The Effect of Student Performance Data on Special Education Instruction

A thesis presented
by

Paul J. Blass, Jr.

to
the School of Professional Studies

In partial fulfillment of the requirements for the degree of
Doctor of Education

In the field of
Education

College of Professional Studies
Northeastern University
Boston, Massachusetts
June 2015
Abstract

The state of New Jersey implemented a new teacher evaluation system that requires student data to be used as one measurement of a teacher’s performance. Up to 30% of a teacher’s annual evaluation score will be derived from quantitative data. One source of data is the student growth objective (SGO) by which teachers develop an assessment to measure their students’ progress. For this phenomenological study the researcher conducted interviews that sought to gain insight into the experiences of four special education teachers as they fulfill the requirement to supply data for their annual evaluation. Because the study involved special education teachers, the researcher utilized a Disabilities Studies in Education (DSE) framework. Disability studies supports the understanding of disabilities within broader social and political circumstances. The primary research question that guided this study was: What are the experiences of middle school special education teachers at a small suburban school who are newly mandated to use standardized student data to inform their pedagogical and curricular approaches? The findings from the study indicate that these teachers did not feel that the SGOs, as implemented, were appropriate for special education students, but the additional focus on data caused the teachers to pay closer attention to state curricular standards.

*Keywords*: Student Growth Percentiles, Student Growth Objectives, Special Education
Acknowledgements

A doctoral thesis, as well as doctoral studies in general, take the proverbial village to help with their completion. With all due respect to the professors of Northeastern University, this paper would not have been possible without the understanding and patience of my wife Laura and son Ryan. Even though she never fully understood why I wanted to complete a doctoral program, my wife supported and encouraged me throughout the process. I would never have finished without her pushing me. I also realize that I was able to devote the time I needed to research and writing because my wife took up the slack for me in other areas. She carried a greater burden so I could accomplish this. I’m fortunate that I never missed any big family events while tending to school work, but these studies did cause me to miss some typical family time, such as sitting on the couch watching television. Though he pointed out such absences, I know my son recognized how important these studies were to me and encouraged me in his own way. My wife and son were often impromptu editors as I needed to check if something “sounded right,” but my “go to” editor was my wife’s mom, Lorraine Natali, who was always happy to review a chapter.

Special thanks goes to Dr. Sara Ewell, my advisor, whose feedback helped guide me into creating a study that not only meant something to me but also had practical applications to my own school district. Dr. Chris Unger provided insights to help make the phemenological study stronger. Dr. Cherie Lombardo, my outside reader, is a building administrator who helped shape chapter five, particularly the recommendations portion. Lastly, I must thank classmate Susan Vatalaro, who was my sounding board throughout most of the program.
## Table of Contents

Abstract .......................................................................................................................... 3

Acknowledgements ......................................................................................................... 4

List of Figures .................................................................................................................. 6

Chapter 1: Introduction .................................................................................................... 7
  Research Problem ........................................................................................................... 7
  Background of Special Education Legislation and Regulation .................................... 8
  Purpose .......................................................................................................................... 11
  Significance of the Problem ......................................................................................... 11
  Positionality Statement ................................................................................................. 12
    Testing bias. ................................................................................................................ 13
    Special education and testing. .................................................................................... 14
    Study participants. ...................................................................................................... 15
    Conclusion. ................................................................................................................ 15
  Research Question ........................................................................................................ 16
  Theoretical Framework .................................................................................................. 16

Chapter 2: Literature Review ............................................................................................ 18
  Organizational Viewpoints ............................................................................................ 19
  Data and Teacher Evaluations ...................................................................................... 20
  Value Added Models .................................................................................................... 23
  Student Growth Percentiles (SGPs) ............................................................................. 26
  Student Growth Objectives (SGOs) ............................................................................. 28
  Utilizing Models ........................................................................................................... 29
  Using Data ................................................................................................................... 32
  Special Education and Data .......................................................................................... 33
  Special Education and SGOs/SGPs .............................................................................. 34
  Conclusion .................................................................................................................... 34

Chapter 3: Research Methods ........................................................................................... 36
  Purpose ........................................................................................................................ 36
  Central Research Question ........................................................................................... 36
  Methodology ............................................................................................................... 37
  Questions and Methodology ......................................................................................... 40
  Researcher Bias ........................................................................................................... 41
  Study Participants ....................................................................................................... 41
  Sampling ....................................................................................................................... 42
  Data Collection ............................................................................................................ 43
    Data storage .............................................................................................................. 44
  Data Analysis Process .................................................................................................. 44
  IPA and Validity .......................................................................................................... 46
  Trustworthiness ........................................................................................................... 47
Chapter 4: Findings .................................................................................................................. 51
  Participant Profiles ............................................................................................................. 51
    Lee ................................................................................................................................... 51
    Alex ................................................................................................................................. 52
    Murphy ............................................................................................................................ 52
    Robin ................................................................................................................................. 52
  Review of the Purpose and Research Questions ................................................................ 53
  Theme One: Confusion and Anxiety Over the Development Process .............................. 54
    Minimum score ............................................................................................................... 58
    Attention to standards ..................................................................................................... 59
  Theme Two: Misgivings Over the Process With Special Needs Students ...................... 61
    Minimum score ............................................................................................................... 63
  Theme Three: Acceptance ............................................................................................... 64
  Conclusion .......................................................................................................................... 67

Chapter 5: Conclusions, Implications, and Recommendations ........................................... 68
  Special Education and Student Growth Percentiles and Objectives ............................... 70
  A “Numbers Game” ........................................................................................................... 73
  Focusing Standards ......................................................................................................... 76
  Conclusions ....................................................................................................................... 77
    Framework ....................................................................................................................... 77
  Implications for Practice ................................................................................................. 79
  Research Implications ..................................................................................................... 80
  Limitations of the Study ................................................................................................... 81
  Application of Results ...................................................................................................... 82
    1. The district should provide more time to develop appropriate SGOs ...................... 82
    2. Investigate mandatory minimums ............................................................................. 82
    3. The district should provide better explanation of how SGOs are developed .......... 83

References ............................................................................................................................ 84

Appendix A: Recruitment Letter ......................................................................................... 92

Appendix B: Consent Form ................................................................................................. 93

Appendix C: Interview Protocol ........................................................................................ 95

List of Figures

Figure 1. Interaction between teachers, data, instruction, and state mandates ................. 39
Chapter 1: Introduction

Research Problem

The New Jersey State Board of Education established that 70-80% of a teacher’s evaluation score is based on administrative observations, while the remainder is based on student testing (New Jersey Department of Education, 2014). Baker et al. (2010) indicated that no model for data collection, however, can be expected to definitively determine the extent by which an individual teacher contributed to a students’ academic growth. Despite this, the state department of education uses such measurements to try and improve student achievement. How teachers and administrators utilize the collected data impact its usefulness and must be investigated.

The need to comply with Race to the Top (RTTT) achievement mandates has prompted districts to measure their students’ success in a more comprehensive manner. RTTT is a federal program that provides competitive grants to states that promote educational reforms including performance-based teacher evaluations (United States Department of Education, 2014). The state established certain parameters as to what data could be employed. In areas currently tested by the state – mathematics and English – a growth model must be used (New Jersey Department of Education, 2014). Growth models evaluate student achievement over time and predict the future achievement of the student. Betebenner (2011) argued that using growth models to determine if a student is progressing towards proficiency is limiting. He cautioned that student growth percentiles (SGPs) are descriptive and should not be used to associate causation. SGPs account for student differences by burying those differences amongst the entire sample. Haertel (2009) noted that it is sometimes assumed that students with similar backgrounds (i.e., receiving free lunch) will earn similar scores and that growth in learning would therefore be normatively comparable. Haertel (2009) further indicated that SGP models presuppose that students with the
same initial test score have the same aptitude for learning, as well as the same out-of-school supports for learning. Since this is not the case, Haertel wondered if the controls for prior achievement in SGPs were insufficient and questioned whether students should be expected to grow at the same pace. Goldhaber, Walch, Gabele, and Schools (2012) concluded that SGPs are good for reviewing student achievement, but they should not be used for teacher performance. Yet, this is what the New Jersey Department of Education proposed. Because of the state’s position, teachers and administrators must understand how data affects classroom instruction.

**Background of Special Education Legislation and Regulation**

Students with overt disabilities (e.g., visual impairment, mental impairments) have a longer history of formal education in the United States than those with learning disabilities. The first free school for the deaf was established in Hartford, CT in 1817. For the first 44 years of the country this was the only institution for those with disabilities (Langworthy, 1990). By 1832, Samuel Gridley Howe helped establish the Perkins School for the Blind, the first school of its kind in America (Perkins, n.d.). Students with mental impairments had their own schools by the end of the 19th century. For example, in 1888, the New Jersey Home for the Education and Care of Feebleminded Children, now the Vineland Training School, opened and was the first of its kind (Vineland, n.d.). Separate facilities were the norm for those with disabilities. That would change as compulsory education laws greatly increased the number of students in public schools, including those with disabilities. Schools had the challenge of addressing the needs of a larger and more diverse student population. By the 1930s there were 164,000 students with disabilities in public schools (Langworthy, 1990). The education of these children, particularly those with more severe disabilities, was limited. It took judicial rulings and legislative mandates to improve the opportunities for these students.
In the landmark case *Brown v. Board of Education of Topeka* (1954), the United States Supreme Court held that “separate facilities are inherently unequal.” While the case specifically applied to schools segregated by race, its interpretation has since grown to include segregation of all kinds. At the heart of the *Brown* decision is the Fourteenth Amendment’s guarantee of equal protection. Thus, if states have laws providing education to their citizenry, then they must provide it for all of their citizens, including students with special needs. Congress recognized the unique nature of children with special needs by adding Title VI to the Elementary and Secondary Education Act of 1965, which created the Bureau of Education for the Handicapped (now called the Office of Special Education Programs). While it did not mandate the education of students with disabilities, it signaled the federal government’s interest in that goal.

In 1975 Congress passed legislation that greatly increased the role of the federal government in special education, P.L. 94-142, *The Education for all Handicapped Children Act (EAHCA)*. The act required states to provide ‘a free and appropriate public education’ for all students with disabilities between the ages of 3 and 21. Students were guaranteed an education in the least restrictive environment (LRE); that is, as close as possible to their non-disabled peers. Zigmond, Kloo, and Volonino (2009) described this as the most “normal” education setting as possible (p. 190). Each student was to have their own individualized education program (IEP) developed through a nondiscriminatory evaluation (Yell, 1998).

In 1990, the EAHCA was amended, and the name changed to the *Individuals with Disabilities Education Act (IDEA)*. This iteration added students with autism and traumatic brain injury as separate and distinct groups. IDEA also required that by age 16 IEP teams establish plans for helping transition students out of the school environment (Yell, 1998). In 1997, IDEA was reauthorized. One change linked participation in state tests to the goal of
increasing student participation in the general education curriculum. This led to the requirement that students with special needs be tested as their peers or with an alternate assessment (Defur, 2002). The importance of IDEA in helping to provide education to those with disabilities cannot be understated.

The most recent legislation to impact special education includes 2001’s Elementary and Secondary Education Act (ESEA), more commonly known as No Child Left Behind (NCLB), and 2004’s revision of the Individuals With Disabilities Education Improvement Act (IDEA). NCLB is part of the trend toward universal standards and accountability, setting the laudable goal that all children be proficient in reading and math by 2014. The alignment of IDEA and NCLB mandated that students with disabilities be included in state assessments of academic achievement. This change was meant to address the persistent problem of educational equity (Eiseman & Ferretti, 2010) and to assure that students with disabilities attained the benefits intended for all students (Wakeman, Browder, Meier, and McColl, 2007). However, McLaughlin (2010) and others feel that students with disabilities are being treated unjustly when held to universal standards.

While research on the use of student performance data is extensive (e.g., Betebenner, 2011; Haertel, 2009; Herman, Yamashiro, Lefkowitz, & Trusela, 2008; Hess & Fullerton, 2010; Pomplun, 2009), little has been dedicated specifically to the field of special education (Holdheide, Browder, Warren, Buzick, & Jones, 2012). Understanding whether a more systematic gathering of data as part of the new observation process mandated by the state of New Jersey can contribute to student success will help teachers with their delivery of instruction.
Purpose

The purpose of this study was to understand the relationship between student assessment and the teaching practices of special education teachers during the implementation of a new evaluation system in New Jersey. Specifically, it examined whether or not special education teachers in grades 6-8 utilize the student performance data that they must now collect under state mandates to change their instruction.

The use of student performance data has been increasingly advocated, particularly since the passage of No Child Left Behind in 2001. As new research helps perfect its application, data will become part of a teacher’s repertoire to improve classroom lessons. School personnel must be trained in the collection and examination of data use to fully understand its meanings and possibilities. This examination of how testing data is used could determine if data collection increases the achievement of special needs students and can clarify assessment data use in teacher evaluations. This study benefits teachers in their instruction and administrators in their evaluations. School boards, the ultimate decision maker for teacher retention, could find the research valuable when judging teacher effectiveness. Similarly, teachers’ unions who advocate for their members, could find the research beneficial when defending teachers under review.

Significance of the Problem

The changes prompted by the Race to the Top (RTTT) legislation are part of a larger accountability movement that began after the publication of A Nation at Risk in 1983. RTTT clearly places accountability for the quality of student performance onto teachers. Though research has shown that associating causation with test scores is difficult to measure or justify (Campbell, Kyriekides, Muijs, & Robinson, 2003; Baker et al., 2010) that is what this legislation requires.
In order to receive federal funds as part of RTTT, the state of New Jersey agreed to change how teachers were evaluated (Califati, 2011). The state adopted new procedures for classroom observations as well as overall performance reviews. The summative evaluations must include student performance data. Up to 30% of a teacher’s score is derived from data. The data can be from standardized tests, teacher-made tests, or student portfolios (New Jersey Department of Education, 2014).

To fulfil the state’s mandate administrators and teachers must understand how data is utilized. Mandating the collection of data does not guarantee the application of findings to change teacher performance, though that seems to be the intent of the legislation. The state’s program to reach RTTT goals, titled AchieveNJ for Teachers, was formulated to recognize teachers who are excelling, to identify those who need additional support, and to provide meaningful feedback and professional development to all teachers (New Jersey Department of Education, 2014). If the results indicate that teachers or administrators are not implementing AchieveNJ as intended, the district may need to better inform its staff as to how to properly employ the use of data to meet the intended expectations.

**Positionality Statement**

I have been working with high school special education students for 20 years. I am certified as a social studies teacher and a special education teacher. My teaching certifications have afforded me the opportunity to teach a wide variety of subjects to a diverse student population. I have taught in the capacity of a self-contained resource room teacher, an in-class support teacher, as well as a regular education history teacher. Throughout my career I have seen special education students being held more academically accountable. The expectation is that most of this accountability be measured through standardized tests. Many special education
students do not respond well to the testing environment, with many viewing the entire testing experience negatively. Many students who have approached tests optimistically tire after several hours. They may start filling in answers just to get done. Such experiences have caused me to question high-stakes standardized tests, especially involving students with disabilities. There are several areas where my experiences impact the lens through which I view this topic, and therefore this study.

