EXAMINING CLASSROOM OBSERVATION INSTRUMENTS FOR ENGLISH LANGUAGE LEARNERS: AN EXAMPLE OF PROPOSING CAUSAL STRUCTURE PERTAINING TO PEDAGOGY

A Dissertation

by

TIBERIO GARZA JR.

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Rafael Lara-Alecio
Committee Members,	Beverly J. Irby
	Fuhui Tong
	Sandra Acosta
	Frederic I. Parke
Head of Department,	Victor Willson

May 2015

Major Subject: Educational Psychology

ABSTRACT

The purpose of this dissertation study was to describe teacher pedagogy through the use of systematic review, observation data, and to comment on causal structure. Three objectives were necessary: the assessment of the literature of classroom observation instruments used in observing English language learners, using archival classroom observation data from Project Middle School Science for English Language Learners (MSSELL), and the casual commentary of teachers' pedagogy during Project MSSELL. The Project MSSELL is a randomized, longitudinal, field-based, National Science Foundation (NSF) funded research project (NSF Award No. DRL-0822343; 2009-2010). Included in the project was archived data for the pedagogy of eight grade 5 teachers during a science intervention. The observation protocol, Transitional Bilingual Observation Protocol (TBOP), used in the project is theoretically derived from the transitional bilingual observation model (Lara-Alecio & Parker, 1994), and measures classroom frequency events in four instruction domains: Language of Instruction, Language Content, Communication Mode, and Activity Structure.

The data for this dissertation are taken from both a treatment and control group. The treatment group is comprised of four teachers participating in the intervention associated with Project MSSELL, while the control group is comprised of four teachers not participating in the intervention. By conducting the following: (a) systematic review of classroom observation instruments evaluating classrooms with English language learners (ELL); (b) frequency analysis in classroom events during Project MSSELL; and (c) commentary of the causal inference for the project in relation to teachers' pedagogy,

the researcher further describes the project's feasibility toward pedagogy conducive to ELL academic achievement. Results from this study implicate treatment teachers, when compared to control teachers, focus more on writing as a way of communication between the student and teacher. Additionally, through a systematic review of classroom observations instruments, the researcher highlights the TBOP's strength toward recording pedagogy. Through causal commentary the researcher converts Project MSSELL into a simple causal structure to indicate the causal effects at the local level. The resultant commentary provided further insight into teachers' pedagogy during Project MSSELL.

DEDICATION

He knew my heart, before I knew Him; and though, I thought I knew best for me, He knew better. This path toward earning a doctoral degree was hardly on my radar when I started college. I use to grab for whatever was near me and make that my career; never thinking, what more could I do if I only invested time. He guided me out of failure many times, but those experiences I see would strengthen me when starting this doctoral degree.

When I started my doctoral education, He made sure that was all I could see before me as if a cliff followed behind me. Clearly, there was nowhere worth going but forward and He made that clear and evident, since during early college I struggled with having an aim or long-term goal. Now, I stand at the height of education to see the path I took and acknowledge I was not alone. Therefore, I dedicate my work to my God; and though, my work is hardly a gift worthy to the creator of the universe, I dedicate all that I am to Him.

ACKNOWLEDGEMENTS

In my six years toward earning a doctoral degree and the writing of this document there are many I would like to acknowledge. First, I would like to acknowledge my dissertation chair Dr. Rafael Lara-Alecio who I have worked with for the last eleven years. We first met when I was an undergraduate earning a degree in philosophy and I wanted to join the Bilingual Education Student Organization; which I did and then had the opportunity to be a student worker for the Bilingual Education Program back in 2004. For 11 years, I have learned from him and encouraged by him to keep moving forward.

In my time at the Bilingual Education Program many faculty have also encouraged me and contributed to my academic training such as Drs. Acosta, Rivera, Tong, Pollard-Durodola, Padrón and Walichowski. Even faculty outside my program but within my department, Educational Psychology, invested their time in me such as Drs. Willson, Thompson, Kwok and Yoon. Faculty support did not stop there, faculty outside my department but within the College of Education and Human Development also invested there time to empower my academic training such as Drs. Goodson, Chlup, Irby, and Bozeman. Therefore, I am grateful to them for I see how desired I am from employers responding to the skills I am possess.

As I developed in academic skill, I found myself not only collaborating with educators but with entrepreneurs, state officials and nonprofit agencies. In fact, during the last year and a half I have continued my doctoral studies by working for the Brazos Valley Council on Alcohol and Substance Abuse (BVCASA). I am thankful to the

people of BVCASA for their support and the finance I was given as their evaluator. They tell me, never has such a great evaluator worked at BVCASA and so I am humbled by their kind words. Truly, I would not have had the strength to finish my degree if it was not for the people of BVCASA accepting me into their workplace. I was treated as family while at BVCASA and I would always be there for them in any need.

Additionally, I would like to acknowledge family and friends. Thank you to my family and relatives for their continued encourage and support in helping me stay the course to graduate. Especially, my fiancée, Maggie, who kept believing in me and cheering me on as I accomplished small and great feats toward receiving my Ph.D. Thank you Maggie, woman of noble character and the one I get to spend the rest of my life with. And to my friends at the Bilingual Education Program, International Christian Fellowship, Aggie Corps of Cadets and others I met along the way thank you for sharing your lives with me while at Texas A&M University.

Lastly, I would like to acknowledge the support of the National Science Foundation (NSF). Chapter 3 of my dissertation is archived data from Project Middle School Science for English Language Learners (MSSELL). Although my study was not funded directly, the archived data I was allowed to use was from Project MSSELL which was funded by the NSF. As a result, I thank the National Science Foundation for their contribution.

NOMENCLATURE

MSSELL	Middle School Science for English Language Learners
ELL	English Language Learner
TBOP	Transitional Bilingual Observation Protocol
LOI	Language of Instruction
LC	Language Content
CM	Communication Mode
AS	Activity Structure
COI	Classroom Observation Instrument
NCLB	No Child Left Behind

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

In 2012, English language learners (ELL) represented 21% of the student population in U.S. public schools; 79% of these learners are Spanish speakers (Aud, Hussar, Johnson, Kena, Roth, Manning, Wang, & Zhang, 2012). For stakeholders in public and ELL education, a major concern has been the science achievement gap between ELLs and non-ELLs (National Center for Education Statistics, 2011; National Assessment of Educational Progress, 2009 Science Assessment). In science, as in other content areas, language instruction plays a vital role in learners' performance (August, Branum-Martin, Cardenas-Hagan, & Francis, 2009; Kieffer, Lesaux, Rivera, & Francis, 2009; Lee, Deaktor, Hart, Cuevas, & Enders, 2005). In addition, science education researchers have noted that science has specific registers, that is, linguistic features, such as academic language and syntactical structures (Gee, 2005; Lembke, 1990; Norris & Phillips, 2003; Wellington & Osborne, 2001). These registers are especially challenging for ELL (Ryoo, 2010).

Additionally, Cummins (1986, 2008) noted the importance of social and academic registers for ELL in the continuum of language development. According to Cummins, this continuum is distinguished between Basic Interpersonal Communications Skills (BICS; i.e., social language) and Cognitive-Academic Language Proficiency (CALP; i.e., academic language). BICS and CALP are used to identify categories of English proficiency in the classroom. In fact, Cummins outlined the importance of ELL primary language in a common underlying proficiency model (1986), which illustrated

the relationship between ELL primary language (e.g., Spanish) and second language (e.g., English).

One method to measure the specific registers among ELLs in the classroom is the use of classroom observation instrument (COI). When observing the classroom with a COI the teacher becomes a vital component to ELL learning. As a result, this dissertation study focused on COIs that quantitatively described the teacher's pedagogy toward serving ELLs. COIs that measure pedagogy can provide insight into the decision-making process for teachers to assess instruction conducive to ELL academic achievement.

Quantitative Observation

Although observation studies have flexibility to include different methods of observing (e.g., naturalistic observation), in this dissertation study quantitative observation became the method of observance. By quantitative observation, the intent is to describe literature which explicitly used an observation instrument to code classroom activity. In other words, the quantitative approach for observation in the classroom involved direct and systematic observation in order to measure specific occurrences observed and how those occurrences should be recorded (Medley, 1992). Furthermore, due to the ELL context of this dissertation study, the focus is on quantitative observation studies related to ELL academic achievement.

Observation studies of classrooms with ELLs have gradually led to the development of instruments to quantify pedagogy (Foorman, Goldenberg, Carlson, Saunders, & Pollard-Durodola, 2004; Foorman & Schatschneider, 2003; Gersten &

Baker, 2000; Haager et al., 2003; Irby et al., 2007; Lara-Alecio et al., 2009; Saunders et al., 2006). In fact, a growing number of classroom observation instruments have been complied by Halle, Whittaker, and Anderson (2010) in relation to early language and literacy development. Also, the systematic measure of quantitative observation has led to identifying gaps in instruction for ELLs (Ramírez, Yuen, Ramey, & Pasta, 1991). For example, one of the early observation studies on ELLs concluded instruction for ELLs was not cognitively demanding or interactive during reading instruction (Padrón 1994).

While observation instruments take different approaches in measuring the occurrences in ELL classrooms (e.g., the compendium of classroom observation instruments recorded by Halle, Whittaker, and Anderson, 2010), there are commonalities among the quantitative approach to measurement. For instance, quantitative observation will involve: (a) observation to have a purpose and focus, (b) observed behaviors to be operationally defined, (c) observers to be trained in observation procedures, (d) the recording of the classroom setting and unit of time, (e) a means to record data, (f) process data, and (g) analyze data (Stallings & Mohlman, 1988). In the context of ELL, observation instruments have collected information on student, teacher, or teacher-student interaction behaviors (e.g., Haager, Gersten, Baker, & Graves, 2003; Lara-Alecio & Parker, 1994; Waxman, Tharp, Hilberg, 2004; Waxman, Wang, Lindvall, & Anderson, 1983), comprehensive school reform (e.g., Calderón, Slavin, & Sánchez, 2011; Irby, Tong, Lara-Alecio, Meyer, & Rodríguez, 2007; Lara-Alecio, Tong, Irby, Guerrero, Huerta, & Fan, 2012) and specific content instruction (e.g., Foorman & Schatschneider, 2003, 2004; Saunders Foorman, & Carlson, 2006). Additionally, Snow

(2002) noted that analysis of classroom observation data has led to improving educational research concerning the academic achievement of ELLs.

Significance of Dissertation

In a synthesis on language of reading instruction for ELLs, Slavin and Cheung (2005) stated there were few published randomized control trials (RCTs) present in the field of bilingual education and the same conclusion is still applicable in 2014. In 2010, a report on the Middle School Science for English Language Learners (MSSELL) project was presented before the Research on Learning in Formal and Informal Settings (DRL) branch of the National Science Foundation (NSF), which Lara-Alecio and Irby stated only two experimental studies could be identified in the literature directly related to science for ELLs. The presences of few RCTs highlights the efforts of studies such as Lara-Alecio and Tong (2013), and Irby (2013); where Project Middle School Science for English Language Learners (funded by the NSF, DRL award number 0822343 and 0822153) success toward these learners' acquiring greater science proficiency is already documented in Lara-Alecio et al. (2012).

By specifically examining the teacher's pedagogy during Project MSSELL, in my research, I have made efforts to identify the underlying instructional relationships between the teacher, the ELL, and the COI, which contributed to the academic achievement of ELLs described in Lara-Alecio et al. (2012). The findings of my dissertation contribute to the literature in three ways. First, the findings add a systematic review (Torgerson, 2003) of the literature on instruments for evaluating pedagogy to serve ELLs. In so doing, the review also serves as a compilation of classroom

observation instruments (COIs) identified in ELL research. Second, my examination of archived data from a randomized control trial (RCT) has value toward making casual interferences. Third, the archived data on pedagogy is in the context of science for ELLs and can address current policy and research initiatives to find viable solutions for increasing academic performance in science for ELLs.

