to continue to Phase II.
In all four constructs, Sam and Yolanda had mean scores that were close to the median. Lilly had one of the highest school accountability mean scores and the highest mean score for both the student accountability and the improvement construct. Mick had the highest improvement score, while Evie had the lowest irrelevant score, and Reg had the lowest student accountability score. Thus, the teachers selected represented a full range of conceptions regarding the F&P. In the following paragraphs, I will provide background information for each of the six interviewed teachers.

**Sam**

Sam is an experienced Namaste School District classroom teacher who has taught elementary school for over 15 years. Sam loves to read and enthusiastically shares her love of books with her students. Sam strongly supports that readers grow from reading “just-right-books” (Sam Interview). One of the first things Sam voiced during the interview is how she views the one-on-one time as the most important part of the benchmarking process, “It’s the ultimate in small group instruction.” Sam enjoys pulling the kids out to a different environment where she can really learn about them as individuals and she noted she does not really look at the F&P folder prior to sitting down with the student. Sam received the following scores for the 4 constructs on the survey: improvement scale (43), school accountability (7), student accountability (13), and irrelevant (22). Sam’s conceptions of the F&P align with the median score for each construct, indicating a middle-of-the-road attitude.

**Lilly**

Lilly teaches special education and has been teaching for 15 years. Special education in the Namaste School District utilizes Fountas and Pinnell’s Leveled Literacy Intervention program for teaching reading which prescribes conducting reading records every other day. Although similar, the reading record and the benchmark assessment are designed for different measurements. Lilly tries to be as precise as she can be right from the start of the
benchmarking process, noting, “Our timing is different to benchmark than regular education teachers, coupled with the fact that the F&P is a close fit with what we already do, so I try to gauge closely where the students are before we even start” (Lilly’s Interview). Lilly received the following scores for each construct on the survey: improvement (15), school accountability (12), student accountability (14), and irrelevant (20). Lilly’s construct scores, three of which are higher than the mean of the full group, reveal an impassioned response to the survey questions.

**Mick**

Mick is in the middle of his career and has predominately taught upper elementary grades. Trained as a reading specialist and holding a master’s degree in reading, Mick values the data collection process and noted that even on his team, WHO administers the F&P differently, hold varying attitudes toward it, and utilize the data differently post administration. Mick received the following scores for each construct on the survey: improvement (33), school accountability (8), student accountability (12), and irrelevant (28). Mick’s scores appear to reveal a discerning attitude toward the usefulness of the F&P. For example, his accountability constructs are a bit higher than the mean of the group, but his improvement score is lower than the group’s mean.

**Evie**

Evie teaches upper elementary school and is newer to the Namaste School District than some other teachers who volunteered to be in the study. Evie holds a special education certificate and has between 9 and 15 years of experience at the time of data collection. During her pre-service time, Evie student taught for a whole year and learned how to administer the F&P alongside her mentor teacher. Then, once employed, Evie experienced professional development in the Developmental Reading Assessment (DRA). The DRA and the F&P share similar administration characteristics and so her benchmarking experience predated her
employment at Namaste School District. Evie received the following scores for each construct on the survey: improvement (39), school accountability (3), student accountability (10), and irrelevant (28). Evie’s construct scores reveal a conservative attitude toward the value of the F&P.

**Reg**

Reg is a very experienced teacher with a master’s degree and a reading specialist certification. Proud of his work ethic, love of reading, and helpfulness to all, Reg expressed excitement to share his experiences with the F&P (Reg Interview). He reflected that precision, pride, and empathetical feelings drive his decision-making processes while administering the F&P (Reg Interview). Reg received the following scores for each construct on the survey: improvement (37), school accountability (6), student accountability (7), and irrelevant (17). During the interview, Reg told a story about a former student who was dismissed from reading support but was reentered a few months later. Reg had already indicated to the family that the child would no longer require extra reading support, and he shared in his interview that this discrepancy was embarrassing as the decision was based on inconsistent data use. Reg disagreed with how the classroom teacher recorded the student’s accuracy, because the student was still learning how to read in English and the F&P manual indicates scoring differently in those instances.

**Yolanda**

Yolanda, with over 15 years of teaching experience, currently teaches a primary elementary grade. She has a master’s degree and a reading specialist certificate. Trying to be as efficient as possible when benchmarking, Yolanda isn’t afraid to skip levels and she never has her students read more than two or three books in a sitting. She expressed that since she is often crunched for time, she feels guilty when she must cut students off from their conversation in order to move along the assessment. Yolanda received the following scores
for each construct on the survey: improvement (33), school accountability (8), student accountability (12), and irrelevant (28). Yolanda’s construct scores show an attitude toward using the F&P for accountability purposes. Her improvement score of 33 (group’s mean = 42.19) signifies that she does not see the F&P as a strong tool to improve teaching and learning.

The next section of this chapter defines participants’ conceptions of each of the four constructs of the survey. For each construct subsection, data is described through an agree-disagree lens, by the item mean and median scores, as three-item scaling, and through the lived experiences of the participants generated from their interview data.

**Assessment as Improvement**

The improvement construct on the TCoA-III (Brown, 2002) elicited participants’ latent conceptions of how the F&P BAS improves teaching and learning. Each of the 12 questions offered 6 Likert choices ranging from strongly disagree to strongly agree. This construct, out of the four in the survey, has the most questions, is aggregated into four subscales, and each subscale score is measured by three items. This larger construct, compared to the other three, affords the most variability in responses.

**Teachers’ Conceptions of Improvement Aggregated by Likert Scale Agreement Indicators**

In two recent studies using the TCoA survey (Brown, 2006), factors loaded in such a way that teachers’ conceptions fell naturally into two categories, positive (agree) or negative (disagree). Although my sample was too small to conduct factor analyses, I was able to analyze the reliability of my data in other ways. Brown (2011) reported:

The Improvement conception captures the notion that the purpose of assessment is to improve student learning and teachers’ instruction. Improvement occurs as students use assessments to evaluate, plan, and improve their learning activities and as
teachers interpret student performance as a means of improving instructional activities. (p. 4)

In this study, survey data revealed a division of opinion on the following questions: (a) The F&P provides feedback to students, (b) The F&P provides feedback to students, and (c) The F&P improves student learning. Table 3 provides an overview of what percentage of participants agreed or disagreed with each statement, along with medians and modes for those items.

Table 3

*Teachers’ Conceptions of Improvement Aggregated by Likert Scale Agreement Indicators*

<table>
<thead>
<tr>
<th>Statement</th>
<th>Agree</th>
<th>Disagree</th>
<th>Median</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provides Feedback to Students</td>
<td>46%</td>
<td>55%</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Integrated with Teaching Practice</td>
<td>91%</td>
<td>9%</td>
<td>4.5</td>
<td>5</td>
</tr>
<tr>
<td>Illuminates Learning</td>
<td>82%</td>
<td>18%</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Provides Student Info about Learning Needs</td>
<td>59%</td>
<td>41%</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Measures Higher Order Thinking</td>
<td>73%</td>
<td>27%</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Improves Students’ Learning</td>
<td>55%</td>
<td>46%</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Different Student, Different Instruction</td>
<td>95%</td>
<td>5%</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Determine Student Learning</td>
<td>77%</td>
<td>18%</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Use Results to Modify my Teaching</td>
<td>95%</td>
<td>5%</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Can Be depended on</td>
<td>91%</td>
<td>9%</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Consistent</td>
<td>73%</td>
<td>27%</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Trustworthy</td>
<td>91%</td>
<td>9%</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>
Whilst several statements elicited a 50/50 response, a few questions resulted in a 70/30 and 80/20 split in answers. For example, statements revealing a split in answers included: (a) *The F&P is a way to determine how much students have learned from teaching*, (b) *The F&P assigns a level to student work*, (c) *The F&P results are consistent*, and (d) *The F&P measures higher order thinking skills*.

Analysis showed that 91% of participants found the F&P to be trustworthy; however, the mean and median for this item was 4, indicating that most participants moderately agreed that the F&P is trustworthy. Analysis also showed that 96% of participants agreed that they use the F&P results to modify their instruction. This item received a mean and median of 5 with a lower standard deviation, indicating that most of the teachers mostly agreed that the F&P results are used to modify their instruction.

*Teachers’ Conceptions of Improvement Realized Through Three-Item Scaling*

The four subscales for the improvement construct are: (a) The F&P BAS determines what students have learned (b) The F&P provides feedback to students (c) The F&P affects teachers’ praxis, and (d) The F&P is trustworthy. The mean score for this construct was 42 with a standard deviation of 9. Reliability statistics showed a coefficient alpha of .873 (see Table 4).

**Table 4**

*Complementary Quantitative and Qualitative Data for Improvement*

<table>
<thead>
<tr>
<th>Three item reliability scales</th>
<th>McDonald’s ω</th>
<th>Complementary Qualitative Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid-Trustworthy</td>
<td>0.879</td>
<td>“The assessment is good but there are so many variables in teachers’ administration, so results are different. It depends on who is doing the scoring that’s why we can’t put so much emphasis on it. Teachers already know or they should know about where the kids are before they even do this.”</td>
</tr>
</tbody>
</table>
| Describes/Identifies Learnings | 0.214 | “I am able to observe reading habits and skills that I may not be able to see the classroom in a whole group setting.”

“I am able to ask questions and have one on one discussions about their reading life. It allows a more comfortable reading setting for the student where other students aren’t watching them.” |
| Improves Teaching | 0.766 | “Benchmarking allows me time to better understand how each student is reading. It is a teaching tool, and it allows me to better support a student in literature at his or her level.” |
| Improves Student Learning | 0.763 | “I can write down anecdotal notes regarding concepts/behaviors to target during 1 on 1 or small group work. Similar behaviors among students lend themselves to coordinated small group work.” |
| Aggregated Improvement | 0.873 | * |

In this study, qualitative data was used to explain the quantitative findings. In the next section I will further explain the survey data for the improvement construct through the interview data.

