What Makes Teachers Effective: Investigating the Relationship Between CABAS® Teacher Ranks and Teacher Effectiveness

Sara Silsilah

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY
2019

ABSTRACT

What Makes Teachers Effective: Investigating the Relationship Between Teacher CABAS® Ranks and Teacher Effectiveness

Sara Silsilah

I examined the relationship between teacher effectiveness as measured by the number of learn units students required to meet an objective and the number of competencies mastered within the categories of teacher repertoires composing the CABAS® rank. Twenty preschool teachers participated in the study. A statistical analysis was used to investigate the degree to which these variables negatively correlated with each other. The results showed that the more competencies teachers mastered, the fewer learn units students required to meet an objective. A second experiment was conducted as an experimental analysis of the correlations found in the descriptive analysis. An adapted alternating treatments design was used to analyze the relationship between the number of competencies teachers mastered and the number of learn units their student required to meet an objective. Four teachers and four teacher assistants participated in the study. The teachers and teacher assistants each taught two sight word objectives for a student with bidirectional naming and a student without bidirectional naming. The results did not show a functional relationship between the number of competencies mastered and a lower LUC (learn unit to criterion). Teachers with more competencies mastered did not present fewer learn units for their students to meet an objective when compared to teacher assistants who had fewer competencies mastered. Possible explanations for a lack of a functional relationship found in Experiment 2 are discussed.

TABLE OF CONTENTS

Chapter I (continued)

Chapter III (continued)

APPENDICES

LIST OF TABLES

Table

LIST OF FIGURES

Figure

DEDICATION

This work is dedicated to my remarkable mother,

Hala Aboulfaraj,

for without her unyielding support, unconditional love, and valuable life lessons,

I would not have become the woman I am today, who is capable of

accomplishing the most challenging yet incredibly rewarding achievement of my life,

such as this one.

Chapter I

INTRODUCTION AND REVIEW OF THE LITERATURE

Identifying variables that contribute to teacher effectiveness is a common goal among some educational researchers from a variety of theoretical orientations (i.e., behavior analysts, Greer, 2002). There is a general consensus among these scientists that teacher effectiveness is reflected in student learning (Fenstermacher & Richardson, 2005; Goe, 2007; Greer, 2002). In other words, researchers in both the educational and behavior analytic fields agree that if student learning occurred, the teacher was effective, and if learning did not occur, the teacher was not effective. Although the definition of effectiveness across fields is essentially the same, there is a noticeable difference with regard to its scientific treatment within each camp.

The difference in methodology inherently lies in the epistemology each field adopts to study the phenomenon of learning. On one side of the continuum, educational research is primarily based on cognitive psychology (Calfee, 1981). Within this realm, learning is conceptualized as mental processes that involve restructuring and coding of information by the learner, who plays an active role in the learning process (Ertmer & Newby, 1993). Stated differently, this approach treats learning as a mental construct. On the other hand, learning from a behavioral perspective stems from the application of the principles of behavior to learning (Greer, 2002). From this perspective, learning is defined as the product of the individual’s interaction with his/her environment. This interaction results in changes in observable behavior such as writing or unobservable behavior such as thinking (Vargas, 2013).

Due to differences in theoretical approaches, behavioral and educational researchers have substantially different approaches to identifying variables that contribute to teacher effectiveness. This divergence has given rise to an ongoing debate between the two camps in terms of establishing effective teachers. Some have argued that a cognitive research approach to education does not provide the teacher with the practical tools she needs to be effective with her students (Greer, 1983; Keller, 1978). Others have proposed that behavior analysis is limited to teaching stimulus-response relations (Ertmer & Newby, 1993) and its procedures are too cumbersome to be incorporated into group instructional settings (Brophy, 1983). Granted, since this critique, the behavioral field has evolved over the past decades to involve better procedures for the effective application of behavior analysis to change socially significant behavior in various disciplines (Falcomata, 2015).

Regardless of the debate, there is no denying that behavior analytic research efforts have yielded a plethora of successful instructional practices that have produced significant changes in student outcomes, particularly within the realm of special education (Hall, 1991). Some of these efforts include precision teaching (Lindsley, 1990), Direct Instruction (Engelmann, Becker, Carnine, & Gersten, 1988), and Personalized System of Instruction (Keller, 1968). Consequently, an effective technology of teaching was established, enabling the teacher to function as a strategic scientist of instruction (Greer, 1983, 1991).

Educational researchers have not had the same level of success in the realm of evaluating teacher effectiveness. A number of researchers have admitted that linking teacher performance with student outcomes has been a difficult endeavor (i.e., Berliner, 2018). Perhaps the difficulty arises from asking the wrong questions (Everson, 2017; Greer, 1991), such as *why* are some teachers more effective than others, rather than asking questions related to *what* teachers actually do to contribute to student learning. For example, some researchers investigated differences in teacher ethnicity and its association with student learning (e.g., Dee, 2004); these types of studies may not account for crucial differences in instruction among teachers such as program planning and instructional decision making. A considerable amount of research on teacher quality, as measured by student outcomes, yielded either inconsistencies across studies or small effects with no practical significance (Goe, 2007, p. 2).

In today's political climate surrounding education, where more laws are being implemented to insure all children are learning (e.g., Every Student Succeeds Act, 2015) and millions of dollars are being spent on educational research that may not provide practical tools for the teacher to meet those demands (e.g., Kane, McCaffrey, Miller, & Staiger, 2013), a reconciliation between the two fields is warranted.

Hence, the following literature review and subsequent study are efforts towards bridging the gap between the two fields. This research may pave the way for a united educational science that promotes practices of effective teaching as defined by student learning. These efforts require a consideration of the dissemination questions posed by Baer, Wolf, and Risley (1987), who asked the following:

> When a program is disseminated, should its disseminators require that its procedures be followed faithfully, no matter where or when the program is used? Or should its users be allowed, and even encouraged, to modify those procedures to fit their local situations and contingencies? (We might first ask, functionally, when we have that choice). (p. 321)

Within the context of teacher effectiveness, we can begin to entertain the questions posed by Baer et al. (1987). As behavior analysts, can we offer a tentative framework for establishing effective teachers? If so, can it be modified at a larger scale so that its application is effective and relevant to the context in which it is implemented? To answer those questions, a review of the educational literature today is warranted.

Literature Review

Teacher Effectiveness in Educational Psychology

A large body of educational research has been devoted to putative factors that contribute to teacher effectiveness. Goe (2007) designed an explicit framework for organizing research on teacher quality in order to "make sense of the many ways in which researchers have been measuring teacher quality over the years" (p. 8).

Goe's framework. The framework encompasses four separate teacher components associated with teacher quality. These components are divided into three categories including: (a) *inputs*, teacher qualifications and characteristics; (b) *processes*, teacher practices; and (c) *outcomes*, teacher effectiveness as measured by student achievement. For a graphical display of the framework, refer to Figure 1.

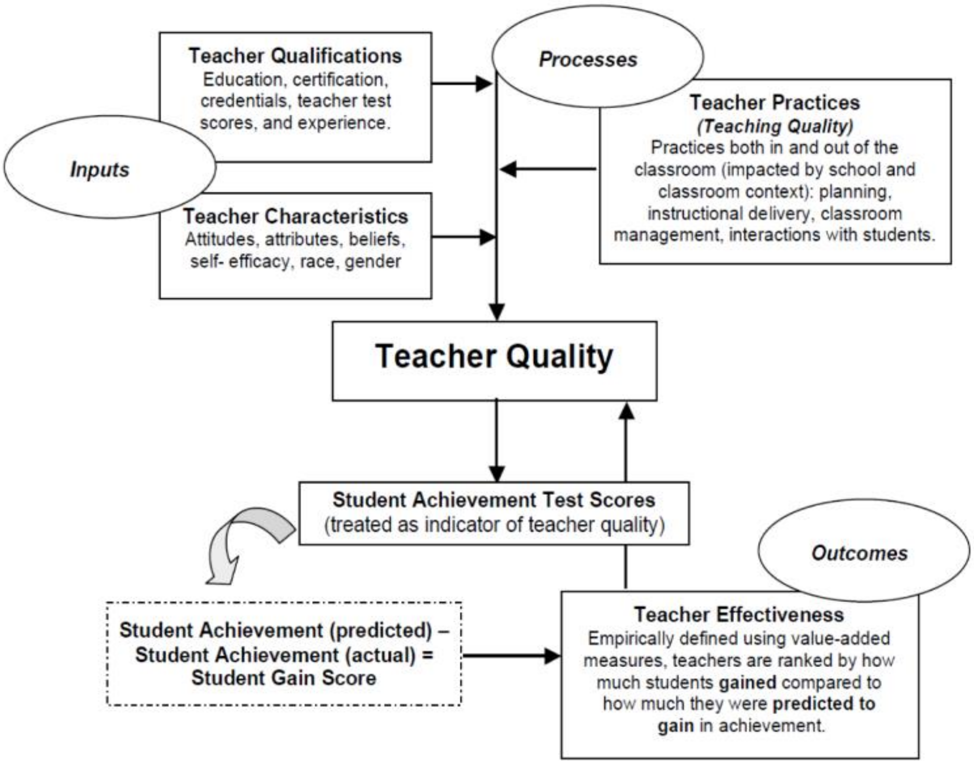

Figure 1. Goe's representation of the teacher quality framework (from Goe, 2007, p. 9)

The above framework was used to analyze research on the relationship between teacher variables and teacher effectiveness from an educational and behavioral standpoint. Analyzing research on teacher effectiveness using this framework may be beneficial in understanding the fundamental differences between the input, process, and outcome in educational and behavioral research.

Teacher Inputs in Educational Psychology

Teacher qualifications. Qualifications refer to teacher credentials such as certification, education, and experience. These variables are a part of a teacher's instructional history, meaning they were acquired through distinct experiences such as attending a university or workshops. Research on these variables mostly revealed inconsistent results in terms of their contribution to student outcomes, except in the subject area of mathematics. A persuasive amount of evidence supports the positive effects of mathematics certification (Goe, 2007; Rice, 2003; Wayne & Youngs, 2003) and degrees/coursework (Wayne & Youngs, 2003) on student mathematical achievement.

Some have proposed that the reason qualifications matter in mathematics is due to the abundance of research on mathematical qualifications, which exceed research efforts in other subject matter (Goe, 2007). Another prevalent finding from this line of research is that teacher experience (i.e., number of years teaching) differentially affects student achievement only in the first few years of teaching; scores begin to level out after that (Hanushek, Kain, O'Brien, & Rivikin, 2005; Rivkin, Hanushek, & Kain, 2005). However, others have found that experience is not associated with student achievement (Gallagher, 2004; Kimball, White, Milanowski, & Borman, 2004).

Teacher characteristics. Teacher characteristics refer to variables that are not acquired through experiences, such as age, gender, and ethnicity. This line of research did not show consistent findings in terms of influencing student achievement (Goe, 2007; Wilson & Folden, 2003). For example, one study found that teacher's race produced differential effects on student achievement (Dee, 2004), while others found that it did not affect student outcomes (Ehrenberg, Goldhaber, & Brewer, 1995).

Collectively, research efforts on teacher inputs have not provided enough evidence to support their association with student achievement, except in mathematics. However, teacher qualifications may be necessary to determine who should be allowed to teach in schools (Goe, 2007). Perhaps they serve as a form of quality control for individuals working with a particularly vulnerable population (i.e., students in K-12). Although it might be important to vet teachers prior to working with student populations, relying on qualifications alone (i.e., certification) may not be enough to produce changes in student learning. Moreover, research on teacher characteristics may not offer any practical significance to the individual teacher since these variables are outside of the teacher's control. Even policymakers may face ethical challenges if they attempt to implement policies that take characteristics into account. For example, it may be unethical to assign teachers and students by their race, as that may result in issues of segregation. Finally, student achievement may be more affected by teacher practices in the classroom, which is discussed the next section.

Teacher Processes in Educational Psychology

For this category of research, I propose separating studies involving teacher practices in the classroom from research conducted on specific instructional practices such as instructional adaptability (Parsons et al., 2017). Due to the magnitude of research available in both categories, only research pertaining to t*eacher performance* is reviewed for the purpose of the current argument.

Teacher performance. Teacher performance is commonly measured by observations in the classroom, principal evaluation, student surveys, self-reporting, and value-added models (Goe, 2008). Most of the studies reviewed measured the association between teacher performance and student achievement using observational rubrics such as Danielson's (1996) *Framework for Teaching (FFT).* Classroom observations are usually conducted by trained observers to evaluate teacher behavior/attitudes using specific rubrics of teacher performance in the classroom. For example, observers scored teacher instruction as satisfactory (0), target for growth (1), proficient (2), and area of strength (Borman & Kimball, 2005).

A review of earlier studies on teacher performance provided some evidence to support the relationship between teacher evaluation as measured by observation systems and student learning. Most of the studies reviewed used observation rubrics based on FFT. Of those studies, some showed small to moderate relationships (Borman & Kimball, 2005; Milanowski, 2004), while others reported significant positive relationship between teacher evaluation scores and student achievement (Holtzapple, 2003). Moreover, some studies showed mixed results across grade levels (Kimball et al., 2004), subject matter (Gallagher, 2004), and school sites (Heneman, Milanowski, Kimball, & Odden, 2006). For example, Kimball et al. (2004) reported significant relationships between teacher evaluations and 5th grade math and reading test scores, as well as reading test scores in the 4th grade; the experimenters did not find significant relationships between 4th grade math scores or any tests in the 3rd grade. One study deviated from the FFT trend and measured teacher performance based on eight classroom observations that were evaluated using 12 teacher performance standards and rubrics (e.g., lesson objectives and student grouping) developed by the experimenters (Schacter & Thum, 2004). The experimenters reported that approximately 84% of the variance in student achievement among teachers could be accounted for by ratings on the targeted dimensions and the indirect influence of classroom composition (i.e., types of students in the classroom). Schacter and Thum (2004) reported promising results; however, the type of students taught by those teachers may have indirectly influenced the variance in teacher scores. Classroom composition may largely influence measures of teacher effectiveness, which is addressed in more detail in subsequent sections.

More recently, a larger-scale study including 3,000 teachers revealed moderate correlations with student achievement when teacher performance was measured using five different observation systems, including FFT among them (Kane, McCaffrey, Miller, & Staiger, 2013). This study was conducted as part of the Measures of Effective Teaching Project (MET), a 45-million-dollar study funded by the Bill and Melinda Gates Foundation to rigorously evaluate measures of effective teaching (Cantrell, 2012). The MET study investigated the following measures of teacher effectiveness, value-added models, student feedback, assessment of content-specific pedagogical content knowledge, and classroom observations (Cantrell, 2012). The project’s major findings revealed that value-added measures (described in detail below) were most predictive of student achievement and that classroom observations provided teachers with important feedback but may not be used to predict student achievement accurately. Furthermore, combining the three main measures of effectiveness, value-added measures, classroom observations, and student surveys provided a superior method for identifying effective teachers (Kane & Staiger, 2012).

Overall, these studies provide important evidence to support the relationship between evaluating teacher performance in the classroom and student achievement. Nonetheless, studies that reported a dissociation between teaching and achievement necessitate further investigation on how or why evaluative measures fail to capture effective teaching. The following section addresses some limitations involving evaluative measures of teacher effectiveness, as measured by student outcomes.

Teaching Outcomes in Educational Psychology: The Case of Faulty Evaluation

A number of educational researchers have proposed that the common measures used to evaluate teachers are flawed (e.g., Darling-Hammond, Amrein-Beardsley, Heartel, & Rothstein, 2012; Everson, 2017). Researchers are finding fault with both value-added (VA) models and teacher performance observations in the classroom, which are the measures most commonly used to assess teacher effectiveness (Berliner, 2018).

Value-added models (VAM). Essentially, VAM “use statistical mixed-model theory and methodologies to conduct multivariate, longitudinal analysis of student achievement to make estimates of school, class size, teacher and other effects” (Wright, Horn, & Sanders, 1997, p. 57). In layman’s terms, the model measures student achievement by comparing students’ actual gains with their predicted gains based on their previous achievement scores.

The validity of VAM has been brought into question by a number of educational and measurement researchers (e.g., Darling-Hammond, 2015; Haertel, 2013). Policymakers use those data for accountability purposes to make high-stakes decisions (e.g., determining if a teacher is eligible to keep teaching) (Everson, 2017). For example, in 2011, the Houston Independent School District dismissed 221 teachers because their VA scores did not represent adequate student gains (Amrein-Beardsley, Collins, Holloway-Libell, & Paufler, 2016). That number is very alarming considering the current controversy regarding the validity of VA measures in the educational literature.

Multiple arguments are circulating among educational researchers as to why VAM should be considered with caution when making high-stakes decisions. Some have even suggested that judging teacher competency based on VAM evaluation may be a “fatal flaw” (Berliner, 2015, p. 1). There are three recurring arguments concerning VAM; those include within-teacher reliability, sorting bias, and extraneous test scores, among others.

Within-teacher reliability (Stability). Many researchers have found that VA models show instability in teacher performance from one year to the other (e.g., Goldhaber & Hansen, 2013). VA scores can show a teacher being effective one year and not effective the next year or vice versa. This variability in scores was also observed when different types of statistical models (e.g., Briggs & Domingue, 2011) and achievement tests were used (Lockwood, McCaffrey, Hamilton, Stecher, Le, & Martinez, 2007). The lack of stability in VA scores led researchers to question their validity in terms of evaluating teachers. Some have suggested that the instability of scores from one year to the other may be due to differences in the demographics of students being taught each year (e.g., Berliner, 2015). Differences in scores due to differences in student population pose a threat to the validity of VA scores because the models may be more sensitive to student characteristics than teacher effects (Everson, 2017). Ideally, VA models can control for student variables if students were randomly assigned to teachers. Unfortunately, random assignment rarely ever happens in schools; hence, it is difficult to control for student characteristics (Baker et al., 2010). The nonrandom assignment of students increases the vulnerability of VA models to sorting bias, which is another common problem of this type of evaluation measurement.

Sorting bias. There is evidence to support the notion that VA scores are sensitive to the students being taught, even after controlling for student characteristics (Koedel & Jullian, 2011;

Newton, Darling-Hammond, Haertel, & Thomas, 2010). For example, Newton et al. (2010) found that after statistically controlling for student characteristics (i.e., demographics and prior test scores), effectiveness scores positively or negatively correlated with teacher ratings depending on the number of advantaged or disadvantaged (i.e., eligible for free or reduced lunch) students in the teacher's classroom, respectively. This finding illustrates the argument that VAM do not control for all student variables, despite the common misconception that they do. Student characteristics may play a significant role in VAM scores because the measurement tools (i.e., standardized test scores) were designed to capture student performance in a particular subject area rather than the effect of teacher practices on student learning.

Extraneous test scores. Performance on a standardized achievement test (SAT) for reading and math is often used as the dependent variable of teacher effectiveness within VA models (Berliner, 2018). Unfortunately, these tests were not specifically designed to detect teacher effects (Berliner, 2018; Goe, 2007). Berliner (2018) asserted that "not a single standardized achievement test has ever shown that its items are instructionally sensitive" (p. 9). In other words, performance on SAT may not reflect the quality of instruction in the classroom, whether it was good or bad. In addition, due to some state laws mandating states to test grade-level content strictly, these tests do not accurately show learning gains for many of the students functioning above or below grade level (Darling-Hammond, 2015). For example, gifted students' posttests may show little academic gains because they had significantly high scores on their pretests, so there was little room for improvement (i.e., a ceiling effect). Therefore, a lot of learning gains may be unaccounted for (Goe, 2007). Another reason why test scores may not reflect a true measure of effectiveness is due to other noisy variables that may affect that score. Haertel (2013) proposed that only about 10% of the variance in pre- and post-student test score gains may be attributed to teacher differences. The rest of the variance may be attributed to other school factors (i.e., peer influences), out-of-school factors (i.e., family background), and random variation (i.e., factors not accounted for in the model). There is some evidence to support this distribution in variance (Fantuzzo, LeBoeuf, & Rouse, 2014; Goldhaber, Brewer, & Anderson, 1999; Nye, Konstantopoulos, & Hedges, 2004). For example, Goldhaber et al. (1999) placed the variance in test scores attributed to family background at about 60%, while school effects collectively explained 21% of the variance, including teacher, class, and school variables. Upon reviewing some studies examining the influence of out-of-school factors on student test scores, Heartel (2013) proposed that teacher effects may account for approximately 10% of the variance in test scores, a small percentage when compared to the 90% of variables that may or may not be known. Thus, the odds that these tests will accurately detect positive or negative teacher effects may be quite slim (Berliner, 2018; Heartel, 2013).

Based on the evidence reviewed thus far, it is safe to conclude that VAM may not be appropriate to evaluate teacher performance, at least not as the sole measure of evaluation. This conclusion coincides with a statement released by the American Educational Research Association (AERA) in 2015. AERA released the statement to inform those who use VAM or those considering their use as an evaluation measure about their “scientific and technical limitations” (p. 448).

In addition to limitations found within VAMs, researchers have also found some drawbacks with teacher observation systems in the classroom, another common method for evaluating teachers.

Teacher observation systems. Teacher observation systems (OS) are used to evaluate teacher practices in the classroom once or a few times a year, depending on whether the teacher is a novice or veteran (i.e., tenured). Observations are usually conducted by a highly trained observer or a principal using an observation instrument to code teacher behavior (Goe, 2008). There are many available instruments. One of the most commonly used instruments is FFT (Danielson, 1996), mentioned earlier in reviewing teacher performance studies. Another common instrument is the Classroom Assessment Scoring System (CLASS; Goe, 2008; Pianta, La Paro, & Hamre, 2007).

FFT observation system. FFT was first published in 1996 by Charlotte Danielson and her colleagues (Danielson, 2013). The framework was developed as an extension of the Praxis III: Classroom Performance Assessment, which is an observation protocol used to evaluate first-year teachers. FFT extended the Praxis III protocol to encompass teaching skills required by experienced teachers as well as novice teachers (Danielson, 2013). The framework is based on a constructivist approach to learning, whereby the learner derives the concept he/she is learning based on his/her own experiences. In other words, learners "construct" their own knowledge for themselves (Danielson, 2007). FFT is not subject-specific; it is a tool that can be used across subjects and grade levels. It includes four domains of teaching consisting of: (a) planning and preparation, (b) the classroom environment, (c) instruction, and (d) professional responsibilities. These domains are broken up further into 22 components which include 76 smaller elements. Evaluators using FFT can rate each of the 76 elements as unsatisfactory, basic, proficient, or distinguished. FFT is one of the most frequently used observation tools for evaluation and professional development purposes (Schoenfeld et al., 2018).