Dissertation Overview

The purpose of my dissertation study was to investigate teachers' pedagogy during grade 5 science instruction with students identified as ELLs. Using archived data from Project MSSELL, I analyzed data collected from the Transitional Bilingual Observation Protocol (TBOP; i.e., the COI) for grade 5. Project MSSELL was a 3-year study, funded by the National Science Foundation (NSF), to increase science and English achievement for students in grade 5 and 6 (Lara-Alecio et al., 2012).

The research questions guiding this dissertation served three individual studies: (a) a systematic review of literature for classroom observation instruments used in classrooms with ELLs; (b) a frequency assessment of classroom observation data from teachers' pedagogy; and (c) a commentary on the causal inference for Project MSSELL in relation to pedagogy. Within each of these studies, the following research questions were addressed:

1. How many classroom observation instruments can be identified from the literature via a systematic review that are focused on classroom pedagogy for English language learners in U.S. classrooms?

2. How do authors describe their classroom observation instruments from theoretical development to application?

3. What were the pedagogical differences of treatment and control teachers during Project Middle School Science for English Language Learners (MSSELL)?

4. What pedagogical differences were displayed from treatment and control teachers as recorded from the Transitional Bilingual Observational Protocol during Project MSSELL?

5. Using Pearl's (2009) causality framework, how can a causal structure for teachers' pedagogy during Project MSSELL be conceived, and what would be the structure's appearance?

Unlike the traditional five chapter dissertation, I used a three-article format. In Chapter 1, I provided a general overview and rationale for the dissertation. Chapters 2, 3, and 4 were written as journal articles and are self-contained studies. In Chapter 5, I provided a synthesis of conclusions across Chapters 2 through 4.

Chapter Overview

My dissertation follows a three-article format with Chapter 1 acting as the overview. As a result, research questions and methodology are strategically placed across Chapters 2, 3, and 4. In Chapter 2, the research questions were used to focus on the literature. The review of literature involved the quantitative method of systematic review in order to reduce researcher bias, provide an account of critical appraisal, and illustrate a systematic process for replicability. More specifically, the systematic review

serves as a compilation of current instruments for classroom observations that serve ELLs. In Chapter 3, I use research questions to guide the discovery for examining teachers' pedagogy during science instruction in both treatment and control conditions. The observation instrument used to record teacher pedagogy was the TBOP. The comparison of treatment and control groups serve to illustrate influences from Project MSSELL. In Chapter 4, I present commentary on the causal structure for teacher pedagogy within the context of Project MSSELL.

Chapter 2

Chapter 2 is a systematic literature review with intentions to rename the chapter, *In Search of ELL Classroom Observation Instruments: A Systematic Review of pre K–12 English Language Learner Classroom Observation Instruments*, and submit to the *Educational Review* journal. Several journals focus on reviews of education research; however, as this chapter also serves as a comprehensive compilation of instruments serving ELLs, I decided *Educational Review* was best suited as a journal for publishing a systematic review with a compilation component.

The systematic review procedures I followed were defined by Torgerson (2003), which resulted in my systematic review with the following characteristics: "the application of strategies that limit bias in the assembly, critical appraisal, and synthesis of all relevant studies on a specific topic" (Porta, 2008, p. 217). The use of a systematic review can improve practice by evaluating quality in studies and synthesizing evidence in order to inform policy decisions (Saini & Shlonsky, 2012). In other words, in using systematic review protocols, researchers document search procedures, report specific

eligibility criteria, and produce replicable studies (Littell, Corcoran, & Pillai, 2008; Torgerson, 2003).

The systematic review was initiated with five electronic search engines ProQuest, EBSCOhost, SAGE, Web of Science, and Scopus. Search terms were arranged in a Boolean search with classroom observation instrument and English language learner. There were several variations of the two search terms, which included the expansion of terms with the asterisk function. The asterisk function included all other search terms with alternate endings such as -ing, -ly, -ment, -cy, and -s. For the search term classroom observation instrument the following terms and asterisks were used: classroom observation technique, classroom observation method*, classroom observation measure*, classroom observation scale, classroom observation protocol, classroom observation tool, classroom observation device, and classroom observation rating. For the search term English language learner the following was used: ELL, English language learn*, Limited English Proficien*, LEP, At risk, English learner, Bilingual Education, Second language acquisition, Language of Instruction, ESL, English as a second language, Second language learning, and bilingual. These descriptors returned 40 studies, covering the years 1983 to 2012. For inclusion, studies were considered relevant if there were enough information describing the use of a classroom observation instrument with specific observable outcomes for ELLs in PreK–12 U.S. classrooms. Studies were excluded if the sample of ELLs had disabilities or the observation protocol was screening for gifted ELLs.

Chapter 3

Chapter 3 with intentions to rename the chapter, *Observing Instruction Patterns: An Observation Study of Grade 5 Teachers during a Literacy-Integrated Science Intervention,* has potential in the *Journal of Adolescence*. Chapter 3 investigates observation data from Project MSSELL, which contains archival data of teachers' pedagogy during science instruction for students in grade 5. Classroom events were counted using the TBOP. Both treatment and control classrooms were evaluated for differences in teachers' pedagogy. As a result, Chapter 3 would fit well with the *Journal of Adolescence* in illustrating the impact Project MSSELL had to alter teachers' pedagogy for ELL academic achievement.

The observation study in this dissertation is based on archival data from Project MSSELL. During Project MSSELL, observation of classroom instruction took the form of systematic observation. According to Stallings and Mohlman (1988) systematic observation is defined to have the following components: "(a) a purpose and setting for observation, (b) operational definitions for all observed behaviors, (c) training procedures for observers, (d) a specific focus of observation, (e) a unit of time, (f) an observation schedule, and (g) a method to record, process, and analyze data" (p. 460-71). Researchers involved with Project MSSELL conducted systematic observation through the means of the Transitional Bilingual Observation Protocol (TBOP). Systematic observation analyzed in this study drew from archival data recorded by the TBOP. The TBOP is a classroom observation instrument with explicitly formulated rules for recording classroom behavior and developed by Lara-Alecio and Parker (1994).

In Chapter 3 of my dissertation, I describe classroom observation findings during Project MSSELL, which have the potential to move forward education research concerning the academic achievement of English language learners (Snow, 2002). Specifically, my data describes classroom observations of teachers in grade 5 classrooms during science instruction. The data were taken using the TBOP. The data were represented in nominal scale, indicating each of the TBOP measures are "mutually exclusive and identical in dispersion from the mode, such data only allow for counting" (Thompson 2006, p. 16). Therefore, the way the data are examined is by noting the number of times TBOP instruction events occur during classroom instruction. As a result, a chi-squared test of homogeneity was used to identify differences between teachers in the treatment and control groups. Also, frequency counts were examined over time to observe pedagogy from all teachers, regardless of condition placement.

Chapter 4

Chapter 4 is entitled, *Causal Commentary of Teacher Pedagogy during Project MSSELL*, and will be submitted to the *Journal of Causal Inference*. This chapter is a commentary on the causal inference made from the interpretation of teacher pedagogy during Project MSSELL. Project MSSELL is a randomized control trial study which illustrated positive intervention characteristics that were successful in increasing ELL academic achievement related to science and English proficiency (Lara-Alecio et al., 2012). Therefore, there is evidence to suggest Project MSSELL had a causal inference relationship between ELL academic achievement and intended project objectives.

In this chapter, I comment on the causal inference associated to teacher pedagogy as they relate to ELL instruction. The basis for causal inference is grounded in Pearl's (2009) framework for constructing causal structures. Using Pearl's framework I began to comment on the causal structure between teacher pedagogy and ELL receiving instruction conducive to their academic achievement. The context of causal inference commented in Chapter 4 is within the parameter of Project MSSELL. First, to infer causation "A variable X is said to have a causal influence on a variable Y if a directed path from X to Y exists in every minimal structure consistent with the data (Pearl, 2009, Definition 2.3.1)" (p. 45). In relation to Project MSSELL, because I focus on a specific aspect of Project MSSELL (i.e., teacher pedagogy) all the variables are observed and mention of unobserved variables during Project MSSELL is not entirely discussed in Chapter 4. Therefore, because I can know all described variables used during Project MSSELL from literature and since Definition 2.3.1 assumes all variables are observed, I comment on causation related to teachers' pedagogy in the sense that the variables are already known.

Pearl (2009) described causal structure beginning with three variables. As a result, I derived three variables from Project MSSELL pertaining to teacher pedagogy in the form of teacher professional development, ELL and non ELL academic achievement. From the teacher, ELL and non ELL measured by the three variables (i.e. teacher professional development, ELL and non ELL academic achievement), I begin to illustrate the influence Project MSSELL had on the teacher as professional development turned into instruction awareness toward ELL and non ELL academic achievement.

Although, in Chapter 4 I do not provide descriptive information on ELL academic performance (i.e., academic achievement), this chapter takes the reader up to the point of ELLs and non ELLs receiving instruction as delivered by Project MSSELL. The reader is referred to Lara-Alecio et al. (2012) in order to observe the ELL academic performance (i.e., academic achievement) during Project MSSELL. I provided a commentary on causation in Chapter 4 of the observed variables by deconstructing the Project MSSELL model into three variables.

Chapter 5

Lastly, Chapter 5 dissertation findings are brought together through synthesis in order to describe the collective meaning of Chapters 2, 3, and 4. In Chapter 2, I examined the prevalence of classroom observation instruments (COIs) relevant to serving the ELL classroom in the research literature. Through the findings of Chapter 2, COIs explicitly designed for serving the English language learner (ELL) classroom became apparent. A COI from Chapter 2 is then described in Chapter 3, where the intent was to illustrate findings of an instrument explicitly designed for serving the ELL classroom (i.e., the TBOP). Pedagogy recorded to serve the ELL classroom was collected with the TBOP, which occurred among treatment and control teachers. Looking at teachers' pedagogy from both conditions (i.e., treatment and control), in Chapter 4 I sought to comment on the causal structure of teacher pedagogy during Project MSSELL. The combined description of Chapters 2, 3, 4 and how they relate to one another is necessary to inform the reader of the collective impact.

CHAPTER II

CLASSROOM OBSERVATION INSTRUMENTS FOUND DURING ENGLISH LANGUAGE LEARNER INSTRUCTION: A SYSTEMATIC REVIEW

Researchers have described the field of bilingual education as lacking instruments for measuring instructional events and language of instruction in the classroom (Irby, Tong, Lara-Alecio, Meyer, & Rodríguez, 2007). Irby et al. (2007) also added that a daily observation measure of the opportunity for students to learn in bilingual classrooms is missing. The lack of classroom observation instruments (COIs) for the bilingual education classroom is an indication there is little information gathered from classroom settings with English language learners (ELLs). However, compiling COIs in this study can inform researchers of what is currently available. The use of COIs has been documented and recognized as important for furthering the field of bilingual education (Snow, 2002); and by using this study to inform researchers, the field of bilingual education can move forward. Therefore, this systematic review first compiles what COIs are used in bilingual classrooms and describes their psychometric properties by addressing two research questions: How many classroom observation instruments can be identified from the literature via a systematic review that observes English language learners in the United States classroom; and how do authors describe their classroom observation instruments from theoretical development to application?