**Teachers’ Conceptions of Improvement as Realized Through Interview Data**

The complexity of education theoretical framework recognizes the researcher’s role within the context of the phenomenon not just as a spectator but as a contributing stimulus of the dynamic system. Thus, interview data is contextually bound and affected by every detail of each interview including interviewer’s and interviewees lived-experiences, personal biases, motivations, personalities, and discourse patterns. Interviews provided a safe space for teachers to explain their views regarding how the F&P improves teaching and learning. Specifically, teachers’ conceptions for improvement were grouped into two themes: (a) grouping for book clubs and (b) benefits of one-on-one time.
**Grouping for Book Clubs.** The Namaste School District uses the Lucy Calkin’s Units of Study as the core resource to support the implementation of the reading curriculum. Book clubs are an integral instructional component of the implementation of the reading curriculum within Namaste School District and teachers in Grades 2-8 form book clubs with the students within their English Language Arts classes at least once or twice a year. Teachers in Grades 2-8 use the Fountas and Pinnell data to inform their decision making as they form their book clubs. This emerging theme is supported with participants’ own words. Yolanda noted that she uses the F&P data when it is fresh to design groups: strategy groups, reading groups, and book club groups. She noted that the information she gleans from the F&P is soon outdated and therefore most effective right away (Yolanda Interview). Similar to Yolanda, Sam said she also uses the F&P level for determining what books she will offer up for book clubs. Sam reported how important it is to her to “confirm, confirm, confirm” the levels in which students are reading. In her classroom she has a saying, you can’t just read, “you have to talk about it.” If students are motivated to “read up”, she will allow them to try if they can talk about what they are reading. Sam reflected upon the fact that when the Namaste School District first began to utilize the Reader’s Workshop model, she was much stricter: “I used to get sick over fighting with students and parents about the books students could choose to read based on their benchmarking scores. Now, I’ve relaxed a whole lot and if students can prove to me that they grasp the story, I am fine with what they choose” (Sam Interview). Mick supported Sam’s views on the role of motivation in using F&P data for improvement:

I use clubs all year, whereas like typically, I think most people only do book clubs when the unit is a book club. In a normal year, I utilize book clubs for the entire year, even if I am doing the non-fiction unit, I have them doing some type of book club because I feel like it helps them. Just with motivation, like if they’re in the
nonfiction unit, they still read their fiction books and it’s done on a more low-key level when I’m teaching other things, right?

Mick’s statement, “In a normal year,” refers to the differences during this particular school year as a result of the COVID-19 pandemic. Specifically, some students engaged in their education virtually while others came to school two days a week. In a normal year, Mick’s use of book clubs throughout the year, although different from the Units of Study curriculum, keeps students reading, and talking about text. When the students are in the book club unit, they are held responsible to complete certain tasks such as posting, writing about their reading, and meeting with their club on scheduled days. Thus, some students may prefer to read fictional works so to keep the motivation to read level high, Mick supports the reading of fiction embedded in the book club frame because students enjoy reading and talking about the same book. In addition to using F&P data to form book clubs, teachers indicated that they appreciated the one-on-one time that the F&P assessment affords.

**Benefits of One-on-One Time.** The F&P is administered one student at a time which provides teachers with an opportunity to focus on the specific literacy skills of each individual child. While Sam was indifferent toward some components of the F&P such as using the historical data on file, overall, she enjoyed the one-on-one time with students:

> It’s just you and the child, you get background information, you get to see how they are thinking without other kids around. Quite honestly, I think you kind of heighten or increase the bond between you and the student. It’s intimate, you put them at ease, and you know, they get to pop out of [the] classroom and you know, it’s fun. You learn things that you sometimes wouldn’t learn about them in front of other kids.

(Sam Interview)

Sam expressed that she uses the F&P information to improve both the students’ reading lives and her relationship with students. Sam connects the data directly to the Reading Workshop
model and she has seen how putting kids in the ‘right’ leveled book, can change their lives. She said, “It’s like helping them to get a foothold or a toehold into, you know, really understanding a book.” Sam also noted that the relationship between the teacher and the student is the most important characteristic of benchmarking.

The expertise that teachers have in regard to encouraging students to want to read and providing the best book choices particularly presented to each student, is the strength of the relationships that teachers build with students, not the tool that is used to assess them. The strength of teaching comes to fruition when the relationships that are established bridge the learner to the learning in all of its complexities. Reg told a story of how, in her opinion, a former student’s educational trajectory was altered in a surprising way based on F&P results.

Reg shared about an experience with a former student, who was new to the country and learning English. The relationship that was established through the reading support lens created a deep understanding of the student’s reading ability. To her credit, she had a bright student who was learning English and new to the country, and within six months to a year the child was reading grade-level text. His parents were proud of him and wondered if their son would continue to need reading support. Reg explained that she did not think the child would qualify for reading support again, yet, the fifth-grade teacher’s impression of the student’s reading skills, based on how his teacher had administered the F&P, was markedly different than Reg’s interpretation of his reading ability.

It has been established that assessment results will differ when human engagement is involved (Brown, 2008; Remesal, 2011). Had there been clearer expectations and professional development for that fifth-grade teacher, the trajectory for the student may have been different which is evidence to explain why teachers are not confident in using the F&P results to measure the quality of the school or to hold the school accountable for students’ reading success. In addition to improvement, the results of this
study indicated teachers generally conceived that the F&P is not useful as a measure of school accountability.

**Assessment as School Accountability**

In Pennsylvania, school accountability is established by a statewide accountability system which is aligned to Every Student Succeeds Act (ESSA, 2015) federal requirement. In addition to the federal accountability markers, Pennsylvania uses the Career Readiness Benchmark and Chronic Absenteeism (PDE, 2021) to measure accountability. The four federal indicators plus the two self-imposed accountability indicators measure how well schools are performing. The term school accountability in this study, brought forth a variety of teachers’ latent conceptions.

*Teachers’ Conceptions of School Accountability Aggregated by Likert*

**Scale Agreement Indicators**

Assessment holds schools accountable because it (a) is a good way to evaluate a school, (b) provides information on how well schools are doing, and (c) is an accurate indicator of a school’s quality (Brown, 2008). Table 5 displays the results for the item-by-item Likert scale analysis for the School Accountability construct. Interestingly, 80% of participants agreed that the F&P provides information on how well the four elementary schools in the Namaste School District teach reading. However, participants did not agree that the F&P is a good way to evaluate schools, nor did they agree that it is an accurate indicator of a school’s quality. The median and mode are identical indicating a strong similarity in teachers’ latent conceptions.

**Table 5**

*Teachers’ Conceptions of School Accountability Aggregated by the F&P Likert Scale*

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Agree</th>
<th>Disagree</th>
<th>Median</th>
<th>Mode</th>
</tr>
</thead>
</table>
A Good Way to Evaluate School | 20% | 80% | 2 | 2
---|---|---|---|---
Provides Info How Well the School Teaches | 80% | 20% | 4 | 4
Reading
Is an Accurate Indicator of School’s Quality | 30% | 70% | 2 | 2

**Teachers’ Conceptions of School Accountability Realized Through Three-Item Scaling**

The three subscales for this construct are: The F&P (a) is a good way to evaluate a school, (b) provides information on how well schools are doing, and (c) is an accurate indicator of a school’s quality. The mean score for this construct was 7.5 with a standard deviation of 5.2. Reliability statistics showed $w = 0.771$. This data supports that teachers do not think the F&P can be used to judge a school’s accountability. Table 6 provides an overview of the reliability data and the complementary qualitative data for the School Accountability construct.

**Table 6**

*Complementary Quantitative and Qualitative Data for School Accountability*

<table>
<thead>
<tr>
<th>Three item reliability scales</th>
<th>McDonald’s $w$</th>
<th>Complementary Qualitative Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>School Accountability</td>
<td>0.771</td>
<td>It is not an assessment tool. It should not be used as one. Due to the variables associated with benchmark administration, the results are not fully valid or reliable (participant open-ended).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Our district does not use the F&amp;P as recommended by the publisher, lack of continuous professional development, and lack of discussion of the results (participant open-ended).</td>
</tr>
</tbody>
</table>
Teachers’ Conceptions of School Accountability as Realized Through Interviews

Teachers within the Namaste School District receive the same guidelines and protocols specific to the administration of the Fountas and Pinnell assessment. These guidelines and protocols are established at the district level. Thinking about school accountability as district accountability makes more sense in this particular context. To that point, Lilly reported the district’s responsibility is to provide an equitable testing situation for each student. For example, some students take the F&P in a quiet location, while others are in a noisier environment without privacy. Lilly’s special education students, some of which have emotional needs, are not always ready to test on the day that Lilly has a substitute. Lilly noted, “They need to feel calm and confident, otherwise I’m going to get very little [data]” (Lilly Interview). To continue, Lilly expressed her concern for the lack of accommodations or modifications built into the assessment to suit the needs of her students:

I feel like the accuracy and the fluency, are very obvious to score, they get it right or wrong. But there’s a lot more subjectivity to the comprehension and I honestly think that it falls on Fountas and Pinnell because the guidance is not as clear as it should be. Like I’ve had other assessments where like there’s an actual rubric that like really follows along with what that is. So that’s one criticism, another is I think that a whole lot of decisions are sometimes pending on one assessment. The setup of the F&P is difficult for my kids, I know they know what’s going on, but it is difficult for them to retrieve and I’ve had some friends who have an IEP goal for testing accommodations, and it doesn’t happen. (Lilly Interview)

Thus, through Lilly’s lens, the F&P is not an accurate measure of school accountability as students’ IEP accommodations are not always followed, leading to potentially inaccurate results.
While schools are held accountable by local, state, and federal agencies, parents also play a role in holding schools accountable for their child’s academic development. However, this assessment is difficult to use to hold schools’ accountable as there are too many variables impacting the results. For example, Evie expressed that the F&P could not be used to hold schools accountable, “I would caution [using the F&P for school accountability] because of the amount of different teacher interpretations of it. A teacher can interpret it differently. I would caution the use of it as an evaluation of the school” (Evie interview).

Yolanda supported Lilly’s views regarding the subjectivity of the F&P as she shared some of the thoughts that run through her head before and during assessment. For instance, “Do I know this story? Wait, did he read that word correctly? Is her prosody on point? Is the student getting tired?” (Yolanda Interview). These in-the-moment decisions coupled with a lack of time create an environment where Yolanda viewed that using the data from this assessment to hold schools accountable was unrealistic. Many of these indicators also supported why teachers’ conceptions of using the F&P for student accountability varied.

Assessment as Student Accountability

The student accountability construct is built around the premise that students are held accountable by the grades and test scores they earn. This construct has three items, which can be viewed in Table 7. The mean score for this construct was 11 with a standard deviation of 1.77. The minimum score was 7 and the maximum scale score was 14 out of a possible 18.