Classroom Assessment Scoring System (CLASS). CLASS is an observational tool developed to assess teacher interactions with their students in preschool through 3rd grade classrooms. It encompasses three major domains: emotional support, classroom organization, and instructional support. Each domain is broken down further to incorporate components of the classroom that are related to student learning. For example, the Instructional Support domain may include these dimensions, concept development, and quality of feedback. The Emotional Support domain may include teacher sensitivity (e.g., acknowledges emotions) and establishing a positive climate, and the Classroom Management domain may include behavioral management (e.g., redirecting inappropriate behavior) and productivity (e.g., majority of classroom time is devoted to instructional tasks). CLASS involves observation cycles, whereby each cycle consists of 20 minutes of observation and note-taking, followed by 10 minutes of scoring (Pianta et al., 2007).

Limitations of Observational Systems

The premise of observational systems to evaluate teachers is promising, since the assessments are used to measure teacher behavior to evaluate effectiveness rather than a child's arbitrary test score. These systems are especially promising when they are designed to evaluate teacher interactions with their students, such as the CLASS system. However, there are still some drawbacks to using observation tools that are based on a cognitive approach to learning and teaching. Some of the most common issues addressed regarding observational systems include lack of stability in observation scores and poor correlations between observation scores and measures of student achievement (i.e., test scores).

Stability of ratings. Some developers of observational systems and researchers stress the importance of sufficiently training raters so that they can reliably code teacher behavior in the classrooms (Goe, 2008; Pianta et al., 2007). For example, CLASS developers require users to obtain certification in scoring and coding the instrument through workshops or online courses (Klette & Blikstad-Balas, 2018). However, some observational systems may be vulnerable to instability in scores across raters, lessons, and observational instruments (Hill, Charalambos, & Kraft, 2012).

Some have argued that instability in scores may be subject to the evaluator's personal judgments regarding good teaching; thus, observers may rate certain teachers differently (Goe, 2007). Others have argued that the instability may be due to factors affecting teachers within any given time or context, such as personal life events, time of day, time of year, student absences, or the weather (Berliner, 2018). One study examined the effects of extraneous factors that may affect CLASS scores using data across 3 years (Buell, Han, & Vukelich, 2017). The experimenters examined the variance in scores across contexts (i.e., large group), students (i.e., gender and age), and seasons (i.e., Fall and Spring). They found that CLASS scores differ across seasons, whereby higher scores were observed in the Spring and dipped back down in the subsequent Fall season. Furthermore, the experimenters found a relationship between CLASS scores and the number of males in the classroom. The percentage of males in the classroom significantly negatively correlated with classroom organization and instructional support. This finding provides some evidence to support the vulnerability of observational systems to extraneous variables that are unrelated to teaching quality such as gender.

Correlation with other evaluation measures. Often the validity of evaluative measures such as observation systems and VAMs is investigated by examining the extent they correlate with one another. Logically, these measures should moderately correlate with one another since they are measuring the same construct (i.e., teacher effectiveness), yet there is no evidence to support this notion (Berliner, 2018). A number of studies have shown that significant positive correlations between observational instruments and student test scores are elusive (e.g., Grossman, Cohen, Ronfeldt, & Brown, 2014; Strunk, Weinstein, & Makkonen, 2014). Contrary to those findings, the MET study reported that observation systems such as FFT and CLASS were significantly associated with math and reading test scores (Kane & Staiger, 2012). The experimenters reported that students taught by teachers ranking below the 25th percentile on their observation scores fell behind by approximately 1 month of schooling in math when compared to peers with comparable characteristics. Another compelling finding was that students of teachers ranked above the 75th percentile gained 1.5 months of schooling when compared to their peers. However, these results were not replicated with English language arts (ELA) scores; results showed smaller teacher effects, which the experimenters attributed to the nature of the tests, not the teachers.

The studies reported thus far provide mixed evidence to support the relationship between teacher performance in the classroom (i.e., observational scores) and student achievement (i.e., VAM scores). These results may implicate a disconnect between what is measured in the classroom and student achievement, not because these two variables are not associated, but because the tests used to measure students' achievement may be unrelated to what the teacher is evaluated on in the observation systems. For example, a teacher may be ranked highly on the CLASS measure for maintaining an organized classroom and providing emotional/instructional support such as frequently responding to student-initiated interactions as well as identifying individual student needs and modifying instruction accordingly, while an item on a 5th grade ELA test may ask "Which detail is important to include in a summary of the article?" (ELA, 2017, p. 12). Due to the general pedagogical nature of CLASS, there are no items on the rubrics that measure effectiveness in terms of teaching the summarization skills necessary to answer the question on the ELA test. Hence, observation systems such as the CLASS cannot be deemed invalid if they do not correlate with VAM scores because they do not measure the same constructs. Perhaps observation systems that align with the Common Core standards (e.g., Danielson, 2013) by subject matter will be more associated with achievement since the constructs measured will be associated with each other.

Other important considerations. The large number of items assessed within observational systems (e.g., 76 elements in FFT and 43 subcomponents in CLASS) may be another reason why there is instability in scores, because not every skill on those instruments may be required in every lesson, subject, or objective. For example, one possible item that can be scored on the CLASS instrument is creativity, which involves brainstorming, planning, and producing. This component may not be relevant in a lesson on adding two-digit numbers because a specific formula is to be learned, not created. Thus, the teacher may get a low score not because of her teaching skills but because the item was irrelevant to her lesson. This component may be relevant, however, in a writing class, so a teacher's score may change according to the subject taught.

Another important issue to consider with regard to the validity of instruments is the use of rating scales to measure teacher behavior, which is present in both FFT and CLASS. Rating scales do not directly measure teacher behavior in the classroom; they are scores given based on inferences made by the observer. Simply stated, a score on a rating scale is perhaps a better measure of the observer's judgements rather than teacher performance in the classroom. In a review and discussion on using the Likert scales in evaluation systems, McGreal (1990) warned against the use of Likert scales in evaluating teachers due to their high-inference nature. He also recommended using low-inference measurement such as "recordings of events that factually describe what occurred" (p. 53). Unfortunately, his recommendation continues to be ignored, as evidenced by the frequent and continued use of Likert scales in measures of evaluation. Even his own colleague Charlotte Danielson, with whom he wrote a book on teacher evaluation, did not heed his warning (Danielson & McGreal, 2000). FFT was revised three times since it first came out in 1996, to include changes such as Common Core state standards (Danielson, 2013), but the Likert scale perseveres. Classroom observations may provide a more accurate representation of teacher performance. However, using a Likert scale to code teacher behavior may result in speculative representations of teacher performance since they were not direct measures of what the teacher did in the classroom.

Nonetheless, observation systems such as FFT and CLASS may provide important insight into teacher competencies that may contribute to effectiveness. For instance, evaluating teachers on competencies inherent in the three domains of the CLASS system could potentially identify effective teacher competencies as measured by student outcomes and provide constructive feedback for teachers that need extra support. The framework may be promising since it provides definitions and examples for each dimension specified, yet more evidence is needed to support the effects of this framework on student learning (Perlman et al., 2016).

Reviewing limitations of teacher evaluation systems from a behavior analytic lens may provide some useful comparisons that lead to identifying important and/or missing factors in teacher effectiveness research across both behavior analytic and educational literature.

Examining Teacher Evaluation Measures: A Behavioral Perspective

From a behavior analytic perspective, the source of the limitations discussed above may be due to the deductive premise underlying operational definitions of teacher performance. Meaning, operational definitions of effective teacher behavior are based on logical reasoning rather than empiricism. That is not to say that evaluation measures such as FFT and CLASS lack operational definitions supported by evidence; rather, this argument is meant to highlight the susceptibility of those definitions to high levels of inference that may potentially lead to unreliable measurement. To illustrate this argument, two examples of operational definitions regarding establishing a positive classroom environment are presented in Table 1. One example illustrates a deductive approach to operational definitions, another illustrates an inductive approach to operational definitions.

Table 1

Examples of Deductive and Inductive Operational Definitions of Teacher Competencies

Operational Definition of Positive Classroom Environment	Underlying Approach	Measurement Example
Considers the emotional and social tone of the classroom. The enthusiasm, enjoyment, and respect displayed during interactions between the teacher and children and among children. (CLASS; La Paro et al., 2004, p. 415).	Deductive: based on logical description of a positive climate = high inference variable.	(1, 2) = low (3, 4, 5) = moderate (6, 7) = High Scores averaged across four observations to represent average teacher score.
To reinforce positively, contingently, and frequently for social and academic responding (Greer, 2008, p. 44).	Inductive: based on observable description of a positive classroom environment = low inference variable.	Teacher presented 4 approvals per minute to students playing appropriately in the free-play setting during a 5-minute observation period.

The examples above may illustrate some of the differences in the treatment of operational definitions in educational psychology and behavior analysis in relation to teacher competencies. On one hand, the CLASS definition of positive climate can be interpreted in seven different ways by one observer; those interpretations may vary from day to day or observer to observer. On the other hand, delivery of reinforcement can be measured in one way by a number of observers—that is, the number of instances teachers provided positive reinforcement to the student in any specific form (e.g., a pat on the back, a “good job!” or bubbles), a number which is not affected by interpretations but by direct observation and recording. Despite the significant differences in definitions and measurement, both approaches measuring student-teacher interactions are geared towards the same constructs of teacher effectiveness, including but not limited to a positive classroom environment. What the behavior analytic perspective may offer educational researchers and practitioners are examples of the functional and parsimonious nature of inductive operational definitions to measure teacher behavior and student learning (e.g., conducting 5-10-min. classroom observations to measure direct instances of two teacher behaviors such as providing positive reinforcement and presenting learning opportunities, versus conducting four 30-min. observation sessions to compile a representative teacher score, by measuring 12 dimensions of teaching within a 20-min. interval). These examples may provide a solution for the proposed need for more parsimonious observational systems reported in one MET report that examined the reliability and effects of five different observational systems (Kane & Staiger, 2012).

Examining the differences between the treatment of operational definitions in educational psychology and behavioral analysis warrants a review of operationalism and its role in psychological practices.

Operationalism in Psychology

The adoption of operationalism from physics to psychology resulted in a heated debate among scientists in terms of identifying the parameters of operational definitions (Chang, 2009). The controversy led to a symposium in the *Psychological Review*, designed to answer critical questions about operationalism proposed by E. G. Boring (1945), such as “what is the purpose of operational definitions?” (Langfeld, 1945, p. 241). What resulted from this symposium were two distinct positions regarding the treatment of operational definitions of psychological terms, the logical position and the behavioral position. Of particular interest is how each camp proposed the establishment of the meaning of those terms through empirical verification.

The logical position. A definition is established in a "semantical metalanguage…. Quite generally, an exact logical analysis of the meaning of scientific terms requires the use of apparatus of syntax and semantics" (Feigl, 1945, p. 252). Stated differently, operational definitions are established via the careful ordering of words and language meaning.

It would seem that the above position gives meaning to a word (i.e., psychological term), by associating it to more words (i.e., operational definition). Thus, what this process implies is that meaning depends more on the language formulated by the scientist rather than an empirical account of those definitions (Moore, 1985). To include empiricism in the process of generating meaning, logical positivists suggested examining the validity of these psychological terms through observable measurement (Feigl, 1945). Skinner (1945) argued against operationalizing psychological terms in such a manner as it allows room for mentalistic constructs based on extraneous factors (i.e., the "logic" of a scientist) to be incorporated in psychology.

The behavioral position. Skinner (1945) did not offer an explanation for what constitutes an operational definition. What he proposed was a careful analysis of how psychological terms are constructed. Skinner suggested doing so by examining the antecedent stimuli (i.e., the event that occasioned the behavior) of the scientist when specific verbal behavior was established (i.e., operational definitions), in addition to examining the consequences from the verbal community that control it (i.e., reinforcement/punishment). Thus, the generation of terms from a logical approach is based on deductive reasoning, whereas a behavioral approach is based on inductive reasoning, whereby terms are derived from data that have shown a consistent pattern of the phenomenon of interest (e.g., positive reinforcement).

Implications on Educational Practices

This distinction in epistemology is possibly hindering educational research. Educational literature is replete with examples of attempting to measure concepts that are deemed “logically” important. For example, as a product of such reasoning, observational tools are established to measure concepts that may or may not be valid in terms of identifying effective teaching. These tools may be comprised of whatever the researcher hypothetically reasoned to be important for measuring teachers—hence, the 76 subcomponents of FFT and other tools like it that are influenced by extraneous variables.

This distinction in epistemology may also be how behavior analysts established measures that were more sensitive to detecting changes in student learning as a result of teacher practices (i.e., Greer, McCorkle, & Williams, 1989). Measures of effective teaching are not deduced in a behavior analyst’s mind. Rather, principles of behavior derived from controlled experimentation are applied to socially significant behaviors such as what the teacher needs to do to be effective. For example, Albers and Greer (1991) found that student correct math responses increased three times more than baseline levels as a result of increasing three-term contingency trials delivered by the teacher. The trials essentially involved a response from the teacher (e.g., “What is 12 x 2”), followed by an opportunity to respond by the student (e.g., “24”), followed by another response from the teacher (e.g., “excellent”). These trials are now referred to as learn units, which some argue may be “a fundamental measure of pedagogy” (Greer & McDonough, 1999), and others have provided evidence to support its effective use in teaching new repertoires (e.g., Bahadorian, Tam, Greer, & Rousseau, 2006). The three-term contingency was derived from past experimental behavior research (Skinner, 1938) and then applied to socially significant contexts.

This is only one example of the importance of generating operational definitions of effective teacher behavior that are derived from empirical evidence rather than logic.

Skinner (1984) suggested a pragmatic approach for analyzing truth of scientific concepts, whereby truth is determined based on the extent of the success of the scientists in achieving their planned outcome by applying their concepts. He proposed that success rather than agreement will yield satisfactory results.

Shift in Epistemology

Considering the arguments presented thus far, we may conclude that educational researchers need a change in methodology and possibly in epistemology. Even some educational researchers have begun to shy away from their own methods, to incorporate more behaviorally-based measures. For example, Darling-Hammond and her colleagues (2010) developed the Performance Assessment for California Teachers (PACT) in an effort to establish better measures of effectiveness and teacher training. Performance assessments also received the attention of Stanford University and the American Association of Colleges for Teacher Education, which led to the development of the national test edTPA and its application in 2013. Educational researchers in the Teacher Performance Assessment Consortium promoted the use of these assessments because they are a direct measure of what the teachers actually do to produce student learning. For instance, teachers are required to submit samples of students' work to show learning gains. Components measured in these tests are similar to the effective teacher repertoires established within the parameters of behavior analysis such as collecting evidence on student learning (Greer et al., 1989). Performance assessment measures are a step forward in the right direction in relation to evaluating actual teacher performance. Teacher competencies inherent in these assessments are in line with some of the key elements of effective teaching in behavior analysis such as analyzing one's own teaching to improve further instruction.

Teacher Effectiveness in Behavior Analysis

Behavior analysis consists of two branches, including the experimental analysis of behavior (EAB) and applied behavior analysis (ABA). EAB involves the investigation of basic principles of behavioral change, whereas ABA consists of developing technology derived from the basic principles to change socially significant behavior (Cooper, Heron, & Heward, 2007). There are some common misconceptions about ABA that may hinder its dissemination in education. One such misconception is that ABA is an intervention for specific populations such as individuals with autism or developmental disabilities (Falcomata, 2015). Another misconception is that ABA is not suitable for teaching more complex repertoires such as language development and problem solving (Ertmer & Newby, 1993). The following section reveals the falseness of these misconceptions by reviewing research that has contributed to the development of an effective system of instruction established within the realm of ABA.

There have been a number of successful research and practical efforts regarding effective instruction in ABA (e.g., The Morningside Model of Generative Instruction; Johnson & Layng, 1994). Nevertheless, one program in particular provides an expansive approach to instruction which includes the role of the student, the teacher, parents, and supervisors as well. This program can be referred to as the Comprehensive Application of Applied Behavior Analysis to Schooling (CABAS®) (Selinske, Greer, & Lodhi, 1991). Since the CABAS® model is an all-encompassing system of instruction, teacher variables within Goe's framework of teacher quality are addressed within that model.

The CABAS® Approach

CABAS® was established by R. D. Greer and colleagues at Teachers College, Columbia University; it is a technology of instruction derived from a substantial body of research on "effective educational practices" (Greer, 1997, p. 60). The system was developed to address America's educational crisis (Greer, 1991) brought to light by the Coleman report (Coleman, 1966). CABAS® is comprised of interrelated components that work together whereby change in one component of the system affects the other. At the heart of the system is the student; all other components in the system are geared towards ensuring that the student is learning (Greer, 1997).

There is evidence to support the effectiveness of the CABAS® model on student learning (Greer et al., 1989; Selinske et al., 1991). A critical component of the model is the teacher functioning as a strategic scientist of instruction (Greer, 2002). The repertoires of a teacher functioning as a strategic scientist are addressed in the subsequent section.

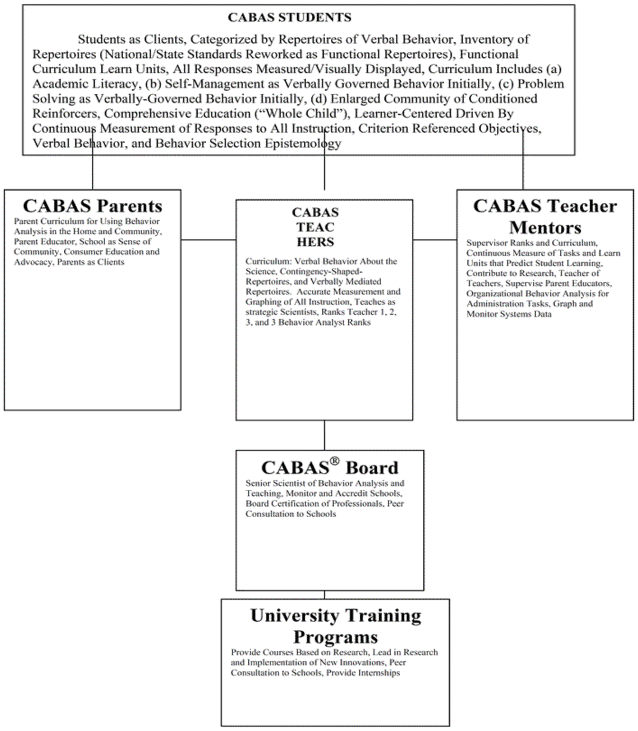

Figure 2. Outlines of the components of the model
(retrieved from Greer, Keohane, & Healy, 2002)

Teacher Inputs in Behavior Analysis

Within the CABAS® model, teacher inputs include specific teacher repertoires rather than teacher credentials or characteristics. Repertoires of a teacher scientist are divided into three categories, including contingency-shaped behaviors, verbal behavior about the science, and verbally mediated repertoires (Greer, 2002). The following descriptions of these repertoires were retrieved from *Designing Teaching Strategies* (Greer, 2002, pp. 43-53).

Contingency-shaped repertoires (CS). CS repertoires are behaviors that the teacher learns through direct contact with the teaching environment. In other words, they are acquired in the classroom through direct training and practice. CS repertoires of a teacher scientist are not to be mistaken for behaviors that a teacher learns through trial and error. Rather, they are repertoires that initially require specific training and supervision, until the teacher performs them fluently and accurately without having to think about them. Some of these repertoires include positively reinforcing social and academic behaviors frequently (e.g., “I like how you shared your toy, Toby!”) and recording student responses accurately (e.g., recording a (+) for correct responses and a (-) for incorrect responses).

Verbal behavior about the science (VB). This repertoire refers to the vocabulary pertaining to scientific research practices as well as instructional practices derived from previous scientific research. The collective vocabulary forms a “community of verbal behavior” (p. 48), whereby the community consists of individuals who are versed in the vocabulary of the science. Learning the vocabulary allows teachers to practice according to the science, define teaching practices and research procedures using scientific vocabulary, and draw from scientific literature instructional solutions to solve learning problems and aid learning. These repertoires are learned through readings and mastery of written or vocal quizzes.

Verbally mediated repertoire (VM). The VM repertoire involves the strategic analysis of instructional problems. The teacher draws from both CS and VB repertoires to solve instructional problems within the parameters of the science. This repertoire enables the teacher to individualize instruction and apply tactics from the science when applicable.

Greer (2002) proposed that all three repertoires are necessary for a teacher to function as a strategic scientist of instruction. These repertoires are taught using the CABAS® ranks, which range in difficulty from one rank to the next (i.e., Teacher I, Teacher II, and Master Teacher). Each rank contains 10 modules for each teacher repertoire outlined above, with subcomponents within each module. Table 2 shows an example of components associated with each rank and module component.

Table 2

Examples of Teacher Rank Competencies Within Each Repertoire

	Teacher I	Teacher II	Master Teacher
CS	Candidate was accountable for 2 objectives achieved for 1 student(s) with learn units to criterion (LUC) appropriate for the student level of verbal behavior.	Candidate was accountable for 7 objectives achieved for 5 student(s) with learn units to criterion (LUC) appropriate for the student level of verbal behavior.	Candidate conducts 25 graph checks for accuracy of decisions made by junior teachers or teaching assistants
VB	Mastery of basic concepts, chapter 1 and 2 (Cooper, Heron, & Heward, 2007).	Mastery of teacher repertoires and Analysis, chapter 3 and 4 (Greer, 2002).	Group of 5 conceptual paper summaries devoted to curriculum design
VM	Completion of one data collection showing an AB or functional relationship (ABAB) in APA style, edited to criterion	Written description and rationale for 10 new performance tactics	Experimental analysis of the effects of a tactic to solve a class-wide or school-wide problem.

Teacher Processes in Behavior Analysis

In addition to the repertoires mentioned above, there are specific practices derived from instructional research that the teacher implements in the classroom. One such tactic is the learn unit.

Learn unit (LU). A LU can be defined as a set of interlocking responses between a student and a teacher, which contribute to learning for both parties. The LU typically involves three components: the presentation of a direction from the teacher, a student response, and a consequence from the teacher. However, the LU may include more than three interactions between the student and the teacher. For example, incorrect student responses are followed by further consequences from the teacher to ensure that the student can emit the target behavior independently. Therefore, the LU may be more than a three-term contingency (Greer, 2002). There are four major functions of the learn unit: (a) it directly measures student learning, (b) it helps teachers monitor student progress and adjusting instruction accordingly, (c) it trains new teachers on the components of effective instruction, and (d) the context of the LU provides information that may help the teacher better identify and solve learning problems (Greer, 2002). In addition to delivering LUs, the teacher also graphs correct responses to LU presentation and makes instructional decisions based on a visual analysis of the data. These decisions are made in accordance with a specific algorithm referred to as the decision tree protocol (Greer, 2002).

Decision tree protocol. The decision tree protocol is comprised of a set of questions and rules that enable the identification of learning problems and the application of the appropriate tactics from the literature. The protocol is comprised of two steps. The first step is identifying when there is a decision opportunity by visually analyzing student data. The second step involves making an accurate decision based on a series of strategic questions to address the source of the problem (Greer, 2002). There is evidence to show that student learning significantly increases when teachers use the decision tree protocol to solve instructional problems (Keohane, 1997). The decision protocol and graphical displays of student responses to LUs are important components within the CABAS® model; however, these components are accurate under the assumption that the teacher is delivering errorless LUs. Inaccurate LUs may skew student data, which could result in decisions made based on teacher errors rather than student performance. Therefore, ensuring that teachers deliver accurate LU presentations to their students is necessary for an accurate analysis of learning and learning problems. Research has shown that teachers' accurate presentations of LUs increases with the application of the Teacher Performance Rate and Accuracy Scale (TPRA), which in turn increases student correct responses to LU presentations (Ingham & Greer, 1992).