The existing variation within bilingual classrooms does not lend itself to universalizing a one-size-fits-all COI. For instance, bilingual classroom settings are usually identified through program-level implementation. For example, the field of

bilingual education recognizes the following programs for the acquisition of English: (a) late-exit, (b) early-exit, (c) maintenance, (d) the 50/50 or 90/10 classroom model, (e) English immersion, (f) one-way immersion, and (g) two-way immersion or dual language. However, the No Child Left Behind (NCLB) Act of 2001 did not make the distinction between bilingual programs and simply required education programs to incorporate language instruction (Wright, 2005). Additionally, teaching students in their native language was optional. Added program variation among bilingual programs is exacerbated by the schools' judgment on what is needed concerning bilingual education as it pertains to NCLB (Benavides, 2004).

Instruction designed in a classroom setting for ELLs usually begins in an English as a Second Language (ESL) instructional model; that is, if schools do not have the means to initiate the use of bilingual education settings (i.e., having access to bilingual teachers; Texas Education Code, 1996). What advances individual bilingual classrooms to comprehensive bilingual programs are school-level initiatives oriented to improving educational services for ELLs. In fact, not all bilingual education programs are said to produce the same student-level achievement across classroom settings. For instance, bilingual education settings in the context of dual language immersion are recognized as most effective (Thomas & Collier, 2002). Barriers to bilingual program implementation are expressed by Lara-Alecio, Tong, Irby, and Mathes (2009) suggesting the inconsistency in starting bilingual programs and in defining the specific type of bilingual program have traditionally held back the advancement of the field and students' achievement.

There are clear characteristics of what bilingual education programs should look like, yet schools can mistakenly alter the necessary components of these programs, leading to incorrect labeling. The inconsistency of implementation and incorrect labeling produce challenges for COIs to adjust for program errors. Also, the variation in bilingual settings (e g., late exit, early exit, maintenance, or English immersion) adds yet another twist to the complexity of COIs. However, researchers have adapted to multiple bilingual settings and programs by focusing observation measurements on learning objectives. Such examples are seen when district-wide or school-wide personnel construct classroom observation instruments for assessing their programs. However, the dangers of such practice allude to the validity and reliability concerns for COIs created by district officials apart from researcher consultation.

Despite the variance in bilingual education settings, ELLs are known to be in mainstream classrooms. If the population of ELLs continues to increase in the United States as described by Aud, Hussar, Johnson, Kena, Roth, Manning, Wang, and Zhang (2012), then the need for teacher training will become even more necessary to meet the academic needs of these learners. In order to observe the effectiveness of teachers in different educational settings (i.e., bilingual education settings as well as mainstream classrooms) where ELLs are present, classroom observation instruments become a necessary tool for researcher and practitioner assessment of teachers. COIs for observing ELLs could imply that the classroom setting is a bilingual classroom or program, yet if ELLs were found in mainstream classrooms then COIs require the versatility to function across numerous classroom settings. Since bilingual classrooms are unlike mainstream

classrooms and more likely to differ in classroom instruction, the common mentally is to think instruction for ELLs do not help mainstream students. However, instruction for ELLs was documented in literature as assisting both ELLs and mainstream students to succeed academically (e.g., Lara-Alecio et al., 2012). As a result, the importance in COIs to record instructional events with ELLs in mind becomes beneficial in recognizing academic achievement among ELLs, especially Hispanic ELLs.

The efforts of education entities (e.g., RAND Corporation, Educational Testing Services, SEDL, and Pearson) to create teacher effectiveness measures also illustrate the necessity in accounting for teachers' pedagogy as influential to student academic achievement. The need for teachers to impact ELLs through their practices has been presented in research findings (Calderón, Slavin, & Sánchez, 2011). In fact, policymakers also recognize pedagogy as directional for education reform (Blank & Pechman, 1995; Mayer 1999). Similarly, the Standards Performance Continuum (SPC), one of the reviewed COI, used standards of effective pedagogy in its development (Doherty, Hilberg, Epaloose, & Tharp, 2002).

In retrospect, teacher effectiveness is not a new topic but rather one filled with constructive information for mainstream education, which can be utilized to inform and clarify pedagogy. Evaluating the effectiveness of teachers continues to be a sensitive topic, due to the potential of teacher devaluation and the loss of employment if teachers are labeled as ineffective. New measures of teacher effectiveness, however, have surfaced such as value-added modeling (VAM; Braun, 2005). VAM is useful in evaluating the educational growth of students during their time with teachers. Braun

(2005) described VAM as the growth of student test gains attributable to the teacher's overall effectiveness. However, Braun further asserts there are strengths and weakness in using VAM for teacher evaluation. Despite trends in VAM use to measure teacher effectiveness, the use of observation data is still necessary in examining how pedagogy can be effective toward ELL academic learning and achievement.

Conceptual Framework

The purpose of this study was to systematically identify COIs used in observing teachers instruct ELLs in U.S. classrooms. To do so, a logic model was constructed by the author and depicted in Figure 1. The logic model guided the process in identifying COIs where researchers described the use of a COI in order to observe teachers pedagogy toward ELLs. The systematic approach first started broadly by bringing together literature where a COI was used in a classroom setting with ELLs. Then, COIs were examined for their exclusivity and intention to measure ELLs activity in the classroom.

The logic model guided the rationale for identifying COIs specific to ELLs. The general process involved the collection of broad literature and then utilizing raters to reduce literature to specific COIs designed for capturing ELL activity in the classroom. For example, in Figure 1 studies that passed systematic review criteria were then assessed by raters to examine the purpose and intent of COIs. The flow of arrows in Figure 1 help to illustrate the process to determine the total number of COIs in this study. Raters with a pool of studies that passed systematic review criteria, then used a coding sheet to determine COIs that were related to serving the ELL classroom.

Additional pursuit for more COIs continued through ancestry search (i.e., looking at the references of included studies for additional COIs), which studies were then examined by raters for exclusivity for serving the ELL classroom. Through the process described in Figure 1, COIs that explicitly described how ELL activity was captured became known as COIs deemed specific to serving the ELL classroom (i.e., COIs with purpose and specific intention for serving the ELL classroom).

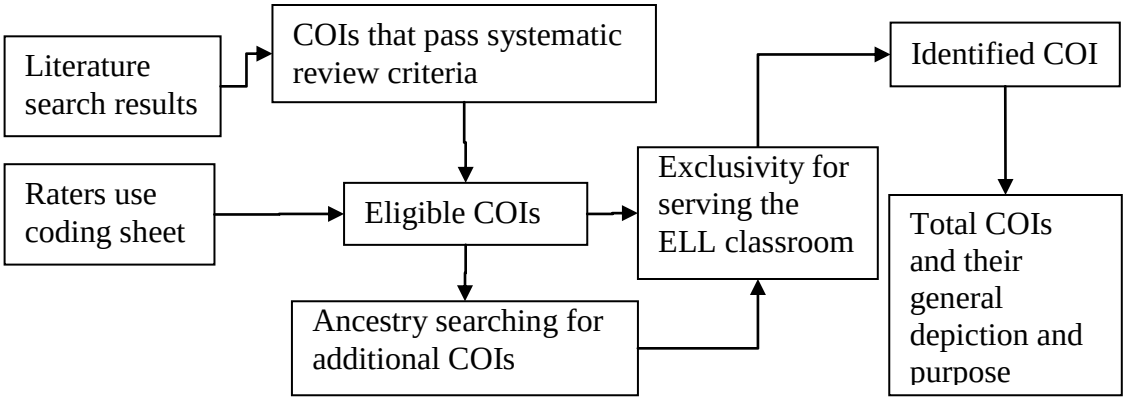

Figure 1. Logic model for collecting classroom observation instruments used with English language learners.

By following exclusion criteria and the coding sheet for COI, the logic model aided in depicting a process for identifying COIs used in conjunction among ELLs present in the U.S. classroom. The research questions (RQ) guided by the logic model are the following:

1. How many classroom observation instruments can be identified from the literature via a systematic review that are focused on classroom pedagogy for English language learners in U.S. classrooms?

2. How do authors describe their classroom observation instruments from theoretical development to application?

Literature Review

Research related to ELL in the classroom increased with the passage of the No Child Left Behind (NCLB) 2002 mandate. Before NCLB there were no specifically tailored and widely-accepted classroom observation instruments (COIs) for capturing the unique aspects of English language learners (ELLs) during instruction. Rather, the intent was to observe program effectiveness by recording the progress of program level objectives. For example, one of the early program level COIs seen in use among ELL was the Stallings COI, which was originally known as the Classroom Observation Instrument by the Stanford Research Institute (SRI) and also called the SRI Classroom Observation Instrument (COI). However, the Stallings COI was geared toward program level measures instead of recording specific ELL measures to enhance ELL academic achievement (Stallings, 1973). The result of a COI focused on program level effectiveness and not the individual students in the classroom would lead to a reduction of the Stallings COI to effectively record ELL explicit measures. Also, because the Stallings COI focused on program-wide measures it was described as having extensive classroom variables and involved several days of training compared to newer COIs explicitly focused on serving the ELL classroom (Waxman, Tharp, & Hilberg, 2004).

However, before NCLB (2002) there was no federal-level movement like NCLB prompting researchers to assess and evaluate the schooling experience of ELLs. After NCLB (2002), COIs became specific and focused on observable measures geared toward ELL academic achievement in the classroom.

The progress of quantitative observation has gradually gotten better in quantifying teachers' pedagogy during classrooms with ELLs (e.g., Foorman et al., 2004; Foorman & Schatschneider, 2003; Gersten & Baker, 2000; Haager et al., 2003; Irby et al., 2007; Lara-Alecio et al., 2009; Saunders et al., 2006). In fact, quantified observation studies have uncovered instructional deficits in classrooms with ELL (Ramírez et al., 1991). For instance, one of the early observation studies with ELL concluded that instruction for ELL was not cognitively demanding or interactive during reading instruction (Padrón, 1994). As a result, quantified observation findings have led researchers to move the field of bilingual education forward by revealing the nature of instruction provided to ELL.

Authors have taken different approaches and recorded different measures in order to construct COIs that are both systematic and quantitative in nature (Waxman, Tharp, & Hilberg, 2004). Different observation studies are worth mentioning due to their contribution to observation methodology in bilingual education (i.e., COIs identified in this study). Observation instruments with a quantitative paradigm are known to measure student and teacher behaviors individually as well as teacher-student interaction behaviors (e.g., Haager, Gersten, Baker, & Graves, 2003; Waxman, Tharp, Hilberg, 2004; Waxman, Wang, Lindvall, & Anderson, 1983). Broadening the scope of

quantitative measure, COIs have also been derived through comprehensive school reform-like interventions and content-specific evaluation of instruction (e.g., Calderón, Slavin, & Sánchez, 2011; Foorman & Schatschneider, 2003, 2004; Irby et al., 2007; Lara-Alecio, et al., 2012; Saunders et al., 2006).

The construction of quantitative COIs has also approached classroom measurement in the form of fidelity checks (i.e., program degree implementation) and program evaluation. For example, the Stallings COI evaluated the degree of program implementation for the National Follow Through Program (1968-1977; Stallings & Freiberg, 1991). Meaning the theory is not located in the COI, but rather in the program. COIs without theory then serve to observe program effectiveness. The Transitional Bilingual Observation Protocol (TBOP; Lara-Alecio & Parker, 1994) has also utilized its instrumentation as a program measure for Project English Language and Literacy Acquisition (ELLA) and Project Middle School Science for English Language Learners (MSSELL) intervention studies (e.g., see Lara-Alecio et al., 2012; Rodriguez, Lara-Alecio, Galloway, & Irby, 2002). Yet, the TBOP was described as having a theoretical foundation from the literature of ELL academic achievement and instruction (Lara-Alecio & Parker, 1994). As a result, the difference between the Stallings COI and the TBOP is the condition of program models guided by COI measurement versus COI measurement guided by practice. However, not all bilingual education settings are explicitly described by school districts as grounded in theory and as mentioned earlier bilingual education programs can incorrectly label themselves (Lara-Alecio et al., 2009).