Teachers’ Conceptions of Student Accountability Realized Through Likert Scale Indicators

Teachers slightly to moderately agreed that the F&P assigns a level to student work, determines if students meet grade level expectations, and the results are used to place students on learning trajectories. Agreement statements include all four levels of
agreement and therefore, are continuous. We see that none of the 22 participants disagreed that the F&P places students on a learning trajectory which is important because if students are placed on a learning trajectory based on their F&P score, echoing the importance of having students read the most difficult text they can handle.

Table 7

*Teachers’ Conceptions of the F&P Aggregated by Likert Scale Agreement Indicators*

<table>
<thead>
<tr>
<th></th>
<th>Agree</th>
<th>Disagree</th>
<th>Median</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assigns a Level to Student Work</td>
<td>70%</td>
<td>30%</td>
<td>3.5</td>
<td>4</td>
</tr>
<tr>
<td>Determines if Students Meet Grade Level</td>
<td>80%</td>
<td>20%</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Places Students on a Learning Trajectory</td>
<td>100%</td>
<td>0%</td>
<td>4.5</td>
<td>4</td>
</tr>
</tbody>
</table>

*Teachers’ Conceptions of Student Accountability Realized Through Three-Item Scaling*

The three subscales items for the student accountability construct are the F&P: (a) assigns a level to student work, (b) determines if students meet grade level expectations, and (c) places students on a learning trajectory. The mean score for this construct was 11.09 with a standard deviation of 1.77. Reliability statistics showed a coefficient $w = 0.701$. Table 8 provides an overview of the reliability data and the complementary qualitative data for the student accountability construct.

Table 8

*Complementary Quantitative and Qualitative Data for Student Accountability*

<table>
<thead>
<tr>
<th>Three item reliability scales</th>
<th>McDonald’s $w$</th>
<th>Complementary Qualitative Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Accountability</td>
<td>0.701</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>“It shows reading strategies or skills that students do or do not know. Gives me insight into students’ approaches: word accuracy, s/c, retell &amp; comprehension” (Open-ended Response).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>“I am able to watch and observe students read and observe reading habits and skills” (Open-ended Response).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>“It shows what types of reading strategies or skills the students do or do not know and gages a general grade- level expectation” (Open-ended Response).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Teachers’ quotes from their open-ended responses revealed how they use the F&P to hold students accountable for their reading work. Frymier (1998) posited that students are responsible for their own learning and teachers are responsible and held accountable “for what they do as teachers but not for what their students do as learners” (xi). The controversy over accountability is beyond the scope of this study but is relevant to consider when interpreting the results. For instance, are teachers accountable for making sure that students are placed on the proper instructional reading level so that their educational trajectories are the most rigorous as possible? Are students accountable for doing their best with every encounter they have with their teachers? Are school districts responsible to use assessment materials exactly as recommended by the publisher?

Teachers indicated that they slightly to moderately agreed that the F&P can be used to hold students accountable; however, the wording of the indicators seems to put the onus on
the F&P, not on the student. For example, one statement read *the F&P can be used to hold students accountable*. This statement puts the action on the assessment, not on the student. In this instance, the wording of Brown’s (2004) survey tool may not be an exact fit for this context, however the concept of student accountability in learning is an ever-revolving conversation.

**Teachers’ Conceptions of Student Accountability Revealed through Interviews**

Interviews revealed that the F&P could, at times, be used to hold students accountable. However, teachers expressed concerns with putting too much weight on the results. For instance, when I asked Evie if she thought the F&P could be used to hold students accountable, she said:

> I feel like it's one piece of the puzzle. And if a student isn't progressing, that's not on them. That's on the supports and the teaching and learning that has happened in between. And so, I don't know that it should be used as an accountability piece for a student, like an end all and be all. I always tell my kids to do their best, I show them and talk to them and things like that, because I think it's valuable to them to kind of understand why we are doing this. But past that I would never say, ‘Oh, compared to last year, I would have thought you made this progress since then.’ I don't think that's fair. (Evie Interview)

Mick agreed that the F&P should not be used to hold students accountable reporting that students are unpredictable:

> They are fatigued, or what was their day like? Is this a day I should be benchmarking them? Should I bring them back a different day because, you know, just their personal life has something going on or they've been out of school, they miss days, things like that. Sometimes I'll wait to benchmark them because I think they're not...their focus might not be there. Sometimes I end assessments because I think their focus isn't there.
and they're not really showing me their what they can do that now. I'll pull them back a different day. Right? (Mick Interview)

Mick’s thoughts capture some of the subjectivity in place regarding testing conditions and timing, and they also address how these conditions can have consequences for the students being assessed.

Alternately, Sam recounted her experiences with students who read up too many levels than they can manage. Sam expressed that she holds her students accountable and instructs them about accountability in the Reader’s Workshop. Sam teaches her students early on in the year that you have to be able to talk about the book. She says, “If a student says, you know the guy with the cloak” but cannot describe Harry Potter in greater detail, then that book is too hard for you right now. Sam shared that she has noticed that this approach encourages some readers to get reading so that they can level up properly.

The final construct, irrelevancy, asks participants to indicate how relevant the F&P assessment is to them. Relevancy is measured by how interfering it is, or how the data collected is not useful. The collective indicated some similar conceptions regarding the F&P’s relevancy and some differences.

**Assessment as Irrelevant**

The F&P as irrelevant construct consists of 9 items that revolve around three themes, the F&P: (a) interferes with teaching, (b) is unfair to students, and (c) is unreliable. Table 10 displays the Likert scale agreement indicators for the irrelevance construct.

**Table 9**

*Teachers’ Conceptions of the F&P Aggregated by Likert Scale Agreement Indicators*
Teachers’ Conceptions of Assessment at Irrelevant Aggregated by Likert Scale Agreement

Indicators

Teachers indicated that they treat the F&P results with caution, they consider the error in test administration, and they rejected the idea that the F&P makes them teach against their beliefs. Teachers indicated their usage of F&P data for developing small groups and book clubs, so their perception of ‘the error in test administration’ lies with other teachers. This means that teachers rely on their own F&P data collection, but not the data supplied by other teachers. The F&P folders move up with the students as they advance in grades. Recording sheets are housed within the folders and students’ instructional and independent levels are written on the outside each time the student sits for the inventory.

Teacher’s Conceptions of Assessment as Irrelevant Realized through Three-item Scaling

The three subscales items for the irrelevant construct are: The F&P (a) interferes with teaching, (b) is unfair to students, and (c) is unreliable. The mean score for this construct was 11.09 with a standard deviation of 1.77. Reliability statistics showed a
coefficient $w = 0.701$. Table 11 provides an overview of the reliability data and the complementary qualitative data for the irrelevant construct.

Table 10

**Complementary Quantitative and Qualitative Data for Irrelevant**

<table>
<thead>
<tr>
<th>Three item reliability scales</th>
<th>McDonald’s $w$</th>
<th>Complementary Qualitative Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bad</td>
<td>0.248</td>
<td>It takes longer than I would expect due to differences in fluency, book length, and student factors. The other challenge is having to test some in the room while other students are working (Open-ended Response).</td>
</tr>
<tr>
<td>Ignore Results</td>
<td>0.756</td>
<td>*It is not an assessment tool (Open-ended Response).</td>
</tr>
<tr>
<td>Inaccurate</td>
<td>0.790</td>
<td>It is an imprecise measure of students’ ability. The results are extremely subjective. Doing it in the classroom seems to not be valid (Open-ended Response).</td>
</tr>
<tr>
<td>Aggregated Irrelevant</td>
<td>0.804</td>
<td>*</td>
</tr>
</tbody>
</table>

Teachers’ responses to the F&P as irrelevant construct revealed that while these teachers do not interpret the assessment as unfair to students, they expressed that they do not trust the results provided from other teachers. They noted inaccuracy and inconsistency with the results, yet they use the results to create book clubs and reading groups.

**Teachers Conceptions of Irrelevancy Revealed Through Interviews**

Teachers resoundingly agreed that the time spent with students and the data that they collect during administration of the F&P is relevant. Participants described targeted irrelevant
conceptions such as the time it takes to administer, inconsistent results, and the misuse of the tool at the district level. Lilly reported that the time it takes to administer outweighs the benefits:

For me, I feel like in special education we spend a whole lot of time assessing, which I know is important to show progress. But it’s at the expense of instruction and I struggle as an educator with that. I need them reading every day. I want them reading, reading time is precious, and I hate to lose it to say we’re going to do, you know, some sort of assessment that takes significant time so it’s a balance. (Lilly Interview)

On the contrary, Mick shared a concern with the trustworthiness of F&P data received from other teachers and voiced it was worth the time investment to see students’ data for himself:

You will see differences in how people administer it and how they utilize it in the classroom. I know there’s a difference just on my team. I know I value it more than some of the members of team. I like it, I like to analyze the data and the decisions that I make will benefit my students. Right? So, like, I don’t really look at what previous teachers say. I’ll look at whatever record and I’ll see if they jotted any notes down. Right. But from what I’ve seen, like we’ve had teachers that say coming into fourth grade, the students were at U and I’m feeling a little skeptical about it. Right. So, like I might go back because I’m the one using the information and by the end of the year, I have been reading with the students so much that I might give a pass on a few questions in the spring because I know they know it. (Mick Interview)

Thus, Mick relied on the inconsistency theme of the irrelevancy construct as he decided which data to use for students.

Adding to the difficulty of using F&P data alone for instructional decisions, Yolanda reported that her younger students have difficulty with specific assessment terminology that
is embedded in higher-level benchmarking texts, and she sees that as a barrier to accurate assessment results. From her perspective, it is not fair for her younger students to be held to the same standards as an older student who may have been taught the literary terms needed to answer a question. For example, *figurative language* and *descriptive language* are two concepts embedded within the comprehension conversation rubric, yet younger students may have difficulty demonstrating their comprehension of the text when it is interwoven with technical terms. Yolanda also mentioned how her young students want to talk about the text and they read in a meandering way, which is what she wants for them. For Yolanda, the stress, the rush, and the unfairness deem the F&P as irrelevant (Yolanda Interview).

**Teachers’ in-the-Moment Decisions During Administration**

Teachers operationalize the administration of the F&P in a variety of ways. The following protocols were reported by participants in the open-ended response section of the survey. Some teachers utilize the F&P word list as a guide for themselves to determine an appropriate text level to start the assessment. In contrast, others reported using the historical reading record data in students’ F&P folders. Yet, additional teachers said they use their anecdotal class notes to determine where to begin. This variety shows how differently teachers operationalize the way that they begin their F&P administration. For example, Sam, one the interview participants, does not use the historical data in the folder because he does not find it to be valid.