The CABAS® components mentioned above were designed to insure that students are learning and are doing so at a satisfactory rate. The outcome measure used to assess the rate of students' learning and, in turn, teacher effectiveness is the number of learn units required by the student to meet an objective, or the number of LUs to criterion.

Teaching Outcomes in Behavior Analysis

According to the CABAS® model of instruction, a teacher's learn units to criterion (LUC), number of teaching trials to meet an objective, is one of the most important, if not the most important indicator of teacher effectiveness (Singer-Dudek, Speckman, & Nuzzolo, 2010). Essentially, the fewer trials delivered to teach an objective, the more effective a teacher is. A low LUC may indicate that a teacher is delivering significantly effective learn units (i.e., teaching opportunities) to her students as evidenced by the number of curricular objectives they achieve.

There is some evidence to support the idea that teacher effectiveness as defined by LUC significantly correlates with the CABAS® teacher ranks (Sherzo, 2010). Scherzo (2010) found that teacher ranks significantly negatively correlated with students' LUC, whereas teacher's level of burnout, self-efficacy, and relationship with students did not. In other words, the higher the teacher rank, the fewer learn units her students required to meet an objective. Although this finding may be intriguing, we have yet to identify which module components play a significant role in teacher effectiveness. It is possible that the three module components are inseparable and codependent on one another. For example, one cannot come up with a verbally mediated solution to a learning problem if he/she is unable to tact (i.e., label/identify) the problem. The following scenario further illustrates this point. A teacher may not know how to create a motivating operation for a mand (i.e., request) through deprivation if she cannot identify that mands only occur when a motivating operation is present (i.e., wanting the item). We might find said teacher saying, "The student just does not like anything; I don't know how to teach her mands if she never wants anything." On the other hand, it is also possible that one component—for example, contingency-shaped behavior (i.e., teacher learned behavior in the classroom)—plays a more significant role within the parameters of creating effective teachers. Stated differently, it may not really matter what vocabulary a teacher knows, only that she is able to teach her students specific objectives perhaps through proper training on classroom practices and decision making. Whether these statements are true or false, there is no empirical evidence to support either claim. Therefore, it is worth exploring the components that contribute to the development of effective teachers.

The purpose of the current study was to extend the findings of previous research on the relationship between teacher ranks and LUC (Scherzo, 2010) by isolating the components within the teacher ranks to determine if they each significantly correlate with LUC. In addition, if there is a significant correlation for each module component, would there be a difference in the degree of significance? Knowing which components within the ranks contribute to teacher effectiveness can positively inform teacher training practices. For example, teacher trainers may choose to focus on the components that significantly negatively correlate with LUC.

Chapter II

EXPERIMENT 1

Method

Participants

The participants were recruited from two private, publicly funded preschools which incorporated the CABAS® model of instruction. A brief presentation was administered to potential participants, which involved a description of the study and its requirements. A packet was then distributed to the participants, including a written description of the study, a consent form, and a brief questionnaire (see Appendix A). Twenty teachers chose to participate in the study. Table 3 shows a summary of the participants' credentials and classroom ratio.

Table 3

Teacher Credentials and Student Level of VBD

Credential	Frequency	Percentage
CABAS® Rank		
No Rank	8	40
Teacher I	6	30
Teacher II	5	25
Master Teacher	1	5
Education Level		
Bachelor's Degree	3	15
Master's Degree	16	80
Doctoral Degree	1	5
Certification Status		
Certified	14	70
Not Certified	6	30
Student Level of VBD		
Pre-foundational	8	40
Independent	9	45
Bidirectional	3	15

Note. Rank held refers to the Rank awarded by the CABAS® advisory board.

The participants were all females and held the position of head teacher, meaning they were responsible for designing their students' educational programs, monitoring student progress, setting a classroom schedule, and training/monitoring other staff in their classroom (i.e., teacher assistants and master students).

Procedure

Data collection. Archival data were collected for the 2017-2018 academic school year, which included approximately 39 weeks. The teachers handed in copies of their personal graphs that were publicly posted outside of their classroom. The experimenter transferred those data on to an Excel sheet for analysis; three participants submitted their data on Excel sheets (15%); and one participant submitted her weekly classroom summary sheets, which included LUC data.

In addition, teachers submitted copies of their Teacher I and Teacher II CABAS® ranks. Teachers working on the Master Teacher Rank or a higher rank were not required to submit them since those ranks involve more complex components and do not require a supervisor's signature; rather, the CABAS® advisory board reviews those ranks to determine if they meet a set standard to award the rank.

Measures

Dependent measures. The following measures were modified from Scherzo's (2010) original procedure. The participants' mean LUC was calculated for the 2017-2018 academic school year, whereas Scherzo collected data across 10 days of instruction and then calculated LUC by dividing the total number of LU presented by the total number of objectives achieved.

Teacher LUC. Teacher LUC was calculated by dividing the sum of the teacher's weekly LUC by the total number of weeks applicable (i.e., weeks data were available). The mean was calculated to aggregate the number of LUC across weeks into a single digit for analysis. It is important to note that actual LUC data may be a more accurate measure of effectiveness than the mean of LUC because it is a direct measure rather than an aggregated one. This would have involved calculating the sum of total LU presented that year, divided by the sum of objectives achieved. However, some of the participants included assessment data to the total number of learn units presented each day; hence, computing actual LUC with assessment data may not have been an accurate representation of teachers' actual performance. Calculating LUC with assessment data may have resulted in teachers having higher LUC due to unconsequated assessment trials which do not reflect quality of teaching.

CABAS® rank competencies. The experimenter counted the total number of competencies mastered on Teacher I and Teacher II ranks, as evidenced by the dated signatures of their supervisors. For teachers working on the Master Teacher rank, the number of competencies mastered were reported by the teacher on the questionnaire administered at the onset of the study. Data were collected or reported on the number of competencies mastered within each module, including VB, CS, and VM components. The sum of competencies mastered was then calculated, for a total number of competencies mastered within each module. Furthermore, the number of components on previously completed ranks were added to the number of competencies mastered within the participants' current rank, for a total number of competencies mastered within each module and across all ranks. Refer to Table 4 for a summary of competencies mastered by each participant.

Table 4

Number of Competencies Mastered Across Ranks and Module Categories

	Total Number of Competencies Mastered across all Ranks			
	VB	CS	VM	Total
Participant 1	20	63	67	150
Participant 2	30	66	81	177
Participant 3	30	53	85	176
Participant 4	5	30	11	46
Participant 5	30	70	85	185
Participant 6	8	15	11	34
Participant 7	30	68	83	181
Participant 8	30	70	85	185
Participant 9	10	30	37	77
Participant 10	6	28	17	51
Participant 11	18	54	54	126
Participant 12	18	54	64	136
Participant 13	5	13	13	31
Participant 14	14	55	51	120
Participant 15	10	20	32	62
Participant 16	12	41	42	95
Participant 17	30	70	85	185
Participant 18	10	29	23	62
Participant 19	6	27	15	48
Participant 20	10	30	37	77

Teacher credentials. Information on teacher credentials was gathered using the questionnaire distributed at the onset of the study. The teachers were asked to report their level of education, certification status, work experience, and CABAS® Rank/s awarded.

Students' level of verbal behavior development (VBD). Students' level of VBD was categorized into three different levels of verbal functioning as defined by Greer (2018). These categories included: (a) pre-foundational level of VBD, (b) independent level of VBD, and (c) bidirectional level of VBD.

Pre-foundational level of VBD. Students functioning at this level of VBD have one or more pre-foundational cusps in their repertoire. For a detailed list of each cusp and capability pertaining to each level of VBD, refer to Table 5 (modified form Greer, 2018, p. 14).

Independent level of VBD. At the independent level of VBD, speaker and listener responses are separate repertoires that develop through different experiences and at different rates throughout the ontogeny of the individual (Greer & Speckman, 2009). For example, a listener can point to pictures of a dog, but cannot tact (i.e., label) a dog walking next to him/her or a picture of a dog; thus, his/her listener and speaker repertoires are independent of each other. In most cases, speaker responses do not develop in the absence of a fluent listener repertoire (Greer & Keohane, 2009).

Students functioning at this level of VBD were either categorized as functioning at the independent listener level or speaker level of VBD, or both. Students functioning at a listener level of VBD have all pre-foundational cusps as well as listener literacy (i.e., discriminating between auditory speech sounds; Greer & Ross, 2008) and auditory matching (i.e., matching sounds/words/phrases-to-sample by pointing; Choi, Greer, & Keohane, 2015). Students functioning at the speaker level of VBD have the cusp transformation of establishing operations across mands and tacts in their repertoire, meaning if they can tact (i.e., label) an item, they can also mand (i.e., request) for the item. For example, a student learning to request bubbles can also label a picture of bubbles.

Bidirectional level of VBD. Students functioning at this level have all pre-foundational cusps in repertoire, in addition to most independent listener and speaker repertoires. These students can also be described as having cusps and capabilities that involve the joining of listener and speaker repertoires (Greer, 2018). These cusps include: the speaker component of naming (i.e., learning operants as a listener results in learning those operants as a speaker in the absence of explicit instruction; Horne & Lowe, 1996); bidirectional naming (i.e., learning operants as a listener results in learning those operants as a speaker and vice versa; Miguel, 2016); say-do correspondence (i.e., individuals' own verbal behavior controls their overt behavior; Baer, Detrich, & Weninger, 1988); and self-talk (i.e., the individual rotates listener and speaker roles when "emitting verbal behavior to themselves"; Lodhi & Greer, 1989, p. 353).

The categories stated above were generally assigned to describe the verbal behavior level of functioning that students in 12:1:2, 8:1:2, and 6:1:2 ratios tend to exhibit since students at the school were placed in classrooms according to their level of VBD. Thus, teachers who taught a class of 12 were categorized as having students who could be described as functioning at the bidirectional level of VBD. Additionally, teachers who taught a classroom of 8 were categorized as having students who could be described as functioning at the independent listener or speaker level of VBD, and teachers who taught a class of 6 were categorized as having students who could be described as functioning at the pre-foundational level of VBD. However, these categories should not be interpreted to imply that every student in those classrooms had each cusp and capability pertaining to each level of VBD. These categories can be interpreted as such; some students in those classrooms may have had cusps and capabilities pertaining to their category of VBD and others may have been working towards attaining them.

Inter-rater Agreement

Inter-rater reliability was conducted for the number of rank components completed by the participants. An independent rater counted the number of signatures recorded within each module category (i.e., VB, CS, and VM). Agreement was conducted for 20% of teacher ranks, with 100% agreement. Agreement was also conducted for participants' LUC presented on the teacher's individual graphs. An independent rater visually analyzed the graph and recorded LUC for each week available. Agreement was conducted for 20% of LUC graphs, with a 100% agreement.

Results

Correlations

Research question 1. Is there a negative relationship between teacher LUC and the total number of competencies mastered across ranks? A Pearson's product-moment correlation coefficient was used to test the relationship between teacher LUC and the total number of competencies mastered. There was a significant negative correlation between total number of competencies mastered and teacher LUC, $r(18) = -.538$, $p = .014$. This finding indicated that the more modules completed by the teacher, the fewer LUs her students required to meet an objective. Refer to Figure 3 for a visual representation of the negative correlation between teacher LUC and number of modules completed.

Table 5

Verbal Behavioral Developmental Cusps and Capabilities (modified from Greer, 2018, p. 14)

Levels of Verbal Behavior	Verbal Behavioral Developmental Cusps and Cusps as Learning Capabilities
Pre-Foundational	• Instructional control • Conditioned reinforcement for observing voices • Conditioned reinforcement for observing faces • Conditioned reinforcement for observing 2D and 3D stimuli • Capacity for sameness across the sense • Generalized imitation
Independent Listener	• Generalized matching • Basic listener literacy • Auditory match-to-sample selection response
Independent Speaker	• Parroting • Echoic-to-mand • Echoic-to-tact • Independent mands • Independent tacts • Transformation of establishing operations across mands and tacts
Bidirectional	• Say-do correspondence • Self-talk • Unidirectional Naming • Bidirectional Naming
Foundational Reader Writer	• Conditioned reinforcement for observing books • Naming accrues from listening to stories read aloud by others • Print transcription • Dictation
Basic Reader	• Textually responding to rate • Responding to own textual responses as a listener • Reading governs own responding • Textually responding joins the naming capability • Conditioned reinforcement for textually responding to printed stimuli
Basic Writer	• Joint stimulus control across saying and writing • Technical writing that precisely affects the reader’s behavior • Aesthetic writing that affects the reader’s emotions
Self-Editor	• Joining of the reader-writer cusps and capabilities
Verbally Mediated	• Textually responding to complex operations • Technical writing to govern the complex operations of others

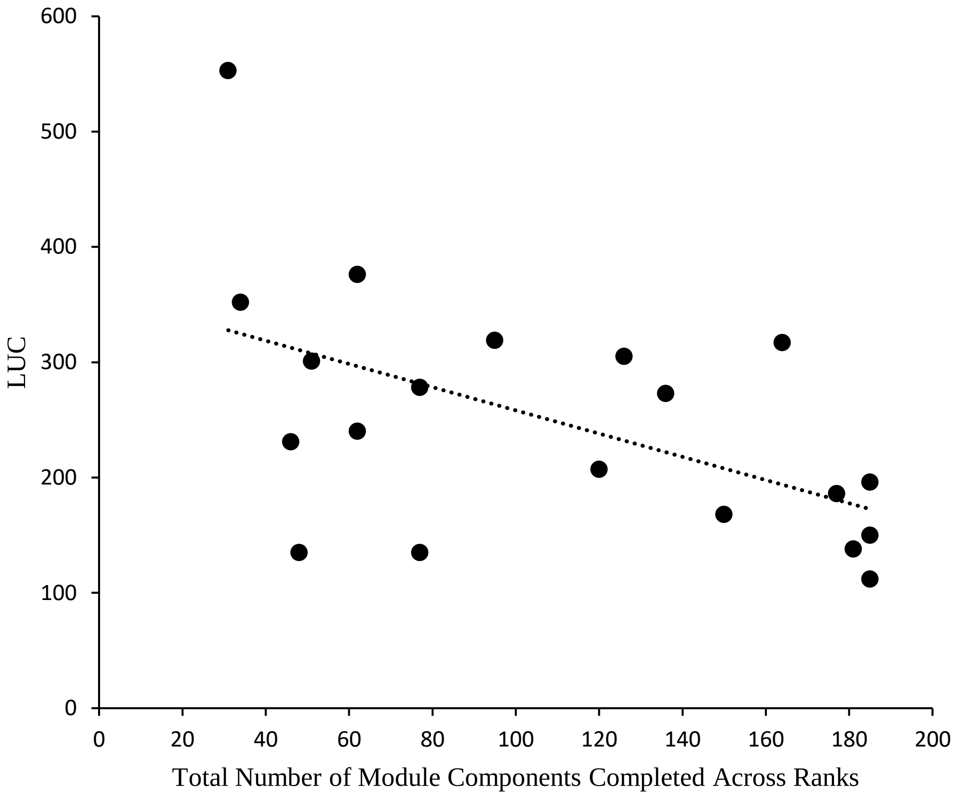

Figure 3. The negative correlation between teacher LUC and the total number of signatures completed across ranks

Research question 2. Is there a negative correlation between teacher LUC and competencies mastered within each module type (i.e., VB, CS, and VM)? A Pearson's product-moment correlation coefficient was used to test the relationship between teacher LUC and the number of competencies mastered within each module type including VB, CS, and VM repertoires. There was a significant negative correlation between teacher LUC and VB modules, *r*(18) = -.468, *p* = .038, as well as VM modules and teacher LUC, *r*(18) = -.510, *p* = .022.

Additionally, there was a stronger negative correlation between CS competencies mastered and teacher LUC, *r*(18) = -.587, *p* = .006, when compared to the other module types (i.e., VB and VM modules). Based on these findings, we may conclude that the more modules teachers complete, the lower their LUC. Moreover, the stronger relationship found between teacher CS repertoires and LUC may indicate that the more practical skills the teacher learns in the classroom, the more effective she will be as measured by a lower LUC. Refer to Figure 4 for a visual representation of the negative correlation between LUC and number of competencies mastered within each module type (i.e., VB, CS, and VM).

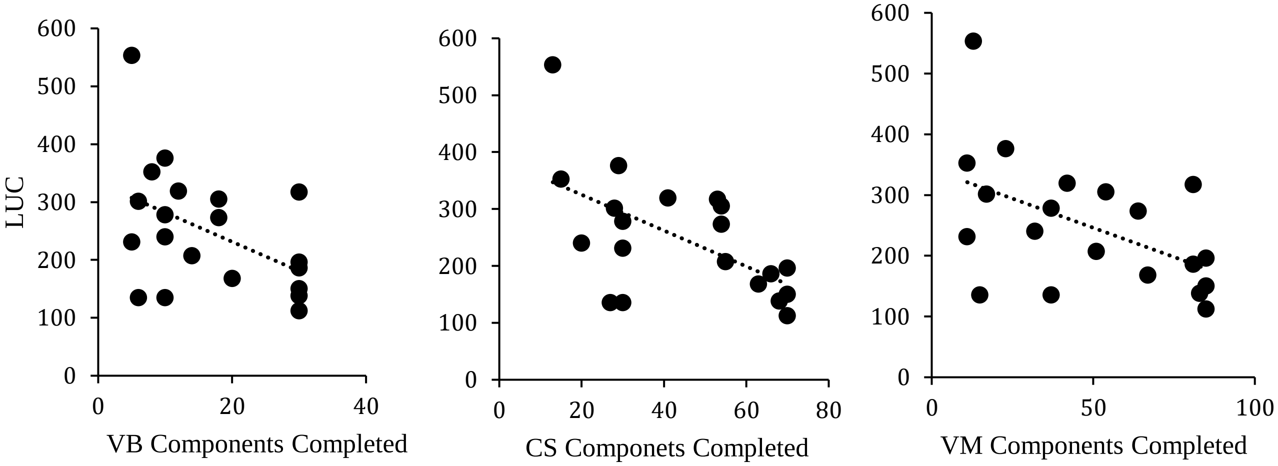

Figure 4. The negative correlation between teacher LUC and the total number of module competencies mastered within each rank category

Research question 3. Is there a negative correlation between teacher LUC and teacher credentials? A Spearman's nonparametric rank-order analyses were used to investigate the relationship between teacher LUC, education, and certification status. The results showed no significant relationship between teacher education, $r_s(18) = -.191$, $p = .434$, or certification status, $r_s(18) = -.065$, $p = .790$. with teacher LUC. Furthermore, a Pearson's product-moment correlation coefficient was used to test the relationship between number of years teaching with teacher LUC. The results showed no significant correlation between number of years teaching and teacher LUC, $r(18) = -.100$, $p = .684$. These findings indicated that teacher credentials and years of experience are not associated with teacher effectiveness.

Research question 4. Is there a negative correlation between teacher LUC and student level of VBD? A Spearman's nonparametric rank-order analysis was used to test whether the student's level of VBD was associated with both measures of teacher LUC. There were strong negative correlations between students' level of VBD and LUC, $r_s(18) = -.583$, $p = .007$. This finding showed that students with a higher level of VBD required fewer learn units to meet an objective. Refer to Figure 5 or a visual representation of the correlation between teacher LUC and student level of VBD.

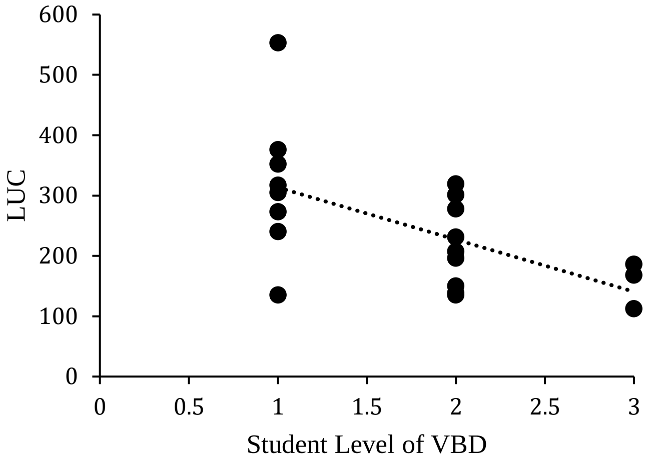

Figure 5. The negative correlation between teacher LUC and student level of VBD

Research question 5. Is there a negative correlation between teacher LUC and the total number CS competencies mastered across teacher ranks when controlling for student level of VBD?

To further examine the significant negative relationship between CS repertoires and teacher LUC, a partial correlation was used to control for student level of VBD. There was a moderate correlation between teacher LUC and the number of CS components completed, $r(18) = -.384$, $p = .052$, when controlling for student level of VBD. These results showed that CS competencies mastered were slightly associated with teacher LUC after controlling for student level of VBD. Thus, student level of VBD may explain more of the variance in teacher LUC than CS competencies mastered, meaning teachers who provide instruction for students with a higher level of VBD are more likely to have a lower LUC and teachers responsible for the instruction of students with lower level VBD are likely to have a higher LUC. It may be possible that the absence of a significant relationship in the above correlation is due to a statistical power issue (i.e., Type II error) of a small sample size. In other words, the sample size may be too small to detect statistically significant effects.

Table 6

Correlations and Descriptive Statistics

	LUC	Total Components	VB	CS	VM	Education	Years Teaching	Certification	Student level of VBD
LUC	1								
Total Components	-.538*	1							
VB Modules	-.468*	.973**	1						
CS Modules	-.587**	.974**	.910**	1					
VM Modules	-.510*	.993**	.974**	.943**	1				
Education	.002	.156	.141	.095	.156	1			
Number of years teaching	-.204	-.102	-.155	-.090	-.088	-.277	1		
Certification	.170	.142	.183	.085	.104	.380	.057	1	
Student level of VBD	-.583**	.507*	.409	.629**	.474*	.072	.209	.216	1
Mean	248.6	109.60	16.60	44.30	48.70	3.90	7.16	.70	1.75
SD	107.20	57.31	9.90	19.99	28.41	.447	6.48	.470	.716
Range	112-553	31-185	5-30	13-70	11-85	3-5	1-12	0-1	1-3

Note. This table was modeled after Scherzo's (2010) descriptive and correlations table (p. 62). A Spearman's rho correlation was used for teacher certification, education, and student level of VBD.

* Correlation is significant at the .05 level

** Correlation is significant at the .01 level

Discussion

This study was designed to investigate the relationship between teacher CABAS® rank competencies mastered and teacher effectiveness as measured by LUC. Specifically, this study was aimed to investigate the association between the competencies mastered within the three teacher repertoires composing the ranks (i.e., VB, CS, and VM repertoires).