Effective program models identified for ELL instruction are dual language programs (Thomas & Collier, 2002). Dual language programs are described as increasing the supportive environmental factors that are conducive to ELL academic achievement (Thomas & Collier, 2003). Such factors can exhibit a positive value added to the ELL's language and culture. Generally, program design models compared to no program in place are more likely to establish structure in the environment for creating supportive instruction for ELL. As a result, schools with programs that foster ELL academic achievement should be recognized for their efforts in providing educational services to ELLs. Another, area of interest in supporting the academic achievement of ELL is to focus on the teacher's method of instruction.

Teachers are viewed as the main instructional component in the classroom, meaning instruction does not happen unless teachers are present. Additionally, teachers have the potential to choose to seek out effective instructional strategies for fostering ELL academic achievement. When teachers actively practice instructional strategies to empower ELL academic achievement, then teachers are working in a combined effort to alter ELL academic performance levels. Although, the use of best practices or effective strategies can be beneficial to ELL academic achievement, the use of effective strategies does not work alone to describe or contribute to the overall academic achievement gains among ELLs. However, when using effective strategies in pedagogy for ELLs in the context of dual language programs, then the supportive conditions from dual language programs work in concert with teachers' use of effective strategies to optimally provide for ELL academic achievement.

Therefore, when there are several conditions occurring in the classroom that have the potential to contribute to ELL academic achievement, then reasonably researchers can take observation measurements. Hence, the development and use of classroom observation instruments (COIs) among educational researchers has resulted from recording observation measures in order to uncover further insight into the events occurring in the classroom. For instance, COIs can take a multi-measure approach to include teacher and student classroom behaviors.

However, COIs that measure many classroom events can become costly and represent an exploratory approach to classroom observation research due to the broad approach to include numerous classroom events. Efficiency related to COI cost of implementation and measurement should follow more targeted objectives to record the most essential classroom data apart from all possible data. Another approach to optimal cost savings of COIs creation involves the construction of COIs from other COIs (e.g., ELLE–Early Language and Literacy Environment, B-TBS–Bilingual Teacher Behavior Rating Scale). However, building COIs from other COIs is not only about cost, but rather an expansion of the original COI's capability to measure further classroom events (e.g., Timed Observations of Student Engagement/Language [TO/SEL]; Foorman, Goldenberg, Carlson, Saunders, & Pollard-Durodola, 2004). When COIs build from other COIs it's important to address the theoretical underpinnings of the original instrument and how they align to new COI. As a result, this systematic review compiles all COIs described in the context of classrooms with ELL present and then examines each COI to conclude which COIs were explicitly designed to record ELL activity in the

classroom (i.e., serve the ELL classroom). As a byproduct of identifying the exclusivity of each COI, this systematic review also described the theoretical background to applied practice of each COI.

Method

This study includes a systematic review design, identified as standard-based, explicit, and replicable (Torgerson, 2003). In fact, this systematic review sought to resemble characteristics of systematic reviews defined by Porta (2008)–"the application of strategies that limit bias in the assembly, critical appraisal, and synthesis of all relevant studies on a specific topic" (p. 217). Because systematic reviews improve practice by evaluating study quality and inform policy decisions by synthesizing evidence (Saini & Shlonsky, 2012), this study sought to establish viable COIs in the field of ELL instruction. In other words by taking a systematic review approach, systematic reviewers can follow the document search procedures for COIs in an ELL context, report specific eligibility criteria, and produce replicable studies (Littell, Corcoran, & Pillai, 2008; Torgerson, 2003), thereby furthering and establishing the cluster of viable COIs for the field.

Search Method

The search method involved a multi-step process to rigorously search and filter studies to the outcome of included studies. Additional search features also included expansion techniques to incorporate addition classroom observation instruments (COIs) that may have been over looked (e.g., ancestry searching focused in identifying additional COIs). The search method used to collect COIs occurred in three stages and

took place in December 2014: (a) searching extensively through all available databases, (b) identifying of COIs from included studies (i.e., studies fulfilling systematic review criteria), and (c) conducting an ancestry search through included studies for the inclusion of additional COIs from literature.

The use of three search engines (i.e., ProQuest, EBSCOhost, and SAGE) was employed to assess database search results with the most records. For instance, the ProQuest search engine had access to 133 databases December 2014, while EBSCOhost had access to 164 databases. The SAGE search engine only searched through Sage publications and hence was only a search through one database. Across all three search engines the keywords (i.e., search terms in a Boolean structure) were expanded in conjunction with each search engine's thesaurus. As a result, thesaurus availability among the three search engines aided in expanding and adding search terms necessary to capture the extent of literature related to COIs in the context of ELLs. The total number of studies from the implemented Boolean search terms of COI and ELL was 203 (e.g., see Figure 2).

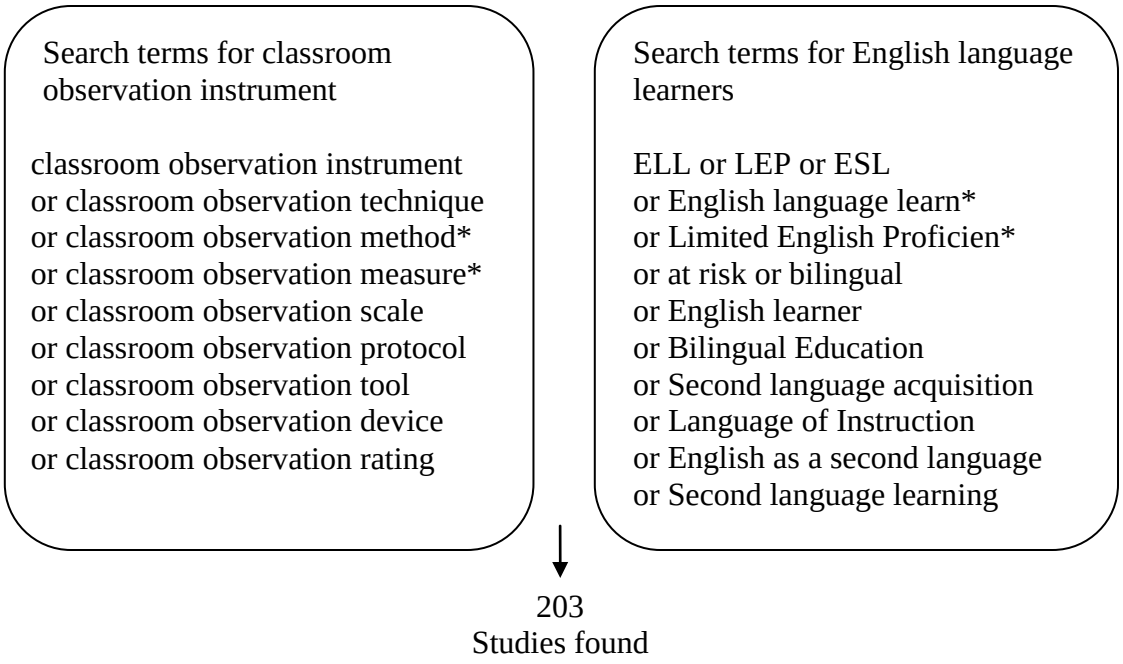

Figure 2. The Boolean search result of classroom observation instruments and English language learners. The * means further spellings of the root word were included

Sample

Using keywords with the ProQuest search engine resulted in 82 databases with at least one record (i.e., a study found to match keywords but not examined against any study criteria). Of the 82 databases, 727 records were found. Excluding duplicated records among the 82 databases, the total sample for the ProQuest search engine was 387 records. The 387 records then underwent established title and abstract criteria; meaning, records that passed through the title and abstract criteria were then known as studies. In other words, if a record title or abstract described the record as not conducted in the United States and not an ELL sample between Prekindergarten through grade 12, then

the record was excluded. Otherwise, a record was then identified as a study and preceded to the next level of screening, which was an examination of the full-text against further systematic review criteria. If a study then met systematic review criteria, then the study was referred to as included study.

The EBSCOhost and SAGE databases also went through the same title and abstract examination process as ProQuest. Using all available EBSCOhost databases with search terms resulted in 212 records across 30 databases. Filtering out duplicate records among EBSCOhost and comparing records with ProQuest results resulted in 54 new records. Study examination into title and abstract screening reduced the sample of records to 17 studies (i.e., records that passed the title and abstract screen). As for Education: A SAGE Full-Text Collection (1847–Aug 2013) database the same Boolean search command found 38 records. Cross-referencing SAGE results for duplicates with EBSCOhost and ProQuest identified nine records as duplicates, leaving 29 new records. Through abstract and title screening, the 29 records then became known as 22 studies ready for full-text examination among raters with the use of a coding sheet.

Coding Sheet

All 203 studies prepared for full-text review were written in English and ranged between the years of 1983 and 2012. Before the coding sheet (see Appendix C) was applied toward full-text review, the studies per database were clustered as followed: 25 studies for ProQuest; 98 for EBSCOhost; and 22 for SAGE database (Education: A SAGE Full-Text Collection). Applying the coding sheet involved two raters and reliability was based on percent agreement leading to consensus. After full-text coding

the results were as follows: 13 studies for ProQuest; 7 studies for EBSCOhost; and 2 studies for SAGE. Also, see Figure 2, a flowchart of the study filtering process (n = 22) and COI sample total (N=37)

The screening process of 22 included studies resulted in 18 COIs which displayed assessment toward recording English language learner (ELL) classroom events. Compilation of COIs involved a two-step approach: (a) searching for additional classroom observation instruments described in each study; and (b) searching through included study references (i.e., ancestry searching) for additional COIs. From Figure 3, 18 COIs are identified with 19 additional COIs through ancestry searching. The coding sheet was necessary to make distinctions between COIs that had been used to observe ELLs in the classroom and COIs explicitly described by authors as specifically designed for serving the ELL classroom. The reason for this distinction was to exclude COIs that measured classroom observation events in which ELLs were present but made no explicit effort to describe outcomes explicitly for ELLs. In order to make this distinction, two raters examined the connection between research questions and stated purpose of included studies to observe the rationale for COI's usage in relation to the targeted classroom sample.

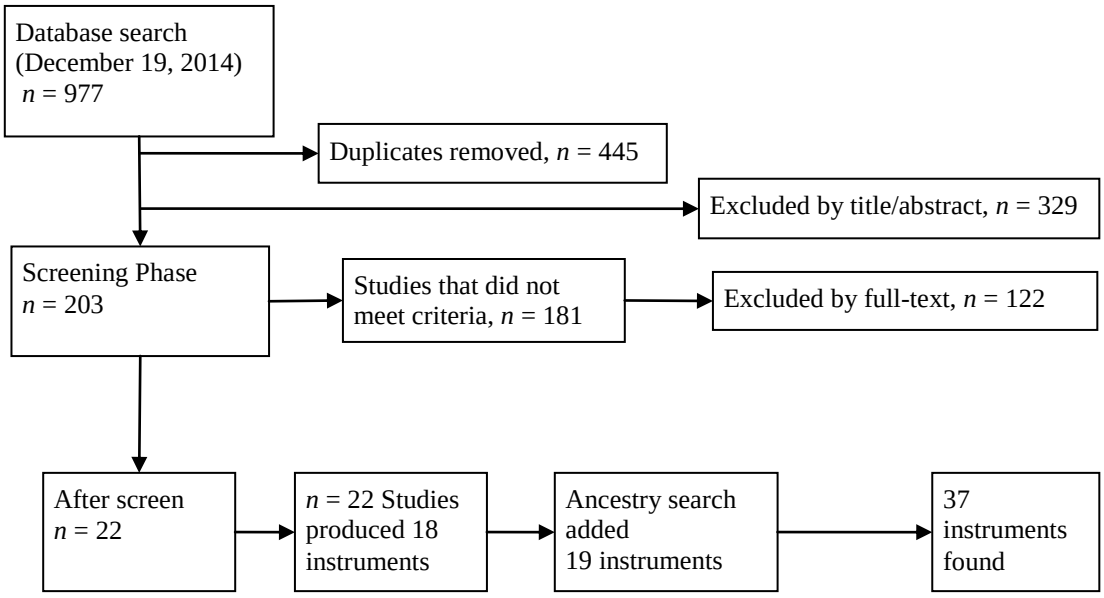

Figure 3. Process for selecting studies and number of classroom observation instruments identified.