Another indication of where teachers feel validity is threatened, is with the degree to which students’ background knowledge increases their reading level score. Teachers acknowledged that the awareness of a student’s background knowledge may impact where they begin the assessment. For example, if the student and her family are outdoorsy, they may have more familiarity with fishing. Thus, the teacher chooses not to use the *Fishing Smarts*
by Luka Berman story, as students can rely too heavily on their background knowledge when decoding and answering comprehension questions.

On the other hand, too little background knowledge can also be a barrier to obtaining accurate results on the F&P. For instance, Yolanda shared, “If I know the student is learning English and may not be aware of the text’s subject, like a gopher or a merry-go-round, I may show the student an image, just tell them what it is, or choose a different book to use.” This example captures one of the limitations of the using the F&P kit to monitor students’ growth up the reading ladder, there are only two books at each designated level. This lack of variety in the texts available at each level can complicate the process for teachers.

Teachers base some of their assessment decision-making on the book topic. For example, A Weighty Decision is about an older student who contemplates taking pills to lose weight for wrestling. When choosing to use this book, teachers might need to consider the sensitivity of the topic as part of their pre-assessment planning. Some teachers do not mind if a student rereads a story that they read the year before or even two years before, while other teachers want to see the student read something they have never seen before. According to F&P a benchmark assessment is completed with an unseen, brand new text for the student. Therefore, engaging with a student while they reread a story enables the evaluator to capture reading behaviors, the assessment validity is damaged. Within the cycle of the teaching and learning system, that students’ reading level as indicated on the data sheets looks just like the level indicator for the next student on the list. The numbers or levels do not tell the whole story of the reader, yet at times, decisions are made that place the student on a learning trajectory based on that indicated F&P level. For instance, one teacher shared:

My knowledge of a student’s reading level, background and interests, informs my administration of the F&P by allowing the process to be more succinct. If I already
know about where they are at, I can easily choose a level. If I know their primary interest, I can choose a fiction/nonfiction text first, to ease them into the testing environment. If I know they are hesitant readers, I may start with an easier text then the level I am confident they are reading at, because it builds confidence in the environment. (Yolanda Interview)

In contrast, a few teachers said that knowledge of the student does not affect the way they administer the assessment at all. For example, one teacher noted, “The F&P provides a script we are expected to use so there is little influence over the assessment with relation to students’ personal interests” (Open-ended response). In contrast, Mick reported:

I decide to start on a fiction or nonfiction in order to avoid text where they have a lot of prior knowledge. For example, we study tsunamis as part of our reading nonfiction unit, I don’t like to use those stories to test students because I feel the results might not be an accurate reflection of what the student can do as a reader. (Interview)

Other considerations during the F&P administration included, but were not limited to, how much scaffolding should be given and how much prompting to give to a quiet student, learning support student, or an English language learner. Specifically, some of the subjectivity in the assessments are in when teachers code the little errors that do not change the meaning of the story differently. Some teachers mark all errors, some do not. These testaments’ answer one of the research questions posed in this study, how do teachers operationalize their conceptions of assessment while administering the F&P? Teachers operationalize the F&P assessment in many different ways.

I coded for the four constructs of the survey tool (Yin, 2018) and then I recoded the qualitative data to capture emerging themes. Time was an emergent theme. All six of the heterogenic interview participants talked about how much time it takes to administer the F&P. Many of the other participants in their open-ended responses referred to time as one of
the challenges of the F&P. Teachers reported that even though they read one on one with students and they may not find the student’s instructional level or they do find the first instructional level and they know the student can manage a higher leveled text, they do not have the time to continue with that student. One participant said:

It takes so much time, and we don’t have a great amount of time it's unfair to students because one student may get a large chunk of time in a quiet spot whereas another student might have to read in the classroom with other students present. I think of the district is going to continue to use the data the way that they are then they need to provide adequate time quiet space for every child. (Lilly Interview)

In the same vein, a teacher said, “It’s not easy. In one day, I listen to 9 students read two stories and I am not even finding both the independent and instructional levels. It’s not fair to the kid who has to then read in the classroom, it seems not to be valid.”

Summary

In this chapter, I discussed my findings of the case study. The case studied was bounded by the four elementary schools in the Namaste School District and the 22 participants who volunteered to engage in this work. I examined the teachers’ conceptions of the F&P, their administration of the F&P, and their operationalization of their conceptions of the F&P during administration. Results showed that teachers held complex views that comprised of several of the studies constructs. The four guiding principles of the TCoA-III survey tool illuminated teachers’ latent conceptions and the qualitative data, explained, deepened, and allowed for a cataloging of data from this context. Results are bound to this location, time, and participants and although are not generalizable to other locations, some of the learnings from the study are. The next chapter describes in more detail the study’s results, limitations of the study, implications for change, and recommendations for future research.
Chapter V: Discussion

Through this study, I examined K-5 teachers’ conceptions of the Fountas and Pinnell Benchmark Reading Assessment Systems 1-3 (F&P). The mid-Atlantic school district, which served as the setting for this study, is comprised of four elementary schools that feed into one middle and one high school. The workshop model instructional framework is used to support the implementation of the elementary reading and writing curriculum. One of the major tenants of reader’s workshop is for students to read books at their independent and instructional reading levels (Calkins, 2018). Books read during independent reading time or within the book club unit(s) are chosen for their “just-rightness” [independent level] whereas books read with the teacher during small group instruction are chosen because their text difficulty is a match with students’ identified instructional reading levels.

In this context, the F&P, which is designed to identify students’ instructional reading level, is administered by classroom teachers in Fall and Spring. After teachers collect the data, it is transferred to a data warehouse. The data warehouse includes demographic data, grades, medical history, assessment scores, and other pertinent longitudinal data for each student in the district. Data points such as MAP, PSSA, and grades are available for parents to view but the F&P data is not available for parents to access. Teachers and parents have the opportunity to discuss individual student reading progress during parent-teacher conferences. In this setting, it is not unusual for parents to ask about reading levels so that they can support their child’s literacy life at home. The district guidelines encourage teachers to discuss an individual student’s strengths and needs as a reader as well as share examples of the types of books students are reading. If asked about a child’s reading level, teachers are advised to discuss F&P reading levels within bands (i.e., This student is most often reading books that fall within the T, U, V text band).
Teachers serve as facilitators and data collectors during test administration in that they choose which text level to begin with and which genre (e.g., fiction or non-fiction) students will read. While research has well-documented that teachers’ conceptions of assessment impact decision-making processes (Brown, 2008; Deluca et al., 2013; EunMi Haslam, 2018; Flores et al., 1991; Remesal, 2011; Sanchez, 2017; Xu & Brown, 2016), to my knowledge this is the first study of a case to examine teachers’ conceptions of the F&P. The results of this study showed how teachers’ latent conceptions of the F&P can be made visible and because of their impact in the classroom are an important topic of study. These results contributed to other researchers’ findings which call for school districts to include the recognition of the roles that, beliefs, bias, and context have on their educational system.

**Summary of the Study**

According to Yin (2018), case study is defined as “an empirical inquiry that investigates a contemporary phenomenon (the ‘case’) in depth and within its real-world context” (p. 16). The examination of how teachers’ operationalized administration of the F&P demarked one of Yin’s (2018) recommendations for determining if your research question is a good fit for this method of data collection. Another indication of fit for this study was the question, what are teachers’ conceptions of the F&P. Thus, case study data gathered with a sequential explanatory mixed methods design, using the complexity of education theory as the guiding theoretical framework, collected data from 22 (17%) K-5 teachers in the Namaste School District during the year of the Covid-19 pandemic. Quantitative data was collected first using the TCoA-III survey tool, proven reliable, and vetted by two retired Namaste schoolteachers. This 27-item, positively-packed, 6 choice Likert scale survey made visible teachers’ latent conceptions of the F&P. I added two open-ended questions at the end of the survey to capture qualitative data from all of the participants.
Quantitative data collected in Qualtrics, was uploaded to SPSS and Jamovi; it was analyzed and used to invite 6 teachers, to continue to Phase II of the study, semi-structured interviews. Phase II participants, purposively chosen to create a heterogenic group, all agreed to continue on and were interviewed separately in one, 30-minute Zoom meeting. Qualitative data from the interviews was cleaned, transcribed, and sent back to participants for member checking. Qualitative data was coded using the a priori conceptions of the TCoA-III survey tool, followed by the coding of emergent themes. Data from all three collections were compared, triangulated, complemented, and presented even when data was dissimilar.

**Application of Theoretical Framework to Findings**

The framework for this study was represented by the four interconnected spheres (See Figure 2) nested within the context of the case under study. Each of the spheres: Teachers’ Conception of Assessment, Researcher, and the F&P are proximal to the complexity of education theory to illustrate the interconnectedness of all things. The linking cords represent the tangential interactions that occur within relationship in the system as a whole. The complexity of education theory supported the study of the case this process embarked upon to examine. By setting the boundary of the case, I created “the condition of possibility for a rule or a law to exist when a boundary is not naturally given as in the case with natural complex systems” (Osberg et al., 2018, p. 220). Binding a case is a tool that allows researchers not to identify reality or to generalize to the larger system, but to inform, so that the researcher may manipulate their environment so that they can live in it and change it; it is only through interacting with our world that knowledge emerges” (Osberg et al., 2018, p. 220). The story that emerged from the study of this case [teachers’ conceptions of the F&P] was completed with the hope that I have clearly told the teachers’ stories of their lived experiences.
Figure 2. Theoretical Framework

*Note:* The dark rectangle represents the Namaste School District, the boundaries of the study. The spheres inside the rectangle, contain the complexity of each system. The connecting links represent the tangential interactions that occur within and between the complex systems. The complexity theory is in the center as the best place to capture the simultaneity of the phenomena under study.

*Theoretical Framework Applied to Teachers’ Conception of Assessment*
The examination of teachers’ conceptions through the lens of the complexity theory gave me permission to create space for emergent discoveries. Davis (2008) showed how complexity and hermeneutics are oriented “by the entangled questions, ‘What is it that we believe?’ ‘How is it that we came to think this way?’” (p. 62). One participant responded to the open-ended question ‘What are the benefits of the F&P? in this way:

Benchmarking allows me time to better understand how each student is reading. It is a teaching tool, and it allows me to better support a student in literature at his or her level. One of the challenges, it is not an assessment tool. It should not be used as one. Due to the variables associated with benchmark administration, the results are not fully valid or reliable (open-ended survey response).