The first research question was designed to investigate the relationship between the total number of competencies mastered across ranks and LUC. The results showed that the total number of competencies mastered was significantly associated with effective teaching as measured by a lower LUC. These results replicated the findings of Scherzo (2010), despite differences in measurement of LUC and rank. Scherzo calculated LUC by dividing the total number of learn units delivered by the teacher to one student, divided by the total number of objectives achieved by that student across 10 days of instruction. In the current study, LUC was measured by calculating the mean LUC of each teacher across an entire academic year while working with different students in her classroom. This method of measurement may have been a better representation of the teacher's LUC since it was measured for a longer period of time. Rank was also measured differently from Scherzo's study. Scherzo measured rank as a categorical variable, which may not have accounted for additional competencies mastered beyond the rank category reported by each teacher. In the current study, the number of competencies mastered were measured to account for all competencies mastered by each teacher across ranks. This may have also been a better representation of teacher repertoires as it accounted for each competency mastered by the teacher on past and current ranks, rather than an aggregated representation of their repertoires, whereby the total number of competencies mastered are masked by one category—the rank they completed.

The second research question was designed to extend the findings from the first one by pinpointing which category of competencies within the rank was most associated with LUC. The results showed that the number of competencies mastered within all three repertoires was significantly associated with establishing more effective teachers as evidenced by a lower LUC. Although all three categories were associated with a lower LUC, the results showed that CS competencies mastered were more strongly associated with teacher effectiveness than VB and VM competencies. This finding should not come as a surprise since CS repertoires are behaviors that are directly emitted by the teachers to occasion changes in student behavior. Although knowing the scientific tacts (i.e., labels) of those teaching behaviors may aid teachers in practicing them, a teacher's scientific knowledge alone cannot objectively occasion change in student learning. For example, presenting accurate LUs results in observable changes in student learning, whereas mastering a quiz on measuring behavior (Cooper et al., 2007, Chapter 4) does not have the same direct effect. Nevertheless, mastering the quiz on measuring behavior may affect the accuracy of recording student responses to LU presentations. Thus, teacher knowledge may mediate the relationship between CS teacher behaviors and student learning. Although this may seem like a reasonable relationship between repertoires, we may not have enough evidence to support the relationship between teacher knowledge affecting teacher practices and, in turn, student learning. This study provided some evidence of the association as indicated by the significant correlations found across repertoires; still, it is not enough to be certain that this meditative relationship exists or to explain the extent of its significance on student learning. This analysis may also be applied to VM repertoires. However, in the case of VM repertoires, both the teachers' VB and CS repertoires together may mediate the relationship between VM repertoires and student learning. Regardless of the lack of evidence to support the interdependent nature of these repertoires on student learning, this study provided some evidence to support the importance of establishing all three repertoires for effective teaching. These results may also lead us to question the need to investigate the interdependent relationship between these repertoires. Do we have to know the nuances behind how these repertoires affect each other? Or is it enough to conclude that they are all practically significant in terms of effective teaching; hence, they should all be part of a teacher's collective repertoire and end the narrative there.

The third research question was designed to investigate the relationship between teacher effectiveness and teacher credentials, including level of education, certification status, and number of years teaching (i.e., experience). The results showed that these variables were not associated with teacher effectiveness as measured by lower LUC. This finding is in line with past research on teacher effectiveness, whereby inconsistent evidence was reported between teacher qualifications such as degree level and achievement gains (Goe, 2007). In addition, this finding may lead us to question: If education level and certification status (measures of teacher knowledge) were not associated with LUC, then why was there an association with VB competencies mastered, which is also considered a measure of teacher knowledge? One explanation is that the content targeted in the VB modules of the rank are supported by research about components of effective teaching, such as behavioral principles associated with learning, and tactics that occasion student learning such as the LU. Mastery of this content is not measured in teacher certification tests, nor is it taught in teacher preparation programs that do not incorporate a behavior analytic perspective on teaching and learning. Thus, these findings may suggest the importance of teaching the content of a strategic science of instruction to mastery in teacher preparation programs.

The fourth question was designed to investigate the relationship between student level of VBD and teacher LUC. The results showed that student level of VBD was significantly associated with teacher LUC. Specifically, teachers of students with a higher level of VBD tended to have lower LUC, and teachers of students with a lower VB tended to have a higher LUC as they may be more challenging to teach. This finding was in line with teacher effectiveness research that reported variability in teacher effectiveness scores, whereby achievement scores show a teacher being effective in one year but not the next and vice versa (Berliner, 2015).

The fourth question was also designed to investigate the relationship between CS competencies and LUC while controlling for student level of VBD. The results showed that when controlling for student level of VBD, there was a moderate negative relationship between CS competencies and LUC. This finding was in line with previous research in the education literature showing the confounding role of student variables on teacher evaluation (e.g., Newton et al., 2012). One explanation for the lack of a significant relationship between CS competencies and LUC may have been the small sample size used for the analysis, resulting in a limited analysis due to low statistical power (restriction in range). Another possibility is that the measure of student level of VBD used may not have been an accurate representation of the variance in student cusps and capabilities inherent in the teacher classrooms. The measure may have been inaccurate because students were assigned a general VBD category according to the level of classroom support needed (i.e., classroom ratio); students in the CABAS schools are assigned to most supported and least supported classrooms, according to their level of VBD. Thus, the VBD category assigned to them may have been a better representation of the classroom support they needed rather than their actual level of VBD and the number of cusps/capabilities in their repertoire. An inadequate representation of student level of VBD may have limited the range of the analysis.

Collectively, these findings spark an empirical need to conduct further analyses of the correlations found between teacher competencies in the rank and student learning, while accounting for more appropriate measures of the range of student level of VBD. There were, however, limitations to this study that are worth noting. Besides the obvious limitation of the small sample size, a considerable limitation was the differences present among the individual programming structure each teacher incorporated when designing STOs that led to the attainment of long-term objectives (LTOs). LTOs were chosen for instruction by the teachers after students were assessed at the beginning of the school year using the CABAS® International Curriculum and Inventory of Repertoires for Children from Pre-School through Kindergarten (Greer & McCorkle, 2009). Once LTOs were chosen, the teacher designed the sequence of STOs to be taught and determined mastery criteria for those objectives. The criteria set to determine mastery may have varied considerably across teachers. Thus, some teachers may have set mastery criterion at 90% accurate responding observed in one session, and other teachers may have set the criterion at 90% across two consecutive sessions of observation, depending on their student level of VBD. Thus, LUC may have varied due to the mastery criteria set across teachers, which was a variable not accounted for in this study.

In addition, the LTOs targeted for instruction may have also contributed to the variance in teacher LUC. The range of LTOs selected for instruction may have ranged from students working on addition and subtraction (higher level of VBD), while other students worked on responding to their name and sitting still during instruction (lower level of VBD). Hence, the substantial variety of LTOs and related STOs may have also contributed to the variance in LUC, which was not accounted for.

A second experiment was conducted to address the limitations of the first study and to extend the findings of the correlations found in the descriptive analysis between teacher competencies mastered and teacher LUC through experimental analysis. I investigated whether teachers' competencies mastered within the rank will affect teacher LUC when controlling for category of instruction, difficulty of instructional objectives, number of objectives taught, and whether the student had bidirectional naming in their repertoire, which was a more specific measure of the student's level of VBD.

Chapter III

EXPERIMENT 2

Methods

Participants

All participants were recruited from a private, publicly funded preschool which incorporated the CABAS® model of instruction. First, a brief presentation was administered to potential participants, which involved a description of the study and its requirements. Next, the head teachers who chose to participate in the study selected a supporting teacher of a different rank and two student participants from their own classrooms. The experimenter then distributed a packet to the teacher participants, including a written description of the study (see Appendix B), instructional materials, and a questionnaire (same as Experiment I).

Teachers. Eight teachers and eight teacher assistants were recruited for the study. The participants were grouped into dyads whereby one participant in the dyad had a higher CABAS® rank than the other or more rank competencies mastered if they were working on the same rank. For example, if Teacher A held a Teacher II rank, then Teacher Assistant A held a Teacher I rank, or if Teacher B and Teacher Assistant B were working on the same rank (i.e., Teacher I), then Teacher Assistant B had fewer competencies mastered in that rank. Each dyad taught in the same classroom, meaning that the teachers were assigned to teach in the same classroom by the directors of the school. Classroom assignments were distributed at the onset of the academic school year.

Teachers were recruited based on their ranks in order to investigate the relationship between student rate of learning and teacher CABAS ranks. For more information on the teachers' level of education, ranks, and components completed, refer to Table 7.

Table 7

Number of Competencies Mastered Across Ranks and Module Categories

Participant	# of VB Modules Completed	# of CS Modules Completed	# of VM Modules Completed
Dyad A			
Teacher A	22	62	77
Teacher Assistant A	10	20	20
Dyad B			
Teacher B	8	28	18
Teacher Assistant B	2	27	9
Dyad C			
Teacher C	30	61	85
Teacher Assistant C	10	23	23
Dyad D			
Teacher D	28	68	83
Teacher Assistant D	10	17	23

Students. Each teacher dyad was assigned two students from their own classroom. One student met the probe criterion for the cusp bidirectional naming and one student did not. Students were recruited based on the presence or absence of bidirectional naming to control for possible accelerated rates of learning that are typical for students who demonstrate bidirectional naming as evidenced by prior research (Greer, Corwin, & Buttigieg, 2011).

Setting and Materials

The study took place in the participants' classroom, at a privately owned, publicly funded preschool that incorporated the CABAS® model of instruction. During probe and intervention sessions, the teacher sat next to the student in child-sized chairs, while instruction with other students was taking place at adjacent tables.

Target stimuli for instructional programs were presented on 3x5 index cards. The instructional set included four exemplars of each target stimulus. Exemplars varied in color and font type. In addition, data sheets were used to collect data on correct and incorrect responses emitted by the students. Data collected during intervention sessions were plotted on a line graph.

Dolch sight words were targeted for instruction across participants. Four different objectives were designed for each program. Each objective included four different sight words or letter sounds. Objectives were then divided between the two teachers whereby one teacher taught two objectives per program and per student, with a total of eight objectives per teacher. Programs were selected by the teachers to insure objectives matched the students' level of VBD and prerequisite skills. The teacher assistants were not involved with choosing the programs as it is the responsibility of the teachers to design choose and design curricular objectives for students in the classroom.

Sight word instruction. During instruction, the teacher presented the student with a target word, then waited 3 seconds for the student to respond vocally. If the student's response matched the target word sound printed on the card, then the teacher delivered praise to the student (i.e., Good! That is "and."). If the student did not respond within the allotted time or if the student emitted a word that did not match the target presented, then the teacher delivered a correction procedure. During the correction procedure, the teacher modeled the correct response, then provided the student with a second independent opportunity to respond; if the student emitted the correct response, then the teacher moved on to the next LU. If the student emitted an incorrect response the second time, the teacher once more modeled the correct response, and provided the student with a third opportunity to respond. If a correct response was emitted, the teacher presented the next LU. If the student emitted a third incorrect response or did not respond, the teacher moved on to the next LU. Praise was withheld throughout the course of the correction procedure. However, approvals continued to be delivered to the student for responses not related to target instruction (i.e., I like how you’re sitting nicely). Refer to Table 8 for a list of short-term objectives (STOs) taught by each participant.

Table 8

List of Sight Words Taught by Teachers and Teacher Assistants

	Student w/BiN		Student no/BiN	
Participant	STO 1	STO 2	STO 1	STO 2
Dyad A				
Teacher A	cold, found which, always	made, these, sleep, before	be, ate, new, came	all, did, our, good
Teacher Assistant A	mice, green those, around	your, would, their, upon	no, but, get, four	are, eat, now, like
Dyad B				
Teacher B	and, the, find, said	an, let, put, live	and, the, find, red	is, see, one, go
Teacher Assistant B	big, run, help, come	by, had, may, once	big, run, help, two	to, can, you, up
Dyad C				
Teacher C	**an, let put, live**	**of, has fly, just**	**by, had may, once**	**as, him old, give**
Teacher Assistant C	**by, had may, once**	**as, him old, give**	**an, let put, live**	**of, has fly, just**
Dyad D				
Teacher D	**we, run, you, it**	**two, and, go, can**	**for, to, up, see**	**red, one, the, in**
Teacher Assistant D	**for, to, up, see**	**red, one, the, in**	**we, run, you, it**	**two, and, go, can**

Note. Objectives in bold were counterbalanced across students to control for operant effects. Objectives were counterbalanced when possible, depending on correct responding on baseline probes.

In Table 8, words in bold were counterbalanced across teacher and teacher assistants to control for operant effects. Stated differently, certain words may have the tendency to be acquired faster than others for reasons outside of the experimenter's control (e.g., conditioned reinforcement for words beginning with the letter A may result in higher rates of acquisition for words beginning with A). Thus, counterbalancing those words across teachers by having both the teacher and teacher assistant teach the same words but to different students may control for those unwanted operant effects.

Measures

Dependent variable. The total number of learn units required to meet an instructional objective (i.e., learn units to criterion, LUC) were measured across teachers, instructional programs, and students with and without bidirectional naming. LUC were measured by calculating the total number of learn units presented within each condition after an objective was met. Mastery criterion was set at 90% correct responding observed in one session. For example, if a student required five sessions to meet an objective, and each session involved the presentation of 20 learn units, then LUC amounted to 100 LUs. In addition to the criterion referenced above, which is referred to as *macro criterion*, an additional mastery criterion was established for a more sensitive measure of LUC. This additional measure is referred to as *micro criterion*, which was the number of learning opportunities required for a child-participant to respond correctly on five consecutive opportunities. In other words, a target was considered mastered if the student emitted five consecutive correct responses for that word. LUC was then be calculated by finding the sum of opportunities presented before mastery criterion was achieved for each target word within each objective. Micro criterion may provide a more sensitive measure for the rate of acquiring sight words, since 20 LU sessions assess mastery only after trials in multiples of twenties (e.g., 20, 40, 60, etc.). The experimenter conducted micro criteria analysis after data collection was complete and macro criteria were achieved for all objectives taught across the four dyads.

Independent variable. The independent measure of the study was the different ranks held by each teacher. LUC were compared between conditions taught by a teacher with a higher rank and a teacher with a lower rank or no rank. All other variables pertaining to objectives and programs were held constant, including level of difficulty, type of program taught, and number of operants targeted within each objective. Data on CABAS® ranks were reported by the teacher in the questionnaire administered at the onset of the study. In addition, teachers submitted copies of their ranks to the experimenter.

Bidirectional naming probes. Data on the presence or absence of naming for each student were collected from the teacher's records prior to the onset of the study. The most recent naming probes were used, which were no more than 2 months old.

Baseline probes. Unconsequated probe sessions were conducted on potential operants targeted for instruction for both academic and verbal behavior programs. Operants not in the student's repertoire were targeted for instruction. For example, if a student responded "tiger" when presented with a picture of a lion, then the operant lion was targeted for instruction. Correct and incorrect responses were recorded in the same manner as mentioned above.

Responses to learn unit presentations. The teachers collected data on correct and incorrect responses emitted during instructional sessions. Once the session was completed, the teacher graphed the total number of correct responses on a line graph with a scale of 20. A plus (+) was recorded for correct responses and a (-) was recorded for incorrect responses.

Experimental Design

An adapted alternating treatment design (Sindelar, Rosenberg, & Wilson, 1985) was used to compare LUC across teachers. The design involved teaching two instructional objectives that were part of the same response class. A response class is a group of behaviors that have the same effect on the environment (Catania, 1979). In counterbalanced and consecutive sessions, teachers of different ranks taught separate but equal difficulty target responses (see selection and equating of targets section below) to a participant. Aside from the teacher running the session, all other instructional variables were held constant, including the type of program, the number of operants targeted for instruction, and the level of difficulty of each instructional set.

The sequence of conditions was counterbalanced across teachers. For example, if Teacher A ran the first session on Tuesday, then Teacher Assistant A ran the first session on Wednesday. In addition, the number of sessions ran by each teacher per day was held constant. For example, if one teacher ran two sessions on one day, the teacher assistant also ran two sessions. Once both teachers achieved criterion on the initial set, instruction began on a second set. Instruction on the second set was terminated once mastery criterion was achieved. Mastery criterion was set at 90% accurate observed in one session.

Selection and Equating of Target Operants

A protocol for equating objectives across teachers was established to insure that selected targets are of equal difficulty. Equating targets across objectives is a crucial component of an adapted multiple treatment design (Sindelar et al., 1985), since the difficulty of targets may result in a faulty comparison of the dependent variable in question. For example, if one teacher was assigned to teach one-syllable words, while the other was assigned to teach two-syllable words, LUC may be lower for the teacher assigned one-syllable words. The above sequence was replicated across objectives and students.

To control for difficulty of objectives across sight word objectives, the following rules were applied to each objective: (a) equal number of syllables (Haq & Kodak, 2015); (b) equal number of letters per word; (c) no words that start with the same letter were assigned to the same objective (Kennedy, Itkonen, & Lindquist, 1994); and (d) no words that were phonemically or visually similar were included in the same objective (Schnell, Vladescue, Kodak, & Nottingham, 2018).

Interobserver Agreement (IOA)

Trial-by-trial IOA was conducted for baseline and intervention sessions. IOA was calculated by the experimenter using TPRA observations. The observer sat in close proximity to the teacher and recorded data on teacher antecedents, student behavior, and teacher consequences. TPRAs were conducted for 50% of IOA sessions. IOA was calculated by dividing the total number of trials the observers agreed upon by the total number of trials presented.

For baseline sessions, IOA was conducted for 40% of sessions with 100% agreement. For instructional sessions, IOA was conducted for 26% of sessions with a mean agreement of 99% and a range of 95-100.

Results

Figure 6 shows LUC for Dyad A as per the macro criterion established at the onset of the study (i.e., 90% x 1). The results showed that there were no differences in LUC across teachers for the student with BiN on the first objectives taught. Both teachers presented 40 LU before mastery criterion was achieved. For the second objective taught, Teacher Assistant A presented 20 more LU (60) than Teacher A (40) to achieve mastery criterion. For the student with BiN, there were no differences in LUC across teachers for both objectives taught. Both teachers presented 60 LU for the first objective taught, and 80 LU for the second objective taught.

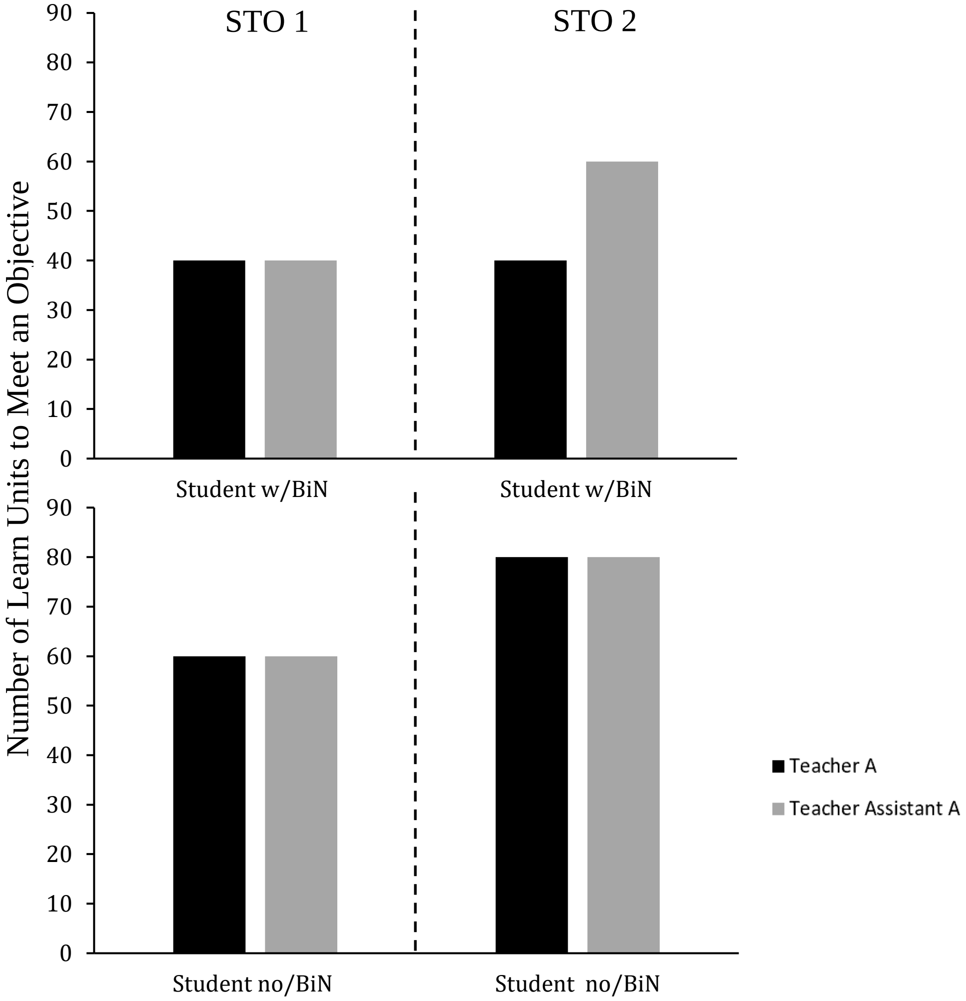

Figure 6. The number of LUC for Teacher A and Teacher Assistant A for two sight word objectives taught for a student with BiN and a student without BiN
The bar represents the total number of LUs presented to achieve macro criterion.

Figure 7 shows LUC for Dyad A as per the micro criterion established to further analyze differences in the rate of acquisition (i.e., five consecutive correct responses per target word). The results showed that there was a slight difference in LUC across participants on the first objective taught for the student with BiN. Teacher A's LUC was 31 and Teacher Assistant A's LUC was 32 (1 LU difference). There was also a slight difference in LUC for the second objective taught for the student with BiN. Teacher A's LUC was 38 and Teacher Assistant A's LUC was 41 (3 LU difference). For the student without BiN, there was a slight difference in LUC across participants for the first objective taught. Teacher A's LUC was 44 and Teacher Assistant A's LUC was 47 (3 LU difference). As for the second objective taught, there was a slightly greater difference in LUC across participants. Teacher Assistant A presented 18 more LU (66 LU) than Teacher A (48 LU) to achieve micro criterion across target words taught.