Inclusion Criteria

The search command to uncover literature on COIs was left intentionally broad and allowed for studies to describe classroom observation instruments in generic terms (i.e., *classroom observation instrument*). Next, distinctions between generically named and officially named classroom observation instruments were established. For instance, COIs with capitalized names was reasoned to indicate separation from generic COIs (i.e., non-capitalized names). In other words, by observing a capitalized COI name or COI name in italics the COI was reasoned as unique and illustrated the author's value in naming the observation instrument. Such characteristics of how COIs were portrayed within studies, was the approach for identifying classroom observation instruments.

Raters had to examine if studies explicitly described the presence of English language learners in the sample. Although, ELLs tended to be Hispanic in the U.S., studies did not have to describe ELLs as Hispanic to be included. As a result, studies merely had to describe samples as consisting of English language learners (ELLs). The ELL sample had to be from pre K to grade 12 and conducted in the United States. Additionally, for mixed samples (i.e., which included non-ELLs and ELLs) the ELLs had to equal an amount greater than or equal to 50% for the study to be included.

Exclusion Criteria

Studies were included if authors described the use of a classroom observation instrument, which collected data on English language learners (i.e., empirical). As a result, commentary or book and study summaries were excluded. The only studies that were not counted were any original studies that described the creation or validation of the COI. If a study did discuss the creation of a COI but also conducted an empirical study, then the study was included.

Study articles were excluded if the study focused on outcome results for ELLs with disabilities, or if the sample was a mixture of ELLs and non-ELLs that resulted in less than 50% being ELLs. Also, studies that were focused on protocols on giftedness of ELLs, was also excluded (e.g., Ramos, 2010). The reason for excluding giftedness observation protocols and ELLs with disabilities was to not deviate from the general ELL population. In addition, studies were also excluded when the sample involved gifted students, special education students, African American students, and if teachers were the only sample apart from students.

Methodological Quality Measure

The Methodological Quality Questionnaire (MQQ; Appendix A) results were based on included studies (i.e., $n = 22$). The purpose of the MQQ was to gain an estimate of methodological quality present among included studies where COIs were identified. The MQQ developed by Acosta, Garza, Hsu, and Goodson (under review), was modified slightly to evaluate the implications and policy criteria (i.e., MQQ criteria 8 and 9) related to COIs. See Appendix A for MQQ components applied to included studies. Additionally, the MQQ had field experience and was not a new methodological instrument (Acosta & Garza, 2011; Acosta, Goltz, Goodson, Padrón, Garza, & Johnston, under review; Garza & Acosta 2013). Assessment for methodological quality was necessary to examine the trustworthiness of included studies. In retrospect, the effort to assess methodological quality among studies has not always been readily apparent as investigated in Garza and Acosta (2013). Therefore, assessment of methodological quality among included studies demonstrates additional efforts to describe the literature.

Results

In this systematic review, I addressed two research questions and examined the methodological quality of included studies. The search year-range was defined by the latest and earliest studies, which resulted between 1973 and 2012. The ends of the range are captured in studies by Stallings (1973) and Padrón, Waxman, Yuan-Hsuan, Meng-Fen, and Michko (2012). The included study sample of 22 was also examined for methodological quality with the MQQ (Methodological Quality Questionnaire). The average MQQ score was 22.84 (SD = 2.74), with the maximum attainable score set at 27

and the median at 23. Inter-rater reliability was also acceptable at 84.8% agreement or 0.536 Kappa. MQQ results in this systematic review are also comparable to MQQ acceptance values found in other systematic reviews with MQQ results (i.e., Garza & Acosta, 2013). As a result, the level of methodological quality among included studies was generally high.

Results are illustrated in Table 1 with the number of studies that had English language learners present during the use of a classroom observation instrument. However, Table 1 was constructed to provide a general overview of the broad literature where COIs were used among ELLs. Additionally, the purpose of Table 1 was to identify COIs that passed criteria established by systematic review protocols. Description of each COI begun with the first column noting authors of each COI, followed by the theoretical underpinning or origin of how each COI was created. There are three columns which provide a brief description of 37 COIs from theory to framework and application (i.e., origin, framework, and measurement columns). The last column served to make rater judgments based on how explicit authors described their COI as designed serving the ELL classroom. As a result, the last column of Table 1 (i.e., ELL purposeful) begun to separate COIs that happened to be used among ELLs and those COIs that were specifically created for serving the ELL classroom.

Table 1

Matrix of Classroom Observation Instruments Reported in Studies with English Language Learners

Classroom observation instrument with citation	Origin	Framework	Measurement	ELL purposeful
Activity Setting Observation System (ASOS; Rivera, Tharp, Youpa, Dalton, Guardino, & Lasky, 1999)	Activity setting categories based in sociocultural theory (e.g., Tharp 2005)	Sociocultural framework	Seven basic categories of activity setting	Yes
Bilingual Teacher Behavior Rating Scale (B-TBRS; Landry et al., 2001)	An adaptation from TBRS (Landry, Crawford, Gunnewig, & Swank, 2001)	Teacher Behavior Rating Scale (TBRS)	Examines the quality and quantity of instruction practices in relation to language use	Yes
Classroom Assessment of Supports for Emergent Bilingual Acquisition (CASEBA; Freedson, Figueras-Daniel, & Frede, 2009)	Based on the Support for Early Language Learners Classroom Assessment (SELLCA; National Institute for Early Education Research, 2005)	Framework to assess teacher and classroom supports for both first and second language acquisition (Castro, Espinosa, & Páez, 2011)	Assesses the level of support of the social, cognitive, and linguistic development of English language learners	Yes
Code for Interactive Recording of Children's Learning Environments (CIRCLE; Atwater, Lee, Motagna, Reynolds, & Tapia, 2009)	Based on the "context of children's classroom activities, the behavior of teachers and other adults in the classroom, and the child's engagement with people and objects" (Halle et al., 2010, p. 113)	Teacher-child interaction in a pre-school setting	Documents the ecological and behavioral features	No

Table 1 Continued

Classroom observation instrument with citation	Origin	Framework	Measurement	ELL purposeful
Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008)	Developmental theory (Foundationally built from the Observational Record of Classroom Environments–ORCE, NICHD Early Child Care Research Network, 1996)	Three domains consisting of ten dimensions (Pianta, LaParo, & Hamre, 2011)	Assesses classroom quality across grade levels, content areas, and instruction support	No
Classroom observation form (Servin, 1983)	Dissertation which was exploratory and descriptive of events occurring in the classroom related to language	Records amount of Spanish versus English spoken; and amount and type of corrective feedback to students' Spanish speech	Measured the verbal behavior of bilingual teachers	Yes
Classroom observation guidelines or observation scales (Luykx & Lee, 2007)	Instructional congruence framework (Lee & Fradd, 1998)	Academic content is meaningful when oriented to students' linguistic and cultural experiences and relevance to their lives (Lee, Maerten-Rivera, Penfield, LeRoy, & Secada, 2008)	Scales–to measure instruction congruence	Yes
Classroom Language and Literacy Environment Observation (CLEO; Holland-Coviello, 2005)	Research on pre-school language and literacy classroom environments affecting children's learning	Aspects of environments for children's language and literacy development	Measures the "quantity and quality of teacher language input, language and literacy teaching, and children's access to literacy materials in the classroom" (Halle et al., 2010, p. 118).	Partial

Table 1 Continued

Classroom observation instrument with citation	Origin	Framework	Measurement	ELL purposeful
Classroom Observation Measure (COM; Ross & Smith, 1996)	Developed at University of Memphis–Ross, Smith, Lohr, & McNelis, 1994	Thirty three classroom indicators coupled in six areas to witness instruction processes or strategies during the teacher's instruction	Measures of instruction strategies or processes during the teacher's instruction	Yes
Classroom Observation Schedule (COS; Padrón, Waxman, & Huang, 1999; Waxman & Padrón, 2004)	Classroom instruction and activities are mediated by student attitudes and perceptions (Anderson, 1987; Doyle, 1977)	Student-mediating paradigm (Schunk, 1992; Weinstein, 1989)	Records student behavior during the instruction learning process (Waxman, Wang, Lindvall, & Anderson, 1990a)	Yes
Collaborative Strategic Reading Intervention Validity Checklist (CSRIVC; Vaughn, Hughes, Schumm, & Klingner, 1998)	Collaborative Strategic Reading (CSR) (Klingner & Vaughn, 1996)	Observation checklist (Vaughn, Mathes, Linan-Thompson, Cirino, Carlson, & Pollard-Durodola, …Francis, 2006)	Checklist–to measure fidelity of collaborative strategic reading implementation	No
Dual Language Activity Setting Observation System (DLASOS; Rivera & Tharp, 2010)	Derived from the ASOS and based on CREDE Standards	Modeled specifically for dual language programs	Provides a measurement for the teacher to assess their level of meeting criteria standards	Yes
Early Childhood Classroom Observation Measure (ECCOM; Stipek & Byler, 2004)	Social-constructivist theoretical orientation	Constructivist (child-centered) and Didactic (teacher-centered) instruction approaches	Assesses the nature and quality of academic instruction on a 1 to 5 scale, and social climate	No
Early Language and Literacy Classroom Observation (ELLCO) Toolkit, Research Edition (Smith, Dickinson, Sangeorge, & Anastasopoulos, 2002)	Based on research on early language and literacy development	Modeled after classroom language and literacy activities and resources	Classroom Environment Checklist (25 items); 2 observation rating systems and 2 observation checklists	Partial

Table 1 Continued

Classroom observation instrument with citation	Origin	Framework	Measurement	ELL purposeful
Early Language & Literacy Classroom Observation: Addendum for English Language Learners (ELLCO-ELL or ELLCO-A; Castro, 2005)	"Assesses how classroom practices are addressing the particular needs of English language learners" (Halle et al., 2010, p. 169)	"Modeled to obtain information about specific classroom practices related to promoting language and literacy development among children who are English language learners" (Castro, 2005, p. 2)	"Designed to examine classroom and instructional factors that affect the experiences of English language learners in early childhood prekindergarten settings" (Halle et al., 2010, p. 168)	Yes
English Language Learner Classroom Observation Instrument (ELLCOI; Baker, Gersten, Goldenberg, Graves, & Haager, 1999; Gersten, Baker, Haager, & Graves, 2005; Haager et al., 2003)	California Reading and Language Arts Framework on instruction practices are linked to English Learners' achievement growth in reading	Framework of measurement validity (Messick 1989, 1995) and additions by Gersten, Keating, and Irvin (1995) & Gersten and Baker (2002)	Instrument–measures instruction quality on several dimensions for ELL	Yes
Early Language & Literacy Classroom Observation (ELLCO) Pre-K Tool (Smith, Brady, & Anastasopoulos, 2008)	"Based on data and feedback from use of the original ELLCO Toolkit, Research Edition" (Halle et al., 2010, p. 164)	Center-based classroom focused on early language and literacy development for 3- to 5-year olds	Includes a teacher interview and observation instrument that addresses classroom structure, curriculum, the language environment, books and book reading, and print and early writing	Partial
Early Language and Literacy Environment (ELLE; created by Mathematica–Atkins-Burnett et al., 2010)	Adapted scales from: ELLCO Research Edition is descriptive; the ELLCO Addendum is an addition; and CHELLO is based on ecological psychology (focused on family/friend/neighbor care)	Adapted from the ELLCO Tool–Research Edition; ELLCO Addendum (Castro 2005); CHELLO–Child Home Early Language and Literacy Observation (Neuman et al., 2007)	Measure of language and literacy support in the environment's materials and activities	Partial