**Testing bias.**

It is not unusual to hear experienced teachers lament about new requirements mandated by the state or federal governments. Changes resulting from RTTT are no exception. Using student data for high-stakes reasons such as teacher retention does not seem right to me as an individual working within these policies. In efforts to remain impartial throughout this study, I can state that I do not like this mandate. I believe this approach attributes the cause of a poor education too narrowly on teachers while ignoring other causes. It is interesting to note that while some policy makers want student test results used to evaluate teachers, Taylor and Tyler (2011) use the reverse approach. They demonstrate that test scores are shown to improve because of rigorous teacher observation. It would seem the traditional methods of teacher evaluation do not need extensive overhauling but rather a revisiting of current methodologies.

In order to implement a high-stakes policy based on student test results, Taylor and Tyler (2011) wrote that districts must first implement observation/evaluation systems that include the supports necessary to improve teachers. I believe this viewpoint takes a more nurturing approach to teachers rather than a punitive approach whereby standardized tests are used to find the ‘bad’ teachers. Throughout the research process, I curbed my feeling regarding state tests to not bias my results. I see value in teachers using data to help guide what they do in the
classroom, but I am concerned with the high-stakes use of data. I do not want to see teachers defending their job because scores did not improve at the predetermined rate.

Damien Betebenner (2011) introduced student growth percentiles (SGP) as a “normative conceptualization of student growth” (p. 44). The changes in student performance are compared to a normative group to determine how typical the change is. He cautioned, however, that SGPs are for descriptive purposes and should not be used to associate causation. Personally, I question why, when the developer of SGPs stated that they should not be used for causation, the state chose to do so. I have severe reservations about using data in such a manner. As a public school employee, I accept that the state, at least for the time being, requires the use of student data as a quality measure. I am hopeful that if data becomes an integral part of the school culture that it be used beneficially and not punitively. Yet, this research does not center on the appropriateness of testing data, but instead seeks to understand its use.

**Special education and testing.**

Special education has come a long way since EAHCA became law. Legal challenges have helped move students with disabilities away from segregated schools and into mainstream settings. The inclusion movement has benefited special education students. Yet, there seems to be an inherent contradiction between the expectations of an individualized education program (IEP) described in special education law and the requirements in NCLB or RTTT that all children be assessed using the same instruments, at least initially, as their peers. Maintaining high standards is important, as is challenging students, but in doing so one wonders if the system is ignoring individuality. With limited research on using student data on students with special needs, a greater understanding of how data can be gathered and beneficially utilized is needed. I
understand that I approach all changes in special education cautiously. I wonder how any change will affect the students both academically and socially.

**Study participants.**

The participants who took part in this qualitative study were teachers within the special education department of a suburban New Jersey school district. The district is relatively small, with approximately 32 special education teachers. The teachers often know each other, working across levels and between buildings because of joint workshops and IEP meetings. The researcher knew almost all of them by sight and several on a personal level. I have worked in the district for 20 years and served for as union president for 3 years, so it would be fair to say that even those teachers who do not know me knew me by reputation. The small, close nature of the district was a concern during the research process. An individual’s answers could make his or her identity known, altering how participants might choose to answer a question. To limit the possibility of identification, the descriptions of participants are brief and intentionally vague.

**Conclusion.**

I recognize that I am critical of those who attack teachers for poor student performance, mainly because that belief ignores other causes, particularly socio-economic factors (e.g. see Schafer et al., 2012). I also believe that this approach has led to high-stakes testing such as that found in RTTT. Administrators, after all, are the ones who hire and observe teachers. Part of their job is to nurture teachers, and if a teacher does not meet the necessary criteria, then replace that person. While I believe there is a time and place to use student data, I do not see its high-stakes use as a solution to schools’ problems, and it is possible that the data could adversely affect good teachers.
Research Question

What are the experiences of middle school special education teachers at a small suburban school who are newly mandated to use standardized student data to inform their pedagogical and curricular approaches? This question is the crux of the research. The state has made data gathering part of the teacher evaluation process, and there is the assumption that the data will help teachers with their instruction. It is imperative to know how teachers view the situation. This study attempted to examine whether they felt the data gathering has proven beneficial or whether they viewed it as a waste of time. At one extreme, all teachers use data to modify their lessons to meet the needs of their students. At the other extreme, teachers obey the law and gather the data but do not use it to change their teaching approach. Understanding the teachers’ attitudes can help district administrators, and by extension state officials, assure that the intention of data gathering is fulfilled.

Theoretical Framework

This research utilized a Disabilities Studies in Education (DSE) framework. This relatively new field originated in the broader area of Disability Studies. The essence of this approach is that the understanding of disability occurs through “human expectations and interactions in social contexts” (Baglieri, Valle, Connor, & Gallagher, 2011, p. 275). Its aim is to expand the “understandings of the daily experiences of people with disabilities” and “work to create and sustain inclusive and accessible schools” (Connor, Gabel, Gallagher, & Morton, 2008, p. 442). Watson (2012) asserted that the “voices” of children with disabilities were often excluded (p. 193). While this research did not directly examine students, it explored how the needs of classified students are met by their teachers under AchieveNJ. DSE can be used to investigate whether a policy constructs an inclusive environment (Tregaskis, 2006). In this case,
the researcher asked teachers if the use of data, such as SGPs and SGOs, was conducive to the special education setting.

By the mid-1990s, research on disability changed from a focus on impairments to a focus on reducing the disadvantages faced by those with disabilities (Watson, 2012). Disability studies evolved from the limiting view that disability was a medical problem to the broader idea that disability evolves from the “complex interaction” between the person with a disability and the “complete physical, human-build, attitudinal and social environment” (Watson, 2012, p. 198). The research is motivated by a desire to change individuals with special needs from “objects of study” to the subject of study and to “present them as active agents” (Watson, 2012, p. 192).

Disability studies scholars often use qualitative research to explore issues related to those with disabilities. Cosier (2012) wrote that a DSE framework should include one of five goals, which she outlined. Of these, this thesis set out to meet Cosier’s goal of “acknowledging that disability is situated in political and social spheres” (p. 82). The students whose SGO/SGP progress is being examined have been classified as needing special education services through a process outlined in the Individuals with Disabilities Education Act. Their classroom placement was determined at their annual IEP meeting. Their teachers must meet goals outlined in the IEPs. It is within these parameters that SGOS and SGPs, a politically imposed requirement, are administered. Disability studies promotes “an understanding of disability within a sociopolitical context” whereby scholars question current educational practices (Cosier, 2012, p. 83). Through this lens, this thesis questioned the impact of SGOS and SGPs within the special education classroom.
Chapter 2: Literature Review

In 1975, Congress passed legislation that greatly increased the role of the federal government in special education, P.L. 94-142, the Education for All Handicapped Children Act (EAHCA). This act required states to provide a free and appropriate public education for all students with disabilities between the ages of 3 and 21. Students were guaranteed an education in the least restrictive environment (LRE); that is, as close as possible to their non-disabled peers or, as described by Zigmond, Kloo, and Volonino (2009), the most “normal” education setting as possible (p. 190). Each student would also have their own individualized education program (IEP) developed through a nondiscriminatory evaluation (Yell, 1998).

In 1990, the EAHCA was amended and the name changed to the Individuals with Disabilities Education Act (IDEA). This act added students with autism and traumatic brain injury as separate and distinct groups. It also required that, by age 16, plans be established to help transition students out of the school environment (Yell, 1998). The policies established helped create a somewhat unique, sometimes separate, educational system. The passage of NCLB helped limit the unique nature of special education by requiring the inclusion of all students in state testing. New Jersey’s changes in teacher evaluation also required that all teachers, including special education teachers, utilize student data as part of their evaluation procedure.

By October 31 of each year, New Jersey teachers are required to develop a student growth objective (SGO) by which they will measure their students’ progress (New Jersey Department of Education, 2014). The intent is for teachers to use the data they gather for their SGO to help determine the needs of their students and adjust their instruction accordingly.
New Jersey is not exceptional in its revamping of teacher evaluations. Motivated by the money available through the federal government’s Race to the Top program, many states rushed to require data components in teacher evaluations to qualify for grant money. The first monies were awarded in April 2010; New Jersey qualified in the third round of funding (United States Department of Education, 2014).

There seems consensus amongst reformers that schools and teachers need to use data, whether for teacher evaluation and/or instructional guidance. Understanding teachers’ familiarity and experience with data will be vital in implementing any data-use requirements.

**Organizational Viewpoints**

The topic of teacher reform is clearly part of the political agenda (White House Profile, 2009). As a result, numerous institutes have issued policy papers regarding the use of testing to measure student and/or teacher performance. Several of these organizations (e.g., the Center for Education Data & Research, the National Center for Analysis of Longitudinal Data in Education Research, and the National Center for the Improvement of Educational Assessment) specifically focus on improving the utilization of educational data and assessment. These institutes suggest that data can be a useful tool and advocate some form of its use.

Organizations not specifically founded for educational purposes have also added to the testing literature. One example is the Economic Policy Institute (EPI), created in 1986 as a non-profit, non-partisan think tank designed to expand the discussions regarding economic policy to include the needs of low and middle-income workers. EPI reviewed studies on the effectiveness of using student test scores to evaluate teachers. They recognize that testing data can be one piece of information used for teacher evaluations, but they conclude that it is unwise to place too heavy of an emphasis on student scores to make judgments about teachers. Test results do not
provide a complete understanding of the contributions teacher do in the classroom that contribute to student learning (Baker et al., 2010).

Institutes with direct educational ties have also provided their perspectives. The Council of Chief State School Officers is a non-political, non-profit organization of public education officials that provides leadership and advocacy on major educational issues. In their study, Goldschmidt, Choi, & Beaudoin (2012) conclude that no single model works better than the rest. They also explain that context matters. Because of testing procedures, student characteristics, and school characteristics, testing results can differ from state to state even when the same model is used due to testing procedures. The lack of consistent results is disconcerting since teachers’ job retention will be partially based on student test results.

**Data and Teacher Evaluations**

There is no clear consensus as to what data should be part of a districts’ repertoire to evaluate teachers. Milanowski (2004) concludes that a rigorous teacher evaluation system can be “substantially related” to student achievement and can provide evidence for performance evaluations, including retention (p. 34). His study utilized the Teacher Evaluation System designed by the Cincinnati Public Schools. The correlation Milanowski (2004) finds is relatively small; 9%-16% of variance in student performance could be connected to variation in teacher performance as measured on the evaluation instrument. His caveat that high correlations are unlikely because of “error in measuring teacher performance, error in measuring student performance, lack of alignment between the curriculum taught by the teachers and the student tests, and the role of student motivation” is revealing (p. 50). Milanowski (2004) demonstrates that even well-planned evaluation instruments must be used with caution. The push for reform may be moving faster than the research on its effectiveness.
Taylor and Tyler (2011) also examined Cincinnati’s Teacher Evaluation System. The authors conclude that rigorous classroom observations, in this case based on the Danielson model, can improve teacher performance. This improvement is then reflected in student test scores, as much as a 10% standard deviation increase in mathematics (p. 29). They speculate that some of the improvement is due to feedback provided during the evaluation. The authors caution that the instrument developed by the Cincinnati School District is particularly involved, and that other districts would need to have systems with a similarly structured feedback.

Yet, districts that experiment and vary evaluation components to meet their needs could have results that measure differences in teacher effectiveness. Tyler, Taylor, Kane, and Wooten (2010) believe that relating classroom practices examined through observations to achievement growth can offer insight as to which classroom practices may be essential in increasing student achievement. Knowing which practices are most successful helps administrators support the growth of teachers in these areas.

Wiliam (2013) believes that assessment for instructional guidance is important to improve schooling, but he contends there is a problem with assessments. He asserts that no single test score does more than reveal whether or not a student has demonstrated an acceptable level of achievement. Test results do not provide guidance. Tests only confirm that action is needed, presumably more and better instruction. Wiliam (2013) argues that the “grain size” of an assessment varies depending on the subject, noting that there would never be enough time to check for understanding of all content (p. 57). Wiliam references Brown’s (1992) study which estimates that 300 assessments would be necessary to cover an average middle school math curriculum. Wiliam (2013) wonders if, in such a scenario, teachers would have the time or instructional strategies to use all the possible data to improve schooling. He advocates a culture
of decision-driven data collection – one where the data are collected “only after a clear theory of how they are to be used has been developed” (p. 58). Such an approach assures that the data is usable. This study worked from the assumption that RTTT’s approach is the opposite – data is collected to help make decisions.

Campbell et al. (2003) believe that parents expect teachers to be evaluated based on their ability to present clear instruction. However, emphasizing instruction significantly weakens the argument that evaluation instruments reflect broader areas such as classroom environment and professional responsibilities. Politicians prefer a simple model that a lay person can easily understand when expressed in sound bites and through campaign promises. Yet, the state-endorsed models are more complex (e.g., those developed by Danielson or Marzano). There seems to be a disconnect between what parents and politicians expect and the expectations proposed by recognized evaluators.

Campbell et al. (2003) also wonder how to judge teachers’ values. Values such as holding high expectations for pupils or having a commitment to the subject matter can affect students. Values are inherently subjective. Subjective constructs can be problematic when trying to fit them into an easily measurable score. Hess and Fullerton (2010) believe the data most useful to parents and policy makers is often simplistic recounts of achievement scores, which offer little contribution to understanding why results were achieved. By focusing so much on student achievement, many employees are either excused from results-driven accountability or held accountable for items for which they have limited control. Thus, the literature seems to suggest caution as to how data is utilized.

Ravitch (2013) contends that evaluating teachers through test scores undermines professionalism. Baker et al., (2010), among others, believe it unwise for states to weigh as
much as 50% of a teacher’s evaluation on scores from existing tests of basic mathematics and language arts skills. Mathematics and language arts teachers can have test scores more easily assigned to them compared to teachers in non-tested areas, but doing so ignores the fact that there are other factors that influence student learning gains. Such factors include the impact of previous teachers or current teachers in non-tested subjects. Curriculum materials, class size, and remedial supports can all impact student achievement. Factors outside of school such as family resource, student health, and family mobility also influence student test score gains. Schools that utilize team teaching, for example, will have a difficult time isolating an individual teacher’s effect on a student (Baker et al., 2010).

Ravitch (2013) asserts that a test should only be used for its designed purpose. Tests that measure student performance in comparison to a norm are not designed to measure teacher performance or quality. Basing a teacher’s performance on heavily weighted assessments can, because of outside factors, result in the maintenance of a weaker teacher or the removal of an effective one. Thus, this practice undermines one stated purpose of testing students.