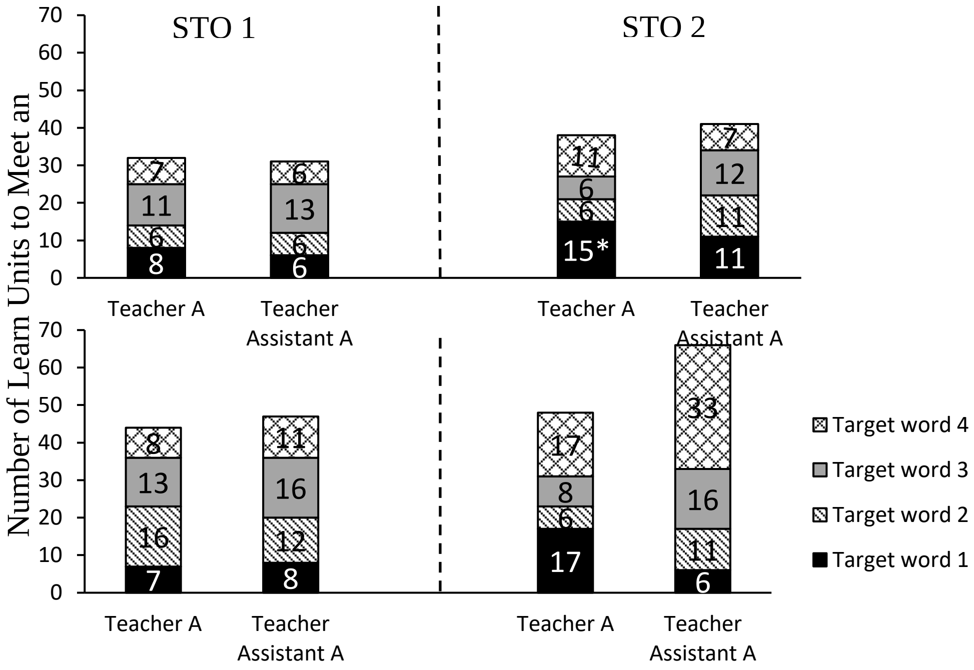

Figure 7. The number of LUC for Teacher A and Teacher Assistant A, for two sight word objectives taught for a student with BiN and a student without BiN
The bar represents the total number of LUs presented to achieve micro criterion. The asterisk (*) indicates that micro criterion was not achieved for that word.

Figure 8 shows LUC for Dyad B as per the macro criterion established at the onset of the study (i.e., 90% x 1). For the student with BiN, the results showed no differences in LUC for both objectives taught. Both teachers presented 40 LU for the first objective and 60 for the second objective before mastery criterion was achieved. For the student with no BiN. Teacher B presented 40 LU more than Teacher Assistant B (60 LU) before mastery criterion was achieved for the first objective taught. On the second objective taught, both teachers presented 60 LU before mastery criterion was achieved.

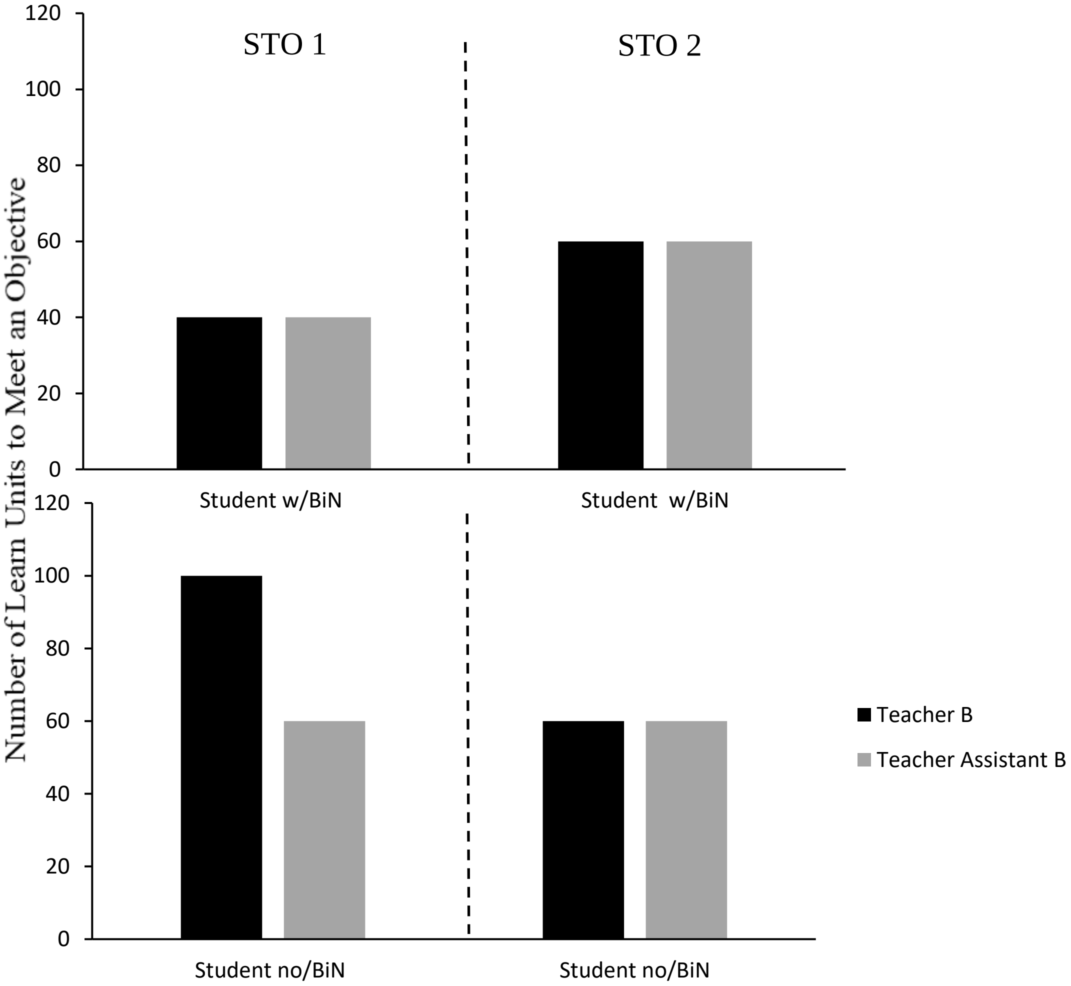

Figure 8. The number of LUC for Teacher B and Teacher Assistant B for two sight word objectives taught for a student with BiN and a student without BiN
The bar represents the total number of LUs presented to achieve macro criterion.

Figure 9 shows LUC for Dyad B as per the micro criterion established to further analyze differences in the rate of acquisition (i.e., five consecutive correct responses per target word). The results showed that there was a slight difference in LUC for the first objective taught for the student with BiN across participants. Teacher B's LUC was 22 and Teacher Assistant B's LUC was 26 (4 LU difference). There was also a slight difference in LUC for the second objective taught. Teacher B's LUC was 26 and Teacher Assistant B's LUC was 20 (6 LU difference). For the student without BiN, there was a slight difference in LUC across participants for the first objective taught. Teacher B's LUC was 24 and Teacher Assistant B's LUC was 30 (6 LU difference). There was also a slight difference in LUC for the second objective taught. Teacher B's LUC was 27 and Teacher Assistant B's LUC was 20 (7 LU difference).

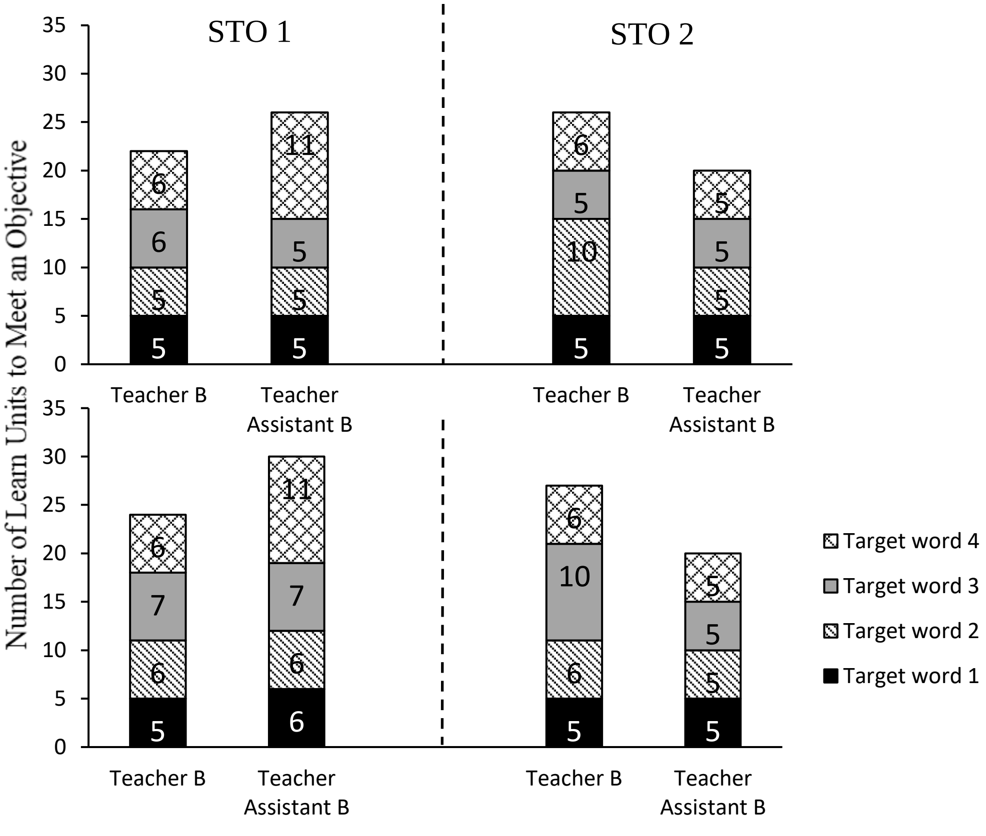

Figure 9. The number of LUC for Teacher B and Teacher Assistant B for two sight word objectives taught for a student with BiN and a student without BiN
The bar represents the total number of LUs presented to achieve micro criterion.

Figure 10 shows LUC for Dyad C as per the macro criterion established at the onset of the study (i.e., 90% x 1). For the student with BiN, the results showed that there were no differences in LUC for both objectives taught across teachers. Both teachers presented 40 LU before mastery criterion was achieved. For the student without BiN, Teacher C presented 20 more LU (180) than Teacher Assistant C (160 LU) before mastery criterion was achieved for the first objective taught. On the second objective taught, Teacher Assistant C presented 20 more LU (60) than Teacher C (40).

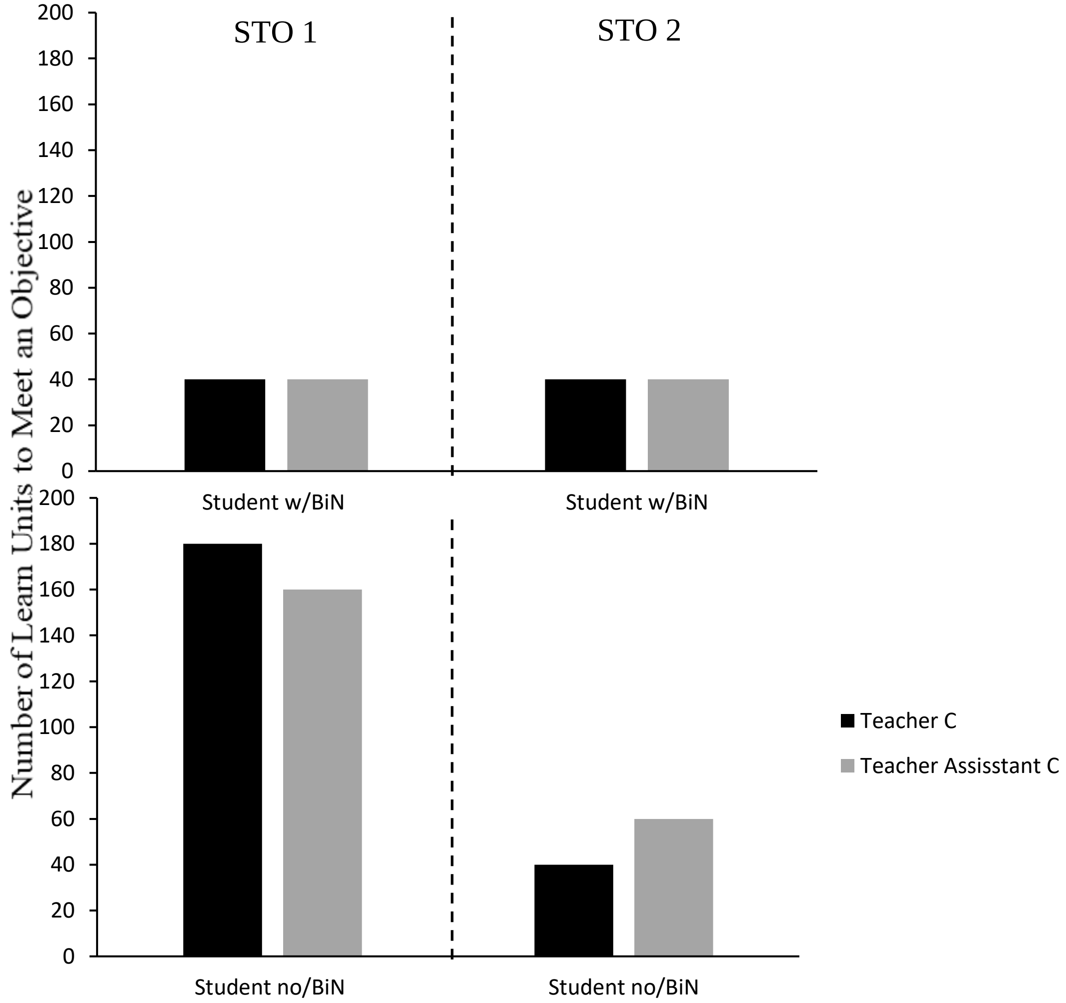

Figure 10. The number of LUC for Teacher C and Teacher Assistant C for two sight word objectives taught for a student with BiN and a student without BiN
The bar represents the total number of LUs presented to achieve macro criterion.

Figure 11 shows LUC for Dyad C as per the micro criterion established to further analyze differences in the rate of acquisition (i.e., five consecutive correct responses per target word). The results showed a slight difference in LUC across participants for the first objective taught for the student with BiN. Teacher C's LUC was 38 and Teacher Assistant C's LUC was 35 (3 LU difference). As for the second objective taught, there were no difference in LUC, both participants presented 26 LU before mastery criterion was achieved. For the student without BiN, there was a greater difference in LUC across teachers for the first objective taught. Teacher C presented 47 more LU (157 LU) than Teacher Assistant C (110 LU). For the second objective taught, there was a slight difference in LUC across teachers. Teacher C presented 36 LU and Teacher Assistant C presented 40 LU (4 LU difference).

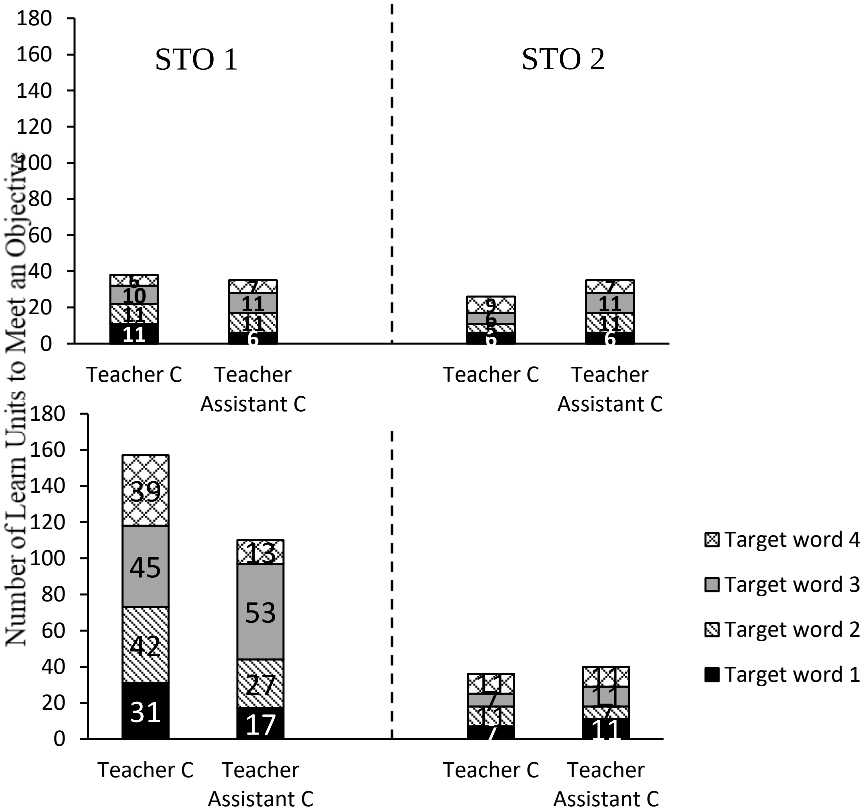

Figure 11. The number of LUC for Teacher C and Teacher Assistant C for two sight word objectives taught for a student with BiN and a student without BiN
The bar represents the total number of LUs presented to achieve micro criterion.

Figure 12 shows LUC for Dyad D as per the macro criterion established at the onset of the study (i.e., 90% x 1). The results showed that for the student with BiN, Teacher D presented 20 more LU (60) than Teacher Assistant D (40) for the first objective taught to achieve mastery criterion. For the second objective taught, Teacher Assistant D presented 40 more LU than Teacher D. For the student without BiN, there were no differences in LUC across teachers for both objectives taught. Both teachers presented 100 LU to achieve mastery criterion for the first objective, and 60 LU for the second objective

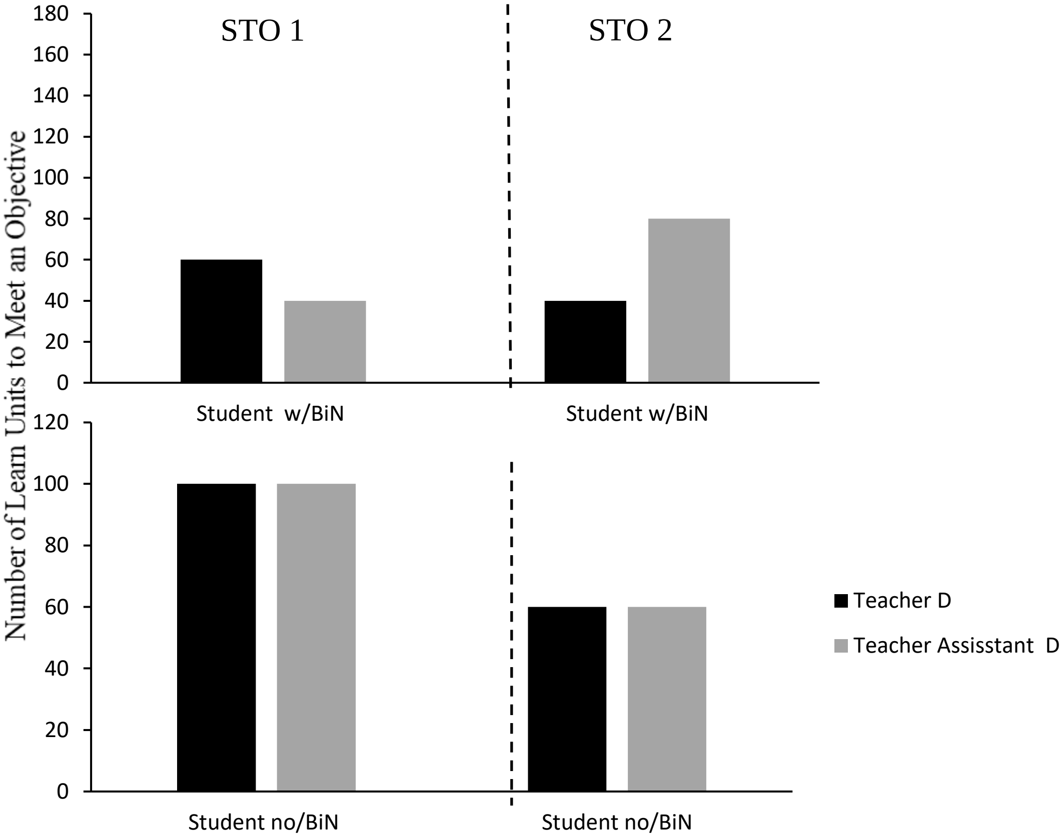

Figure 12. The number of LUC for Teacher D and Teacher Assistant D for two sight word objectives taught for a student with BiN and a student without BiN
The bar represents the total number of LUs presented to achieve macro criterion.

Figure 13 shows LUC for Dyad C as per the micro criterion established to further analyze differences in the rate of acquisition (i.e., five consecutive correct responses per target word). The results showed a noticeable difference in LUC across participants for the first objective taught for the student with BiN. Teacher D's LUC was 49 and Teacher Assistant D's LUC was 31 (18 LU difference). As for the second objective taught, there was a greater difference in LUC, Teacher D's LUC was 29 and Teacher Assistant D's LUC was 59 (30 LU difference). For the student without BiN, there was a slight difference in LUC across teachers for the first objective taught. Teacher D presented 68 LU and Teacher Assistant D presented 64 LU (4 LU difference). For the second objective taught, there was a slightly greater difference in LUC across teachers. Teacher D presented 60 LU and Teacher Assistant D presented 41 LU (19 LU difference).

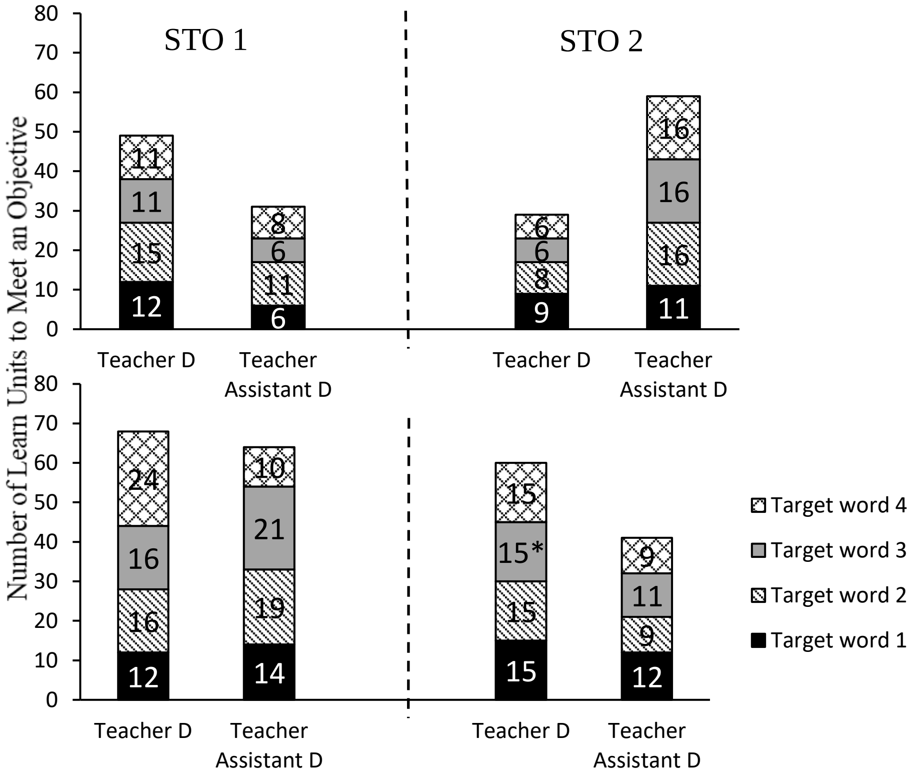

Figure 13. The number of LUC for Teacher D and Teacher Assistant D for two sight word objectives taught for a student with BiN and a student without BiN
The bar represents the total number of LUs presented to achieve micro criterion.
The asterisk (*) indicates that micro criterion was not achieved for that word.