Table 1 Continued

Classroom observation instrument with citation	Origin	Framework	Measurement	ELL purposeful
Early Literacy Observation Tool (E-LOT; Grehan, Smith, & Ross, 2004)	"Successor of the Literacy Observation Tool (LOT)" (Halle et al., 2010, p. 172)	"Aligned to the National Reading Panel and National Research Council findings and captures all essential components of the Early Reading First Program" (Halle et al., 2010, p. 172)	"Designed to measure research-based instructional practices, student activities, and environmental settings in early childhood classrooms where teachers are engaged in teaching the foundations of reading and other literacy processes" (Halle et al., 2010, p. 172).	Partial
Expository Reading Comprehension (ERC) observation instrument (James-Burdumy et al., 2009)	Frequency of instruction behaviors reading experts have deemed vital for reading comprehension	Observation checklist for instruction practices developed by James-Burdumy et al. (2009)	Instrument–records frequency of instruction behaviors for reading comprehension	No
Intercultural Development Inventory (IDI; Hammer & Bennett 1998)	"Theoretical possibilities of integrating culture and language learning" (CARLA, n.d.)	Bennett model of intercultural sensitivity–describes the ways in which people construe cultural differences	Students' intercultural development	No
Language Interaction Snapshot with End-of-Visit Ratings (LISn+EVR; Atkins-Burnett, Sprachman, & Caspe, 2010)	LISn and EVR are descriptive and rooted in English language learner research	LISn–Language Interaction Snapshot (Sprachman, Caspe, & Atkins-Burnett, 2008); EVR–End-of-visit	LISn–Examines language interactions of an individual child; EVR–collects data on instruction related to language and literacy development and classroom organization and management	Yes

Table 1 Continued

Classroom observation instrument with citation	Origin	Framework	Measurement	ELL purposeful
Local Systematic Change Classroom Observation Protocol (LSC COP; Horizon Research Inc., 2000)	Designed for the National Science Foundation's Local Systemic Change Through Teacher Enhancement Program	Contextual Background and Activities section–collects descriptive information; followed by a rating system	Protocol–to measure quality of an observed science or math lesson	No
Observation Measures of Language and Literacy (OMLIT; Goodson, Layzer, Smith, & Rimdzius, 2006)	"Research on the acquisition of English of English language learners informed the development of the OMLIT" (Halle et al., 2010, p. 218)	Combined OMLIT measures "provide an in-depth assessment of the quality of the language and literacy activities in the classroom" (Halle et al., 2010, p. 217)	"A battery of measures to address the need for research-based, reliable and valid measures of the instructional practices and environmental supports for language and literacy in early childhood classrooms" (Abt Associates, undated, p. 1)	Partial
Opportunity To Learn/Academic Language Exposure (OTL/ALE) survey (Martinez, Bailey, Kerr, Huang, & Beauregard, 2010)	Informed by OTL and ALE frameworks and instruction practices at NCES and CRESST (Borko, Stecher, Alonzo, Moncure, & McClam, 2005; Boscardin, Aguirre-Muñoz, Chinen, Leon, & Shin, 2004; Brewer & Stasz, 1996; NCES, 2006)	Four-dimension OTL (Opportunity To Learn) model (Stevens 1993); Three-dimension ALE (Academic Language Exposure; Bailey, Butler, Stevens, & Lord, 2007)	Opportunity to learn and academic language exposure measures for English language learners	Yes
Supports for Early Literacy Assessment (SELA; Smith, Davidson, Weisenfeld, & Katsaros, 2001)	Principles of emergent literacy and language–such as children's awareness of print, interest in reading and writing, oral language development, and phonological awareness	Center-based preschool settings (e.g., Head Start) for young children's language and literacy development	Measures the quality of young children's support in language and literacy development	Partial

Table 1 Continued

Classroom observation instrument with citation	Origin	Framework	Measurement	ELL purposeful
Supports for English Language Learners Classroom Assessment (SELLCA; National Institute for Early Education Research, 2005)	Effective strategies to support English language development in both classroom and parent activities English language development	Involves the teacher's awareness of each child's cultural background, encourages parents to participate, encourages use of native language, and supports	Assesses the degree of support for language and literacy development among English language learners	Yes
Sheltered Instruction Observation Protocol (SIOP; Echevarria, Vogt, & Short, 2012)	"Principles from English as a second language and bilingual education research" (Waxman, et al., 2004, p. 11)	Sheltered Instruction Model	Measures the extent of sheltered instruction implemented	Yes
School Observation Measure (SOM; Ross, Smith, & Alberg, 1999)	Drawn from surveys/discussions with policy makers, researchers, teachers, and administrators. (Ross et al., 2004)	A whole school (random visit) observation model	Measures the frequency of 24 instruction strategies during observations	Yes
Standards Performance Continuum (SPC; Tharp & Gallimore, 1988; Tharp, Estrada, Dalton, & Yamauchi, 2000)	Sociocultural perspective (Vygotsky, 1978)	Five pedagogy standards (Tharp Estrada, Dalton, & Yamauchi, 2000)	Teacher performance based on Standards for Effective Pedagogy	Yes
Support for Social-Emotional Growth Assessment (SSEGA; Smith, 2004)	Supports related to social-emotional growth in the context of classroom environment (as well as classroom routines) and teacher behavior	Teacher and child interactions in preschool classrooms	Documents effective classroom events related to children's social-emotional growth, teacher behavior, and academic classroom routines and activities (e.g., unrushed transitions)	No

Table 1 Continued

Classroom observation instrument with citation	Origin	Framework	Measurement	ELL purposeful
Stanford Research Institute (SRI) Classroom Observation Instrument or Stallings Classroom Observation Instrument (COI; Stallings, 1973)	Specifically designed for the evaluation of National Head Start and Follow Through Planned Variation programs	Multi-tool strategy: Classroom Summary Information (CSI); Physical Environment Information (PEI); Classroom Checklist (CCL); Five Minute Observation (FMO)–Flanders Interaction System (1969)	Instrument and adapted forms–evaluate a wide variety of educational components significant to the Follow Through sponsors	No
Transitional Bilingual Observation Protocol (TBOP; Lara-Alecio & Parker, 1994)	Four Dimensional Transitional Bilingual Pedagogical Model (Lara-Alecio & Parker, 1994)	Transitional Bilingual Observation Protocol (Lara-Alecio & Parker, 1994; Bruce, 1995)	Describes the pedagogical occurrences by four instruction domains in the classrooms for English language learners	Yes
Teaching For Meaning Classroom Observation Form (TFM; Knight & Ackerman, 1997)	Teaching for meaning (Knapp, Shields, & Turnbull, 1995; Knapp & Adelman, 1995; Tharp et al., 2000)	Developed to evaluate teaching for meaning in the Connections projects (Knight & Smith, 2004)	Instrument–measures teaching for meaning components and student engagement	Yes
Timed Observations of Student Engagement/Language (TO/SEL; Foorman, Goldenberg, Carlson, Saunders, & Pollard-Durodola, 2004)	From the Time Observation of Student Engagement (TOSE), Foorman and Schatchneider (2003)	Developed for an NICHD-funded study on early reading interventions (K-4) and was given a language component	Time-sampling instrument on instruction components, student engagement, and language	Partial
Teacher Roles Observation Schedule (TROS; Waxman, Wang, Lindvall, & Anderson, 1990b)	Descriptive of the nature and pattern of teacher instruction behaviors	Teacher centered observation	Records instruction settings, interactions, and content	No

Table 1 Continued

Classroom observation instrument with citation	Origin	Framework	Measurement	ELL purposeful
Quality of Early Childhood Care Settings: Caregiver Rating Scale (QUEST; Goodson, Layzer, & Layzer, 2005)	Current practices on children's development and learning– cognitive, language and early literacy, emotional, social, and physical developments	Best practices for center-based care for children aged 0 to 5 development and learning	Rating scale on caregiver warmth/responsiveness and ability to support child development	Partial

Note. The ELL purposeful column relates to how raters scored each classroom observation instrument as specifically geared toward serving the English language learner classroom.

Rater judgment was used to determine the outcome for the ELL purposeful column and involved two raters whom examined literature in order to decide the authenticity of COIs described as explicitly created for serving the ELL classroom. This further examination of COIs was necessary, because the systematic review merely brought forth literature which described COIs in the context of ELLs. However, not all COIs used in the context of ELLs are considered COIs for ELL assessment in the classroom. As a result, the last column (i.e., ELL purposeful) in Table 1 served to make the distinction between authors that described their COI as explicitly created for capturing ELL activity in the classroom (i.e., serving the ELL classroom). In the proceeding tables this systematic review continues the examination process of determining which COIs are suitable for capturing ELL activity in the classroom.

Additionally, eight COIs were difficult to determine if they specifically were intended to measure ELL classroom activity; due to COIs focusing on language and literacy but not exclusively mentioning how they measured ELL classroom events. Academic achievement for ELLs is related to language and literacy development, yet among raters the task was to determine if explicit descriptions were present from included studies which depicted COIs as created for serving the ELL classroom. COIs that were difficult to determine whether they were serving the ELL classroom (i.e., labeled as ELL purposeful) were deemed "partial" in Table 1 and required further review of ancestry literature.

The author then cross-referenced with how other authors viewed COIs in Table 1, which also served to determine the final outcome to the last column (i.e., ELL

purposeful column in Table 1). One such cross-referenced source used in reaching consensus of whether COI were created for serving the ELL classroom was a compendium prepared by Child Trends (Halle, Whittaker, & Anderson, 2010). In Halle, Whittaker, and Anderson (2010) the authors sought to compile a comprehensive list of early childhood observation instruments, of which could be examined for any mention of instruments with serving the ELL classroom. Furthermore, attempts to contact researchers in determining hard-to-examine COIs that could play a role in explicitly and exclusively serving the ELL classroom was part of the examination process.

In order to address the notion that not all identified COIs are explicitly for the use of serving the ELL classroom related variables during instruction, an additional filtering process was necessary. As a next step, Figure 4 illustrates the process of taking Table 1 results and determining which COIs were for serving the ELL classroom. The logic driving the process to reduce the broadly accepted COIs is depicted in Figure 4 and involved the taking of studies included in the systematic review and gradually examining the 37 COIs for explicit descriptions of COIs used in classrooms serving ELLs. In the midsection of Figure 4 are COIs that could be classified as such due to their emphasis in coding language and literacy development.

The judgment used by raters to the extent of determining if COIs were for serving the ELL classroom (i.e., the ELL purposeful classification in Table 1) was based on author descriptions of their COI and cross-referencing with additional sources. Of the 37 COIs found through systematic review criteria, ten COIs were not considered to serve the ELL classroom (i.e., researchers did not explicitly describe their COI as intended in

the use of classrooms serving ELLs). Another eight COIs were considered related to descriptions of COIs use in classrooms serving ELLs or partially as described in Table 1 (i.e., first related to literacy and language development, but little or no mention of ELL). As a result, 19 of the 37 COIs found through systematic review passed additional rater examination.