Research in the use of data for teacher evaluations seems to provide more questions than answers. Even well-developed plans like Cincinnati’s Teacher Evaluation System prove problematic. The greatest concerns arise over quantifying factors other than teacher instruction that impact student learning and test scores.

**Value Added Models**

The intent of Value Added Modeling (VAM) is to determine the extent by which some entity, such as teachers or schools, has affected change in students. The results are compared across students, allowing summaries associated with each teacher to show whether one’s students are performing above or below expectations. The “value added” component refers to
yielding more growth than expected when accounting for the characteristics of the students or school citations. One way VAM is used is to measure the impact a teacher’s background has on students. Wayne and Youngs (2003) conclude that there was enough evidence to indicate that the college rankings of the institution a teacher attended and the teacher’s scores on accreditation tests impact student learning. Teachers who attended better colleges and/or earned higher test scores had students that learned more. VAMs try to account for these types of characteristics.

Harris and Sass (2009) examined one Florida school district to determine what teacher characteristics helped improve student performance. Their findings suggest that attributes like teacher intelligence, subject knowledge, and teaching skills help to make a ‘good’ teacher, while personality traits like caring, motivation, and enthusiasm are not related to improvements in student achievement. The authors conclude that value-added measures, gathered over several years of test data, can be a good predictor of future teacher performance. If only one year of information is being utilized, a principal’s evaluation is just as predictive.

There are many different versions of a VAM. Ehlert, Koedel, Parson, and Podgursky (2013) compared three growth models. They conclude that a two-step VAM is the most effective. This model adjusts for differences in test score performance between students with different traits and in varying school environments before estimating the school effects. This approach enables a better accounting of student differences and has the potential to “overcorrect” for student disadvantage (p. 10). It helps to assure that growth rankings are proportional to the evaluation sample.

Schafer et al. (2012) question the use of VAM. Less than half of teachers fall into areas that are covered by most state tests. In addition, other personnel, such as librarians, aides, and counselors, can have an impact on student success. Other variables such as socioeconomic
status, individual aptitude, home environment, and per-pupil expenditure, can significantly impact on a student’s progress. Because of such concerns, Schafer et al. (2012) deem the use of VAMs in high-stakes decision making as a “very high risk” (p. 3). Stronge, Ward, and Grant (2011) also believe that such high-stakes proposals be pursued with “extraordinary care” (p. 351). They add that if VAM data is used, it should be compiled from multiple years and that prior achievement and socioeconomic status be controlled, else the effects of teachers on student achievement could be shrouded by student-level variables.

In a joint briefing, the American Educational Research Association and the National Academy of Education (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2011) caution against the use of VAMs. They conclude that VAMs are unstable, as results vary from class to class and from year to year. In addition, they found that teacher ratings are “significantly affected” by differences found within the students assigned to them (p. 5). Darling-Hammond, Amrein-Beardsley, Haertel, and Rothstein (2011) note that teachers with large numbers of students with special needs have lower gains than the same teachers when they teach other students. The briefing notes that VAMs do not “disentangle” the other influences on student progress, such as the impact of earlier teachers on current performance (p. 6). Ravitch (2013) goes further, attacking Value Added Modeling by calling it “bad science” and describing the method as “inaccurate, unstable, and unreliable” (p. 113).

Based on the current literature, it appears that VAMs are problematic. They attempt to place a quantitative value onto a frequently subjective characterization of performance. For example, Wayne and Young’s (2003) assertion that teachers who attend better colleges have students that learn more seems to ignore the socio-economic conditions of the students as well as the socio-economic status of the teachers themselves. The challenge for reformers has been to
develop an objective assessment of what contributions a single teacher makes on student learning.

**Student Growth Percentiles (SGPs)**

The need to comply with NCLB achievement mandates prompted states/districts to measure their students’ success. The New Jersey Department of Education (2014) permits the use of student growth percentiles (SGPs). Understanding the intentions and limitations of this method may help educators develop a clearer insight into their use.

Models for accountability fall into two broad categories: status or growth. Models that focus on status qualify student performance based on current achievement, examining that performance at a given time without accounting for other variables. Typically, student performance is judged using a criterion-reference test that qualifies a student as proficient or not proficient. Betebenner (2009) describes such models as demanding since they require all students to reach an acceptable level of performance without condition.

Conditional achievement models, or growth models, evaluate student achievement over a longer time period and predict the future achievement of the student. Betebenner (2009) argues that using growth models to determine if a student is progressing towards proficiency is limiting. He introduced student growth percentiles (SGP) as a “normative conceptualization of student growth” (p. 44). The changes in student performance are compared to a normative group to determine how typical the change is. He cautions, however, that SGPs are for descriptive purposes only, and should not be used to associate causation. He compares SGPs to a pediatrician’s growth chart. Growth charts let parents know how their child’s height compares to similar peers but does not explain why their child might be taller or shorter than others. SGP
results are similarly easy for parents to interpret but are not designed to assign a reason for the scores.

Goldhaber, Walch, Gabele, and Schools (2012) indicate that SGPs do not explicitly account for differences in student background but may implicitly account for such differences by comparing students that are similar to each other in terms of baseline achievement. SGPs account for student differences by burying those differences amongst the entire sample. It is assumed that students with similar backgrounds (i.e., those who receive free lunch) will earn similar scores and that their growth in learning would be normatively comparable. Haertel (2009) cautioned that SGP models presuppose that students with the same initial test score have the same aptitude for learning as well as the same out-of-school supports for learning. Since this is not the case, he wonders if the controls for prior achievement in SGPs may be insufficient and whether students should be expected to grow at the same pace. Goldhaber et al. (2012) conclude that SGPs are good for reviewing student achievement, but they should not be used for teacher performance. Yet, this is what some departments of education propose to do.

The city of Denver, Colorado implemented a comprehensive reform of how teachers are paid, called Denver’s Professional Compensation System for Teachers (ProComp). The four-year pilot implemented what is commonly called ‘merit pay’. Teachers could receive bonuses or permanent salary increases based on their performance. The central piece of ProComp is based on SGPs. Interestingly, one category of SGP incentives can only be used by math, literacy, and language arts teachers because it is tied to state testing in those areas. Also of interest, teachers who do not meet expectations can lose a previous salary increase, though “consideration” is given to the characteristics of that teacher’s students and the “extent of contact secondary students have with other teachers” (DeGrow, 2007, p. 8). It is noteworthy that one of the most
comprehensive data-based programs in the country recognizes the limitations of such data. Yet, despite the limitations expressed in the research, states like New Jersey are moving ahead with high-stakes testing.

**Student Growth Objectives (SGOs)**

Not all content areas (e.g., art, music, history, and foreign language) are annually tested using standardized assessments. Without such instruments, SGPs cannot be calculated. To meet the evaluation requirements of these teachers Student Growth Objectives (SGOs), sometimes called Student Learning Objectives (SLOs), are substituted. The SGO process begins when a teacher or teachers identifies an expected learning outcome for a group of students over a period of time, whether in an individual class, across many classes, or throughout a district. The teacher(s) ascertains the students’ current level of performance before determining an appropriate goal to be accomplished. Within the New Jersey system, teachers collaborate with their principal to establish goals that are rigorous but attainable (New Jersey Department of Education, 2014). SGOs are typically developed using local measures such as teacher-made tests and portfolios.

The research on SGOs is limited. Most studies analyze the use of SGOs for teacher compensation (see e.g., Brodsky, DeCesare, & Kramer-Wine, 2010). Goldhaber and Walch (2012) analyzed ProComp, the Denver Public Schools pay for performance program, and found that SGOs corresponded to teacher performance, as measured by student gains on the Colorado state test. The SGOs themselves could not be based on state test scores, but the state test scores of students undergoing SGOs improved because of the efforts teachers made at meeting the SGO goals. Some research has addressed the proportion of teachers achieving SGO objectives (Gill, Bruch, & Booker, 2013). Goldhaber and Walch (2012) refer to a report by the Austin School
District, written by Schmitt and Ibanez (2011), that suggests that teachers who had students reach their SGOs were more likely to increase their scores on state tests. Meanwhile, Balch and Springer (2015) found “no significant relationship between the percentage of SGOs a teacher meets and the teacher’s effectiveness” (p. 120).

Holdheide et al. (2012) explain a significant challenge to using SGOs is that they may not be comparable across classrooms if the established procedures for developing the SGOs are not standardized. The authors caution that attaching SGOs to high-stakes decisions like the retention of teachers could incentivize teachers to set easily obtainable goals. New Jersey principal’s evaluations are now partially based on an average of the teachers’ SGOs (New Jersey Department of Education, 2014), which could encourage them to accept easily obtainable goals from their teachers. Teachers need training to set appropriate growth targets, interpret student data, and modify their instruction. They also need time to complete these steps (Holdheide et al., 2012).

**Utilizing Models**

In a four-year comparison study, Goldschmidt et al. (2012) conclude that no growth model is clearly better than others in all situations and circumstances. According to their study, “context matters” (p. 9). That is, different models address different questions and the state in which tests are administered affects how the model works. Goldschmidt et al. (2012) reveal that the model chosen can lead to different inferences regarding a school. It is possible for one school to be given high marks using one model and average marks using another when the models are fundamentally different. Districts need to be certain as to what they hope to achieve when using test scores, lest the results they receive give an erroneous impression. That being
said the models Goldschmidt et al. (2012) examined revealed reasonably consistent results and showed no such extremes.

When examining consistency across years, Goldschmidt et al. (2012) found that SGP based tests were more stable than gain-based models, which typically rank students as proficient or not. No such relationship was established with VAM. The study did note that SGPs are less reliable when schools are small. Grady, Lewis, and Gao (2010) examined the effect of size on SGPs. They find that when sample size is particularly small (i.e., 500) the SGP estimates are negatively biased. They add that Betebenner’s suggested minimum sample size of 7,000 is conservative. They recommend a large sample size if SGPs are used for important personnel decisions. States like New Jersey plan to compare students from across the state, creating sample sizes well above the minimum recommendation.

SGPs and VAMs are least influenced by student background compared to other models. When examining different levels of school (elementary, middle, and high), Goldschmidt et al. (2012) concluded that more data provided better results; that is, students who had results from multiple grades provided more reliable results. New Jersey students, for example, are tested each year from grade 3 to 8 but not again until grade 11. Because of that, Goldschmidt et al. (2012) recommend making different decisions on how to attribute growth at the high school level.

As previously mentioned, testing models fall into two broad categories: status and growth. Goldschmidt et al. (2012) caution districts about using models that measure both of these areas. Doing so may unintentionally disadvantage schools with particularly high or low scores. In some cases, the growth indicators may counter the status indicators resulting in the school being classified as average when exemplary or failing would be more appropriate.
Pomplun (2009) investigated several growth tests and compared them to the students’ classroom grades. He concludes that the tests were a valid indicator of student success but with some interesting idiosyncrasies. Mathematics results were valid at the class level, not the individual student level. Honors students gained more than standard students, who had gained more than students in developmental courses. Students earning As and Bs in class demonstrated greater growth than students earning Ds or Fs. One explanation is that the test content is not aligned with the course curriculum, which would vary by level. Pomplun (2009) cautions districts about using growth scores, but concedes that if they are used, districts should make sure their curriculum aligns with the test.

Haertel (2009) asserts that student achievement data can provide a useful tool for improving schools; however, they are “far from a ready made solution” (p. 20). Technical experts, policy makers, administrators, and teachers must collaborate to ensure proper implementation of the collected data. State departments of education may decide which model to use, but local districts ensure that the teachers and parents understand what data is collected, how it is interpreted, and for what purpose it is utilized. It seems that public relations plays a part in data use as districts try to make all stakeholders understand the results.

Ysseldyke et al. (2004) analyzed the effects of high-stakes and large-scale assessments on students with disabilities and concluded that there have been benefits to including these students in the same process as their non-disabled peers. They found increased participation in assessments by students with disabilities, as required under NCLB. There has also been an increase in expectations for such students, a likely response to including them in the testing. Consequently IEP development now reflects higher standards for students with disabilities.
(Ysseldyke et al., 2004). Their study does not discuss whether or not testing should be part of a teacher’s evaluation.

**Using Data**

Herman et al. (2008) report that teachers need training to use data. Similarly, principals, who are frequently the conduit of information, need a clear understanding of data and its potential use. Herman et al.’s (2008) study, while limited, indicates that data use in each participating school varied but in all cases was underutilized. This is consistent with Hess and Fullerton (2010), who believe that the new “buzzword” of data-driven instruction does not mesh with the cultures of most school districts (p. 6). Administrators and teachers have had little exposure to fact-based decision making. Districts will have to allocate time and resources to familiarize their staff with how data can be used to improve instruction.

Lange, Range, and Welsh (2012) would concur that teachers need help in using data. They maintain that data-driven decisions are part of a process that involves collecting, analyzing, and using data. This process should be cyclical with a design rooted in focused inquiry. To accomplish this, teachers need help from their school leaders. Lange, Range, and Welsh (2012) advocate that long-term change can only occur when the school culture understands and values the use of data. There must be a plan in place for the collection and storage of data. The authors promote leadership teams that would foster the use of data amongst all staff. In addition, the data should be presented in a way that “eliminates blame and punitive action” while using data as a tool to improve student learning (p. 5). Within the scope of this thesis, there are obvious difficulties using data to improve learning if data is also used as a way to blame or punish teachers.
Districts are now required to collect student data more than ever before. Some states will utilize this data to make personnel decisions. The complexity of the data could easily overwhelm educators, much less persons outside the field. Care should be given as to how data will be presented and utilized. It has become imperative for administrators and teachers to learn how to use this collected data to change classroom performance.

**Special Education and Data**

Holdheide et al.’s (2012) summative report notes that not enough is known about the quality of academic growth models for students with disabilities. They suggest that most models fail to address the “unique challenges in accurately measuring achievement growth of students with disabilities” (p. 1). They caution the use of testing data to make high-stakes decisions until further research is conducted.

Special education teachers are part of the approximately 69% of teachers that work in a field not covered by state testing (Gould, 2013). These teachers cannot use value-added or student growth percentile models. New Jersey has determined that teachers in non-tested areas may use student growth objectives. A problem with such an approach is the availability of an appropriate pretest that measures “the student knowledge and aptitude” prior to instruction (Meyer, Christian, Gawade, & Wang, 2013, p. 5). At this point the state permits teachers to develop their own assessments, with administrative approval.

Gould (2013) suggests that authentic assessments are a good alternative for students with disabilities. Their teachers can adapt the assessment to the students’ ability, though having differing expectations can result in a problem with scoring reliability. Reliability can be increased with the use of a rubric and multiple scorers. In addition, teachers and students could
“establish and understand task criteria” before beginning, alleviating anxiety associated with testing (Gould, 2013, p. 29).