Figure 14 represents the differences in acquisition rates for students with BiN and students without BiN. For students of Dyad A, the student with BiN had a LUC of 45 and student without BiN had LUC of 70 (25 LU difference). For students of Dyad B, the student with BiN had a LUC of 50 and student without BiN had LUC of 70 (20 LU difference). For students of Dyad C, the student with BiN had a LUC of 40 and student without BiN had LUC of 100 (60 LU difference). For students of Dyad D, the student with BiN had a LUC of 55 and student without BiN had LUC of 88 (33 LU difference).

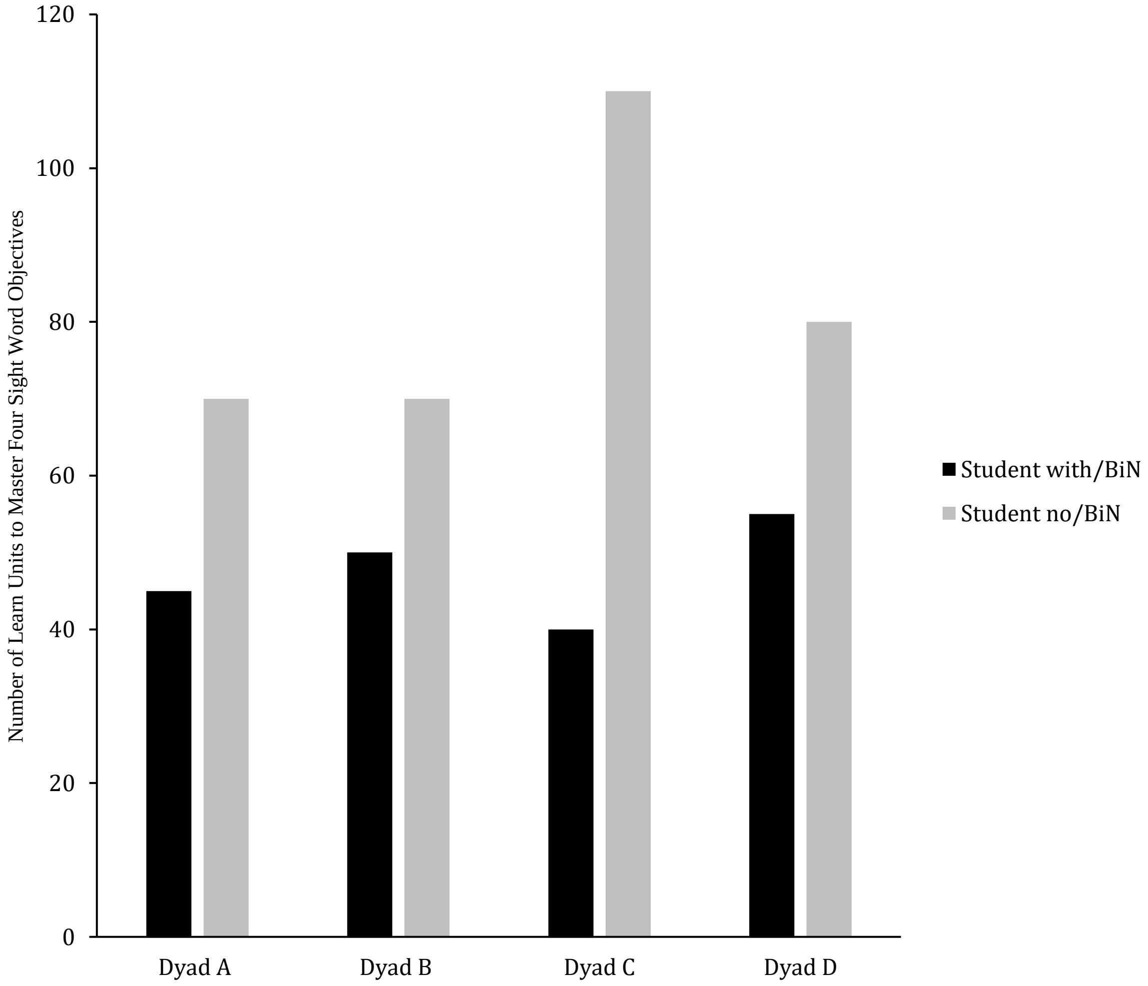

Figure 14. The differences in LUC for students with BiN and students without BiN The black bar represents LUC for the student with BiN and the grey bar represents LUC for the student without BiN.

Figure 15 shows teachers' and teacher assistants' LUC as per macro and micro criteria across the four sight word objectives taught. Teacher A's LUC as per macro criteria was 55 and Teacher Assistant A's LUC was 60. As per micro criteria, Teacher A's LUC was 40 and Teacher Assistant A's LUC was 46. Teacher B's LUC as per macro criteria was 65 and Teacher Assistant B's LUC was 55. As per micro criteria, Teacher B's LUC was 25 and Teacher Assistant B's LUC was 24. Teacher C's LUC as per macro criteria was 75 and Teacher Assistant C's LUC was 75. As per micro criteria, Teacher C's LUC was 64 and Teacher Assistant C's LUC was 53. Teacher D's LUC as per macro criteria was 65 and Teacher Assistant D's LUC was 70. As per micro criteria, Teacher D's LUC was 51 and Teacher Assistant D's LUC was 48.

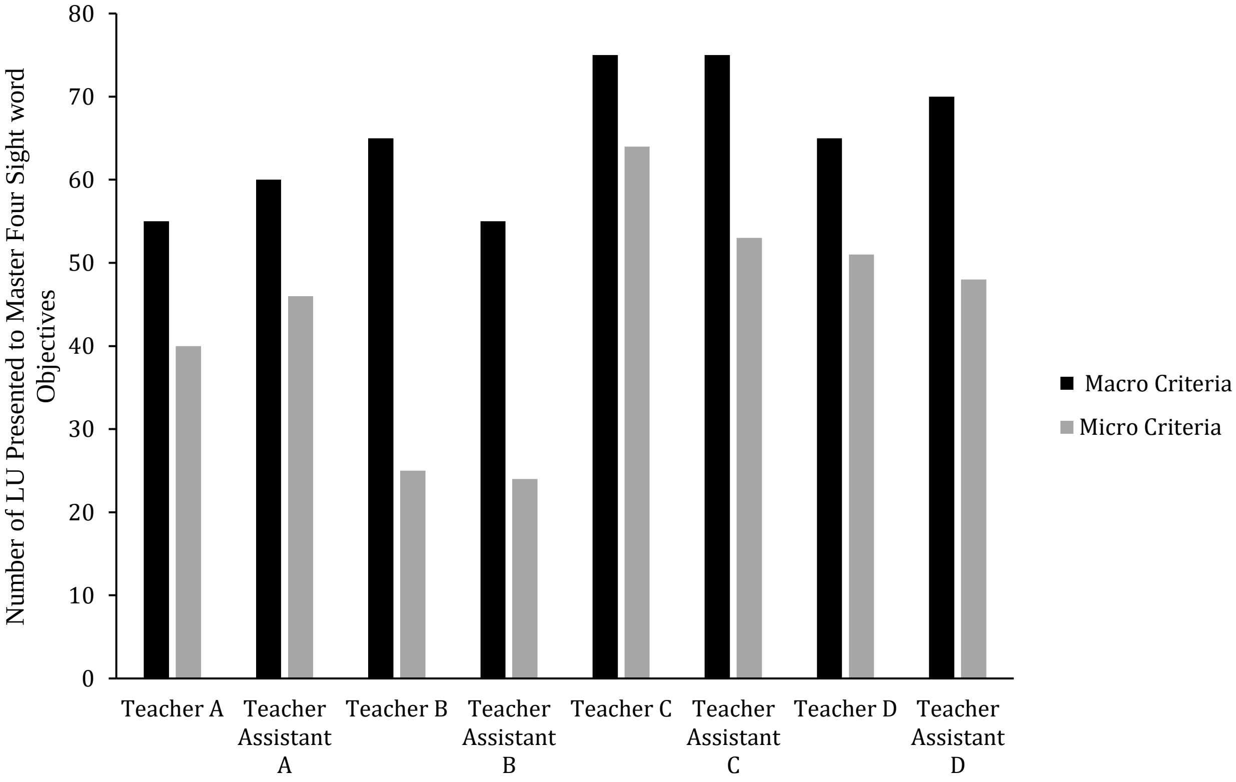

Figure 15. The teachers' and teacher assistants' LUC across four sight word objectives taught as per macro and micro criteria
The grey bar represents macro criteria LUC and the grey bar represents micro criteria LUC.

Discussion

The purpose of the second experiment was to conduct an experimental analysis of the correlations found in the descriptive study between teacher competencies and teacher effectiveness as measured by LUC. The study was designed to compare LUC for teachers with more competencies mastered and teacher assistants with less competencies mastered in order to investigate whether teachers with more competencies mastered will have a lower LUC. The results of the study did not show a functional relationship between teacher LUC for sight word instruction and the number of competencies mastered. An objective taught by a teacher with more competencies mastered did not consistently result in lower LUC when compared to the teacher assistant with fewer competencies mastered. Moreover, these results did not replicate the findings found in the correlations conducted in the statistical analysis. To further analyze these findings, the differences in LUC were examined across conditions and dyads to determine if there was a notable variance in LUC.

When analyzing the number of sessions required to meet macro criterion (i.e., 90% x 1), three experimental conditions out of 16 conditions (19% of conditions) showed the teachers having lower LUC and three conditions showed the teacher assistants having a lower LUC (19% of conditions). Additionally, there were no differences in LUC for the remainder of experimental conditions (63% of sessions). An analysis of LUC with micro criteria revealed eight experimental conditions showing teachers as having lower LUC (50% of conditions) and eight conditions showing teacher assistants as having lower LUC (50% of sessions). Thus, an overall analysis of LUC across all 32 conditions, which included those analyzed using macro and micro criteria, revealed that both teachers and teacher assistants had a lower LUC for 25% of conditions. Therefore, 50% of conditions showed no differences in LUC between teachers and teacher assistant, and 50% of conditions showed that lower LUC was equally varied across teachers and teacher assistants.

There may be multiple explanations for the absence of a functional relationship between teacher competencies mastered and LUC in the second experiment. One explanation is that both teachers and teacher assistants were explicitly trained on delivering accurate LU, as evidenced by the number of CS competencies each teacher and teacher assistant mastered. CS modules consist of three components: (a) mastery of competencies on the matrix accompanying the rank; (b) accountability for achieving LUC that are appropriate for the student's level of VBD; and (c) presentation of LU at an adequate rate for the student's level of VBD. The matrix accompanying the rank is composed of additional competencies pertaining to classroom management, conducting and recording assessment data, and errorless TPRA observations for programs targeting verbal behavior, protocols to induce verbal behavior cusps/capabilities, academic literacy, and self-management. All participants in the study mastered at least 17 CS competencies and 85 at most. Consequently, we may conclude that mastering as few as 17 CS competencies can result in effective teaching as measured by student learning of objectives. Furthermore, Teacher Assistant D, who had the fewest number of CS competencies mastered, did not exhibit the highest LUC; this finding may provide further evidence to support that teachers who were trained on delivering accurate presentations of LU will be effective in increasing student correct responding and mastery of objectives (Albers & Greer, 1991).

Another explanation for a lack of a functional relationship may be due to the population of student participants selected for the study. Students were selected from classrooms that can be described as functioning at the bidirectional and independent listener/speaker level of VBD, which tend to have lower LUC than students functioning at the pre-foundational level of VBD, as evidenced by the correlation found in the descriptive analysis. Therefore, student participants may have had sufficient cusps/capabilities and prerequisite skills to learn sight words without encountering acquisition problems that required teachers' VM expertise. Across all conditions of the study, there was only one VM decision made by Teacher D on the first STO taught for the student without BiN. The teacher identified a problem with the antecedent portion of the LU and implemented a stimulus prompt tactic to address the problem. Implementation of the tactic did not result in lower LUC for that teacher; however, there was a 20% increase in correct responding from the previous session. There was also a 30% increase in correct responding during Teacher Assistant D's session, who did not implement the tactic during her session. This increase may have been due to a spillover effect from Teacher D's session. The decision to implement the tactic may have also established the appropriate stimulus control for responding to sight words, as evidenced by a 25% decrease in the student's LUC for the second objective taught across both teachers. This finding showed that teacher's VM competencies may lead to instructional decisions that accelerate student rate of learning, a prime measure of effective teaching (Singer-Dudek et al., 2010). Unfortunately, beyond the decision made by Teacher D, there was no further evidence to support the relationship between teacher VM competencies and LUC. Perhaps replicating the study with students functioning at the pre-foundational level will yield more opportunities for teacher VM decisions, since they may encounter more learning difficulties that require strategic analysis of the LU in context and the selection of an appropriate tactic to resolve the problem. Replicating the study with students functioning at the pre-foundational level of VBD may show a difference in LUC between teachers with more competencies mastered and those with fewer competencies mastered.

Another finding worth noting is that students with BiN consistently learned faster (i.e., lower LUC) than students without BiN; refer to Figure 14 for representation of LUC across students. This finding is in line with past research showing that students with BiN learn from instructional presentations and at a faster rate than students without BiN (Greer et al., 2011). We may conclude from these findings and past research that students with BiN will learn, regardless of the quality of instruction, because they can learn from teacher presentations alone (i.e., a demonstration on the chalk board of how to add 1+1), whereas students without BiN may need explicit instruction and perhaps some additional tactics to learn 1+1, which may or may not be offered in a general education or inclusion classroom. Thus, establishing BiN for students who do not have it may provide them with the tools needed to learn, regardless of the quality of instruction available to them. These results also extended the findings from the descriptive analysis with regard to the association between student level of VBD and LUC. Students with BiN may be considered as functioning at a higher level of VBD; thus, teachers of those students will naturally have lower LUC.

Limitations. The experiment was designed to control for confounding variables such as setting events by having the teacher and teacher assistant run an equal number of sessions on the same day. A possible limitation to the experimental design may have been the number of sessions that were not conducted on the same day across teachers (4% of sessions); two sessions for Dyads B and D were not conducted on the same day as their teacher assistants. Whether this error was confounding to the results is unclear; no variability in LUC was observed due to these sessions.

Another limitation involves the analysis of LUC with micro criterion since it was added as a subsequent analysis after data collection for the study was complete; two teachers, Teacher

A and Teacher D, did not teach one of the target words to mastery. This limitation may have underestimated their number of LUC according to the micro criteria. Specifically, it may have hindered the analysis of one objective for Dyad A since Teacher A's data showed a slightly lower LUC than Teacher Assistant A. Teacher D had a higher LUC than Teacher Assistant D, despite not teaching one word to mastery for that objective, so this limitation did not hinder Dyad D's analysis of LUC.

A final limitation may have been that the experimenter was too technological in vocal and written instructions when presenting the study to participating teachers. Since there were a number of rules to follow within the sequence of the design (i.e., run the same number of sessions per day), the participants may have refrained from deviating from the rules to the extent that the majority of teachers did not make any decisions or changes during instruction beyond presenting accurate LU. This limitation may be due to the teachers' rule-governed behavior established as a result of their instructional history with rule following. In other words, they may have learned that following the rules leads to reinforcement and deviating from the rules leads to punishment; hence, they may be motivated to follow the rules more often than not. Teachers and teacher assistants may have been reluctant to make changes to the objectives, even though there were no rules against making changes to the program or making decisions, if needed. The experimenter only stated that if a decision opportunity arises, the teacher and teacher assistant should not collaborate during the analysis of the problem and the decision to be made.

Only Dyad B made changes beyond the scripted objectives given to them by the experimenter; they included echoic prompts for the first few sessions and faded those prompts in subsequent sessions, a tactic that is common practice in their classroom (i.e., CS behavior). Teacher B sequenced all new instructional objectives to include echoic prompts. Using echoic prompts may explain why Dyad B had a considerably lower LUC than other dyads (refer to Figure 15) as per micro criterion analysis. It is important to note macro and micro criteria for Dyad B were determined for sessions that were independent of echoic prompts; therefore, low LUC was not due to prompting. Why Dyad B deviated from its rule-governed counterparts may not be clear; perhaps it is a case of CS behavior triumphing over rule-governed behavior.

We cannot be certain that the other teachers or teacher assistant would have done anything differently had the experimenter assured them that this was their independent teaching project, and so they were free to make any changes as long as they did not deviate from the rules. Nonetheless, it is worth considering the possible limitations of teacher rule-governed behavior in subsequent studies designed to measure teacher effectiveness.

General Discussion

This study was designed to investigate the relationship between teacher CABAS® ranks and teacher effectiveness as measured by LUC. The results showed that the number of competencies mastered within the three repertoires composing the rank were significantly associated with effective teaching, as evidenced by lower LUC in the first study. However, after accounting for student level of VBD, there was a moderate correlation between CS competencies and LUC. Although the correlation was moderate, it was bordering on significances, suggesting that a bigger sample size may have resulted in a higher correlation. The second experiment did not provide further evidence to support the role of teacher competencies on LUC. However, the results from the experimental analysis provided some evidence to support that CS competencies mastered contribute to effective teaching if we consider measuring effectiveness as mastery of objectives taught rather than rate of acquisition (i.e., LUC). This analysis may lead us to question different ways of measuring effectiveness including mastering short-term and long-term objectives as well as rate of acquiring those objectives. It may also lead us to question the contexts of when each measure of effectiveness is appropriate. For example, what measure of effectiveness is more appropriate for students with BiN, rate of acquisition, or the number of long-term objectives achieved per academic year, or both? Within the context of teaching students with BiN, both measures may be equally important. Rate of acquisition provides a measure of whether the teacher is effective in modifying instruction to match the students VBD level, such as incorporating instructional demonstration LU within their curricular programming, which has been shown to increase student with BiN rate of learning (Frank, 2018; Greer et al., 2011; Hranchuk, 2016). In addition to the rate of acquisition, the number of objectives taught can also be an important measure of teacher effectiveness to assess the teacher's capacity for covering instructional material within a given school year. There might also be an association between modifying instruction to increase acquisition rates and the number of objectives taught since the faster the students acquire objectives, the faster new objectives can be introduced. Therefore, both effectiveness measures may be equally important for all students of varying levels of VBD because the faster they can learn, the more we can teach.

Overall, these studies provided preliminary evidence to support the importance of operationally defining and establishing teacher repertoires for effective teaching. To further analyze the relationship between teacher competencies and teacher effectiveness, a closer look at each teacher repertoire composing the rank is warranted.

At first glance, it may seem that the ranks are a unique interpretation of what teachers need to have in order to be effective. However, educational researchers have investigated the effects of these repertoires on student learning to some degree. Darling-Hammond (2016), in a review of teaching and teacher evaluation, brought to light lines of research that contributed to the knowledge of teaching. Of particular importance, Darling-Hammond noted research efforts on teacher thinking and decision making (Clark, 1983), pedagogical content knowledge (Shulman, 1986), and teachers' practical knowledge (Clandinin, 1985; Grimmett & Mackinnon, 1992). Interestingly, each of the research efforts outlined above coincide with each of the teacher repertoires that compose the CABAS® rank. The next section addresses these bodies of research in line with each teacher repertoire composing the ranks.

Teacher Thinking or Verbally Mediated Repertoires

Some refer to teacher thinking and decision making as the "hidden side" of the teaching profession (Jackson, 1966). In other words, these are teacher processes that are not directly observable but control teacher practices in the classroom. This hidden process may be comparable to VM repertoires, which are defined as "verbal directions from the science control the analytic operations that a teacher applies to a particular learning difficulty experienced by a student" (Greer, 2002, p. 57). Clark (1983) suggested that research on teacher thinking has an inherently different goal than teacher effectiveness research because it aims to identify the inner workings of a teacher's mind on factors such as planning and decision making, rather than measuring observable behavior of effective teachers. He suggested that these lines of research do not have a competitive relationship, but a complementary one. Jackson (1966) also suggested that both lines of research are necessary to understand teaching. The results of the current study supported the notion that teacher thinking (i.e., VM repertoires) and teacher practices (i.e., CS repertoires) are intertwined. The number of competencies mastered within each of those domains were significantly associated with teacher effectiveness. In other words, teachers may become more effective (i.e., have a lower LUC) as they know more about the science by mastering unit quizzes (i.e., VB components), in addition to receiving more errorless TPRAs (i.e., CS components). Errorless TPRAs indicated that a teacher was delivering LUs without errors, which involved presenting the student with unambiguous directions, followed by the delivery of an appropriate consequence (i.e., reinforcement or correction procedure). An errorless TPRA is also characterized by recording student correct and incorrect responses to LU presentations with 100% accuracy. Therefore, if a teacher did not record student responses with 100% accuracy as per data collected by the observer conducting the TPRA, then it would not be considered errorless even if LU presentations were free from error.

Educational researchers on teacher thinking have taken different approaches. Some researchers focused on the planning aspect of teaching, meaning the procedures teachers undertake prior to teaching, such as organizing materials or designing the sequence of objectives to be taught. Some researchers investigated the nature and quality of teacher planning (Clark, 1983), while others examined the relationship between teacher planning and teacher decision making during instruction (Morine-Dershimer, 1979). Although informative, planning research may not provide enough evidence to support the notion that teacher planning leads to student achievement, as it is mostly ethnographic and descriptive in nature. A more promising line of research, in terms of student learning, addresses teachers' instructional decision making based on curriculum-based measurement (CBM) (Deno, 1985). CBM research is somewhat similar to the treatment of VM repertoires within the CABAS® approach to instruction (Keohane & Greer, 2005).

Decision-making Research: CBM vs. CABAS®

In both CBM and CABAS® models of instruction, instructional decision making is based on a graphical display of student progress, whereby teachers are required to make an instructional change when students' data show they are not learning at a satisfactory rate based on specific rules of data interpretation. However, there are some significant differences between CBM research and the decision tree protocol that should be noted. These differences include: rules for identifying decision opportunities, the frequency of data collection on student progress, frequency of analyzing student data, and how long-term objectives are chosen for instruction. For example, typical data collection in CBM protocols involves recording data on student performance twice a week and analyzing the data every 3-4 weeks (Stecker, Fuchs, & Fuchs, 2005), whereas as data collection and analysis within the CABAS® model occurs daily on all behavior targeted for instruction (Greer, 2002). Although these differences may be significant, there is evidence to support the use of monitoring student learning and modifying instruction accordingly on student achievement, in both CBM research (i.e., Fuchs, Deno, & Mirkin, 1984) and behavior analytic research (Keohane & Greer, 2005).