This highly educated, experienced teacher, who has been with the Namaste School District for quite a long time, showed us through these words how complex this phenomenon is. We can see what the participant believes, but how they came to think this way is literally too complex to discover. If we imagine the participant’s words within a sphere connected tangentially to everything else that may be included in that mini-system, and how those concepts interact or rub against their closest neighbors, one can almost see the weaving of ideas like a crocheted blanket. This representation of an interpretation should not be seen as a simple reduction and decomposition of one’s thoughts or intentions (Davis, 2008).

Complexity theory does not deconstruct things to analyze the parts, rather, it considers but does not identify, the wholeness because of the parts.

**Theoretical Framework Applied to the Researcher**

The examination of my location within the study of the case is another identifying factor of the complexity of education theory. This theory recognizes the researcher as an affective and effective condition or variable. Davis (2008) strongly suggested that complexivists have a moral obligation to address these questions:
• How am I complicit in affecting or hoping to affect the phenomenon that I study?
• How is this research *educational*—that is, how does it educate?
• How might this research be taken up?
• How might I represent/present these interpretations? (p. 64).

I have examined my motives and concluded that I am interested in teachers’ conceptions of the F&P more than my conceptions of the F&P. For clarity, I see it as a packaged, thought-provoking tool that may indicate learning opportunities for students. Currently, I think the assessment tool would best be utilized in a dynamic way or when teachers do not have other means of identifying reader behaviors for teaching purposes.

To answer Davis’s (2008) second question, I do consider this research educational because it provided new information to researchers, teachers, the Teaching and Learning department in the setting of the study, and to myself. This data showed that teachers’ hold multiple, conflicting views the F&P and they were shown to affect and effect [not a surprise] how they operationalized the administration of the F&P. Differences in administration styles reflected teachers embodied lived experiences, the contexts in which they worked, and local policies regarding technique, data collection parameters, and how the data was used post-test. These results bring forth evidence that benchmark reading assessment results are affected by many factors which, due to their ability to impact the F&P test administration and results lead to the conclusion that this assessment is unreliable and perhaps invalid when used differently than its creators intended. Further discussion of these emerging conclusions appears in this chapter under implications and future research recommendations.

**Theoretical Framework Applied to the F&P**

The complexity of education theoretical framework encouraged seeing the parts of a system not as something put together to create the whole, rather, the whole is the synergy between the parts. Just as participants had varying conceptions of the F&P, teachers had an
equal amount of variability in what they deemed the most important reading behaviors. For instance, 78% of participants indicated that fluency (words per minute) was important to them, 67% said ability to answer questions without too much hesitation, ability to answer without using the book, and seeming interested or not too bothered by the experience were important, 44% of participants indicated that reading with expression and perseverance-exhibiting characteristics of a learner were most important, 33% of participants said the ability to retell the story without prompting and fixing errors was important, and 22% of participants indicated that accuracy was one of the most important reading behaviors students’ exhibit while being tested with the F&P. It was not within the scope of this study, due to the Covid-19 pandemic, to examine the relationship between identified important reading behaviors and F&P administration and/or how those conceptions impact guided reading lessons or incidental teacher-student interactions.

Examining teachers conceptions of the F&P and how those conceptions operationalized teachers’ decision making processes revealed a variety of microsystems. Complexity theory posits the phenomenon of emergent ideas which emanates from the edge of chaos (Davis & Sumara, 2006)

**Limitations of the Study**

This study is limited by the context, researcher’s lens, participants’ viewpoints, and restrictions placed upon the study as a result of the Covid-19 restrictions. Other limitations include the survey tool I chose to use the interview structure, and the 30-minute timeframe I placed upon the interviews. I initially thought that using Zoom would feel less intimate and uncomfortable to the interviewees, but that did not seem to be the case for the participants in this study.

Another limitation of the study may lie within the TCoA-III survey tool which measures responses to four constructs for the purpose of assessment: improvement, student
accountability, school accountability and irrelevance. However, given the limited scope of this tool, it may not have captured other purposes for assessment.

**Consequences of the Pandemic**

This chapter includes a discussion of those invisible [latent] variables as they came to be identified through the survey and the interviews. This study’s design was altered dramatically because of the pandemic, specifically: (a) schools moved to virtual formats, where teachers and students quickly converted [in two weeks] to remote teaching and learning conditions, and (b) plans were altered to meet the IRB’s evolving pandemic regulations. As the only researcher collecting data for the study, it was important to take time to reflect upon my presence within the study and how that affected my design and analysis choices. For example, in addition to the preset themes determined for coding, new themes emerged from the transcription analyses. One of those emergent themes was time.

**Emergent Themes**

Time was the most recurring emergent theme along with the greatness of one-on-one time spent with kids. My role as a building reading specialist, places me in a position to grasp how time is a factor for teachers when they are benchmarking, so I was not surprised to see it emerge. Another emergent theme was how teachers’ use the data post-assessment to inform their decisions for grouping students into book clubs and reading groups which was also expected. The theoretical framework, places me in the middle of the study of the case therefore requires reflection and an understanding that my role is not to come to conclusions. Thus, complexity according to Osberg et al. (2008):

suggests a temporal epistemology which implies that the quest for knowledge is not in order that we may develop more accurate understandings of a finished reality as it is rather the quest for knowledge is about finding more and more complex and creative ways of interacting with our reality through doing this through intervening in our own
realities we find out how to create more complex realities with which we can interact in yet more complex and creative ways the point is that from a complex systems perspective there are no final solutions only ongoing interactions leading to increasingly more complex interactions and solutions. (p. 215).

The *a priori* themes for coding purposes mirrored the four constructs in the survey. They are assessment is used for improvement (*I use the results to inform my teaching*), school accountability (*The F&P is a good way to evaluate how well the school teaches reading*), student accountability (*The F&P determines what students have learned*), and irrelevancy (*I consider the error in test administration*). A third emergent theme was test-equity. Subthemes under test-equity included: (a) testing conditions, (b) data use, (c) time, and (d) content of the books.

I aimed to bring forward the examination of teachers’ conceptions of the F&P to enrich my understanding of the phenomena of how conceptions impact the administration of the F&P assessment. I used the following research questions to facilitate the data collection.

What are teachers’ conceptions of the F&P? (1) In what ways do teachers operationalize the administration of the *Fountas and Pinnell Benchmark Reading Systems 1-3*? and (2) How do teachers’ conceptions of assessment inform their administration of benchmark reading assessments?

**Overview of Study Results**

The study of the case was examined through an explanatory sequential mixed methods study, bound within the Namaste School District’s four elementary schools. This study examined teachers’ conceptions of the F&P, and teachers’ conceptions of assessment ranged considerably and as did their benchmarking techniques and lived experiences. The results from this study aligned with other researchers’ examinations of teachers’ conceptions of assessment (Brown, 2008; Deluca et al., 2013; EunMi Haslam, 2018; Flores et al., 1991;
Remesal, 2011; Sanchez, 2017; Xu & Brown, 2016). Two of the main differences between this study and others that utilized the TCoA-II were the sample size [small] and the targeted focus to one assessment in particular.

First and foremost, participants agreed that having one-on-one time with their students is a powerful experience. Sam [Phase II participant] reported, “I enjoy it it's one on one time with kids in a different environment. It's the most focused assessment that we do; you and the child bonding - if you do it right, it improves your relationship.” Participants saw the value in time spent with children when the setting was quiet, private, non-rushed, and not too long. Participants’ views regarding time spent with students was not reflective, nor would we expect it to be, of the collective of Namaste elementary teachers as evidenced by participants’ remarks. For example, Evie said, “For some people they don’t see the value of the one on one, having that time listening to them listening to them problem solve, jotting down things to work on later” (Evie interview). Evie continued:

The time that it takes to do the assessment in a classroom is a hard thing. It requires a lot of time it's not fair to other kids in the classroom or the student testing. It needs to be quiet in order for it to be fair. It’s not easy, we have a day. I can read with nine kids. I can listen to them read two books and not even find both independent and instructional levels. In order to be fair to the kids, it needs to be a quiet spot with the teacher, otherwise doing it in the room seems to be not valid. (Evie interview)

Overall, teachers in this study slightly to moderately agreed that the F&P improves teaching and learning. Teachers voiced that the F&P provides: (a) better understanding of students’ scaffolding needs, (b) one-on-one time to work students, (c) insight into students’ reading skills and strategies, (d) a gauge of grade-level expectations, and (e) information for book club designs and small-group lesson development.
**Accountability**

In terms of accountability, teachers more readily indicated that the F&P is better suited to hold students accountable more so than schools. However, one teacher strongly disagreed that the F&P holds students accountable saying, “If a student is not learning how to read, it isn’t their fault, it’s [teachers] our fault. Something needs to be examined with the teaching and learning experience for that child” (Evie Interview).

**It Feels Like a Chore**

Several teachers indicated that this process felt more like a chore than a sound mechanism for identifying students’ instructional reading levels. According to Ravitch (2016), “Accountability makes no sense when it undermines the larger goals of education” (p. 18). The goal of assessment is to improve students’ learning; it is not to prove; it is to improve. According to the National Research Council (2001), high-stakes tests are questioned for their ability to improve learning, and classroom assessments “are not being used to their full potential” (p. 1). Educationists are right to remember that “the meaning of a text is what the reader makes out of what she finds when she reads. Meaning in this sense is not in the text” (Pinar, 1995). Thus, teachers are placed in an interesting position.

**Time**

Teachers expressed that they felt rushed and pressured to move as quickly as possible on benchmarking day. One teacher reported listening to 9 children read 2 stories each, and in some cases she was not able to identify the independent and instructional reading levels for those students which left her feeling overwhelmed (Yolanda Interview). Time was a theme with two faces in this study. Teachers voiced that they felt rushed to see as many students as they could in the limited [their words] amount of time they had yet loved the time they had with students. For Sam, the time with her students was a novel way to create and deepen her relationship with her students. The relationship, according to Sam, is the most important
factor while benchmarking. Sam saw the time together with the student as an opportunity to sow the seeds of love for reading. A self-professed “lover of all things reading workshop” Sam shared that her goal is to spread her love of reading to her students in the hopes that they, too, will become readers.