Value-added models present several challenges for students with disabilities. Holdheide et al. (2012) find that the relatively small size of special education classrooms made the measure of teacher effects on student performance less reliable. In addition, the delivery of special education services varies according to student need, making the accurate attribution of student gains to both general and special education teachers a challenge. Similarly, Holdheide et al. (2012) contend that the “inconsistent use of accommodations” can affect the measure of growth (p. 5). It is not unusual for IEP testing modifications to be changed after a test to better meet a student’s needs on future tests. Thus, the quest to meet students’ needs conflicts with the quest to mathematically-quantify teacher contributions to student learning.

**Special Education and SGOs/SGPs**

Specific research about SGOs and students with special needs is limited. A search of the past ten years using Scholar OneSearch and EBSCO using the key terms “special education and student growth objectives” or “student growth percentiles” and combinations of those words revealed articles on achievement tests, curriculum measures, alternative assessments, and performance pay. Only Holdheide et al.’s (2012) work, discussed above, specifically address student growth objectives. The lack of comprehensive research in this area is disconcerting as school districts across New Jersey, as well as across the country, have or will implement teacher evaluations using SGOs and SGPs.

**Conclusion**

The Council for Exceptional Children (2013) recognize the importance of well-develop teacher evaluation instruments. However, they assert that such instruments must recognize the
uniqueness of a special education teacher’s role. A teacher’s evaluation should include “accurate and reliable indicators” of teacher “contributions to student growth” (p. 74).

Any model used for collecting student data is limited to its designed purpose. The utilization of the data must conform to its limitations. Policy makers need to recognize the complexity of student learning. The factors that shape that process are numerous and varied. No single test could be expected to account for all the variables that affect students. Therefore, no test should be viewed as the absolute indicator of student growth. Because of that fact, the literature seems to be clear that it is unwise to make personnel decisions based largely on growth modes. Unfortunately, some states, including New Jersey, have chosen to do just that.

This study began with the assumption that it will be up to administrators and teachers to move beyond the fear generated by having a large portion of an evaluation tied to one test. They should take information about student success (or failure) and learn from it. This data can be utilized along with a solid, research-based observation instrument to help determine the characteristics/strategies that a teacher has that are working and those that need improvement. The use of data is not an educational fad that will disappear after a few years. As new research helps perfect its implementation, data will become part of a teacher’s ‘bag of tricks’ used to improve what is done in the classroom. School personnel must be trained in data use so that they fully understand its meanings and possibilities. Further research on how growth models can account for influences beyond the school is warranted. Similarly, as more states require data to be used for high-stakes personnel decisions, the reliability of instruments for ascertaining who is a ‘good’ teacher must be corroborated.
Chapter 3: Research Methods

The New Jersey State Board of Education recently established that up to 30% of the teachers’ evaluation score will be based on administrative observations, while the remainder will be based on student testing data (New Jersey Department of Education, 2014). Baker et al., (2010) indicated that no single existing model for data collection can be expected to definitively determine the extent by which an individual teacher contributed to a student’s academic growth. Despite this, the state department of education uses such measurements suggesting that they are tied to improving student achievement. How teachers and administrators utilize the collected data can impact its usefulness and is worthy of investigation.

Purpose

The purpose of this study was to understand the relationship between the collection of student assessment data and the teaching practices of special education teachers during the implantation of a new evaluation system in New Jersey. Specifically, it examined whether or not special education teachers in grades 6 through 8 utilized the student performance data that they must now collect from SGPs and/or SGOs to change their instruction. The success of the reforms, prompted by the federal program Race to the Top (RTTT), will ultimately depend on documenting that students are learning more. Learning how data, an important component in New Jersey’s implementation of RTTT (titled AchieveNJ), fits into the reforms is vital to its implementation.

Central Research Question

The core of the research focused on the question: What are the experiences of middle school special education teachers at a small suburban school who are newly mandated to use standardized student data to inform their pedagogical and curricular approaches? The
question is relevant because the New Jersey Department of Education has made data acquired from SGPs and/or SGOs part of the teacher evaluation process. There is the assumption that the data will help teachers improve their instruction. It is imperative to know how teachers view the use of data. Do they feel that data gathering has proven beneficial or do they view it as a waste of time? At one extreme all teachers will use data to modify their lessons to meet the needs of their students. At the other extreme teachers will obey the law and gather the data but will not use it to change their teaching approach. Understanding the teachers’ attitudes will help district administrators, and by extension state officials, assure that the intention of data gathering is being fulfilled.

**Methodology**

This research utilized an interpretive phenomenological analysis (IPA) design, an approach that concentrates on an individual’s “personal perception or account” of an event (Smith, 1996, p.263). This method explores an experience “in its own terms” (Smith, Flowers, & Larkin, 2009, p. 1). Smith et al. (2009) explain that IPA researchers are concerned with what happens when an important event impacts the typical course of peoples’ lived experiences. For Smith et al. (2009) “an experience” occurs when someone becomes aware of what is happening rather than “unselfconsciously” going through the “everyday flow of experience” (p. 2). This semantic exercise distinguishes IPA as an approach where the researcher must analyze how different individuals view a phenomenon and look for commonalities between the explanations.

In her overview of the interpretivist paradigm, Mack (2010) describes the influence that hermeneutics and phenomenology had on its development. Smith (2010) similarly credits the phenomenological concern with participants’ lived experiences and the hermeneutic concern that
“experience is only accessible through a process of interpretation” by the researcher and participant (p. 186).

Hermeneutics began as a method to interpret the Bible but expanded to include the reading of any text so that the “intention and meaning behind appearances” are understood so that a correct understanding can be made (Moustakas, 1994, p. 9). Hermeneutic theorists examine the methods and purposes of interpretation. They wonder if it is possible to uncover the original intent or meanings of an author. They wonder about the relationship between the context of when a document was written and the context of when it was interpreted (Smith et al., 2009). Smith et al. (2009) assert that an IPA researcher could be engaged in a “double hermeneutic” since the researcher is trying to understand the participant’s understanding of what is happening to them (p. 3). Thus, there is an interaction between an older, original understanding and the possibly newer conception.

Phenomenology studies awareness as experienced from the first-person point of view. The important facets of such experiences are their intentionality and how they are being received (Smith, 2013). Creswell (2013) explains that such an approach enables researchers to examine phenomena in their natural settings where they try to make sense of the meanings people bring to the phenomena. While Douglass and Moustakas (1985) state that phenomenology extracts the “essence of experience” (p. 43).

According to Creswell (2013), phenomenologists concentrate on describing the commonality between participants’ experiences. This thesis, as phenomenological research, took different interpretations shared by special education teachers and provided organization and purpose to the findings. It described what the teachers encountered when using data and then how they utilized it. Such descriptions would not be possible using quantitative methods. Smith
(2007) equates the data collected from personal accounts to that of the historic texts analyzed through a hermeneutic approach. The transcripts from the participants’ interviews become the text the researcher analyzes.

Figure 1. Interaction between teachers, data, instruction, and state mandates.

Figure 1 illustrates the interaction between teachers, data, instruction, and classroom instruction, without excluding the influence of the state mandates. Conceptualizing the major components aided in the design of the study because it provided a better understanding of how the use of data could impact what takes place in the classroom. This study was developed as a reaction to the mandate by the New Jersey Department of Education to use student data in teacher evaluations. The researcher investigated how teachers responded to the new requirement. In order to collect data, students underwent performance testing, such as SGPs and SGOs. Teachers analyzed the
data and possibly altered their instruction techniques. Students were exposed to instruction and responded accordingly. Understanding how teachers use the data to guide their instruction could improve how teachers use data to meet the needs of their students.

Questions and Methodology

The primary research question exemplifies an interpretivist approach. While the state mandated the use of data, teachers determined how to use the data within their own classrooms. Because of the personal nature of implementation and the lack of existing research within this area, a qualitative approach was most appropriate. Understanding how teachers fulfilled their obligation was needed to help understand its effectiveness. Qualitative research is able to provide a depth of data that reveals the experiences of the participants, in this case teachers in one school. Teachers had some flexibility as to what kind of data to use, and this research approach helped accommodate that flexibility as well as the uniqueness of each teacher.

The central question provided the overall guide for the interviews, but there were opportunities for the researcher to ask new questions based on participants’ responses. The researcher developed an interview schedule to help guide the process (Appendix C). Smith et al. (2009) describe a schedule as a set of prepared questions laid out in an order that the researcher anticipates will be the most suitable for the participant. Collectively, these questions revealed how comfortable special education teachers are with using student data, particularly SGPs and SGOs. In addition, discovering the methods in which teachers use the data provided valuable insights for administrators and some best practices for fellow teachers. Since collecting data is now required, understanding how teachers related to the data and whether or not they used the data to alter classroom instruction was imperative in gauging the effectiveness of the new mandate.
Researcher Bias

The researcher is part of the special education department at the center of this research and is also implementing the state requirements within his own classroom. It would be impossible to entirely remove the researcher from the situation, even if another district were used. To help minimize unintended influences, the researcher refrained from revealing his own experiences to those he interviewed and remained unbiased and open to the experiences of the study participants throughout the process.

Study Participants

Individual perspectives are at the core of this study. Understanding the implications of state mandates within a school district and, more specifically, within the special education teacher population were driving factors in this research. The teachers who participated in this study were all members of a Special Education Department located in a New Jersey public school. It is a relatively small district with approximately 32 special education teachers. These teachers tend to know each other across levels and between buildings because of joint workshops and Individualized Education Program (IEP) meetings. The researcher works within the district and knows almost all of the special education teachers by sight and several on a personal level. The researcher has worked in the district for 20 years and believes that those teachers who do not know him well do know who he is. They will be interested in personally helping the researcher and in learning about the results, which may include some best practices to share.

Before conducting the study, the researcher received permission from the superintendent of schools to conduct the study at the study site and to contact teachers. The researcher sent all teachers within the middle school special education department a letter inviting them to participate in a study on the new evaluation system. This recruitment letter is included as
Appendix A. The letter explained the purpose of the study, as recommended by Rubin and Rubin (2012). It also asked the participants to respond if they were willing to participate. The intention was to have at least four participants, but there was no intent to exclude interested participants. Had the number of volunteers exceed four, the researcher would have randomly selected four participants. Four individuals agreed to participate in the study, and no further recruitment took place as a result, having met the minimum desired number of participants.

Interviews were scheduled with participants at a time and place convenient for them. At the first meeting, each received an informed consent form (Appendix B) reiterating the purpose of the study and how their information would be used. It also how the interviews would be recorded and transcribed to assure accuracy. Each teacher who participated was emailed a final copy of the study’s results.

All participants volunteered for this study and could withdraw at any time. The participants were not at risk of physical harm. The researcher’s primary concern involved the small, close nature of the district. It was not beyond the possibility that an individual’s answers could have made his or her identity known. This possibility may have altered how some participants chose to answer the questions. The interviewees were given pseudonyms to protect their identities. In addition, the researcher was careful when describing the interviewees. The wording used to describe the participants was intentionally vague, such as “a middle school special education teacher” without reference to gender, teaching assignment, or specific years of service.

**Sampling**

This study utilized purposeful sampling. As described by Creswell (2012), this is intentionally selecting individuals and sites to learn or understand a phenomenon. The specific
strategy used was typical sampling, which samples individuals that are “typical to those unfamiliar with the situation” (Creswell, 2012, p. 208). This enabled the researcher to understand how an average group of special education teachers used student data in their classrooms. This particular district has little racial variation within its staff, but the selected department’s employees vary in the years of teaching experience.

The researcher selected one suburban, public school district for the study. This district shares characteristics with 87 districts from the same District Factor Group (New Jersey Department of Education, 2010b). The relatively small size of the research site provides greater familiarity between professionals at each grade level and promotes an interactive culture.

The researcher chose to exclusively examine the Special Education Department within the middle school because of their familiarity with using student data. These teachers have used data from psychological reports, learning evaluations, and state testing long before the 2014 state mandate. The department served as a good bellwether on the further implementation of data from the sources prescribed by the state department of education. The teachers received no tangible incentive for their cooperation.

The open-ended interviews afforded the researcher the opportunity to discuss, in detail, how the teachers utilize data. They were semi-structured in format, wherein the researcher comes with prepared questions but plans to ask others as warranted (Rubin & Rubin, 2012).

Data Collection

After receiving permission from the appropriate parties, the researcher sent a letter to each teacher’s school address explaining the study and requesting their participation. (See Appendix A.) The interviews lasted between 75 and 90 minutes. The teachers arranged the interviews at a time convenient to them. The teachers’ responses were digitally recorded for
later transcription. In addition, the researcher took notes during the interview to help him in asking further questions. The questions were intended to provide detail as to how these teachers use data to alter their instructional techniques.

**Data storage.**

All data from the interviews was secured to protect confidentiality. The researcher used a locked cabinet at his place of work and a secure file cabinet at home. All files were stored on a password protected computer. The signed consent forms will be stored for three years at the home of the researcher.

**Data Analysis Process**

Smith et al. (2009) describe a generic process to help analyzing data in IPA research. Their first step is “reading and rereading” and is singled out “to ensure that the participant becomes the focus of analysis” (p. 82). The researcher read and reviewed the transcripts to assure that the voices of the special education teacher participants were not lost.

The second step for Smith et al. (2009), “initial rating,” is very similar to coding (p. 83). Coding is the process by which a researcher segments and labels text to form “descriptions and broad themes” from the data (Creswell, 2012, p. 243). Saldaña (2012) divides coding into two steps: First Cycle and Second Cycle. First Cycle is conducted during the initial stages, while Second Cycle is a way of analyzing and classifying the initial coding. In this case, the researcher chose to combine two methods of First Cycle coding: descriptive and In Vivo. Descriptive coding condenses in a word or phrase the basic topic of the data (Saldaña, 2012). Typically the descriptor is a noun. In Vivo coding also uses a word or short phrase to summarize the statement, though the wording is in the actual language used by the respondent (Saldaña, 2012). First Cycle coding is also similar to the third step described by Smith et al. (2009), “developing
emergent themes” (p. 91). At this stage the analysis moves away from the transcript to the researcher’s notes, where the participants’ experiences become fragmented as the data is reorganized. The researcher took the linear descriptions found in the transcripts and dissected them into broad categories that reflected the participants’ experiences.

For Second Cycle coding the researcher chose to use Focused Coding, where the researcher develops categories based on the relevance and frequency of the previously coded data (Saldaña, 2012). This corresponds with Smith et al.’s (2009) fourth step, “searching for connections across emergent themes” (p. 92). This is a fluid process, where the researcher is “encouraged to explore and innovate” their categorizing (p. 96). The researcher did not have predetermined categories, rather the development of categories as well as commonalities and differences was governed by the participants’ responses.

The researcher consciously chose not to use any computer assistance during coding. The four interviews were transcribed and printed so the researcher could write directly on the transcripts. The researcher read through each interview and underlined words or phrases that seemed relevant to the research question. Next, he re-read the material and used descriptive coding to categorize the underlined material. Lastly, using In Vivo coding, the researcher highlighted phrases where the interviewee’s language seem particularly relevant.