It is also important to note that CBM and CABAS® research on decision making has provided different contributions to the literature. On one hand, the CABAS® model provides a detailed strategic analysis of instructional problems to enable the selection of an accurate and effective solution to the problem from the literature (Greer, 2002). For example, if a student is not learning to identify letter names by pointing, the teacher may ask the following questions: (a) Does the student have the prerequisite skills to do so, such as emitting a point response to known items? or (b) Is the source of the problem phylogenetic in nature, whereby the student cannot physically extend his/her pointer finger to touch the target stimulus? or (c) Is the source of the problem a missing verbal developmental cusp, such as conditioned reinforcement for observing print stimuli? On the contrary, CBM decision-making models do not provide teachers with a strategic algorithm for analyzing learning problems; rather, CBM protocols only provide prompts for *when* the teacher should make an instructional change, but not *how* to decide what change to make (Stecker et al., 2005). This may be equivalent to telling a traveler that he/she is heading towards a dead end and a route change is in order, but not providing the traveler with a map or a GPS system to find his/her destination, leading the traveler towards a journey of trial and error and perhaps an endless guessing game.

Nonetheless, the contribution of CBM research on decision making is worth noting. CBM research addressed the application of a computer-based data collection and monitoring system to cater to teacher satisfaction and willingness to participate in data collection and analysis. For example, Fuchs (1988) found that teachers were more willing to comply with the decision-making protocol when a data management software graphed and provided an analysis of student data, whereby the system alerted the teacher of a decision opportunity. Other CBM research addressed the use of data collection and analysis in general education settings. For example, Fuchs, Fuchs, Hamlett, Phillips, and Bentz (1994) examined the class-wide application of CBM to mathematic instruction. The experimenters found that students performed better on a mathematical task (i.e., a 50-problem measure from a statewide curriculum) when they were assigned teachers in the CBM condition that provided the teachers with both class-wide skill analysis as well as instructional recommendations.

Contrary to CBM research, CABAS® research on decision making is yet to investigate the application of computer-based data collection and analysis systems on student achievement. It may be worth exploring the feasibility of programing the decision tree protocol into a data collection software, while measuring accurate teacher decision making and, in turn, student learning. Moreover, a data software may increase the application of data collection and analysis for teachers and practitioners outside of CABAS® schools and general education classrooms, which may aid in the dissemination of the strategic science of instruction set forth by the model

(Greer, 2002). There is some evidence to support the use of computerized data management in data-driven schools on increasing supervision time and the number of student objectives achieved (Babbitt, 1986). Babbitt (1986) also found that using data management software was both time- and cost-efficient and enabled 7.36% to 15.34% more accurate data calculations. Exploring the role of technology in aiding teachers to be more effective is an area of research certainly worth exploring, particularly within schools dictated by student and teacher data.

In either case, the studies reviewed above as well as the results from the current study revealed the importance of teachers' VM repertoires on effective teaching, as evidenced by increases in the rate of student learning (i.e., lower LUC). Thus, teacher preparation programs should address these repertoires in training prospective teachers and aim to foster a teacher culture that emphasizes the importance of measuring student learning and modifying instruction accordingly.

Pedagogical Content Knowledge or Verbal Behavior About the Science

The term *pedagogical content knowledge* (PCK) was first coined by Lee Shulman (1986) in his presidential address "Those Who Understand: Knowledge Growth in Teaching." Shulman proposed that PCK goes beyond teacher proficiency in subject matter; it is a category of teacher knowledge pertaining to teaching specific subject matter. In other words, he argued that teachers should not only be experts in the subject matter being taught, but also experts on how to teach it. Furthermore, Shulman suggested that PCK involves the following components: (a) providing students with representations of subject matter that are most conducive to comprehension; (b) these representations may be derived from research or wisdom of practice; (c) knowing what makes learning topics easy or difficult, which includes the understanding of the misconceptions and pre-conceived notions that different students may have about the subject matter; and (d) knowing strategies that will aid students in rectifying their misconceptions on a given topic.

Conceptually, PCK (Shulman, 1986) and VB about the science described earlier (Greer, 2002) share some similarities with regard to the role of teachers' pedagogical knowledge. These similarities include having an expert knowledge base on pedagogy of specific subject matter, modifying instruction to cater to student differences, and applying strategies from research to aid learning. There are, however, notable differences between the two concepts. One difference involves the motivation behind developing this repertoire for teachers. Greer (2002) proposed the VB about the science is an important teacher repertoire for the following reasons: (a) to "engage in the practices of the science"; (b) "to apply the findings of the science (in this case to teaching); (c) "to analyze events scientifically"; and (d) "to communicate…with other scientist practitioners" (p. 48). Greer (2002) also suggested that VB about the science is a prerequisite repertoire that is necessary for the development VM repertoires, which depends on the extensiveness of the teachers VB repertoire.

Although Shulman also emphasized the strategic application of pedagogical knowledge to learning problems, his suggestion for the application of CPK did not seem to go beyond expert knowledge of teaching and rectifying student misconceptions, which may be categorized as a problem in the students' instructional history interfering with learning new content—only one of many possible contexts of learning problems (Greer, 2002). Learning problems may stem from a variety of different sources, not just a student's instructional history; these may include motivational issues, phylogenetic limitations, and/or missing prerequisite skills, among others (Greer, 2002).

Theoretically, the researchers mentioned above proposed the potential importance and functions of teacher pedagogical knowledge for strategic teaching. However, empirical evidence supporting the role of teacher knowledge on student learning is somewhat limited in both PCK and CABAS® literature. PCK research is mostly ethnographic in nature (Phillips, 2014), involving qualitative data collection methods such as open-ended interviewing (Porta & Keating, 2008). In addition, a number of researchers examining PCK measured teachers' knowledge while neglecting to include measures of student learning (i.e., Hill et al., 2008; Kleickman et al., 2013). The studies that did measure learning provide some evidence to support that teachers' PCK contribute to student academic achievement (Baumert et al., 2010; Hill, Rowan, & Ball, 2005); however, results from these studies should be interpreted with caution due to missing data and attrition reported by the experimenters.

There is also some evidence to support the role of VB about the science on improving teacher VM repertoires in the CABAS® literature (Keohane & Greer, 2005; Nuzzolo, 2002). For example, in two experiments, Nuzzolo (2002) investigated the effects of direct and observed learn units provided by a supervisor on the emission of scientific tacts by participant teachers in intervention and classroom settings. The experimenter also measured accuracy of decisions based on the visual analysis of student data, the rate of correct and incorrect student responses, and percentage of correct responses. The results showed a functional relationship between supervisor direct and observed learn unit presentations on the emission of accurate scientific tacts emitted by teachers in both intervention and classroom settings. The experimenter also found that increases in accuracy of scientific tacts resulted in increases in the accuracy of data-based decisions teachers made, thus providing evidence to support the interdependent relationship between VB behavior about the science and VM repertoires of teachers.

Furthermore, the experimenter reported significant increases in the percentage of correct responses emitted by the participant teachers’ students; however, the experimenter did not report if correct responding was measured for the same objective taught during baseline conditions. Thus, increases in correct responses may have been a result of LU presentations delivered by the teachers over time. Keohane and Greer (2005) also found that instructing teachers to use verbally governed questions to address instructional problems increased students’ number of objectives achieved. Although the experimenters aimed to manipulate teachers’ VM decisions to examine their effects on student learning, this study provided some evidence showing that increasing teachers’ knowledge about the science (i.e., mastering the verbally governed questions) mediated the relationship between VM competencies and student learning.

The results of the studies summarized above as well as the results from the descriptive analysis provide slight evidence to support the notion that the VB about the science competencies contributes to student learning. Moreover, these results also provided evidence to support the interdependent relationship between VB and VM teacher repertoires; that is, increased acquisition of VB about the science aids in the development of teacher repertoires needed to identify, analyze, and solve instructional problems. These studies are the only research available on the effects of developing teacher VB repertoire on VM repertoires and student learning in the CABAS® literature. Since the first study was correlational in nature and the other studies (Keohane & Greer, 2005; Nuzzolo, 2002) did not intend to directly measure effects of teacher VB competencies on student learning, further evidence is warranted to examine the meditative relationship of teachers VB on their VM repertoire, and analyzing their effects on student rate of learning (i.e., LUC) and possibly an additional measure of learning—the acquisition of long-term objectives. I propose examining the effects of both of these repertoires on student learning because it may be difficult, if not impossible, to isolate the teachers' VB repertoire from VM repertoire. That is, once the teacher acquires knowledge about how to solve certain instructional problems, she will most likely use that knowledge in practice, resulting in the use of VM repertoires.

Teachers' Practical Knowledge or Contingency-shaped Behavior

Some refer to practical knowledge as craft knowledge, which is the intuitive wisdom that is acquired through practicing the craft of teaching and goes beyond the parameters of teaching as an applied science (Grimmett & Mackinnon, 1992). This approach romanticizes teaching by referring to it as an art rather than a science, a position that may have detrimental implications to education. Fortunately, this approach was not popular among educational researchers, as evidenced by the limited research on craft knowledge (Russell, 2015). Another proposition was offered by Clandinin (1985). He dubbed the term *personal practical knowledge* (PPK), a term he derived from conducting narrative research on teachers and established a framework for understanding its meaning and origins. Clandinin proposed that teachers' PPK is a collection of images that teachers accrue through professional and practical experience. These images in turn guide the teachers practice in the classroom. Here is an example from Clandinin's work to illustrate how images are constructed.

> Some experiences have a "watershed" character and form vivid detailed visual images. Aileen's image of a particular child's face is illustrative. She now views her experience with this particular child as a turning point in her teaching, frequently calling his face to mind with other children's problems.

Clandinin offered this narrative approach to teacher research as a way to provide a new perspective on the role of teachers in education, and to elevate their status to a more active role in educational research and practice. The proposition that Clandinin suggested with regard to images guiding teacher practice is somewhat similar to the notion of derived relational responding proposed by relational frame researchers (Hayes, Barnes-Holmes, & Roche, 2001). Hence, it may be possible that these images, or what RFT researchers refer to as relational frames, control or “guide” teacher practices. However, due to the narrative nature of Clandinin’s research, there is little, if any, evidence to support the role of PKK on student learning since research on PPK is mostly interpretive and qualitative in nature (i.e., Wetzel, Hoffman, Roach, & Russell, 2018). These approaches to practical knowledge are distinctly different from the treatment of CS repertoires in behavioral research, perhaps too different in their epistemology and methodology to propose a useful contrast or comparison.

Greer (2002) referred to CS repertoires as being fluent in applying the best practices of teaching as a science, which were derived from rigorous research on pedagogy and learning (i.e., the learn unit, contingent approvals, planned ignoring inappropriate behaviors etc.). Furthermore, Greer suggested that these repertoires are not to be mistaken for behaviors learned through trial and error in the classroom. CS repertoires are established through explicit training by supervisors or teacher mentors, until emitting these behaviors with fluency is established. The results of both experiments in this study supported the role of CS repertoires on effective teaching.

There are some similarities between CS repertoires and the observational rubrics used in observational systems such as the CLASS as well as performance assessment such the edTPA. There are a number of overlapping competencies across these rubrics and the CABAS® rank, such as creating a positive classroom environment, catering instruction to fit individual student needs, and providing adequate feedback. Taken together, these evaluation and training tools provide insight on what could be the most important competencies necessary for teachers to be effective in the classroom. Future research could aim to provide evidence to further support these competencies overlapping across observational systems, performance assessments, and the CABAS® rank.

The research reviewed on teacher repertoires thus far revealed a concerning pattern in relation to measuring teacher behaviors in the absence of student learning (i.e., Clandinin, 1985; Hill et al., 2005; Kleickmann, 2013; Wetzel, 2018). Neglecting student learning in the aforementioned literature may stem from the aversion that some researchers have towards what is commonly known as process-product research (i.e., teacher effectiveness research). Process-product studies can be defined as research investigating teacher behaviors that contribute to student academic achievement (Shulman, 1986), such as direct instruction or time on task. This aversion to process-product research may be due to the perfunctory mandates policymakers established as a result of process-product research, which mostly resulted in behavioral checklists for evaluating teachers that are insensitive to student learning and innovative teaching (Darling-Hammond, 2016). Although process-product research may be methodologically and conceptually flawed (Shulman, 1986), it involves the investigation of what may be the most crucial aspect of teaching—student learning. An aspect that is missing in the majority of qualitative research on teaching such as teacher thinking and PCK and PPK research. That is not to say that quantitative approaches to research (i.e., process-product methods) are superior or should replace qualitative research on teaching. I argue that without a measure of student learning, these research efforts—no matter how informative—may not contribute to the understanding, development, and practice of effective teaching. Studying components of teaching in the absence of student learning may be equivalent to taking data on a teacher giving a meticulously planned and innovative lesson to an empty classroom. Unlike gravity or the sun, teaching is not an independent natural phenomenon. The interdependent relationship between teaching and learning should limit the parameters of research on teaching so that it cannot be studied in the absence of learning. Although learning can occur in the absence of teaching (Greer, 2002), teaching cannot objectively occur in the absence of learners. Thus, I argue that all research on teaching must incorporate variables of student learning if it is to be considered a valid source for informing policy and practice.

Implications and Future Research

The findings of this study provided preliminary evidence to support the application of a strategic science of instruction on effective teaching as measured by student outcomes. This study is also unique in its contribution to the literature as it is the first to examine the association between each teacher repertoire composing the rank and student learning. The findings supported the need to operationally define teaching behaviors that are effective in producing measurable student outcomes. This implication may not be welcomed by educational researchers who argued that process-product research does not address teaching behaviors that evoke meaningful learning that goes beyond rote learning (Darling-Hammond, 2016). However, operationally defining teacher repertoires does not necessarily translate to establishing behavioral checklists, such as the ones designed to evaluate teachers as a result of process-product research (Darling-Hammond, 2016). The operational definitions of these repertoires provided a potential platform for teacher training and effective teaching research, so that measuring teaching becomes more accessible and less of a paradox.

It is also important to note that although this study incorporated what may be considered a valid measure of student learning, including LUC and mastery of objectives, measuring teacher effectiveness is not necessarily limited to these measures alone. Future researchers may also consider measuring effectiveness by the number of long-term objectives students master.

We can conclude from the review of the literature and the preliminary findings from the current study that teaching as a science provides practical validity for all entities involved in the educational context, including teachers, students, and researchers.

Future research should aim to replicate the findings of the descriptive analysis with a larger sample that includes more variability across student level of VBD and the number of competencies mastered by teachers. Moreover, researchers could examine the relationship between competencies mastered and teachers' correct and incorrect decisions as well as the number of errorless TPRA observations teachers receive. Another possibility for future researchers could be to experimentally isolate the type of competencies mastered by teachers to examine the effects of CS, VB, and VM on LUC. Moreover, researchers could examine the effects of only isolating VB and VM competencies by having teachers master an equal number of CS competencies, while having some teachers master only VB competencies and other teachers only master VM competencies. This study may be more feasible since all teachers will be trained on delivering accurate LU and successfully managing student behavior, perhaps allowing a fairer comparison across teachers.

The absence of significant and functional relationships in the current study may be due to extraneous variables that future researchers could control for. One such variable was grouping students with BiN and those without in one category. Differences in acquisition rates, instructional programing, and the form of instruction between students functioning at the bidirectional level of VBD and the pre-foundational level may be too vast to include in the same sample. Thus, analysis of these groups should be conducted separately so that teacher competencies associated with effectively teaching these different student profiles can be identified and improved. Furthermore, the second experiment did not include variables that were sensitive to detect and measure teacher competencies, specifically VM decisions. Some potential variables that may result in more opportunities to measure VM decisions may be the novelty of instructional programs, selecting students functioning at the pre-foundational level of VBD, or conducting the experiment with novice teachers. Finally, future experimenters may also consider increasing the number of objectives taught by each teacher across different types of programming (e.g., verbal behavior programs and other academic programs such as math) for more accurate and stable representations of teacher LUC.

Revisiting Baer et al.'s (1987) Dissemination Questions

As stated previously, this study was an effort to bridge the gap between research in educational psychology and behavior analysis, in order to establish the parameters of quality teaching that serve both teacher and student populations. This goal may be implausible, at least in the near future; nonetheless, behavior analysts should still aim to develop creative solutions for disseminating the science effectively. To entertain this goal in regard to the findings of the current study, we must revisit the questions posed by Baer et al. (1987):

> When a program is disseminated, should its disseminators require that its procedures be followed faithfully, no matter where or when the program is used? Or should its users be allowed, and even encouraged, to modify those procedures to fit their local situations and contingencies? (We might first ask, functionally, when we have that choice). (p. 321)

Thus, the question remains: Can CABAS® ranks be modified at a larger scale so that their application is effective and contextually appropriate for its users? Answering this question requires a thoughtful analysis on the part of behavior analysists of the contingencies that control educational research efforts as well as other entities in the educational community such as policymakers, schools, and teachers. Furthermore, behavior analysts should assimilate research that is purely controlled by the motivation to disseminate the science to address the gap inherent in this line of research (Baer et al., 1987).

REFERENCES

Albers, A. E., & Greer, R. D. (1991). Is the three-term contingency trial a predictor of effective instruction? *Journal of Behavioral Education, 1,* 337-354.

American Educational Research Association (AERA). (2015). AERA statement on use of value-added models (VAM) for the evaluation of educators and educator preparation programs. *Educational Researcher, 44,* 448-452.

Amrein-Beardsley, A., Collins., C., Holloway-Libell, J., & Paufler, N. A. (2016). Everything is bigger (and badder) in Texas: Houston's teacher value-added system. *Teachers College Record*. Retrieved from http://www.tcrecord.org/Content.asp?ContentId=18983

Babbitt, R. L. (1986). *Computerized data management and the distribution and rate of tasks performed by supervisors of a data-based residential educational organization*. (unpublished doctoral dissertation). Columbia University, New York.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1987). Some still-current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 20,* 313-327.

Baer, R. A., Detrich, R., & Weninger, J. M. (1988). On the functional role of the verbalization in correspondence training procedures. *Journal of Applied Behavior Analysis, 21,* 345-356.

Bahadourian, A. J., Tam, K. Y, Greer, R. D., & Rousseau, M. K. (2006). The effects of learn units on student performance in two college courses. *International of Behavioral Consultation and Therapy, 2,* 246-264.

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E, Ladd, H. F., Linn, R. L., Ravitch. D., Rothstein, R., Shavelson, R. J., & Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers*. EPI Briefing Paper #278. Washington, DC: Economic Policy Institute.

Baumert, J., Kunter., M., Blum., W., Brunner, M., Voss, T., Jordan., A., Klusmann, U., Krauss., S., Neubrand., & Tsai., Y-M. Teachers' mathematical knowledge, cognitive activation in classroom and student progress. *American Educational Research Journal, 47,* 133-180.

Benjamin, A. C. (1955). *Operationism.* Springfield, IL: Bannerstone House.

Berliner, D. C. (2015). Exogenous variables and value-added assessments: A fatal fFlaw. *Teachers College Record*, Retrieved from http://www.tcrecord.org/Content.asp?ContentId=010307

Berliner, D. C. (2018). Between Scylla and Charybdis: Reflections on and problems associated with the evaluation of teachers in an era of metrification. *Education Policy Analysis Archives, 26,* 1-29.

Boring, E. G. (1945). The use of operational definitions in science. *Psychological Review, 52,* 243-245.

Borman, G. D., & Kimball, S. M. (2005). Teacher quality and educational equality: Do teachers with higher standards-based evaluation ratings close student achievement gaps? *The Elementary School Journal, 106,* 3-20.

Bridgman, P. W. (1927). *The logic of modern physics*. New York, NY: The Macmillan Co.

Briggs, D., & Domingue, B (2011). *Due diligence and the evaluation of teachers: A review of the value-added analysis underlying the effectiveness rankings of Los Angeles Unified School District teachers by the Los Angeles Times*. Boulder, CO: National Education Policy Centre.

Brophy, J. E. (1983). If only it were true: A response to Greer. *Educational Researcher, 12,* 10-13.

Buell, M., Han, M., & Vukelich, C. (2017). Factors affecting variance in classroom assessment scoring system scores: Season, context, and classroom composition. *Early Child Development and Care, 187,* 1635-1648.

Calfee, R. (1981). Cognitive psychology and educational practice. *Review of Research in Education*, *9,* 3-73.

Cantrell, S. M. (2012). The measures of effective teaching project: An experiment to build evidence and trust. *Education Finance and Policy, 7,* 203-218.

Catania, A. C. (1979). *Learning*. Englewood Cliffs, NJ: Prentice-Hall.

Chang, H. (2009). Operationalism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2009/entries/operationalism/>.

Choi, J., Greer, R. D., & Keohane, D. (2015). The effects of an auditory match-to-sample procedure on listener literacy and choice. *Behavioral Development Bulletin, 20,* 186-206.

Clandinin, J. (1985). Personal practical knowledge: A study of teachers' classroom images. *Curriculum Inquiry, 15,* 361-385.

Clark, C. M. (1983). Research on teacher planning: An inventory of the knowledge base. In D. Smith (Ed.), *Essential knowledge for beginning educators*. Washington, DC: American Association for Colleges of Teacher Education.

Colman, J. S. (1966). *Equality of educational opportunity.* Washington, DC: National Center for Educational Statistics, Washington, DC.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis.* Upper Saddle River, NJ: Pearson/Merrill-Prentice Hall.

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching* (2nd ed.). Washington, DC: Association for Supervision and Curriculum Development.

Danielson, C. (2013). *The framework for teaching evaluation instrument: Instructionally focused edition.* Princeton, NJ: Educational Testing Service.

Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice*. Princeton, NJ: Educational Testing Service.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives, 8*, 1-44.

Darling-Hammond, L. (2010). *Evaluating teacher effectiveness: How teacher performance assessments can measure and improve teaching.* Washington, DC: Center for American Progress.

Darling-Hammond, L. (2015). Can value added add value to teacher evaluation? *Educational Researcher*, *44,* 132-137.

Darling-Hammond, L. (2016). Research on teaching and teacher education and its influences on policy and practice. *Educational Researcher, 45*, 83-91.

Darling-Hammond, L., Amrein-Beardsley, A., Heartel, E. & Rothstein, J. (2012). Evaluating teacher evaluation. *The Phi Delta Kappan, 6,* 8-15.

Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Heilig, J. V. (2005). Does teacher preparation matter? Evidence about teacher, certification, Teach for America, and teacher effectiveness. *Educational Policy Analysis Archives, 13*, 1-51.

Dee, T. S. (2004). The race connection: Are teachers more effective with students who share their ethnicity? *Education Next*, *4,* 52-59.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219-232.

Ehrenberg, R. G., Goldhaber, D. D., & Brewer, D. J. (1995). Do teachers’ race, gender, and ethnicity matter? Evidence from the National Educational Longitudinal Study of 1988. *Industrial and Labor Relations Review, 48,* 547-561.

Elrich, O., & Borich, G. (1979). Occurrence and generalizability of scores on a classroom interaction instrument. *Journal of Educational Measurement, 16,* 1979.