**Irrelevant**

The most salient characteristic of teachers’ conceptions of the F&P as irrelevant was their belief of its inconsistency, unreliability, and trustworthiness. Barnes and Dacey (2014) concluded that “if teachers believe that assessment is irrelevant then it cannot (should not) be used for learning, for student accountability or for school accountability purposes” (p. 291). In the context of this study, teachers predominately did not trust historical information in students’ F&P folders, they did find the information that they collected to be valuable for immediate use to assist them in determining book club placement, independent reading book selection, and for developing reading groups for instruction. For some, the F&P is a confirmatory exercise of what they already knew regarding student reading behaviors based on classroom formative data. Classroom formative data, a powerful tool, when used properly, captures readers’ wherewithal across the day(s) with different kinds of text in situ which is different than reading with students while administering the F&P.

The results of this study demonstrated that improvement, school and student accountability, and irrelevance all played a role in teachers’ operationalization of the F&P. Yet, there were limitations to this study in that data was gathered through the TCoA-III survey tool, the particular open-ended response questions, and the researcher’s line of questioning during the semi-structured interviews. Therefore, these results are bounded within this case context and it is understood that the platform utilized interacted with and developed the data set forth.
Limitations in Methodology

One limitation of the methodology may be the theoretical framework I used, as the complexity of education theory, although not new to researchers, is still an emerging theory in education research. This theory has been criticized for its inability to predict or prescribe and for its transdisciplinarity (Sumara, 2006, p. 8). In addition to the complexity of education theory used as a theoretical framework, case study too is seen as a ‘lesser than’ method of data collection (Yin, 2018). Yin explained that this line of thinking is outdated and in fact case study is now elevated to the category of a Method. According to Yin (2018), “Sources such as the U.S. Government Accountability Office have documented the evaluation applications of the case study method” (p. 5). Thus, the study of a case, as is the case, in this study, brings strength to the design, in my opinion.

Another limitation of my methods may lie in the semi-structured interview technique. Fraenkel et al. (2012) reported that a limitation in this style of interviewing is the researcher’s flexibility. Having flexibility within the interview may prevent the interviewer from touching upon all of the important topics that are in the interview guide, therefore limiting the ability to compare or complement participants’ responses (Fraenkel, 2012, p. 452).

An external audit of my qualitative data coding decisions made by me may have led to different emergent themes. In order to address this limitation, I reviewed my coding decisions with my advisor and realigned parent and child codes to address any patterns she saw that I had yet to address.

Limitations in Analysis

Methodologists in the field of education debate which statistical procedures best fit with small data sets. For this study, I made the conscious decision to represent my data through descriptive statistics, acknowledging that inferential analyses could potentially be misleading. Thus, while a larger sample size would have better allowed for me to detect
differences between subsets of participants based on factors such as gender, literacy
development, and years of teaching, those analyses were beyond the scope of this particular
study. Another limitation in analysis may be the methods I used to choose participants for the
Phase II semi-structured interviews. I used a heterogenic method instead of using a
homogenic method for Phase II because I wanted to capture a broad range of teachers’
conceptions to triangulate with the survey data.

The largest limitation in analysis for this study was my ability to code the six 30-
minute interviews. First round coding entailed reading the transcripts. Then, I read the
transcripts again and cleaned the data. During the third coding experience, I used the TCoA-
III four conceptions as targets. After reading and rereading for the four construct codes, I read
the interview transcripts to look for emerging themes which were captured through my lens
and their words. So, the space in the middle may account for other themes that I did not see.

Limitations in Generalizability

This data is bound in this setting and can be used by the teachers, administrators, and
other stakeholders in the Namaste School District to develop professional programs, consider
existing policies, and increase discourse about the F&P. Furthermore, this study, placed in a
mid-Atlantic public-school system, utilizes the reader’s and writer’s workshop model
instructional framework, the Units of Study (Calkins, 2015) as a core resource, and the F&P.
The ways in which these and other factors interact with one another within this system, are
particular to this context.

Implications for Future Educational Research

This study revealed participants’ conceptions of the F&P, how participants
operationalized the administration of the F&P, and how participants’ conceptions impacted
their decision making processes before, during, and after test administration. This
examination, illuminated through the complexity of education theory, that although the F&P
is a standardized assessment designed for teacher inquiry, the synergy of and between humans is a factor which I feel, should no longer go ignored.

Teachers, however, are only a portion of the complex system and are not in position to do other than directed by their school district. Thus, research regarding school boards’ conceptions of assessment is an area ripe for study. Furthermore, the reliance of school boards on the recommendations of their curriculum departments, calls for researchers to examine how all parties involved choose the assessments that they do because conceptions of assessments matter on every level. The complexity of education theory provides education researchers the wherewithal to examine systems in holistic terms which is a well-used methodology in science and healthcare because the recognition and acceptance of emergence from a complex system is the evidence that the system is alive and functioning whereas non-emergent systems distinguish (Davis & Sumara, 2008).

To that end, action research within the setting may shed additional light on how school district and teachers’ conceptions of the F&P impact their decision-making processes. An ethnographic, longitudinal study outside of pandemic restrictions may provide depth and more insight into this phenomenon. Teachers are the most important factor in the classroom and upwards of the 20% of the teachers in this setting shared their conceptions of the F&P. More research is needed to not only gather teachers’ conceptions of reading inventories but also how the data collected from inventories is utilized as participants in this setting indicated that they disagreed with how it was being used. Researchers can examine the ramifications of administering and using reading inventories differently than the publishers recommendations. We wouldn’t use a hammer to sand a piece of wood so future research into school boards’ approval processes of reading assessment inventories is ripe for examination.

In life, we use tools to deepen our understanding of text such as sticky-notes, other texts, rereading and annotations. Perhaps it is time to allow children to return to the world of
adults by providing them the same tools that we use to deepen our comprehension. For example, teachers might allow students to have a couple of hours with the same text to try to capture, or better yet, teach them how to capture their own cognitive footprints while they create understanding of information that is and is not relevant to them.

Other studies may allow teachers the freedom to bathe in the cycle of teaching and learning with their students as the Common Core Standards specifically and deliberately did not include directions for teachers to follow in their quest to invite and entice students to try their hand at standards acquisition. Let’s see what happens.

The Common Core Standards, the newest assessment standards, and the ESSA call for students to experience more complex test and engage in fewer and more creative assessment tools which are culturally relevant, thorough, and include the student within and along the process. The F&P does not provide information to the student regarding their performance and it was not designed to do so. Assessment is evolving to identify students’ cognitive maneuvers in situ and the instructional level is that which the teacher utilizes which is why the most successful teachers move students beyond grade level expectations. The systems’ framework needs to breathe enough to provide the synergistic complexities’ emergent behaviors space to come to fruition.

In this study, teachers’ conceptions of the F&P were varied and complex. This study is the beginning of a new line of research regarding how teachers’ conceptions impact their decision-making processes while conducting an informal reading inventory. The impact of teachers’ conceptions is not a new line of research, in fact the body of research is deep; however, looking at how the conceptions are operationalized before, during, and after F&P administration, is an area ripe for research. Examining this phenomenon with different methods, and different theoretical frameworks will build upon this study and enrich the data pool. Due to the pandemic, there were several limitations to this study that other researchers
may want to incorporate in future studies. For example, using the ‘think aloud’ method while
answering survey questions would provide the researcher with a clear view of how the
participant is interpreting the question.

Another data collection technique that offers an interesting perspective is to employ
an action research model. Topics for action research may include longitudinal data collection
with a control group and an intervention group with a focused attention to examining how
teachers use the data collected from the F&P, how other districts use the data from the F&P.
How does it look if you take students to the hard level? How does that change their reading
experiences throughout the year? What if that was paired with initial state exam results or
other data that school districts typically collect such as MAP or DIBELS Next data.

Researchers may want to consider enlisting separate focus groups in their data
collection for teachers and students. What are students’ conceptions of the F&P? The F&P is
a performance assessment so teaching students how to practice taking this test, just as they
practice for the PSSA for weeks out of the year, may increase their comfort level and
wherewithal. Students and teachers at times have different discourse patterns which impact
the interaction, the students’ knowledge of what is expected, and how they respond to oral
questioning (Wixson, 2017).

Considering the use of student selected relevant texts may provide researchers with
important information just as using the F&P in a dynamic way. Teachers reported enjoying the
one-on-one time with students and mentioned how powerful it is.

Dynamic Assessment: Increasing Student Agency

Osberg et al. (2008) reported, “The epistemology of emergence [complexity], based on
Deweyan transactional realism, calls for a switch in focus for curricular thinking [and
assessment, I postulate], away from questions about presentation and representation towards
questions about engagement and response” (p. 213). Representational epistemology, that which
Western society schools are built upon, assumes that knowledge exists in a separate space. Complexity proposes a different ethos for Western schools to consider, one that is more concerned with living it out than representing it. One way to transform, in my opinion, the static-time-consuming, subjective results of the F&P [as reported by participants], is to use a dynamic approach to assessment.

Dynamic Assessment (DA) differs from static assessment (measures what the student has gained) in the sense that the proctor utilizes the experience as a teaching tool (leadersproject.org, 2013). DA finds its beginnings rooted in Lev Vygotsky, Reuven Feuerstein's writings (1978), Budoff, and Brown and Campione. Grounded in Sociocultural theory and Constructivism, DA seeks to develop a more in-depth understanding of the learner. It is grounded in the understanding that learning is socio-culturally inclined, occurs with a more knowledgeable other, and is fluid. Like the complexity of education theory, DA seeks to produce change rather than to measure a stable trait or set of traits by making deliberate use of examiner-examinee relationships rather than seeking to limit them. Also, DA embeds intervention within the assessment, observing the effects of the intervention. Shepard (2000) agreed with Vygotsky that DA recognizes the opportunity to teach a reader what she is ready to learn next, whereas, in traditional assessment scenarios, data is collected sans teaching. Dynamic assessment aligns with Aoki’s (2004) philosophy of competence. For as Pinar explained in the introduction to Aoki’s (2004) book, competence can be defined to mean the understanding created by both, teacher, and student, together (p .7). To concur, Minstrell et al. (2011) created a framework for assessment in instruction which considered what teachings have emerged, rather than on the document that espouses what is to be learned. Based on their study, I posit that teachers may consider using dynamic methods of assessment that assess students over a longer period of time, using a variety of everyday texts.
so that the lived experiences and synergistic happenings are valued, expected, and set within districts’ curriculum guides.