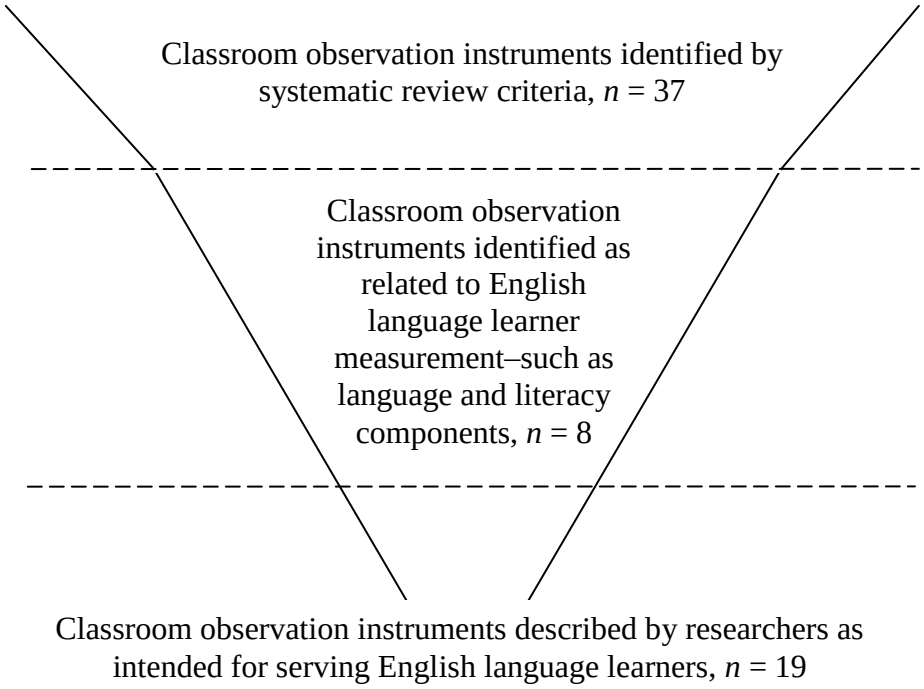

Figure 4. The process used to examine classroom observation instruments for English language learners.

Summary of Research Questions

Utilizing systematic review protocols, the author found 37 COIs, which served as a broad sample of COIs to begin examining for specific descriptions of COIs used in classrooms serving ELLs. As a result, not all 37 COIs are described as specifically intended for pedagogy in the U.S. classroom serving ELLs, nor have all COIs been specifically intended to record ELL events in the classroom. What followed was the systematic output that led to further examination of 19 of 37 COIs described in the range of serving ELLs as illustrated by Figure 4. Also, the 19 COIs are considered potential COIs serving the ELL classroom through cross-referencing with several authors specifically describing the 19 COIs in the context of serving ELLs in the classroom. The second research question elaborates on the theoretical underpinnings and psychometric properties of the 19 COIs examined through the systematic output used in addressing the first research question.

The 37 COIs were examined by how explicit and purposeful authors described their COI as being specifically intended for serving the ELL classroom. In Table 2 highlighted statements are provided to illustrate each author's intent of their COI as not serving ELLs in the classroom. Through literature identified in this systematic review, the examination of each COI was led through rater examination of any explicit description literally stating the author's COI for use for serving ELLs or dual language learners (DLLs). The term DLLs was added because of the exclusivity toward ELLs. As a result, raters investigated for ELL exclusivity as the main criteria for determining

whether COIs were intended to serve ELLs in the classroom (i.e., partially or completely intended for serving the ELL classroom).

Table 2

Classroom Observation Instruments Classified as Not English Language Learner Classroom Purposeful

Classroom instrument	Descriptions depicting classroom observation instruments as not explicit or exclusive to English language learner classroom events
Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008)	Underpinning the entire CLASS tool is the theory that the "primary mechanisms through which children acquire readiness-related competences are social relationships children form with peers, parents, and teachers" (Mashburn & Pianta, 2006); Measurement related to English language learners is provided in a general context
Early Childhood Classroom Observation Measure (ECCOM; Stipek & Byler, 2004)	The ECCOM was developed to "assess the nature and quality of instruction as well as the social climate and management of the classroom" (Stipek & Byler, 2004)
Collaborative Strategic Reading Intervention Validity Checklist (CSRIVC; Vaughn, Hughes, Schumm, & Klingner, 1998)	The CSRIVC was used to assess teacher professional development measures (Hitchcock, Dimino, Kurki, Wilkins, & Gersten, 2011); Also, "the CSRIVC is an observational checklist created by the intervention developers to measure fidelity of Collaborative Strategic Reading implementation" (Vaughn et al., 1998; Hitchcock et al., 2011)
Expository Reading Comprehension (ERC) observation instrument (James-Burdumy et al., 2009)	The ERC is "used to categorize and code teachers' comprehension and vocabulary instruction" (Hitchcock et al., 2011)
Local Systematic Change Classroom Observation Protocol (LSC COP; Horizon Research Inc., 2000)	According to the observation protocol methodology, "the instrument was developed to measure the quality of an observed K-12 science or mathematics classroom lesson by examining the design, implementation, mathematics/science content, and culture of that lesson" (Horizon Research Inc., 2000)
Stallings Classroom Observation Instrument (Stallings, 1973)	The COI "was developed to record classroom occurrences as a way to determine whether there were planned educational variations in the Follow Through programs" (Stallings, 1973); English language learners were present but the COI was focused on program evaluation

Table 2 Continued

Classroom instrument	Descriptions depicting classroom observation instruments as not explicit or exclusive to English language learner classroom events
Teacher Roles Observation Schedule (TROS; Waxman, Wang, Lindvall, & Anderson, 1990b)	The TROS "is a systematic observation schedule designed to document observed teacher behaviors in the context of ongoing classroom instruction-learning processes" (Padrón, 1994); The TROS by itself needs another classroom observation instrument to measure classroom interaction, especially if English language learners are present
Support for Social-Emotional Growth Assessment (SSEGA; Smith, 2004)	According to the researchers "the SSEGA does not have specific items that address language or ethnic diversity" (Halle et al., 2010, p. 295)
Intercultural Development Inventory (IDI; Hammer & Bennett 1998)	The IDI assesses intercultural competence–"the capability to shift cultural perspective and appropriately adapt behavior to cultural differences and commonalities" (IDI, n.d.); IDI measures are specific to culture
Classroom Language and Literacy Environment Observation (CLEO; Holland-Coviello, 2005)*	CLEO "component will be described in terms of their derivation from research linking elements of social interaction and literacy environments with children's emergent literacy development" (Holland-Coviello, 2005, p. 10); No mention of English language learners exclusivity was found, but emergent literacy development is vital to English language learners
Early Literacy Observation Tool (E-LOT; Grehan, Smith, & Ross, 2004)*	The E-Lot is "aligned to the National Reading Panel and National Research Council findings and captures all essential components of the Early Reading First program" (Halle et al., 2010, p. 172); The standards rooted in the E-Lot can be conducive to English language learners reading comprehension and early reading but exclusivity to English language learners was not mentioned

Note. * = possible for English language learner observation because of language and literacy measures, yet not explicitly described in design for English language learners use.

Table 2 includes highlighted statements among 11 of 37 COIs that raters determined as not serving the ELL classroom. The COIs needed to explicitly connect to systematic review criteria by author's describing their COI's purpose as it related to serving the ELL classroom. The highlighted statements in Table 2 served to illustrate reasons to categorize the listed COIs as not serving the ELL classroom (i.e., not serving

the ELL classroom). Rater judgment decisions were also made to not include COIs that could be used for serving the ELL classroom because of strong measures in language and literacy, yet did not specify the COI as being for English language learners (i.e., COIs that have a focus toward language and literacy, but not explicit or exclusive to serving the ELL classroom).

Additionally, illustrated in Table 3 are highlighted statements raters found from 7 of 37 COIs that incorporated measures explicitly for serving the ELL classroom. The highlighted statements for each COI in Table 3 are found in literature by their respective authors that described transparency on how their COI served ELLs in the classroom. However, the following seven COIs are considered partially related to being able to serve the ELL classroom due to author statements only describing parts of their COI as for serving the ELL classroom. First, COIs were partially considered for being able to serve the ELL classroom because initially they mentioned language and literacy development in the presence of classrooms with ELLs (i.e., from Table 1). By examining further the purpose and intent of each of these partially-termed COIs for serving the ELL classroom (i.e., from Table 3), raters determined how close language and literacy development focused COIs came to explicitly stating the entire COI for serving the ELL classroom (i.e., the theory, purpose, or specific instrument components of each COI that measured ELL events during the classroom).

Table 3

Classroom Observation Instruments Identified as Having Explicit Measures Related to Being English Language Learner Purposeful

Classroom observation instrument	Percent serving English language learners	Descriptions depicting classroom observation measures as explicit to serving the English language learner classroom
Supports for Early Literacy Assessment (SELA; Smith, Davidson, Weisenfeld, & Katsaros, 2001)	10% or 2 measures explicitly stated out of 21	The assessment of emergent literacy and language development; 2 items (of 21 items) "assess the extent to which a child's native language is maintained and developed within the classroom setting, and the use of effective strategies to help children understand and acquire English" (Halle, Whittaker, & Anderson, 2010, p. 291)
Observation Measures of Language and Literacy (OMLIT; Goodson, Layzer, Smith, & Rimdzius, 2006)	50% or 3 measures explicitly stated out of 6	Of the 6 OMLIT measures (Classroom Description, Snapshot of Classroom Activities–Snapshot, Read-Aloud Profile–RAP, Classroom Literacy Instruction Profile–CLIP, Quality Rating of Language and Literacy Instruction–QUILL, and the Classroom Literacy Opportunities Checklist–CLOC), the OMLIT-CLOC, OMLIT-Snapshot, and OMLIT-QUILL address culturally and linguistic components
Early Language & Literacy Classroom Observation toolkit (ELLCO; Smith, Dickinson, Sangeorge, & Anastasopoulos, 2002)	21% or 3 measures explicitly stated out of 14	The ELLCO consists of three assessment components, which together "describes the extent to which classrooms provide children optimal support for their language and literacy development" (Halle et al., 2010, p. 158); The classroom observation component has 3 items (Item 12–Recognizing diversity in the classroom, Item 13–Facilitating home support for literacy, and Item 8–Presence of books) related to linguistic and cultural diversity in the classroom (Halle et al., 2010, p. 159)
Early Language & Literacy Classroom Observation–Pre-K (ELLCO; Smith, Brady, & Anastasopoulos, 2008)	5% or 1 measure explicitly stated out of 19	The item Recognizing Diversity in the Classroom documents the teacher's efforts to use children's prior knowledge and interest, make home-school connections for all children, and determine if cultural and linguistic diversity are valued
Quality of Early Childhood Care Settings: Caregiver Rating Scale (QUEST; Goodson, Layzer, & Layzer, 2005)	4% or 3 measures explicitly stated out of 69	Addresses language development and early literacy; Comprised of two measures (The Environment Checklist and the Caregiver Rating Scale), the QUEST Provider Rating measure has 3 items (of 69) that assess the caregiver's approach to supporting English language learners in the group

Table 3 Continued

Classroom observation instrument	Percent serving English language learners	Descriptions depicting classroom observation measures as explicit to serving the English language learner classroom
Code for Interactive Recording of Children's Learning Environments (CIRCLE; Atwater, Lee, Motagna, Reynolds, & Tapia, 2009)	12% or 7 measures explicitly stated out of 57	The CIRCLE notes language used in the classroom, which was considered a simple observation measure from ELL– "Observers note the primary language for each child being observed. Observers also note whether the child uses conventional words in a language other than English or uses sign language" (Halle et al., 2010, p.113)
Opportunity To Learn/ Academic Language Exposure (OTL/ALE) survey (Martinez, Bailey, Kerr, Huang, & Beauregard, 2010)	86% or 6 measures explicitly stated out of 7	As described by, the OTL construct was "complemented with three ELL serving facets of OTL identified by previous research at the National Center for Research on Evaluation, Standards and Student Testing (CRESST; e.g., Boscardin et al., 2004)"; OTL was considered ¾ parts serving the ELL classroom As for ALE researchers say, "our definition of Academic Language Exposure is informed by recent theoretical and empirical work on language acquisition and learning"; ALE was considered able to serve the ELL classroom (3/3)

Note. Percent to measure English language learner observation = an approximate portion of a classroom observation instrument to explicitly measure English language learner activity in the classroom. The percentage is computed by dividing the number of explicitly ELL related measures by the total number of classroom observation measures.