For the Second Cycle coding, In Vivo and descriptive terms were itemized in a chart. The initial words and phrases were categorized into groups. These groups were further narrowed to focus the meanings of the categories and to eliminate any overlap between categories. Once the categorized chart was completed, it became visibly clear which ideas dominated the interviews. This process made the material easier to organize for analysis as commonalities between the interviewees became apparent.
IPA and Validity

Smith et al. (2009) contend that formulaic methods of assessing validity do not fit qualitative research. Rather, qualitative research should be judged in ways more appropriate to its approach. Smith et al. (2009) highlight the principles of Yardley (2000) as appropriate to qualitative research. These principles are broad in nature and involve the overall approach to methodology and analysis rather than a proscribed method of approach. Seemingly, if a researcher does the work correctly, there should be no problems with validity. Yardley’s first principle is sensitivity to context. Smith et al. (2009) assert that good IPA research is sensitive to the data and that the researcher should not lose focus on making sense of the participants’ experiences. A good IPA study will “give participants a voice” (Smith et al., 2009, p. 180).

Yardley’s (2000 as cited by Smith et al., 2009) second principle is commitment and rigour. A good researcher accepts the IPA approach and insures that the participant is comfortable so that the researcher can fully probe the experience and receive ample data to analyze. The analysis must be, according to Smith et al. (2009), “sufficiently interpretative” by moving beyond a retelling of the interview to a true elucidation of what the experience meant to the participants (p. 181).

Yardley’s (2000 as cited by Smith et al., 2009) third principle is transparency and coherence. Transparency involves a careful description of the research process as well as the selection of the participants. Coherence refers to a logical presentation of the arguments.

Yardley’s (2000 as cited by Smith et al., 2009) final principle is impact and importance. The validity of research lies in its revelation of something valuable and applicable. The researcher was confident that this investigation revealed useful information that will help the district in its implementation of the data portion of the new evaluation system.
Trustworthiness

Creswell (2012) prefers a more fixed approach to validating research. Trustworthiness is a more traditional term to describe what Creswell (2013) calls ‘validation.’ Regardless of the terminology, it is an attempt by the researcher to assess the correctness of the findings. Creswell (2013) writes that validation is a strength of qualitative research because the narrative derived from the researcher’s closeness to participants adds to the study’s accuracy. Two methods of validation promoted by Creswell (2013) were utilized in this research: clarifying research bias and member checking.

Clarifying research bias.

By clarifying research bias, the researcher explains their position and any biases or assumptions that might influence the study. The researcher previously indicated his background as a New Jersey teacher experiencing the same use-of-data mandate investigated in the study. Possible conflicts or biases were openly expressed. In this study, being fully removed from the research phenomenon is impossible for the researcher, therefore bracketing was necessary. Bracketing is a concept whereby researchers set aside their own experiences as much as they can to take a new look at the phenomenon being examined. Creswell (2013) adds that “bracketing personal experiences may be difficult for the researcher” (p.83). Machi and McEvoy (2012) assert that researchers can control their biases and opinions and can commit “to being open-minded, skeptical, and considerate of research data” when they rationally identify and confront their views (p. 19). Throughout the process, the researcher clearly noted when ideas were based strictly on personal experience. It is hoped that this forthrightness added to the trustworthiness.
Member checking.

Member checking is a process in which the researcher asks participants to review the findings and interpretations for accuracy and credibility (Creswell, 2012). Lincoln (1986), who uses the word “fairness” to represent the broad idea that information is properly presented, states that member checking enhances the fairness of a study and that its “assiduous use should lead to a common judgment” (p. 13).

In this study, the four teachers who were interviewed had the opportunity to review their transcripts. The individual transcripts were emailed to each participant, and they were asked to provide feedback regarding accuracy or to provide any additional information. Each participant approved their transcripts, which added to the trustworthiness of the study.

External Validity

Creswell (2012) defines threats to external validity as problems that restrict the ability to draw correct inferences from the sample data to other persons or settings. Generalizing this study to other schools was not the intent of the research. While all public schools in New Jersey are impacted by this mandate, other states, even those accepting Race to the Top funding, may have different requirements. Other complications to generalization include the fact that the sample was limited to special education teachers, who often teach in settings unique from their general education peers. However, the best practices derived from this study are relevant to teachers who are required to use student data, as well as those who choose to do so on their own. The approaches offered by the respondents may help teachers use hard data to design lessons appropriate to the needs of their students.
**Protection of Human Subjects**

Creswell (2012) outlines three basic principles when using human subjects. The first is the benevolent treatment of participants. There were no known or obvious risks to the personal well-being of the participants. In addition, there were no foreseeable professional risks, as the local superintendent approved the study and the researcher did not have supervisory oversight of any of the participants.

Creswell’s (2012) second principle is respect for the participants. Participation in this study was voluntary and participants had the right to withdraw at any time. Pseudonyms were used to protect the identity of participants. In fact, only the most general descriptions were used because the small size of the district increased the chance of identifying respondents. Furthermore, the researcher made every effort to keep the respondents informed of the process as well as the progress of the study.

Creswell’s (2012) third principle involves justice. There no known risks to being involved with the study. There are no specific benefits to each participant, though all will share in the results. One goal of the research was to compile a list of best practices for the use of student data. As student data becomes a significant component of a teacher’s evaluation, the need to know how to compile and use data increases. The list of best practices will help all teachers, even those who currently use data, to see other potential methods for using student data. Assisting teachers to better understand their students should improve instruction, which is one goal of the state’s mandate.

In November 2013, the researcher completed the National Institutes of Health Office of Extramural Research training course titled ‘Protecting Human Research Participants.’ The certificate of completion was included with the Institutional Review Board (IRB) application.
The appropriate institutional review boards approved the research prior to data collection as described below.

**Consent Process**

The researcher gained approval to proceed with the study from the IRB. He met with the district superintendent and obtained consent to conduct the research. Participants first learned of the study through their school mail. The researcher notified all special education teachers within one suburban district to expect an envelope in their school mailbox. This envelope contained a full explanation of the research problem and the scope of the research study. It asked for the teachers to respond if they wanted to participate.

Prior to the individual interviews the researcher reviewed the purpose of the study. The researcher addressed any questions. Any perceived risks or concerns was reviewed and addressed by the researcher as recommended by Rubin and Rubin (2013). Participants were reminded that their participation was voluntary and that they could withdraw consent at any time without consequence. The participants read and signed a consent form (see Appendix B).
Chapter 4: Findings

This chapter reports the findings that resulted from interviews conducted with four special education teachers. The catalyst for the investigation was a new law enacted by the New Jersey legislature requiring that student data be used as part of a teacher’s annual evaluation. This data takes two forms: Student Growth Percentiles (SGPs) and Student Growth Objectives (SGOs). SGPs and SGOs are used to judge teacher performance, however, this data can also help teachers understand their students’ strengths and weaknesses, which can guide them to make adjustments in their instruction.

The study was conducted in a small suburban middle school with approximately 350 students. For the purpose of this study, the school is referred to as the Clarksburg Middle School. There are nine special education teachers within the school, providing services in self-contained, resource room, and mainstreamed environments. There are 64 students within the program, the majority of whom (38) are classified as having a “specific learning disability” (R. D’Arecca, personal communication, March 23, 2015).

Participant Profiles

Because of the small, close-knit nature of the district, gender-neutral pseudonyms were used to protect the anonymity of the participants. Likewise, their background information was kept to a minimum so that their identity could be concealed. Brief biographies are included to introduce the participants.

Lee.

“Lee’s” first work experience was in a private, residential school in the 1980s. This job fostered an interest in special education, which eventually led to a teaching job in a private, state run facility for those with severe disabilities. Approximately ten years ago Lee began teaching in
the Clarksburg district. Lee’s experiences within the district have included both self-contained classes and in–class support settings.

**Alex.**

“Alex’s” path toward a teaching career path began during freshman year of high school when a history teacher became an inspiration towards that career path. Before the first year of college, Alex’s summer job was as a one-on-one assistant in a school for those with severe disabilities. This was the spark that ignited a passion for special education. Alex’s first teaching position was in a middle school self-contained behavior class. Alex changed districts and, for two years, taught in a Grade 4-8 resource room before moving to Clarksburg to teach predominately in a replacement setting.

**Murphy.**

“Murphy” began teaching in the late 1980s as a departmentalized special education teacher. The next experience was three years teaching in an elementary communications handicapped class. Murphy left public education for a few years to work in a preschool, eventually becoming its director. Murphy missed teaching and returned to the public school setting, first as a long-term substitute, then as a permanent teacher within Clarksburg.

**Robin.**

“Robin” joked that it was a “bleeding heart” that led to a career in teaching. Robin, whose degree specialization is in emotional disturbances, first began teaching in a program for troubled youth in the early 1980s. Robin moved to Florida and, for several years, taught students with emotional disturbances in a self-contained setting. Robin moved to New Jersey and spent six years in a self-contained behavior class and six years in a resource room.
All four participants are tenured teachers. Each participant is Caucasian; there were no teachers of color within the district at the time of the study. Several of the participants live in Clarksburg and have had their own children attend its schools. Lee and Robin have both taught for over 20 years and have been in the district for at least 10. Alex and Murphy have taught for over 10 years, most of that time within the Clarksburg Public Schools. The four participants have an extensive understanding of students with severe learning disabilities. Most of their experiences involve self-contained special education classes, with less experiences with in-class support regular classroom settings.

Review of the Purpose and Research Questions

The purpose of this study was to understand the relationship between student assessment and the teaching practices of special education teachers during the implementation of a new evaluation system. Specifically, it examined whether or not special education teachers in grades 6-8 utilized the student performance data that they must now collect to change their instruction.

The overarching question that guided the research was: What are the experiences of middle school special education teachers at a small suburban school who are newly mandated to use standardized student data to inform their pedagogical and curricular approaches? Two secondary questions contributed to the research:

1. Do teachers feel the data gathering has proven beneficial?
2. Do teachers feel the use of data from SGPs and SGOs is conducive to the special education setting?

The participants in this study provided candid, personal accounts of their experiences in adjusting to the state data collection mandate. Three distinct overarching themes emerged: confusion and anxiety over the development process, misgivings over the process with special
needs students, and acceptance of the new requirement. Within these themes, two additional subthemes were identified: hostility toward a district-required minimum scale for SGOs and a greater attention to standards.

**Theme One: Confusion and Anxiety Over the Development Process**

All of the participants acknowledged that they were familiar with assessment data to monitor student progress. The most commonly identified form was teacher made tests. Each teacher also referenced state testing, while two referred to MAP testing (Measure of Academic Progress). The new state mandate, however, required using data in a format that was new to each of the teachers—SGPs and SGOs.

Each teacher was asked whether or not they had done their own research on SGPs and SGOs. Three responded that they had done their own research. Lee had done some research and found most of it to be negative, including a lot of “conspiracy theories” as to how the data would be used to go after teachers’ wages. Lee had attended a professional conference that discussed “successes and challenges” of SGOs and how teachers could improve the process. Robin also did research and concluded “what we are being told to do isn’t how I would interpret the information that is out there for SGOs and SGPs.” Robin believed the minimum scores established by the Clarksburg district (discussed in detail below) go against the ideas of measurement inherent in SGOs and SGPs.

Alex admitted to more extensive research on the topic. Alex read research on EngageNY, the state of New York’s educational reforms, and said there was “no correlation between SGPs and teacher performance.” Alex described findings where a teacher could be highly effective one year and the very next year be partially effective because of SGP data. This
research caused Alex to be suspicious of SGPs, especially when used with students with special needs because the tests are “so above their instructional level.”

Three of the teachers felt that the purpose of the SGPs and SGOs was to measure the growth of academic achievement of their students. Alex conceded the purpose of SGPs is to “compare students against the norm” but wondered about the effectiveness of the data because of the students’ abilities. Alex added, “If the test is above the student’s grade level, then it doesn’t matter how they are going to perform because that number is going to be incorrect.” Robin was adamant that any visual test given to students “is a reading test” and was “not actually measuring what they know.”

Murphy believed the SGPs and SGOs were designed to keep teachers “on track” and “accountable.” Murphy did not find fault with that purpose but worried if the process could become “a numbers game.” When questioned, Murphy explained that the numbers used in SGOs could be manipulated to make them “less rigorous” or adjusted to make the score of a teacher fall.

Alex dismissed those who described the SGP as a diagnostic tool. Alex described the SGP as a “school district wide programmatic data thing” that the administration could use to assess and revise curriculum. Alex suggested SGPs could be used to see “trends across grade levels.” It could improve the district as a whole, but there was “no correlation” between a higher SGP and better teaching performance.

The roll out of the new SGP/SGO requirement coincided with a revamping of the district’s teacher evaluation system as well as a conversion to electronic lesson plans. Robin reflected the overall sentiment when stating “it is a shame that we have one more thing being mandated that we have to do.” Lee admitted to being “a little nervous” about the new policy.
Murphy, who was apprehensive about trying anything new, was not resistant to the specific changes but definitely felt stress because “there was so much change going on at once.” The experience was “very overwhelming.” Murphy wondered, “How many things can we deal with as teachers?”

Beyond the general feelings of confusion generated through the implementation of a new requirement, the teachers had a few specific difficulties they encountered while developing their SGP/SGO. The experience of Lee highlighted the angst undergone through the SGO process. During the interview Lee’s tone still exhibited uncertainty about the procedures. Lee explained, “My biggest difficulty was I panicked first.” Calmness would later ensue, but Lee “never really had a clear idea of what was expected.” It was “like a rabbit hole” trying to piece everything together and make it coherent. Lee missed the “old fashioned” teaching series that provided scripted guidance that a teacher could follow until comfortable enough to augment the lessons.

Lee noted that the process was “challenging” and upsetting because “I didn’t feel that anyone in the world could achieve that [the set goals].” Lee lamented that more time was spent “looking at the grade on their test, than on what I’ve actually want them to learn.” Because of the complexity of the process, Lee did not foresee using more data in the classroom, at least not as the “primary” source of information about the students. Lee also speculated that the requirement would become something people “just do” and “forget about it” until there is an email reminding them to “turn it in.” Lee wondered if teachers’ plans were even matching up with the goals and standards.

Both Alex and Murphy complained about the extra time it took to develop the SGOs. When asked about coping with the extra responsibility, Murphy laughed and responded, “I didn’t. I had a nervous breakdown.” Colleagues helped with understanding the process but the
time needed to complete SGOs would take hours after school. Alex explained, “You’re doing extra paperwork to get points on a paper.” The points for the evaluation were equated to how the point system for figure skating had changed over the years. Alex coped by significantly modifying the tests, stating, “It’s all about data manipulation.”

The time dedicated to administering SGOs also meant that time was taken from other lessons. Alex disliked having to cut “more research based lessons” because “your SGO is the end game.” The general consensus of the participants was that teachers must concentrate on the test because its score is a significant portion of how their performance is measured. The reality is that teachers want to “protect” themselves by making sure the students are prepared as possible. This district uses the Danielson (2007) model to evaluate teachers. Alex explained that the model wants cooperative group work and discovery learning, which does not work with SGO/SGP preparation, which tends to be “skill and drill.” Alex saw this contradiction as a “balancing act.”