**Diversity is Everywhere**

Another implication of the study of this case is that although bounded within the confines of the Namaste School District, it would seem irresponsible to disregard the district’s nestedness within the educational system at large. To do this, a complexivist considers the broaden factors of play within the system. Thus, as Osberg et al. (2008) reminded us, children’s’ role in society, beginning with the elites, changed dramatically in the sixteenth and seventeenth century. Prior to this time period, children’s worlds were with the adults. In the nineteenth and twentieth century, life changed for most children as they were removed from the world of adults and placed in school. Children were placed in school so that they could be prepared for real life (Osberg, 2008, p. 216).

**The Power of Relevant Text**

Christ et al.’s (2018) study revealed a 16% increase in connections to text when the text had higher relevance scores deemed by the participants (p. 124). Thus, as current research shows learning occurs in situ and reading comprehension is higher when reading texts deemed as relevant to self, the time may be ripe to not only teach with globally-relevant texts of multi-difficulty, but to observe students’ abilities when they are working with text deemed to be relevant to self. The complexity of education theory espouses how on the edge of chaos lies the beginning of emergent ideas.

**Fountas & Pinnell’s Systems Approach**

Thus, although education is slow to change, there is compelling evidence and enough of it, for educationists, policy developers, school boards, teachers, and students to audit current practices and consider a renewal approach to cleanse and check on the micro-systems within the complex system. Fountas and Pinnell (2018) offered a systems approach to high
literacy outcomes for all children. Their approach includes four elements: (a) a shared vision and set of core values, (b) common goals, common language, and collective responsibility, (c) a high level of teacher expertise, and (d) a culture of continuous professional learning (pp. 7-19). Evidence within this study validates their systems approach and aligns with the theoretical foundations of the reader’s and writer’s workshop model.

**Summary**

Furthermore, readers get bound up by their own interests and it is unfair, according to Hunsberger (1995), to assess a student’s knowledge of a story immediately after they have read the story one time. In addition, readers follow the lead of the author “trustingly and blindly,” so the ability to reread affords the reader a chance to read holistically with control of the text that is not possible on a first read. The older work of Hunsberger (1995) is still relevant today as Shanahan (2020) reminds us, “We want to promote practices and policies that will have the greatest possibility of ensuring equity and excellence in reading” (p. 2).

Not only do educationists want to promote equity and excellence for students, so do the results of this study remind us how important the conceptions of the most important factor of a students’ success in school is the teacher. In support of teachers’ professional growth opportunities, we are reminded by Li (2017):

Societal values, school culture, access to professional development, and interactions with colleagues play a significant role in teacher identity formation and reformation. The interactional work in which teachers engage in their professional contexts display their thinking, actions, and their individual and collective images of being teachers.

(p. 194)

To increase teacher and school board agency, school districts can utilize available surveys and frameworks to assess current assessment programming. For example, Duke University and Achieve partnered to create a process for stakeholders to utilize to make it
easier to move forward with self-examination of school district assessment systems. While the success of this process rests on several factors, one of the most critical is the extent to which district leaders intentionally incorporate the perspectives, experiences and expertise of teachers and other educators across schools. Teachers are the primary administrators, interpreters and users of assessments, and their front-line perspectives are essential for “taking the temperature” on the assessment environment in the district and building the case for action.

Future research could consider examining how teachers’ voices included in school district assessment audits impact the process, and changes to current praxis. In addition, the role of the teacher in assessment decisions and in general as educational leaders in the U.S. needs to be reexamined. As teachers around the world are called to shift into more creative pedagogy and assessment methods, understanding teachers’ conceptions and incorporating their voices in a voluminous way, is an exciting turn for educators (Patston et al., 2017). For Vandayer & Killen (2007) reported “educators cannot use assessment strategies that they do not understand or for which they lack skills, and the effective use of any strategy will be limited by the educators' ability to think about and control what they are doing”.

This study provided answers to the research questions and confirmed that teachers’ conceptions of the F&P do inform the operational decisions made before, during, and after administration. Assessing students’ reading progress dynamically, with student chosen relevant text, within the teaching and learning cycle, begins, in a small way, to diminish the power structures of school and shows respect for students’ as they are. For, as the demographic structure of schools continues to include more students for whom English is not their primary native language, and the debunking of the instructional reading level, fairer assessment techniques may be considered. Assessing reading comprehension is complex and
at times elusive for example Clark, (2019) proclaimed that one of the most utilized techniques, retelling, goes beyond the scope of reading comprehension.

To promote the authentic transfer and knowledge acquisition techniques that humans use in a natural setting, educationists should break away from past practices to more forward thinking methods where the focus of assessment is geared toward deciphering how students learn and leaning in in the moment to inject useful information for the learner.

Moreover, the results of this study extend Kontovourki’s (2012) ethnographic study: reading leveled books in assessment saturated classrooms; a close examination of unmarked processes of assessment. Kontovourki reported, “No assessment is no-stakes and that even unmarked processes of assessment actually have stakes for readers” (p. 170). To concur, O’Reilly et al. (2014) noted that consequences exist on the inferences drawn from assessment scores. Not only do consequences exist on the inferences drawn from assessment scores, inferences and conceptions of benchmark reading assessments impact the results to begin with.

Further research needs to be conducted regarding stakeholders conceptions of benchmark reading assessments with particular care and awareness as to the purpose and consequences of how results are used. As Hoffman (2017) declared, “Literacy is a tool that is useful in solving the challenges we encounter. Literacy is not the challenge or the goal: learning is the goal” (p. 267).
References


Achieve (2012). Feedback loops for common core standards implementation:

Survey tools for state education leaders.

www.achieve.org/files/GuidanceforsurveysFINAL6-25-12-TOSHARev2.pdf


American Library Association (2007). "Interpretations of the Library Bill of Rights”.

http://www.ala.org/advocacy/intfreedom/librarybill/interpretations Document ID:

066677f2-3938-cbb4-7dba-2edff70d279b


García, O., & Traugh, C. (2002). Using descriptive inquiry to transform the education of linguistically diverse U.S. teachers and students. In L. Wei, J. Dewaele & A. Housen (Eds.), *Opportunities and Challenges of (Societal) Bilingualism* (pp. 311-328). Walter de Gruyter.


Haggis, T. (2008). ‘Knowledge must be contextual’: Some possible implications of
complexity and dynamic systems theories for educational research. *Educational Philosophy & Theory, 40*(1), 158-176.


Hicks, M. (2013). Interim assessments – What they are and how to use them to benefit student learning. NWEA. https://www.nwea.org/blog/2013/interim-assessments-use-benefit-student-learning/


Publications, Inc.


Solomon, R. C.


Minitab (February, 2015). Choosing between a nonparametric test and a parametric test. Blogminitab.com


analysis in convergent mixed methods studies. *Methodological Innovations, 10.*


Talbott, E. (Eds.), *The handbook of diversity in special education* (pp. 399-422).

Hoboken, NJ: John Wiley & Sons Inc.


Yazan, B. (2015). Three approaches to case study methods in education: Yin, Merriam, and 
Stake. *The Qualitative Report, 20*(2), 134-152.

https://www.nova.edu/ssss/QR/QR20/2/yazan1.pdf

Yin, R. (2013). How to do better case studies: (with illustrations from 20 exemplary case 
studies). In Bickman, L. & Rog, D. (Eds.), *the SAGE handbook of applied social 
research methods (pp. 254-282)*. SAGE

Publications. http://dx.doi.org/10.4135/9781483348858

Young, V. M., & Kim, D. H. (2010). Using assessments for instructional improvement: A 
literature review. *Education Policy Analysis Archives, 18*(19),

52. https://epaa.asu.edu/ojs/article/view/809

of nonconsciously priming emotion concepts on behavior. *Journal of Personality and 
Social Psychology, 93*(6), 927-939.
Hello Marie

There are 27 items in the TCoA-IIIA, you have 22 people. This violates the expectation of at least 5 cases per variable.

So you are right that standard EFA and CFA techniques are not going to work

At best you might be able to aggregate all the items for irrelevance and improvement into a single factor each and recover that with EFA/CFA

I have never used the partial least squares algorithm but it might work


But here’s what I would do instead

1. Conduct scale reliability estimation for each of the 3 item scales
2. Then repeat using the aggregated improvement and irrelevant scales
3. If these approach defensible (preferably using McDonald’s omega instead of Cronbach’s alpha) then calculate scores. You can calculate this in R or using jamovi/jasp—stupid SPSS still doesn’t do it.

- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach’s alpha. *Psychometrika, 74*(1), 107-120. doi:10.1007/S11336-008-9101-0

And just don’t do EFA/CFA

Good luck

Prof. Gavin T L Brown, PhD
Associate Dean Postgraduate Research (ADPG)
Director Quantitative Data Analysis and Research Unit
Faculty of Education & Social Work
The University of Auckland/ Waipapa Taumata Rau
Tel: +64 9 3737599 ext. 48602
I am working virtually and if I send you an email outside your reasonable work hours, please note that I am doing so because it works for me in managing my work/life balance. I do not expect that you will read, respond or take action outside of your normal working hours.

From: Derby, Marie N <MD686656@wcupa.edu>
Sent: Saturday, 5 December 2020 9:19 AM
To: Gavin Brown <gt.brown@auckland.ac.nz>
Subject: TCoA-lll abridged version

Hello Dr. Brown,

We have corresponded in the recent past about your TCoA survey. I asked 138 teachers to participate in my study. Twenty-two participants completed the survey in Qualtrics. I am now using SPSS to analyze the data. It is my understanding that this sample size is too small to run it through an SEM software. Do you know of a software program that mimics SEM for smaller sample sizes? I was really hoping to utilize the survey the way it was intended.

Thank you,

Marie Derby doctoral candidate, West Chester University, Pennsylvania, US
Appendix B

Institutional Review Board Certificate

TO: Marie Derby & Heather Schugar
FROM: Nicole M. Cattano, Ph.D.
Co-Chair, WCU Institutional Review Board (IRB)
DATE: 8/5/2020

Project Title: Teachers' Conceptions of Assessment: How K-5 Teachers Approach Benchmark Reading Assessments: A Mixed-Methods Study
Date of Approval: 8/5/2020

[Expeditied Approval]

This protocol has been approved under the new updated 45 CFR 46 common rule that went into effect January 21, 2019. As a result, this project will not require continuing review. Any revisions to this protocol that are needed will require approval by the WCU IRB. Upon completion of the project, you are expected to submit appropriate closure documentation. Please see www.wcupa.edu/research/irb.aspx for more information.

Any adverse reaction by a research subject is to be reported immediately through the Office of Research and Sponsored Programs via email at irb@wcupa.edu.