Elrich, O., & Shavelson, R. J. (1978). The search for correlations between measures of teacher behavior and student achievement: Measurement problem, conceptualization problem, or both? *Journal of Educational Measurement, 15*, 77-89.

Engelmann, S., Becker, W. C., Carnine, D., & Gersten, R. (1988). The direct instruction follow through model: Design and outcomes. *Education and Treatment of Children, 11,* 303-317.

Ertmer, P. A., & Newby, T. J. (1993). Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly, 6,* 50-72.

Everson, K. C. (2017). Value-added modeling and educational accountability: Are we answering the real questions? *Review of Educational Literature, 87,* 35-70.

Every Student Succeeds Act of 2015, Pub. L. No. 114-95, 1802 Stat. 129 (2015).

Falcomata, T. S. (2015). Defining features of applied behavior analysis. In H. S. Roane (Ed.), *Clinical and organizational applications of applied behavior analysis* (pp. 1-18). doi: 10.1016/B978-0-12-420249-8.00001-0

Fantuzzo, J. W., LeBoeuf, W. A., & Rouse, H. L. (2014). An investigation of the relationship between school concentrations of student risk factors and student educational well-being. *Educational Researcher, 43*, 25-36.

Feigl, H. (1945). Operationalism and scientific method. *Psychological Review, 52*, 250-259.

Fenstermacher, G. D., & Richardson, V. (2005). On making determinations of quality in teaching. *Teachers College Record, 107,* 186-213.

Frank, M. R. (2018). *The effect of bidirectional and unidirectional naming on learning in new ways and the relation between bidirectional naming and basic relational concepts for preschool students* (Order No. 10808429). Available from ProQuest Dissertations & Theses Global. (2039532063). Retrieved from https://tc.idm.oclc.org/login?url= https://search-proquest-com.tc.idm.oclc.org/docview/2039532063?accountid=14258

Frank, P. G. (1956). *The validation of scientific theories*. Boston, MA: Beacon Press.

Fuchs, L. S. (1988). Effects of computer-managed instruction on teachers' implementation of systematic monitoring programs and student achievement. *The Journal of Educational Research, 81*, 294-304.

Fuchs, L. S. (1988). Effects of computer-managed instruction on teachers' implementation of systematic monitoring programs and student achievement. *The Journal of Educational Research, 81*, 294-304.

Fuchs, L. S., Deno, S. L., & Mirkin, P. K. (1984). The effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal, 21*, 449-460.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., & Bentz, J. (1994). Classwide curriculum-based measurement: Helping general educators meet the challenge of student diversity. *Exceptional Children, 60,* 518-537.

Gallagher, H. A. (2004). Vaughn Elementary's innovative teacher evaluation system: Are teacher evaluation scores related to growth in student achievement? *Peabody Journal of Education, 79,* 79-107.

Goe, D. (2008). *Approaches to evaluating teacher effectiveness: A research synthesis.* Washington, DC: National Comprehensive Center for Teaching Quality.

Goe, L. (2007). *The link between teacher quality and student outcomes: A research synthesis.* Washington, DC: National Comprehensive Center for Teacher Quality.

Goe, L., Bell, C., & Little, O. (2008). *Approaches to evaluating teacher effectiveness: A research Synthesis*. Washington, DC: National Comprehensive Center for Teacher Quality.

Goldhaber, D. & Hansen, M. (2013). Is it just a bad class? Assessing the long-term stability of estimated teacher performance. *Economica, 80,* 589-612.

Goldhaber, D. D., Brewer, D. J., & Anderson, D. J. (1999). A three-way error components analysis of educational productivity. *Education Economics, 7,* 199-208.

Greer, A. B. (2018). *Vocal and non-vocal verbal behavior between mothers and their children diagnosed with autism spectrum disorder* (unpublished doctoral dissertation). Columbia University, New York.

Greer, R. D. (1983). Contingencies of the science and technology of teaching and prebehavioristic research practices in education. *Educational Researcher, 12,* 2-9.

Greer, R. D. (1991). The teacher as strategic scientist: A solution to our educational crisis? *Behavior and Social issues, 1,* 25-41.

Greer, R. D. (1997). The comprehensive application of behavior analysis to schooling (CABAS®). *Behavior and Social Issues, 7,* 59-63.

Greer, R. D. (2002). *Designing teaching strategies: An applied behavior analysis systems approach*. San Diego, CA: Academic Press.

Greer, R. D., Corwin, A., & Buttigieg, S. (2011). The effects of the verbal developmental capability of naming on how children can be taught. *Acta de Investigacion Psicologia, 1,* 23-54.

Greer, R. D., & Keohane, D. D. (2009). The evolution of verbal behavior in children. *The Journal of Speech and Language Pathology and Applied Behavior Analysis*, *4*, 10-39.

Greer, R. D., Keohane, D. D., & Healy, O. (2002). Quality and comprehensive applications of behavior analysis to schooling. *The Behavior Analyst Today, 3*, 120-132.

Greer, R. D., & McCorkle, N. P. (2009). *International curriculum and inventory of repertoires for children from pre-school through kindergarten*. Yonkers, NY: Fred S. Keller School and CABAS®.

Greer, R. D., McCorkle, N., & Williams, G. (1989). A sustained analysis of the behaviors of schooling. *Behavioral Residential Treatment*, *4*, 113-141.

Greer, R. D., & McDonough, S. H. (1999). Is the learn unict a fundamental measure of pedagogy? *The Behavior Analyst, 22*, 5-16.

Greer, R. D., & Ross, D. E. (2008). *Verbal behavior analysis: Inducing and expanding new verbal capabilities in children with language delays*. Boston, MA: Pearson Education.

Greer, R. D., & Speckman, J. M. (2009). The integration of speaker and listener responses: A theory of verbal development. *The Psychological Record, 59*, 449-488.

Grimmett, P. P., & Mackinnon, A. M. (1992). Craft knowledge and the education of teachers. In G. Grant (Ed.), *Review of research in education*. Washington, DC: American Educational Research Association.

Grossman, P., Cohen, J., Ronfeldt, M., & Brown, L. (2014). The test matters: The relationship between classroom observation scores and teacher value added on multiple types of assessment. *Educational Researcher, 43*, 293-303.

Haertel, E. H. (2013). Reliability and validity of inferences about teachers based on student test scores. The 14th William H. Angoff Memorial Lecture., Princeton, NJ: Educational Testing Service.

Hall, R. V. (1991). Behavior analysis and education: An unfulfilled dream. *Journal of Behavioral Education*, *1*, 305-316.

Hanushek, E. A., Kian, J. F., O'Brien, D. M., & Rivkin, S. G. (2005). *The market for teacher quality* (Working Paper No. 11154). Cambridge, MA: National Bureau for Economic Research.

Haq, S. S., & Kodak, T. (2015). Evaluating the effects of massed and distributed practice on acquisition and maintenance of tacts and textual behavior with typically developing children. *Journal of Applied Behavior Analysis, 48*, 85-95.

Harris, D., N., Ingle, W. K., & Rutledge, A, A. (2014). How teacher evaluation methods matter for accountability: A comprehensive analysis of teacher effectiveness ratings by principles and teacher value-added measures. *American Educational Research Journal, 51,* 73-112.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. New York, NY: Kluwer Academic/Plenum Publishers.

Heneman, H. G., Milanowski, A., Kimball, S., & Odden, A. (2006). Standards-based teacher evaluation as a foundation for knowledge-and skill based pay. CPRE Policy Briefs. Retrieved from http://repository.upenn.edu/cpre_policybriefs/33

Hill, C. H., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems a case for the generalizability study. *Educational Researcher, 41,* 56-64.

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction, 26,* 430-511.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal, 42,* 371-406.

Holtzapple, E. (2003). Criterion-related validity evidence for a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education, 17,* 207-219.

Horne, P. J., & Lowe, C. F. (1996). On the origins of naming and other symbolic behavior. *Journal of the Experimental Analysis of Behavior, 65,* 185-241.

Hranchuk, K. (2016). *Instructional demonstrations are more effective than consequences alone for children with naming* (Order No. 10096871). Available from ProQuest Dissertations & Theses Global. (1781222296). Retrieved from https://tc.idm.oclc.org/login?url=https://search-proquest-com.tc.idm.oclc.org/docview/1781222296?accountid=1425

Ingham, P., & Greer, R. D. (1992). Changes in student and teacher responses in observed and generalized settings as a function of supervisor observation. *Journal of Applied Behavior Analysis, 25,* 153-164.

Jackson, P. W. (1966). *The way teaching is*. Washington, DC: National Education Association, Washington, DC. Center for the Study of Instruction.

Johnson, K. R., & Layng, T. V. J. (1994). The Morningside model of generative instruction. In R. Gardner, D. Sainato, J. Cooper, T. Heron, W. Heward, J. Eshleman, & T. Grossi (Eds.), *Behavior analysis in education: Focus on measurably superior instruction*. Belmont, CA: Brooks/Cole.

Kane, T. J., & Staiger, D. O. (2012). Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. MET Project Research Paper.

Kane, T. J., McCaffrey, D. F., Miller, T, & Staiger, D. O. (2013). Have we identified effective teachers? Validating measure of effective teaching using random assignment. MET Project Research Paper.

Keller, F. S. (1968). Good-bye teacher. *Journal of Applied Behavior Analysis, 1,* 79-89.

Keller, F. S. (1978). Instructional technology and educational reform: 1977. *Behavior Analyst, 1,* 48-52.

Kenedy, C. H., Itkonen, T., & Lindquist, K. (1994). Nodality effects during equivalence class formation: An extension to sight-word reading and concept development. *Journal of Applied Behavior Analysis, 27*, 673-683.

Keohane, D. (1997). *A functional relationship between teachers' use of scientific rule governed strategies and student learning*. (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (Order No. 9723806).

Keohane, D., & Greer, R. D. (2005). Teachers' use of a verbally governed algorithm and student learning. *International Journal of Behavioral Consultation and Therapy, 3*, 252-271.

Kimball, S. M., White, B., Milanowski, A. T., & Borman, G. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education, 79*, 54-78.

Kleikman, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S., & Baumert, J. (2013). Teachers' content knowledge and pedagogical knowledge: The role of structural differences in teacher education. *Journal of Teacher Education, 64,* 90-106.

Klette, K., & Blikstad-Ballas (2018). Observation manuals as lenses to classroom teaching: Pitfalls and possibilities. *European Educational Research Journal, 17,* 129-146.

Koedel, C., & Julian, R. B. (2011). Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein Critique. *Education Finance and Policy, 6*, 18-24.

La Paro, K. M., Pianta, R. C., & Stuhlman, M. (2004). The classroom assessment scoring system: Findings from the prekindergarten year. *The Elementary School Journal*, 104, 409-426.

Langfeld, H. S. (1945). Symposium on operationism: Introduction. *Psychological Review, 52,* 241-242.

Lindsley, O. R. (1990). Precision teaching: By teachers for children. *Teaching Exceptional Children, 22,* 10-15.

Lockwood, J. R., McCaffrey, D. F., Hamilton, L. S., Stecher, B., Le, V., & Martinez, J. F. (2007). The sensitivity of value-added teacher effect estimates to different mathematics achievement measures. *Journal of Educational Measurement, 44,* 47-67.

Lodhi, S., & Greer, R. D. (1989). The speaker as listener. *Journal of the Experimental Analysis of Behavior, 51,* 353-359.

McGreal, T. L. (1990). The use of rating scales in teacher evaluation: Concerns and recommendations. *Journal of Personnel Evaluation in Education, 4,* 41-58.

Miguel, C. F. (2016). Common and intraverbal bidirectional naming. *The Analysis of Verbal Behavior, 32*, 125-138.

Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education, 79,* 33-53.

Moore, J. (1985). Some historical and conceptual relations among logical positivism, operationism, and behaviorism. *The Behavior Analyst, 8*, 53-63.

Moore, J. (2008). *Conceptual foundations of radical behaviorism.* Cornwall-on-Hudson, NY: Sloan Publishing.

Morine-Dershimer, G. (1979). *Teacher plan and classroom reality: The South Bay study IV.* (Research Series No. 60). East Lansing, MI: Michigan State University Institute for Research on Teaching.

Newton, X. A., Darling-Hammond, L., Haertel, E, & Thomas, E. (2010). Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts. *Education Policy Analysis Archives, 18,* Retrieved, from http://epaa.asu/ojs/view/810

Nuzzolo, R. (2002). *The effects of direct and observed supervisor learn units on the scientific tacts and instructional effects of teachers*. (unpublished doctoral dissertation). Columbia University, New York.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis, 26*, 237-257.

Parsons, S. A., Vaughn, M., Scales, R. Q., Gallagher, M. A., Parsons, A. W., Davis. S. G., Pierczynski, M., & Allen, M. (2017). Teachers' instructional adaptations: A research synthesis. *Review of Educational Research, 88,* 205-242.

Passmore, J. (1967). Logical positivism. In P. Edwards (Ed.), *The encyclopedia of philosophy* (Vol. 5, 52-57). New York, NY: Macmillan.

Pecheone, R., & Chung, R. (2007). *Technical report of the Performance Assessment for California Teachers (PACT): Summary of validity and reliability studies for the 2003-04 pilot year*. Stanford, CA: Stanford University.

Perlman, M., Falenchuk, O., Fletcher, B., McMullen, E., Beyene, J., & Shah, P. S. (2016). A systematic review and meta-analysis of a measure of staff/child interaction quality (the Classroom Assessment Scoring System) in early childhood education and care settings and child outcomes. *PLoS ONE 11*: e0167660. doi:10.1371/journal.pone.0167660

Phillips, D. C. (2014). *Encyclopedia of educational theory and philosophy*. Thousand Oaks, CA: Sage. SAGE Publications ProQuest Ebook Central, https://ebookcentral-proquest-com.tc.idm.oclc.org/ lib/teacherscollege-ebooks/detail.action?docID=1715010.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2007). *Classroom assessment scoring system* (Observation Protocol Manual). Baltimore: Paul H. Brookes.

Porta, D. D., & Keating, M. (2008). *Approaches and methodologies in the social sciences: A pluralist perspective*. Retrieved from https://ebookcentral-proquest-com.tc.idm.oclc.org

Praetorus, A., Pauli, C., Reusser, K., & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction, 31,* 2-12.

Rice, J. K. (2003). *Teacher quality: Understanding the effectiveness of teacher attributes*. Washington, DC: Economic Policy Institute.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73,* 417-458.

Russell, T. (2014). Teacher craft knowledge. In R. Gunstone R. (Eds.), *Encyclopedia of science education.* Dordrecht, The Netherlands: Springer, Dordrecht.

Schacter, J., & Thum, Y. M. (2004). Paying for high- and low-quality teaching. *Economics of Education Review, 23*, 411-430.

Scherzo, K. (2010). *The role of teacher-student relationships in an early childhood comprehensive application of behavior analysis to schooling (CABAS®) setting*. (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (Order No. 3447922).

Schnell, L. K., & Vladescu, J. C., Kodak, T., Nottingham, C. L. (2018). Comparing procedures on the acquisition and generalization of tacts for children with autism spectrum disorder. *Journal of Applied Behavior Analysis, 51*, 769-783.

Schoenfeld, A. H., Floden, R., Chidiac, F. E., Gillingham, D., Fink, H., Hu, S., Sayavedra, A., Weltman, A., & Zarkh, A. (2018). On classroom observations. *Journal of STEM Education Research, 1,* 34-59.

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in distinguishing meta-analysis results. *Review of Educational Literature, 77,* 454-499.

Selinske, J. E., Greer, R. D., & Lodhi, S. (1991). A functional analysis of the comprehensive application of behavior analysis to schooling. *Journal of Applied Behavior Analysis, 24,* 107-117.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15,* 4-14.

Sindelar, P. T., Rosenberg, M. S., & Wilson, R. J. (1985). An adapted alternating treatments design for instructional research. *Education and Treatment of Children, 8,* 67-76.

Singer-Dudek, J., Speckman J., M., & Nuzzolo, R. (2010). A comprehensive analysis of the CABAS® model of education at the Fred S. Keller school: A twenty-year review. *The Behavior Analyst Today, 11,* 253-265.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. Oxford, UK: Appleton-Century.

Skinner, B. F. (1945). The operational analysis of psychological terms. *Psychological Review, 52,* 270-277.

Skinner, B. F. (1957). *The behavior of organisms: An experimental analysis*. New York, NY: Appleton-Century-Crofts.

Skinner, B. F. (1984). The evolution of behavior. *Journal of the Experimental Analysis of Behavior, 41,* 214-221.

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42,* 795-819.

Stronge, J., H., Ward, T., J., & Grant, L., W. (2011). What makes good teachers good? A cross-case analysis of the connection between teacher effectiveness and student achievement. *Journal of Teacher Education, 62,* 399-355.

Strunk, K. O., Weinstein, T. L., & Makkonen, R. (2014). Sorting out the signal: Do multiple measures of teachers' effectiveness provide consistent information to teachers and principals? *Education Policy Analysis Archives, 22*. http://dx.doi.org/10.14507/epaa.v22.1590

Vargas, J. S. (2013). *Behavior analysis for effective teaching.* New York, NY: Routledge.

Wayne, A. J. & Youngs, P. (2003). Teacher characteristics and student achievement gains: A review. *Review of Educational Research, 73,* 89-122.

Wetzel, M. M., Hoffman, J. V., Roach, A. K., & Russell, K. (2018). Practical knowledge and teacher reflection from a practice-based literacy teacher education program in the first years: A longitudinal study. *Teacher Education Quarterly, 45, 87*-111.

Wilson, S. M., & Floden, R. (2003). *Creating effective teachers: Concise answers for hard Questions (Addendum to the report, Teacher preparation research: Current knowledge, gaps, and recommendations*). Washington, DC: American Association of Colleges for Teacher Education.

Wright, S. P., Horn, S. P., & Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11,* 57-76.

Appendix A: Teacher Packet Distributed in Experiment 1

Teacher Questionnaire

Note: I will be the only one with access to these data, once the responses are logged onto my database, I will shred the questionnaires.

Name: ________________ Classroom #:___________________

1. Circle one. What is your level of Education?

GED Associate Degree Bachelors degree Masters Doctoral

2. How many years have your worked at Keller?

Years __________ Months __________

3. Circle one. Are you a certified teacher?

Yes No

4. If yes. How many years have you been certified?

Years____________ Months__________

5. Have you taught prior to Keller? If so how many years?

Yes No Years_____________

6. What Rank did you hold in the 2017-2018 academic school year?

Teacher I Teacher II Master Teacher Assistant BA

7. Have you completed your Teacher II Rank prior to 2014 revisions?

Yes No

8. If you are working on your **Master Teacher Rank**, how many components have you completed within each module?

#VB_________ #CS_________ #VM_________

Consent Form

Fred S. Keller School Informed Consent Format

WRITTEN DESCRIPTION OF RESEARCH

Study Title: What Makes Teachers Tick: Investigating the Relationship Between Teacher CABAS® Modules Completed and Teacher Effectiveness

Name of the Principal Investigator: Sara Silsilah

Name of Co-Investigator: none

Contact Name and Phone Number for Questions/ Problems:

This is an educational research study. This research study includes only participants who choose to take part. Please take your time to make your decisions. Discuss it with your friends and family.

WHY IS THIS STUDY BEING DONE?

I am interested in finding out which components within the CABAS® ranks are important for teacher effectiveness as measured by learn units to criterion and STO's/LTO's met.

HOW MANY PEOPLE WILL TAKE PART IN THE STUDY?

About 40 teachers.

WHAT IS INVOLVED IN THE STUDY?

1. Fill out a 5 min Questionnaire,
2. Hand in a copy of your current module (e.g., Teacher I or Teacher II),
3. Hand in a copy of your completed module (e.g., Teacher I or Teacher II),
4. Hand in copy of your individual daily/weekly data for the year 2017-2018 (i.e., graphs you hang outside the classroom).

HOW LONG WILL I BE IN THE STUDY?

Approximately 10-15 min.

WHAT ARE THE RISKS OF THE STUDY? (all research has risks, even if they are minimal)

For more information about risks ask the researcher or contact

There are no risks for this study. The information I collect from you will remain confidential. I will be the only one with access to it.

ARE THERE BENEFITS TO TAKING PART IN THE STUDY?

WHAT ABOUT CONFIDENTIALITY?

There are no direct benefits for participating in this study. It will be used for educational purposes.

You will receive no payment for taking part in the study. However, you will receive a chocolate bar for filling out the questionnaire.

WHAT ARE MY RIGHTS AS A PARTICIPANT?

Taking part in this study is voluntary. Refusal to participate will involve no penalty or loss of benefits to which you are otherwise entitled. You may choose not to take part, may leave the study at any time, or not answer research questions, which you consider inappropriate. Leaving the study will not result in any penalty or loss of benefits to which you are entitled. We will tell you about any new information that may affect your welfare or willingness to stay in the study.

WHOM DO I CALL IF I HAVE QUESTIONS OR PROBLEMS?

For questions about the study contact the researcher(s),

Name: Sara Silsilah Phone #: (646)-315-1718

For questions about your rights as a research participant, contact the Fred S. Keller School Institutional Review Board (which is a group of people who review the research to protect your rights at 212 –914 965 1152, 223.

The Institutional Review Board of the Fred S. Keller School has determined that this research meets the criteria for human subjects according to Federal guidelines. -You will get a copy of the approval letter.

CONSENT:

I have read or have had read to me the preceding information describing the study. All my questions have been answered to my satisfaction and this form is being signed voluntarily by me indicating my desire to participate in this study. I am not waving any of my legal rights by signing this form. I understand I will receive a copy of this consent form.

PERMISSION: I ________________________voluntarily approve of my participation.

Signature of Participant ______________ Date_____________

Appendix B: Written Description of the Study

Instructions

1. Baseline sessions (un-consequated): Record (+) for correct responses and (-) for incorrect responses. Use data collection sheet in experiment binder. Conduct for both student w/BiN and student no/BiN.
2. Share baseline data with Sara, who will then assign operants to each condition. (you can send pictures of raw data via text, or email)

Running the Program

- Begin by running STO 1 assigned to each Teacher, with both students.
- Run a session or more daily.
- Run equal number of sessions daily, if teacher 1 runs two sessions in one day, teacher 2 also runs two sessions that same day.
- If one teacher is absent, do not run the session.
- Rotate who runs the first session daily, if Teacher 1 ran a session on Monday first, then Teacher 2 runs a session first on Tuesday.

A. When to stop an STO (for both STO 1 and 2)

- When **criterion** is met with **90%** accuracy across two sessions or **100%** in one session.

OR

- When **12** sessions have been conducted but criterion was not met.

B. When to start STO 2

- When both Teachers are at a stopping point, as explained above.
- Both teachers must start STO 2 at the same time.

C. Decision Analysis

- Use the decision tree protocol to make decisions.
- If needed, tactics should be chosen by the Teacher running the STO.
- Teacher 1 and Teacher 2 should not collaborate on choice of tactic or instructional problems if they arise.

D. Data Collection

- Record data on data sheets in the experiment binder.
- Put data in excel document found in the experiment dropbox.