Robin felt the type of assessments added to the confusion and anxiety. Robin chose to use grade-level tests for the students even though the reading level of the tests was beyond the students’ abilities. This was done to make the SGOs similar to the state tests, which are given by grade level. Robin would prefer to test the students at their level, but teachers are required to select goals and objectives from grade level standards, not “the level that a student might be working.”

One problem expressed by two of the teachers involved the students themselves. When discussing the standardized MAP test, Alex noted that the data “isn’t all the way accurate.” The test involves one day, one “snap shot of what the student can do.” Alex felt the test results could be skewed if the student rushed through the test, or if the student did not give their full effort. In
addition, when teachers are not permitted to modify the test, Alex wondered if the students even understood the questions being asked. Students could receive a lower score because of something beyond their control. Conversely, Alex said there were students who “got lucky on a multiple choice item” and scored higher than they were performing in class.

Murphy agreed that student behavior significantly affected test results. On classroom assessments, Murphy is able to correlate a score to a student’s behavior on that day. Murphy could check to see if a student was “having an off day and then their grades were low, or were they having a good day and their grades were still low.” Murphy joked that a “fifth column” should be added to the SGO chart that stated “Not having a good day.”

Minimum score.

Each of the participants had strong feelings about the new requirement for SGPs and SGOs. Few were positive. Alex was the most blunt, saying “it is a waste of time.” Much of the anger was focused on how the school district chose to categorize the data.

Each teacher developed a SGO in collaboration with a building administrator. Students were divided into three tiers: low, medium, and high. Expectations were set for each of the tiers. The district established that students needed a minimum score of 70% to be deemed successful and count towards the teachers’ SGOs. This process presented challenges to the teachers and highlighted professional differences over the tiering process.

Lee, who was “confused” and “upset” by the whole process, had a “hard time” placing the students into tiers, particularly with sixth grade students who were only in the building a few weeks when the process occurred. Lee was never “clear on the percentages” and how they were determined. Being unfamiliar with the students and not being given “enough” information added to the confusion. Similarly, Murphy joked that “my brain can’t fully understand” how the
students were grouped but conceded that administrators have been “extremely helpful.” Lee added that the administrator was good at helping fill out the form and calculating the scores and percentages. However, Robin found “no flexibility” with the administration because they stuck to the minimum score.

Alex, who felt that SGOs “used correctly” should be part of the teacher evaluation, complained about the mandatory minimum which “forced” students to be placed into the groupings of high, medium, or low. The students set into each group were “based off of percentage numbers and not necessarily what the individual student is.” Alex felt the scores were “arbitrarily set” and that the administrators “pretty much dictate the math.” The way the district developed the target scores was “ridiculous” and “not fair.”

Robin felt that teachers were “pigeon-holed into cut-off numbers” when placing students into groups, which was “influencing not only the information but … the validity of the test.” A “bar was set” and the teachers are “stuck” with it. In fact, Alex felt the process as it is was had become an “exercise in compliance.” It was something teachers must do so that a number could be “put to paper” and then applied to a teacher.

Robin asserted that SGOs were not “equitable depending on the content area that you teach.” Students in science, social studies, or math “are able to learn and spit back” whereas English tests are “all application.” Robin felt this disparity could penalize teachers based on what they teach.

Attention to standards.

Each participant was asked whether there were any events that occurred during the implementation of the new student data requirement that they would describe as positive. Three
of the interviewees believed that they had become more consciously aware of the curriculum standards and expectations.

Alex commented on the broader area of teacher evaluation. Alex felt there was a new focus on what teachers needed to do. There are “clear benchmarks as to what you need to be doing in the classroom. It says it right on the sheet [evaluation form].” Alex felt that the score assigned to a teacher needed evidence supporting it, and that was good because a teacher “can look at things and implement them into your teaching.”

Murphy said that the new use of student data “made me look at what I was teaching and the method that I was teaching it and really became more individualized.” It also made it easier to approach the administration for something because it could be validated with data; it “happened twice this year, which is very nice.” The data also enabled Murphy to be more reflective. Murphy shared how in an in-class support class a recent use of data showed that a majority of students “weren’t successful.” Discussing this with the general education teacher, Murphy indicated that it could be something the teachers had done wrong and that they should “redo this.” Murphy was not sure if this would have happened without the data to help convince the other teacher.

The new data collection requirement caused Robin to “actually look at and use the common core assessments” and to “design my assessments in a different way.” This helped make the tests more useful because they gave more information. Robin was able to “pinpoint” areas of weakness better than with previous assessments. Alex echoed similar ideas stating that “it forced me to take a look at the standards.” Like Robin, this translated into changing how teacher-made tests were constructed. “We took a little bit more careful look as to how we developed assessments for students.” This helped provide a more “authentic assessment” of
what the students are doing. Alex added that even though there is a better understanding of the standards, “it doesn’t mean I’m going to be able to hit them with every single student.”

The one exception was Lee, who believed the new attention on technology was a positive development from the new process. Lee indicated that there was new awareness about “access to computers, computer labs, and equipment.” Lee hoped that with more people “concerned” and with the recognition that improvement was needed that “more will be done to fix it.”

**Theme Two: Misgivings Over the Process With Special Needs Students**

This study interviewed special education teachers to explore how the SGP and SGO processes affect special education programs. Each of the participants indicated that the current practices did not fit well within special education. Murphy expressed concern that the process affects special education “greatly” because “our students’ growth is minimal versus general ed kids” who have “at least six months to a year’s growth.” Murphy gave an example of a student who showed “huge amounts of growth” when remembering to put a capital letter at the beginning of a sentence and a period at the end. Robin noted that many of the students coming into middle school special education read at the second or third grade level. So, even if a year’s growth is achieved, the students will only be reading at the fifth or sixth grade level by the time they are done middle school. Murphy finds no “usefulness” in continually testing such students beyond their ability.

Lee stated, “I have the luxury of creating assessments” in a self-contained classes. These assessments helped to “focus” the instructional needs. Lee was concerned that the plan to reach all the core curriculum standards did not correspond to Lee’s philosophy which was to build a foundation and then move forward. Lee objected to the students having to “find success at their grade level” while still working on the foundation. “I have beginner readers” that had to pass
Lee showed examples of test questions where students had to use precise language and gather relevant information. The students Lee taught had a very difficult time because of the “injuries to their brains.” “Even gathering words to speak” could be challenging, Lee shared. Because of their weaknesses, Lee would “never have picked these goals for my students.” The state picked the goals. Lee wanted to be allowed to use a separate set of goals for the special needs students that “gave me a better gauge of how successful they really were.”

Alex concurred. Alex described the new goals as expecting students to “not only master and remember but to apply it at a higher level.” That higher level of thinking was beyond what many of Alex’s students “are developmentally able to do.”

Murphy was comfortable using informal assessment data and questioned the use of SGOs. When pressed, Murphy recognized that informal methods were subjective and were “more of a gut feeling.” Murphy would teach a lesson and if the “kids didn’t seem to get it” would reteach it later that day or the next “in a new manner.” Recognizing that the students needed a different approach or further reinforcement, Murphy kept trying until the students had achieved success. The informal methods used enabled lessons to be quickly individualized to help students master the material. Alex also “tended to rely” on teacher observation and the “natural skill set” that teachers possess to evaluate their students.

Murphy does not think the efforts to complete SGOs had made a difference. Murphy declared, “I can’t compare last year’s students to this year’s students because they are totally different.” Murphy worried that “in a sense that is what we are doing.” In the “special ed world” what worked with last year’s students may not be useful with this year’s students. Even within the same year there are vast differences. Murphy has two replacement classes that “look nothing
alike.” The structure is the same but what is covered is “totally different” because of the personalities, behavior problems, and learning disabilities.

The current process is “forcing” teachers to place students in the same category because there needs to be high, medium, and low tiers. In order to have enough data for an SGO, Murphy must combine the students from two different classes. Murphy pointed out that no student should be in the high group since “that is why they are in my class.” Yet, students who cannot read on grade level are expected to reach a certain score, rather than have it reflect the students more individually. For example, the SGPs and the PARCC test are “not really testing growth,” as Murphy explained. “That’s testing how well they did on the test.” Nor is there any way to account for student behavior during the test. Murphy raised the issue of scoring teachers based on student test scores when those scores could be influenced by students’ emotionality.

Alex seemed a little more open to using the new type of data within special education settings than the other participants. Alex, who found no value in SGPs, believed SGOs could provide meaningful data. Alex believed standardized tests to be “so above” the instructional level of the students that the data was “unusable.” SGOs, however, could assess the students on what was being taught in class. The students would not be judged against each other, as in a SGP. Rather, students’ individual growth would be assessed. SGO tests could be in many formats. The formats could be “individualized” to meet the needs of each student. Alex felt SGOs were helpful because they utilized “various data points” to provide an “authentic assessment.”

Minimum score.

The confusion and concerns over the district’s required minimum SGO score were highlighted when discussion turned to special education. Despite believing in the value of
SGOs, Alex, as stated previously, did not agree with the mandatory minimum student score of 70%. Alex was more specific when explaining that in a typical special education class there may be one or two students in the “higher group,” while administrators want four or five. Alex felt the administrators want the tiering like “a nice bell curve” but in “the reality” of a special education classroom it “doesn’t always match that way.” Alex provided an example of a hypothetical student that initially scored a 20 on an assessment. Later, if the student scored a 60 on a similar assessment, he suggested, the assumption would be, “that’s a lot of growth.” Yet, the student did not achieve the minimum score of 70 for it to “count for a teacher.” Despite significant growth the teacher seemingly failed this student. From Alex’s point of view, the SGOs were not being used the way they should.

Robin expressed similar concerns. Even with students working significantly below grade level, there is a cut-off score of 70. “In a perfect world,” Robin suggested, those students would be assessed on the appropriate grade level test and then the teacher could decide reasonable but challenging goals for the student even if the target score were below 70%. The whole process “infuriates me” because the school is “not using common sense.” Robin believed that students need to take an assessment at their level, showing their growth and achievement.

Theme Three: Acceptance

All the interviewees accepted the realities of the requirement to use student data as part of the evaluation process. Each made adjustments to make the SGO process align with their own philosophies as well as their own classroom procedures.

The interviews took place in the second year for the SGOs, therefore, teachers were more familiar with what was expected of them. Alex, for example, changed the format used for SGOs from one end-of-year post-test to smaller unit tests. This enabled the measurement of students’
growth throughout the year. It also enabled the process to be as normal as possible so that there is no “special SGO thing.” Alex liked how the process fell in line with what the class was already doing. Robin also had more frequent assessments the second year, usually after each model curriculum unit. When asked if the increased number of assessments helped “reach” more students, Robin emphatically replied, “Absolutely not. My whole job is reaching students, so whether or not I am formally assessing them, I’m assessing them on a daily basis anyway.”

Murphy gave a benchmark test at the beginning of the year and gave a follow-up test mid-year. This was done in a team-taught support class and enabled both teachers to check the progress of the students. In the replacement classes, Murphy described being “more diligent in collecting and recording data” about the students.

Three participants hoped that the process will get easier next year when more student data is available. Lee said, “I can form a base” this year that can be used next year. Alex concurred that teachers were still adjusting their assessments and will continue to do so from year to year. This is necessary because each year the students are different. At the same time, each year gives teachers a greater set of resources, many of them self-made, that they can use.

Lee found it is important to get “very familiar” with the standards so that teachers knew what to focus on and how to organize the course. Lee wondered if teachers in the same subjects would benefit from “shared resources” such as tests. Tests were particularly challenging to make because each question should be connected to one of the core standards.

Alex suggested that teachers should “be careful” about the rigor of their assessment. If the test is too difficult, Alex believed the data “is not going to go your way and it’s going to impact you.” Alex reiterated that if teachers could lower the target score to below 70, a test could be made more rigorous because that would be “more appropriate” for the student. As it is
now, Alex explained, a rigorous test would go over the head of most of the special education students and “failing” grades could hurt the teacher, so there is a practical incentive to keep tests “easy.”

While all four teachers seemed to accept the new SGO mandate, they realized that they needed more time to understand and implement them. Lee believed there needed to be more time for teachers from different grade levels to communicate with one another. This would help them set goals. It would be particularly valuable if the middle school teachers could talk with fifth grade teachers. Lee felt the “personal observations” shared by colleagues would be extremely beneficial. Alex wished for more time to focus on “the instruction piece” of fulfilling the standards.

Murphy spent an enormous amount of time “painstakingly” making six different comprehension tests matched to the reading levels of the students. Because of their uniqueness, the tests needed to be administered individually. This required adjusting classroom procedures to meet the realities of such specialized assessments. Murphy admitted it was “a lot of work,” but it was worth it “to make the SGO valid.” Murphy added, “In good conscience I cannot give them [all] the same test.” Special education students did not have such modifications on the [state] standardized tests, but whose lexicon, the teachers acknowledged, was well above the students’ reading level. Similar to Murphy, Alex made three different tests to assess students. Each test had a different reading level while still “hitting the standards,” even if the standard was “modified.”

Murphy explained that the SGOs made in support classes are set by the general education teacher. Murphy advised modifying tests based on student IEPs. Typical modifications include: extra time, fewer choices, and a different format. One student needed to have each test read to
him. As Murphy explained, the student was a “genius” but “reading was a struggle.” Robin said that teachers need to “modify to the point where your students are comfortable in doing the assessment and feel proud that they’ve achieved something.” Robin’s most important modification is reading. If the students can hear the story and read along, “the difference in their performance is night and day.”

Conclusion

Implementing the new state requirement caused much confusion and anxiety amongst the four interviewees. They did, however, adjust to accept the role of SGPs and SGOs, where they could meet the requirements with the least disruption to their instructional routine as possible.

Requiring special education students to show a growth score with a minimum of 70% concerned all the teachers. Lee was “nervous” but not because of the standards. There had always been standards; “they just didn’t have fancy names.” Lee was nervous about being held accountable. She figured “down the line someone is going to say ‘How could this student have gotten an 87 on this [SGO] when she can’t read?’” Similarly, Alex was bothered by the students with the most severe disabilities having to be assessed using the same standards as other students. Alex would like to be working with those students on more life skills, but the administration will no longer allow that to occur. Murphy believed that having students who are at least 2-3 years behind in reading assessed using grade level tests “doesn’t seem quite legit.” Using grade level standards are “tough” with these students.
Chapter 5: Conclusions, Implications, and Recommendations

The education of children with disabilities is a fairly recent undertaking. Whereas formal schooling, at least for the elite, goes back millennia, a commitment to educating the masses, specifically those with special needs, dates less than two centuries. Yet, since the earliest schools for the blind and deaf, there has been a radical change in how those with disabilities are viewed which has enabled all students, regardless of their challenges, to obtain an education. This change has roots in early educational pioneers as well as in new legal interpretations of full citizenship rights.