Signature:

Co-Chair of WCU IRB

Protocol ID #: 20200805A
This Protocol ID number must be used in all communications about this project with the IRB.

WCU Institutional Review Board (IRB)
IORG#: IORG0004242
IRB#: IRB00005030
FWA#: FWA00014155
Appendix C

Participation Invitation

Title: Teachers’ Conceptions of Assessment: How K-5 Teachers Approach Benchmark Reading Assessments: A Mixed-Methods Study

Investigator(s): Marie Derby MEd; Dr. Heather Schugar

Project Overview:

Participation in this research project is voluntary and is being done by Marie Derby as part of her Doctoral Dissertation. Your identity and participation in the study are completely confidential. This project aims to understand the complexities of how K-5 teachers integrate multiple factors while administering benchmark reading inventories. Your participation will take about 20-60 minutes and can be completed at any time from any location. Your participation will contribute to this exciting, relevant topic. Educational research is the cornerstone of which local, state, and federal policies are built upon. If you would like to take part, West Chester University requires that you agree and sign this consent form.

You may ask Marie Derby or Dr. Heather Schugar, Marie’s advisor, and committee chairperson any questions to help you understand this study. If you do not want to be a part of this study, it will not affect any services from West Chester University. If you choose to be a part of this study, you have the right to change your mind and stop being a part of the study at any time.

1. What is the purpose of this study?
   - This project aims to understand the conceptions of assessment of K-5 teachers and how those conceptions are operationalized during the administration of benchmark reading assessments.

2. If you decide to be a part of this study, you will be asked to do the following:
   - complete a survey Teachers’ Conceptions of Assessment abridged version (Brown, 2006)
   - participate in an interview
   - This study will take about 20-60 minutes of your time.

3. Are there any experimental medical treatments?
   - No

4. Is there any risk to me?
   - Minimal risks include loss of time, and, unlikely, but possible discomfort if you participate in an interview.

5. Is there any benefit to me?
   - This study is not intended to directly benefit you.
   - Findings will be shared with participants
   - Dedicated time devoted to your personal conceptions of assessment
6. **How will you protect my privacy?**
   - Password Protected File/Computer
     - Identifying details are kept confidential and your participation can be kept anonymous. The information will be protected on my personal password-protected laptop. Names will be replaced with numbers for organizational purposes.
     - Your records will be private. Only I, Dr. Heather Schugar, and the IRB will have access to your name and responses.
     - Your name will **not** be used in any reports.
     - To maintain social distancing, interviews will be conducted via the internet using Zoom. The session will be recorded, housed on my personal, password-protected laptop, and discarded once the study has concluded.
     - Interviewees will be able to participate in a place and space that works best for them.
     - Records will be destroyed after manuscript development, but no less than three years after study completion.

7. **Do I get paid to take part in this study?**
   - No

8. **Whom do I contact in case of research-related injury?**
   - For any questions with this study, contact:
     - **Primary Investigator:** Marie Derby MEd at 610-256-1671 or md686656@wcupa.edu
     - **Faculty Sponsor:** Dr. Heather Schugar at 443-812-4489 or hschugar@wcupa.edu

9. **What will you do with my Identifiable Information/Biospecimens?**
   - Not applicable.

For any questions about your rights in this research study, contact the ORSP at 610-436-3557.

I, _________________________________ (your name), have read this form and I understand the statements in this form. I know that if I am uncomfortable with this study, I can stop at any time. I know that it is not possible to know all possible risks in a study, and I think that reasonable safety measures have been taken to decrease any risk.

______________________________

Subject/Participant Signature    Date:______________

______________________________

Witness Signature               Date:______________
Appendix D

Participation Invitation Letter

Dear Invitee,

My name is Marie Derby. I am a doctoral student at West Chester University’s Policy, Planning, and Administration program. I am kindly requesting your participation in a doctoral research study that I am conducting titled: Teachers’ Conceptions of Assessment: How K-5 Teachers Approach Benchmark Reading Assessments: A Mixed-Methods Study. The aim is to examine the relationship between teachers’ conceptions of assessment and their approach to administering the Fountas & Pinnell Benchmark Reading Assessment. The study involves completing Gavin Brown’s online survey tool called Teachers’ Conceptions of Assessment, which takes 20 minutes followed by a few open-ended questions, and participate in a Zoom interview with me from the comfort of your home or classroom which takes about 45 minutes. Participation is completely voluntary, and you may withdraw from the study at any time. The study is completely anonymous; therefore, it does not require you to provide your name or any other identifying information. If you would like to participate in the study, please read the Informed Consent letter below. To begin the study, click the survey link at the end. Your participation in education research is of great importance. As a collective, your anonymous voices will contribute to the body of research that is collecting and examining teachers’ conceptions of assessment. Your voices impact policy and practice. This study has been approved by the West Chester University Institutional Review Board Protocol #20200805A.

Thank you for your time and participation.

Warmly,

Marie Derby MEd, Doctoral Student, West Chester University
Appendix E

Conceptions of Assessment Survey Tool

Conceptions of Assessment III Abridged Survey

This survey asks about your beliefs and understandings about the F&P ASSESSMENT, whatever that term means to you. Please answer the questions using YOUR understanding of assessment while you are administering benchmark reading assessments.

1. Please indicate which of the following assessment PRACTICES you have in mind when you think about assessment.

Please give your rating for each of the following 27 statements based on YOUR opinion about assessment. Indicate how much you agree or disagree with each statement. Use the following rating scale and choose the one response that comes closest to describing your opinion.

➢ Strongly Disagree
➢ Mostly Disagree
➢ Slightly Agree
➢ Moderately Agree
➢ Mostly Agree
➢ Strongly Agree

Note that the ratings are ordered from Disagree on the LEFT to Agree on the RIGHT.
<table>
<thead>
<tr>
<th>Conceptions of Assessment</th>
<th>Strongly Disagree</th>
<th>Mostly Disagree</th>
<th>Slightly Agree</th>
<th>Moderately Agree</th>
<th>Mostly Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>The F&amp;P provides information on how well schools are doing.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. The F&amp;P places students into categories.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. The F&amp;P is a way to determine how much students have learned from teaching.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. The F&amp;P provides feedback to students about their performance.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The F&amp;P is integrated with teaching practice.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. The F&amp;P results are trustworthy.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. The F&amp;P forces teachers to teach in a way against their beliefs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. I conduct the F&amp;P but make little use of the results.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
7. The F&P results should be treated cautiously because of measurement error.

8. The F&P is an accurate indicator of a school’s quality.

9. The F&P is assigning a grade or level to student work.
   The F&P establishes what students have learned.

10. The F&P feeds back to students their learning needs.

11. The F&P information modifies ongoing teaching of students.

12. The F&P results are consistent.

13. The F&P is unfair to students.

14. The F&P results are filed & ignored.

15. Teachers should take into account the error and imprecision in the F&P.

16. The F&P is a good way to evaluate a school.
<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>17.</td>
<td>The F&amp;P determines if students meet qualifications standards.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18.</td>
<td>The F&amp;P measures students’ higher-order thinking skills.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19.</td>
<td>The F&amp;P helps students improve their learning.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20.</td>
<td>The F&amp;P allows different students to get different instruction.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21.</td>
<td>The F&amp;P results can be depended on.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22.</td>
<td>The F&amp;P interferes with teaching</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23.</td>
<td>The F&amp;P has little impact on teaching.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24.</td>
<td>The F&amp;P is an imprecise process.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Would you also provide the following personal information?

A) What is your highest degree?

☐ Bachelor  
☐ Postgraduate Certificate  
☐ Master’s Degree  
☐ Doctoral Degree

B) Do you hold a literacy or reading specialist certification?

☐ yes  
☐ no

B) For how many years have you been teaching?

☐ Less than 2  
☐ Between 2 and 5  
☐ Between 6 and 10  
☐ More than 10

D) Most of your teaching experience

☐ K-3  
☐ 4-5

C) What professional development have you had?

☐ None  
☐ Some hours as part of pre-service teaching  
☐ ½ to 1-day Workshop or Seminar  
☐ Completed undergraduate Paper  
☐ Completed postgraduate Paper  
☐ Other: (give details) ______________________________
Appendix F

Semi-Structured Interview Protocol

<table>
<thead>
<tr>
<th>Research Question:</th>
</tr>
</thead>
<tbody>
<tr>
<td>How do teachers’ conceptions of assessment inform their administration of</td>
</tr>
</tbody>
</table>

Participant Code:
Date:
Start time: End time:
Participant’s location: Researcher’s location:

1. Tell me about your hobbies and things you like to do outside of school?

2. At what point in your life did you decide to become a teacher?
   a. Do you have other teachers in your family?

3. Tell me about your college and teacher prep experience.
   a. What advice do you have for upcoming teachers?
   b. How can teacher prep programs prepare future teachers to handle the demands of assessing students’ reading levels.

4. To get to know you a bit better, tell me how your childhood experiences in school regarding reading assessments were.

5. Moving to you, now, as a teacher in Namaste School District, tell me about the process of administering the Fountas & Pinnell [BAS].
   a. What is the benefit to the student in and after that experience?
   b. How does the experience of administering the F & P [BAS] benefit you, the teacher?

6. Considering what you told me about the benefits of administering the F & P [BAS]....
a. How is the F & P [BAS] an effective method of identifying students’ instructional reading levels?

b. How is the data that you collect during this process utilized within the classroom?

7. Explain if there are any ethical considerations you have in using the F & P [BAS] to identify students’ instructional reading level.

8. Now our focus turns to bias, beliefs, and worldviews. Explain how your bias impacts the administration of the F & P [BAS].
   a. How does your belief system impact the experience of administering the F & P?
   b. How do your worldviews impact the experience of administering the F & P?
   c. How do your conceptions of assessment impact the administration of the F & P?

9. Should local policy change, in your opinion, regarding the administration of the F & P [BAS]?

10. Is there anything that you would like to add to our discussion?

I want to thank you for participating in my study. Remember, you can stop participation at any time. I will keep our interview in a secured location and your identity will be anonymous. Are you willing to meet with me again to review the information that I gathered from our interview? May I, at that time, if needed, ask you a few more clarifying questions? I will transcribe this interview word for word and look for themes while I am coding the data that you have provided for me. I appreciate your willingness to spend time on this topic. Do you have any questions for me? If this process has caused you discomfort or if you have questions about this process, please feel free to reach out to Dr. Heather Schugar at HSchugar@wcupa.edu. She is my advisor on this project.