This inclusion, however, now faces a contradiction. The latest trends for accountability require students with disabilities to be assessed just as their peers. The goal of increasing standards may curtail the development of the students it is intended to help. It is clear that a balance between the needs of students and the needs of society to assure academic rigor is difficult to reconcile.

The idea of improving student performance through teacher evaluation is not new. Yet, the increased momentum requiring the use of student data in the process is. New Jersey has been awarded federal money as part of the Race to the Top Program. This money will help finance educational reforms, including using test scores in teacher evaluations. Districts must now develop more comprehensive teacher evaluation instruments with the underlying hope that more rigorous evaluations will improve classroom instruction. The New Jersey State Board of Education established that up to 80% of an evaluation’s score be based on administrative observations, with the remainder based on student testing (New Jersey Department of Education, 2014).
The state has adopted the Student Growth Percentage (SGP) model. Rather than comparing, for example, all fourth graders to each other, this model compares a student’s academic growth relative to his or her peers who began at the same level. This growth can then be credited, in part, to teacher instruction. For subjects that do not utilize standardized tests, New Jersey permits the substitution of Student Growth Objectives (SGOs). The literature indicates that no existing model, however, can be expected to definitively determine the extent to which an individual teacher contributed to a students’ academic growth (Baker et al., 2010).

The Clarksburg Middle School, the site of this study, is a grade 6-8 school. It has approximately 350 students, 64 are classified as special education students. Nine teachers work in the special education department, four of whom volunteered to be part of this qualitative study. The purpose of this study was to understand the relationship between student assessment and the teaching practices of special education teachers during the implementation of a new evaluation system in New Jersey. There were several themes that emerged from the one-on-one interviews that contribute to a better understanding of what teachers need to help them fulfill their obligation to use student data.

The primary question that guided this study was: What are the experiences of middle school special education teachers at a small suburban school who are newly mandated to use standardized student data to inform their pedagogical and curricular approaches? The interview protocol consisted of 11 questions (included as Appendix C). Because this research used an interpretative phenomenological analysis approach, the researcher often went ‘off script’ to get a better understanding of the participants’ experiences regarding data gathering and implementation.
All four teachers had past experiences that motivated them to not only enter the teaching profession but to concentrate their career within the field of special education. Each of them had taught for over a decade. The teachers had experienced other educational reforms, but the changes implemented under AchieveNJ, the name for educator evaluation in New Jersey, were viewed as a needless intrusion and departure from what they believed should be taking place within their classroom.

**Special Education and Student Growth Percentiles and Objectives**

In the 2014-15 school year, the New Jersey Department of Education required student data to be an integral part of teachers’ evaluations. Teachers in grades 4 to 7 and who taught English and math who have at least 20 students with baseline data were able to use Student Growth Percentiles (SGPs) as part of the process. All teachers, whether they use an SGP or not, must complete at least one Student Growth Objective (SGO); those without SGP data must develop two SGOs (New Jersey Department of Education, 2015). None of the participants in this study met the qualifications for an SGP, so they were each required to complete two SGOs.

The information provided by the state of New Jersey regarding SGOs is straightforward. Though, it would seem, it is intentionally vague to allow for individualized interpretation at the local district level. For teachers, the SGO is intended to provide “a method by which teachers can improve their practice while clearly demonstrating their effectiveness through student progress” (New Jersey Department of Education, 2015, p. 1). AchieveNJ (New Jersey Department of Education, 2015) has four criteria that all SGOs must meet:

- Specific and measurable academic goals that are aligned to state academic standards
- Based on student growth and/or achievement using available student learning data
- Developed by a teacher in consultation with his or her supervisor
New Jersey state code 6A:10-4.2(e)3 explains that SGOs are to be developed as a collaboration between teachers and administrators. In its own guidelines, AchieveNJ recommends that administrators help teachers “take ownership of the SGO process as a powerful way to improve teacher practice and student achievement” (New Jersey Department of Education, 2015, p. 2).

Wolf (2007) believes accountability tests like SGOs are “critical methods for incentivizing desired behavior in the education environment” (p. 692) and should be promoted by public policy. This sentiment was echoed by “Murphy” who wanted “valid” SGOs because that would help “evaluate how I’m being as a teacher.” Murphy admitted that thinking about SGOs is “partially driven” by the desire to score 3s and 4s on the Danielson evaluation rubric and because “I don’t want to be a 2.” Murphy’s “perfectionism” would be reflected by earning the top score of 4 on the evaluation as higher scores suggest higher levels of performance.

Wolf (2007) also believes in more frequent diagnostic tests, especially in special education, which often “eschews” regular assessment (p. 691). He considers “diagnostic” testing to include teacher made tests, unit tests, final exams, and standardized tests not primarily used for accountability. Wolf (2007) adds that students who have trouble learning to read or have a learning disability “probably benefit more than the average student from regular assessments” (p. 698). Testing, according to Wolf (2007), should provide “individual-level data” (p. 692). These beliefs are exemplified by the special education program at Clarksburg Middle School.

Prior to the new state evaluation requirements, the interviewed teachers frequently assessed their students, though the assessments were often subjective. The SGO process has
helped them quantify student performance. The participants typically followed the format of a pre-, or benchmark test, followed later by a post-test, whether at the end of a unit or at a predetermined point, such as mid-year. Marion, DePascale, Domaleski, Gong, and Diaz-Bilello (2012) do not find the pre-/post-test methodology desirable. According to Marion et al. (2012), using a pre-test makes sense “in an instructional context but is fraught with problems” when used to measure teacher performance (p. 5). They believe that “even in the best cases…growth outcomes at the class level are simply not stable” and that developing a reliable pre-/post-test would be “problematic” (p. 6). Marion et al. (2012) assert that because of the difficulty in finding one good assessment in non-standardized test areas (e.g., science and social studies) it does not “make sense” to believe there are two appropriate tests, which would be necessary for a good pre-/post-test format (p.6).

Marion et al. (2012) recommend that the end of a unit or end of the course assessment “of the highest quality available” be used without using the pre/post-test format (p. 9). The results could be analyzed with the most appropriate model available, such as ANCOVA. Yet, the authors concede that most districts do not have the resource ability to analyze data in such a manner. Because of such limitations, they recommend tiering, so that the objectives could be differentiated for students but caution that students could be poorly placed. Tiering takes place in the Clarksburg district.

Marion et al. (2012) prefer the use of commercial assessments “if they meet the critical alignment requirements and users can document that they are appropriate measures of the learning objectives” (p. 8). The challenge in this recommendation is that, as “Lee” admitted, “I’m not really a test maker.”
It is appropriate to question, therefore, if teachers and administrators have enough training to recognize when a test is valid. The importance of this question cannot be understated. Marion et al. (2012) believe it difficult for classroom teachers to develop a pre-test/post-test that would be a reliable instrument to be used in the performance evaluation of teachers.

A “Numbers Game”

Holdheide et al. (2012) postulates that attaching SGOs to a high-stakes decision like the retention of teachers could motivate some teachers to set goals that were easily obtainable. This temptation should be curtailed with administrators’ involvement in the development of SGOs. However, 10% of the score of each building principal’s annual evaluation is based on the average of the teachers’ SGO scores within the building (New Jersey Department of Education, 2014). While this is a small percentage, it cannot be ignored. Some principals may have an incentive to support teachers who set relatively easy SGOs. There is no evidence to suggest administrators in the Clarksburg district have succumbed to the temptation of easier goals, since the district set the minimum passing score for the SGOs at 70%, the same percentage students need to pass an assignment.

The AchieveNJ guidelines do not mention the establishment of a minimum score. The minimum score of 70% established by the Clarksburg administration angered the four interviewees. Tregaskis (2006) believes that disability studies should investigate whether policies like Achieve NJ align with the inclusive approach of special education. Based on the teachers’ feedback, the 70% minimum requirement in an SGO is incompatible with the type of growth typically seen in students with special needs.

While Ysseldyke et al. (2004) believed that raising the expectations for students with disabilities can be positive, none of the interviewed teachers found the “arbitrarily set” minimum
score of 70% to be beneficial to their student population, nor the quality of their instruction. The teachers were not against challenging their students. Rather, the score was viewed as something that did not align with the realities of the pace at which special needs students advance. A hypothetical student could improve 50 points on an assessment, moving from an 18 to a 68 which is a 277% increase, but the teacher would not meet the standard because the student did not earn a 70. “Murphy used an example of a student who advanced to using a capital letter at the beginning of a sentence and the proper punctuation at the end. For the student, the change was “huge,” but measuring that growth, which by grade level standards is small or assumed, would be difficult within the changing evaluation system.

There is legitimate concern, if not outright fear among the participants, of what could happen because of SGO scores. Teachers like Alex who understand the mathematical complexities when developing SGOs admit that the process is about “data manipulation.” Alex advised against making SGOs that are too rigorous, especially in the case of special education when grade-level standards used for the SGOs are above what is “appropriate” for the students.

Upon reflection, the researcher wondered if some SGOs are set artificially low because the teachers are required to use grade-level standards that seem unreasonable for certain students. They could also be set artificially low to make sure the teacher does not score poorly on their evaluation because the students are not assessed at their level. Would the school get a more accurate measurement of growth if the mandatory minimum score were eliminated, especially with special education students?

Meek (2006) emphasizes that students with special needs “are equal” to their non-disabled peers, but they are not the same. Consequently, assessments should be modified to better meet the needs of the individual child. Meek (2006) states that schools need to “sample
and document" the progress of special education students “in a more humane way” (p. 295). She explains that many state tests are written above the stated grade level, which is even more detrimental to special education students, leading to scores that “provide no meaningful information” (p. 296). All four interviewees described assessments developed outside of their replacement or self-contained classes that were written well above the reading level of their students. To help students with special needs the state provides a list of approved testing modifications (New Jersey Department of Education, 2010a). However, the state does not permit the reading passages themselves to be modified, nor does it allow the reading passages on the English Language assessment be read aloud. Teachers can, however, read other portions of tests, assure directions are understood, and record answers the student dictates.

Robin expressed most succinctly what each of the interviewees had conveyed: teachers must “modify, modify, modify” assessments to meet the students’ needs. As long as the state requires standardized tests, students with special needs will need to take them. The best option for teachers is to assure that necessary modifications are incorporated into the students’ IEPs so that the testing reasonably measures the students’ abilities.

The situation is different with SGOs, which are often teacher made. Within their own special education classes, Murphy and Alex spent a lot of time developing several versions of tests so that the reading level matched the students’ present level. This methodology led to a more accurate assessment of the students’ abilities and growth. The SGO procedures could also be adjusted, making them less rigid and daunting for the students.

All four interviewees expressed concerns that students became frustrated with assessments that were written beyond their abilities. Two described how a student’s mood or emotionality could affect testing results. Murphy was particularly concerned that a student’s
behavior could cause the child to do poorly, which would adversely affect the teacher’s evaluation score. Alex expressed similar concerns, believing that students who do not give their complete effort skew the results. Upon reflection, the researcher accepts that students’ behavior and attitude can affect test results, as it aligns with experiences in his own classroom. When tests are a factor in determining whether or not a teacher maintains their position, it becomes disconcerting.

**Focusing Standards**

Defur (2002) believes the current systems of standards “does not acknowledge or value that there are multiple views…of demonstrating one’s competence” (p. 210). The interviewees had similar misgivings. The special education teachers found the grade-level standards to be above the functioning level of their students. They found that their students learned at a slower pace and that, despite the best efforts of previous teachers, their students had not caught up to grade level. The students’ slower pace of learning had a cumulative effect of having students who were several grades behind their peers. Yet, they were judged on the same level.

Lee worried that with the volume of core curriculum standards, a teacher would not be able to provide the foundational instruction necessary for special education students to move ahead at the pace necessary to close the gap with their peers. Yet, students are tested on the grade-level standards. Lee added, “I would never pick these goals for my students.” Lee wished that the state would permit separate goals for students with special needs, goals that would “better gauge” how successful they really were.

Alex expressed trepidation regarding the standards. Clearly, Alex felt that some standards did not align with what the most needy students would require. There seemed to be a
shift away from life skills and more towards academic skills. Alex would prefer to be permitted to use standards that fit the students.

According to Robin, the use of the common core standards “allowed me to design my assessments” in a way that provided more information. This focus on the standards allowed the “pin pointing” of skill deficiencies. Despite this praise, Robin firmly believed the best way to assess students is through daily observation, suggesting that day to day interaction with the students revealed areas of weakness to address.

Conclusions

I began this study to investigate how teachers perceived and then implemented the Student Growth Percentiles or Objectives as a requirement of their evaluation. I wanted to specifically understand whether this requirement “fit” into special education. I found that the teachers, while somewhat annoyed with adding something new into their busy schedules, were not outright opposed to the idea of SGOs. Nor were they doing anything to impede the implementation of the new policy. The general consensus about the implementation of SGOs was negative, although Alex added that if used correctly, SGOs could provide useful data to help guide instruction. The major concerns of the participants revolved around how the district chose to set a mandatory minimum score for the SGOs, something not found in the state guidelines or in the literature.

Framework.

The intent of Disabilities Studies in Education (DSE) is to provide a better understanding of the daily experiences of students with disabilities. DSE can also provide ‘voice’ to those with disabilities (Baglieri et al., 2011). DSE scholars often use qualitative research to explore the issues of disability within sociopolitical contexts (Cosier, 2012). By examining the experiences
of special education teachers within the framework of new data collection and evaluation requirements, this study provides insights and a voice into the impact of this policy.

The researcher questioned whether SGPs and SGOs were conducive to the special education setting. Overwhelmingly, the participants in this study believed that the implementation procedures did not account for the uniqueness of special education. While DSE argues in favor of the social construct of special education (Baglieri et al., 2011) it does not discount that students with disabilities often need modifications that enable them to receive an education that is as similar to that of their non-disabled peers as possible. The teachers felt needlessly hampered by the arbitrary establishment of a minimum passing score of 70%. They believed it was not realistic for special education students who function well below grade level to be held to a specific grade level standard. The Clarksburg district requirement failed to recognize that a student can demonstrate significant growth on an assessment without reaching a “passing” score and the interviewees provided examples to demonstrate this phenomenon.

Some teachers indicated that they would have to set SGOs with lower expectations that would allow their students to “pass.” This goes against the basic principles of DSE, which strives for all students to be treated as similarly as possible. It also goes against a stated intention of SGOs: to help teachers measure student progress and adjust classroom lessons accordingly.

DSE examines how societal barriers can exclude special education students (Watson, 2012) and how classroom placements can limit student learning by curtailing access to the general curriculum (Baglieri, 2011). This study indicates that the focus on high-stakes assessments has caused special education teachers to pay closer attention to the general education standards. Alex disliked how more practical “life skills” were ignored as the focus
shifted to academics. Yet, requiring all students to utilize the same goals limits the exclusionary barriers that special education replacement classes establish.

**Implications for Practice**

There are a number of implications for practice based on this study’s findings. Part of New Jersey teachers’ annual evaluation is now based on student data (New Jersey Department of Education, 2015). Besides use in teacher accountability, the data can be used by teachers to alter their instruction to meet the needs of their students.

The development of SGOs was a confusing and difficult process for the four interviewees. Murphy felt the directions given were “poor.” Murphy was further bother by the fact that “no two buildings in the district” followed the same procedures. Murphy was “blown away” by what was being done in the name of SGOs in other buildings compared to what was expected in the middle school. Lee is seemingly still bothered by the entire SGO process. Such confusion can not be good for the proper implementation of the state mandate. No reform can be expected to be successful if several of the participants do not understand its processes.

Several of the interviewees used a pre-/post-test methodology, which was somewhat discredited by Marion et al. (2012), especially when used for teacher evaluation. Schafer et al. (2012) feel the use of any VAM, including SGOs, to make high-stakes decisions like teacher retention is very risky. Understanding how to develop more reliable SGOs is important for teachers and administrators as this process continues.

It would be advantageous to explore the expertise that individual teachers possess involving student assessment. It would also be beneficial to understand how teachers compile and utilize their student data. Some teachers are just better at using data and are a potential resource to improve the use of student data.
Developing SGOs takes time, particularly if the teacher individualizes tests, a common occurrence in special education. It is not uncommon for students with special needs to read below grade level, which requires individualized tests for some students. Learning more efficient methods to develop SGOs that are both reliable and individualized help both teachers and students.

**Research Implications**

The information gathered from this study indicates a need for future research. The researcher discovered no comprehensive study on the use of SGOs. This small study does little to fill that void. Further research is needed to discover whether or not SGOs, in particular, reflect teacher contributions to students’ knowledge base. Research is also needed to discover whether a particular method for implementing SGOs is better than others.

Current research seems divided over the use of SGOs to measure teacher performance. For example, Schmitt and Ibanez (2011) connected SGOs to increased state test scores, while Balch and Springer (2015) found no relationship between successful SGO scores and teacher effectiveness. This study did not attempt to investigate the application of SGOs as a tool for teacher evaluation. Rather, this study investigated teachers’ experiences with SGOs. How teachers approach SGOs and develop SGOs can have an effect on their success and consequently their scores on annual evaluations.

The results of this study suggest the need for a similar study with a larger sample size, especially one that uses multiple districts. This study only examined special education teachers’ perceptions. Special education is a unique field with unique challenges. Investigating how other districts or states address the challenge of assessing students with special needs on grade-level
would prove helpful. Research in the general education field would be necessary to understand how their students’ needs may be different.

Limitations of the Study

As with all research, this study has limitations. While capturing the experiences of four participants is appropriate for a phenomenological study, it still only represents a small portion of the population. This study examined only one building within one district in a state that has over 600 school districts. It cannot and should not be generalized as a result.

As a teacher within the same district, I have a clear understanding of each interviewees’ experiences because I must complete SGOs as well. It is possible that my own biases have influenced my interpretations of the data, despite my attempts to minimize them. I was equally as frustrated as the interviewees about having something else to do regarding evaluations, but I did not have the same frustration they each had regarding the minimum SGO score.

The fact that the teachers volunteered to participate could have impacted the results. It is possible that they wanted to help out a colleague, but it is also possible that they wanted to have their opinions heard. All of the teachers who volunteered had a decade or more of experience. Seasoned teachers could be more reluctant to change. It is possible that novice teachers would have had different experiences.

It is also important to consider that this study asked teachers to reflect upon an experience that is relatively new. The “growing pains” that the teachers–and administrators–are working through can cast a negative light to the situation. It is possible that the teachers were adverse to more changes dictated by the state, since several major reforms had occurred over the past five years. Recent reforms included changes to the teachers’ pension system and teachers’ medical benefits. The new legislation resulted in more money taken from teachers’ paychecks and to a
feeling that the teaching profession was not respected. The district itself made a change to require electronic lesson plans during this same period. People who would normally accept change may have been more resistant because so much was happening at once. Good (2001) advocates “small wins” as an appropriate approach to reform. Good feels that too often schools try many large reforms at once and then find difficulty in reaching prescribed goals. It is possible that the Clarksburg district is struggling with too many reforms at once.

Application of Results

As a result of the findings of this study, the following recommendations are suggested.

1. **The district should provide more time to develop appropriate SGOs.**

Marion et al. (2012) assert that “substantial professional development” is needed in the areas of setting and evaluating objectives. Alex believed time was needed each year to adjust assessments to meet the needs of the new students. Because the state of New Jersey requires that data be used for teacher evaluations, it is important that teachers understand how to develop assessments that will not only help administrators with their evaluation but also help teachers understand their students.

2. **Investigate mandatory minimums.**

Because all participants found the 70% mandatory minimum SGO score was not conducive to special education settings, the district should investigate changing the standard. Further insights could be gained through a discussion with the entire special education department of the middle school, as well as at the high school and elementary levels. It is possible the other teachers experienced the same difficulty developing SGOs capable of enabling students to reach the required minimum score. It is also possible that the problem is much smaller and was disproportionately represented in the interviews. The state wants SGOs that “provide a method
by which teachers can improve their practice” and their students’ progress, therefore it is important for the district to assure there are procedures help to make that possible (New Jersey Department of Education, 2015). It defeats the purpose if teachers make relatively easy SGOs just to earn a good score on their evaluation.

3. **The district should provide better explanation of how SGOs are developed.**

Murphy’s concern that the district seemed to have different expectations in its different buildings warrants investigation. While it is understandable that there would be some variation between grade levels and subject matters, there should be a basic format by which all SGOs are constructed. Most of the interviewees found the process confusing. Lee and Murphy, for example, had trouble understanding how the calculations were made. It is not a good practice for teachers to lack an understanding of how something that could affect their employment is derived.
References


Marion, S., DePascale, C., Domaleski, C., Gong, B., & Diaz-Bilello, E. (2012). Considerations for analyzing educators’ contributions to student learning in non-tested subjects and grades with a focus on student learning objectives. *Center for Assessment.*


Vineland Training School. (n.d.) History. Retrieved from
http://www.vineland.org/history/trainingschool/history/history.html

Wakeman, S., Browder, D., Meier, I., & McColl, A. (2007). The implications of No Child Left
Behind for students with developmental disabilities. *Mental Retardation and
Developmental Disabilities Research Reviews* 13, 143-150.


White House Profile. (2009). Secretary Arne Duncan. Retrieved from
http://www.whitehouse.gov/blog/author/Secretary%20Arne%20Duncan


Education*, 82(4), 690-702.


Ysseldyke, J., Nelson, J. R., Christenson, S., Johnson, D. R., Dennison, A., Triezenberg, H., ... &
Hawes, M. (2004). What we know and need to know about the consequences of high-
stakes testing for students with disabilities. *Exceptional Children*, 71(1), 75-95.

cclimate of full inclusion. *Exceptionality*, 17, 189-204.
Appendix A: Recruitment Letter

February 18, 2015

Dear special education colleague:
I am writing to tell you about research I am conducting for my doctoral thesis. The topic is “The Effect of Student Performance Data on Special Education Instruction,” and I have chosen the middle school as my focus.
The purpose of this study is to understand the relationship between the collection of student assessment data and the teaching practices of special education teachers during the implementation of the new teacher evaluation system. Specifically, it will examine the experiences of special education teachers in grades 6-8 with the student performance data that they must now collect from Student Growth Percentiles (SGPs) and/or Student Growth Objectives (SGOs). One goal is to develop a list of “best practices” from which all of us could benefit.
I am interested in all of your experiences, and I am hopeful that you would be willing to help.
The research would involve a one-on-one interview; the time and place can be arranged at your convenience. It would take between 60 and 90 minutes. The questions will focus on how you feel about and how you cope with the SGPs and SGOs. You identity will be kept confidential.
Dr. McAleer has approved this study.
It is your decision whether or not to participate. Your involvement is voluntary. Naturally, I would be happy if all of you were willing to contribute.
Please let me know if you are interested by emailing me at blass.p@husky.neu.edu If you do not email me and volunteer, I will assume you are not interested.
If you have questions, please call me on my cell

Thanks,

/s/ Paul Blass
Appendix B: Consent Form

Northeastern University, Department of Professional Studies
Name of Investigator(s): Principal Investigator, Dr. Sara Ewell; Student Researcher, Paul Blass
Title of Project: The Effect of Student Performance Data on Special Education Instruction

Informed Consent to Participate in a Research Study
We are inviting you to take part in a research study. This form will tell you about the study, but the researcher will explain it to you first. You may ask this person any questions that you have. When you are ready to make a decision, you may tell the researcher if you want to participate or not. You do not have to participate if you do not want to. If you decide to participate, the researcher will ask you to sign this statement and will give you a copy to keep.

Why am I being asked to take part in this research study?
We are asking you to be in this study because you are a special education teacher who is now required to gather student data as part of teacher evaluation process.

Why is this research study being done?
The purpose of this study is to understand the relationship between the collection of student assessment data and the teaching practices of special education teachers during the implantation of a new evaluation system in New Jersey. Specifically, this will involve Student Growth Percentiles (SGPs) and Student Growth Objectives (SGOs).

What will I be asked to do?
If you decide to take part in this study, the student researcher will interview you to examine your use of student data. This interview will involve several scripted questions, though it will also include follow-up questions based on the discussion.

Where will this take place and how much of my time will it take?
You will be interviewed at a time and location that is convenient for you. The interview will take between 60 and 90 minutes. The interview will be recorded to assure accuracy. Once the interview has been transcribed, the student researcher will send you a copy through inter-department mail for you to review. The researcher will then briefly meet with you one more time to confirm the accuracy of the transcript and to give you an opportunity to add to any answers.

Will there be any risk or discomfort to me?
Other than “nervousness” from being interviewed we see no foreseeable risk or discomfort.

Will I benefit by being in this research?
There are no direct benefits for you. Any benefits derived will be professional in nature. It is hoped that this research will result in several “best practice” uses of student data that will be shared will all members of the special education department.

Who will see the information about me?
Your part in this study will be confidential. Only the interviewer will see the information about you. No reports or publications will use information that can identify you in any way or any individual as being of this project. Those interviewed will be given pseudonyms to conceal their identity.

All data from the interviews will be secured to protect confidentiality. The researcher has a locked cabinet at his place of work and a secure file cabinet at home. All files will be stored on a password protected computer.

In rare instances, authorized people may request to see research information about you and other people in this study. This is done only to be sure that the research is done properly. We would only permit people who are authorized by organizations such as the Northeastern University Institutional Review Board to see this information. Members of the local school district will not be privy to this information.

### What will happen if I suffer any harm from this research?

No foreseeable harm should come from this research. It is not physical in nature. The district superintendent has given permission for the interviews, and pseudonyms will be used.

### Can I stop my participation in this study?

Your participation in this research is completely voluntary. You do not have to participate if you do not want to and you can refuse to answer any question. Even if you begin the study, you may quit at any time.

### Who can I contact if I have questions or problems?

If you have any questions about this study, please feel free to contact faculty advisor, Dr. Sara Ewell, at

The student researcher, Paul Blass, may be reached at

### Who can I contact about my rights as a participant?

If you have any questions about your rights in this research, you may contact Nan C. Regina, Director, Human Subject Research Protection, 960 Renaissance Park, Northeastern University, Boston, MA 02115. You may call anonymously if you wish.

### I agree to take part in this research.

______________________________________________
Signature of person agreeing to take part

______________________________________________
Date

Printed name of person above

______________________________________________
Signature of person who explained the study to the participant above and obtained consent

______________________________________________
Date

Printed name of person above
Appendix C: Interview Protocol

**Introduction:**

I have invited you to speak with me today because you are a special education teacher who is now required to use student data as part of your annual evaluation.

To begin this process, I would like your permission to record our conversation so that I can ensure its accuracy. The recording will only be for my use. Do I have your permission to do so? [If yes, begin the recording. If not, plan on taking more detailed notes]. I will also take notes during the interview, so do not become nervous when you see me writing. Your confidentiality will be maintained through a pseudonym; I will be the only one to know your identity.

To meet the requirements of Northeastern University I have a consent form to review with you. The highlights of the document are:

1. This study is investigating your use of student data, specifically SGPs and SGOs.
2. All information will remain confidential.
3. Your participation is voluntary.
4. You may stop at any time.

[Give time to read form.]

Do you have any questions about the interview process or the consent form? [If all is acceptable, have participant sign form.]

The interview will last between 60 and 90 minutes. During that time I will ask questions regarding the use of student data both before and after the implementation of the new teacher
evaluation system. It will be most helpful if you provide me with any and all information that you feel is important regarding your experiences. Do you have any questions at this time? [Once any questions are answered begin interview.]

Let’s begin…

Getting to know interviewee:

Please provide me with some brief background. For example, please tell me how you came to be a teacher?

1. Please tell me what “data-driven instruction” means to you.

2. Before the new evaluation requirement, did you use assessment data to drive your instruction?
   a. Types?
   b. What were your views on its usefulness?
   c. What were the challenges?

Transition: “Thank you. I am now going to move onto questions specifically dealing with SGPs and SGOs.”

3. What do you understand the purpose of Student Growth Percentiles (SGPs) and Student Growth Objectives (SGOs) to be?

4. How do you feel about the new requirement to use SGOs and SGPs?

5. Considering the whole experience of the new requirement for student data, were there any events that happened that you would describe as positive?
   a. If yes…What made them positive?
6. What, if any, difficulties did you have with the development of your SGP/SGOs and how did you cope?
   a. If yes…What influenced how you coped with any difficulties?

7. Developing SGPs and SGOs is to be a collaborative process with administrators. Describe your experiences with that process.

Transition: “Thank you. The next questions will focus on your use of SGPs and SGOs.”

8. Have you done any research regarding the implementation of SGPs/SGOs and if so what were your findings?
   a. If yes…Did/will they change your implementation approach?
   b. If yes…Did/will you discuss your findings with your fellow teachers? What advantage do you feel that will have for student growth?

9. What changes, if any, have you implemented in your classroom as a direct result of SGPs/SGOs?
   a. If yes…Were you apprehensive to try something new?
   b. If yes…Did these changes end up helping you “reach” more students?
   c. If no…Why do you think you did not make any changes?

10. Do you find yourself thinking about the SGPs/SGOs beyond the requirement of having “to do” them?
    a. If yes…In what way?
    b. If no…Why do you think that is so?

11. What, if any, advise or “best practices” would you give to colleagues regarding SGPs/SGOs?
    [Probe. Teacher may not recognize the uniqueness of their practice(s).]
Closing:

We have reached the end of the interview, but before I finish I have one more question.

Is there anything else you would like to share?

Examples of prompts for all of the above questions:

- Can you tell me more about that?
- Can you give me an example of that?
- What did you think of that?
- How did you feel about that?
- What did that mean to you?