In summary, Figure 5 provides a visual illustration of COIs and how they relate to serving the ELL classroom and the amount of supportive literature for each by using the name of each COI as the search term. Each COI was individually searched through the ProQuest and EBSCOhost search engines and included all available databases, along with Goggle searches to estimate a general number on the perceived prevalence scale. The number of times each COI appeared in databases combined with Google search

results translated to each COI placement on the prevalence scale. In other words, the prevalence scale was used to provide a general idea of how frequent each COI was present in the literature with added count values from Google search results.

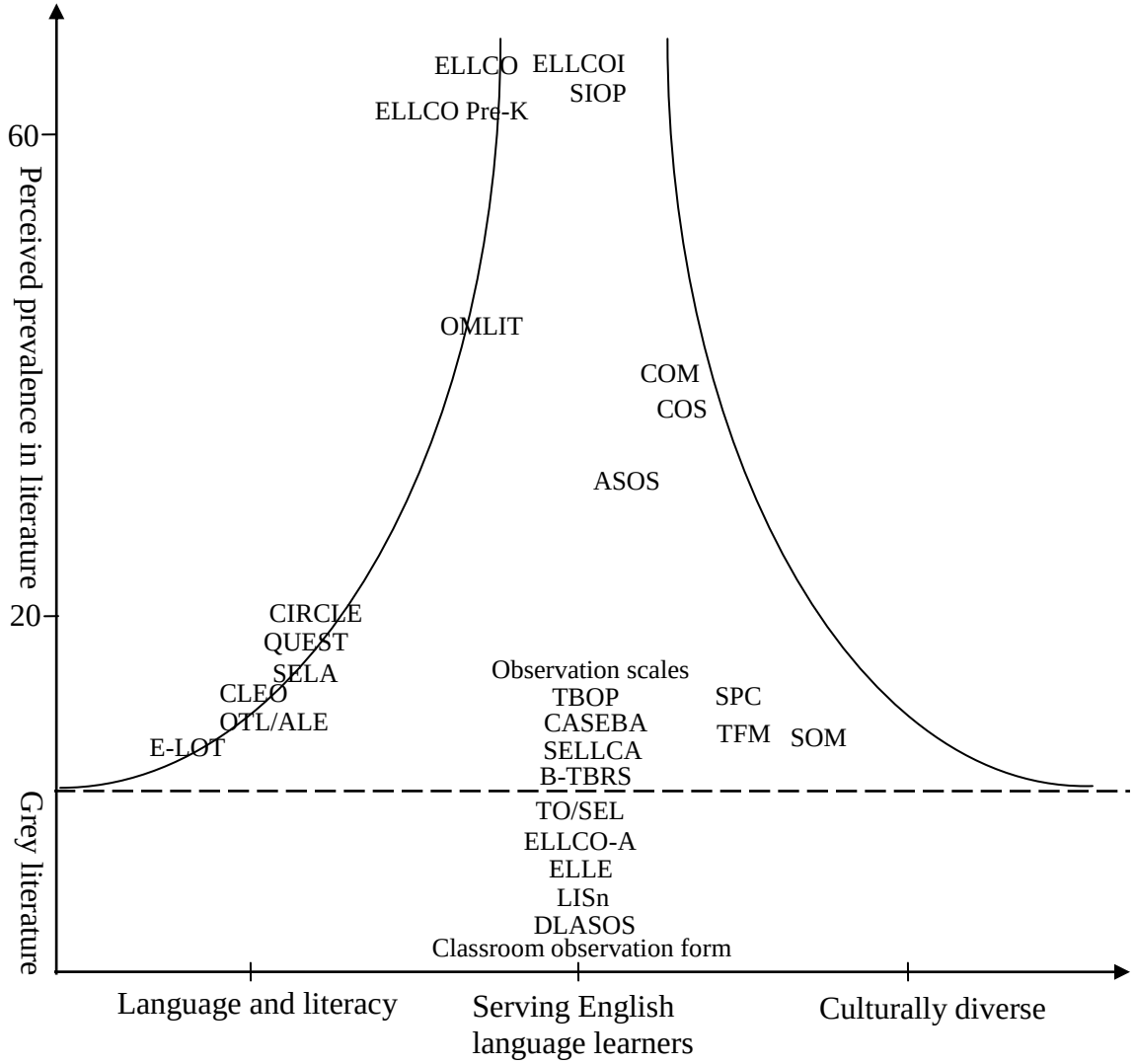

Figure 5. Funnel plot illustrating the prevalence of observation instruments for serving English language learners.

The funnel plot (i.e., Figure 5) displays the outcome of 19 potential COIs for serving the ELL classroom. However, by only considering COIs outside of gray literature (i.e., COIs found in studies that have not gone through the journal peer-review process) and focused on pedagogy for ELLs in the classroom, the results of this systematic review become very small. For instance, the ELLCOI, TBOP, ASOS, B-TBRS and observation scales developed by Luykx and Lee (2007), with SPC, COM and TFM as notable become the handful of COIs serving the ELL classroom. In general, from Figure 5, COIs that could be used in the context of serving the ELL classroom are: ELLCOI, SIOP, ASOS, SELLCA, CASEBA, B-TBRS, TBOP, classroom observation guidelines (i.e., observation scales by Luykx and Lee, 2007), SPC, TFM, COS, COM and SOM. The OTL/ALE instrument is also noteworthy to mention as a potential instrument to use because of the 86% relevance to serving the ELL classroom.

Despite the small handful of COIs serving the ELL classroom, COIs differ in their approach to describe the ELL classroom. For instance, there are different pedagogical and linguistics approaches incorporated among COIs, which determine how the ELL classroom is described. COIs also differ in instruction models and psychometric properties as they describe the ELL classroom. Although COIs vary in measurable outcomes, the results from Figure 5 are arguably considered as having the strongest of intentions for serving the ELL classroom. Alternatively, because of differences in psychometric properties between COI, the reasons for using one COI over another depends on what was the needed ELL classroom description desired. The reason for using a certain COI would ultimately depend on the purpose. As an added clarification of the differences between COIs identified as serving the ELL classroom, Table 4 provides further context of the psychometric properties for COIs labeled as yes or partial to serving the ELL classroom (i.e., as label in the ELL purposeful column in Table 1).

Table 4

Psychometric Properties of Selected Classroom Observation Instruments

Classroom observation instrument	Purpose of instrument	Pedagogical framework	Language framework	Additional psychometric properties
Activity Setting Observation System (ASOS; Rivera et al., 1999; Rivera, Tharp, Youpa, Dalton, Guardino, & Lasky, 2005)	Description from researchers: The ASOS "provides an objective description of the defining attributes of classroom activity settings" (Rivera & Tharp, 2004, p. 208)	Based on Center for Research on Education, Diversity & Excellence (CREDE) standards for effective pedagogy and uses the activity setting as the unit for analysis	Sociocultural theory; Language is a tool for learning	Reliability: Activity Structure (AS) $R = 0.99$ and Cohen's kappa ranges for the following instrument measures: Product = 0.73 to 0.74, Personnel = 0.61 to 1.0, Student Initiative or Choice = 0.79 to 1.0, Joint Productive Activity = 0.59 to 0.70, Modeling/Demonstration = 0.65 to 0.72, Teacher/Student Dialogue = 0.63 to 0.79, Responsive Assistance = 0.79 to 0.87, Contextualization = 0.61 to 0.81, Connected AS = 0.63 to 0.75 Validity: Established by CREDE researchers Training time: Not apparent Adapted from a previous instrument: No Assessment: Activity setting

Table 4 continued

Classroom observation instrument	Purpose of instrument	Pedagogical framework	Language framework	Additional psychometric properties
Bilingual Teacher Behavior Rating Scale (B-TBRS; Landry et al., 2001)	As described by researchers the B-TBRS, "examines the quantity and quality of instructional practices in relation to the language used" (Atkins-Burnett et al., 2010, p. 51).	Based on quality of instructional practices	Language use in the classroom	Reliability: Inter-rater reliabilities for TBRS subscales range from .80 to .98; Total scale with internal consistency of .96 Validity: Has construct validity and stability across time Training time: 2 day minimum Adapted from a previous instrument: Yes, adapted from the Teacher Behavior Rating Scale (Landry et al., 2001) Assessment: Three-point scale (rarely, sometimes, often) to assess quantity of opportunities for learning; and a four-point scale to assess quality of learning opportunities
Classroom Assessment of Supports for Emergent Bilingual Acquisition (CASEBA; Freedson, Figueras-Daniel, & Frede, 2009)	"Designed to assess the degree to which pre-school teachers and classrooms are providing support for the social, cognitive, and linguistic development of English language learners, with a focus on language and literacy"(Freedson, Figueras-Daniel, & Frede, 2009, p. 1)	"The CASEBA is designed for settings with English language learner pre-school students and assesses the teachers' cultural responsiveness" (Halle et al., 2010, p. 62)	"Subscales of the instrument are also suitable to assess supports for language and literacy for all pre-school children" (Halle et al., 2010, p. 62)	Reliability: Not available Validity: Not available Training time: At least 4 days Adapted from a previous instrument: Yes, Support for Early Language Learners Classroom Assessment (SELLCA; National Institute for Early Education Research, 2005) Assessment: 7-point Likert scale, "where 7 indicates that a specific form of support and accompanying practices are present in close to an ideal form, while 1 represents the total absence of any such practices" (Freedson, Figueras-Daniel, & Frede, 2009, p. 1).

Table 4 continued

Classroom observation instrument	Purpose of instrument	Pedagogical framework	Language framework	Additional psychometric properties
Classroom observation form (Servin, 1983)	Dissertation study to observe teacher behavior in Spanish use, the use of linguistic deviations, and the amount and type of corrective feedback	Observation of teacher pedagogy	Observation of language use in the classroom	Reliability: Not available Validity: Not available Training time: Not apparent Adapted from a previous instrument: No, found as dissertation study Assessment: Teacher language use in the classroom
Classroom observation guidelines–based on observation scales developed by Luykx and Lee (2007)	The classroom observation guidelines were used for purposes of instruction congruence fidelity and observation scales were used to "summarize teacher and student behaviors deemed important to establishing instructional congruence" (Drews, 2009, p. 24)	Congruence, "the pedagogical practices that bridge the lives of students with the worlds of science and school in ways that are meant to empower the students and create relevant learning environments" (Drews, 2009, p. 5)	Language is considered inherently tied to the knowledge (or funds of knowledge) students bring to the classroom (Moll, Amanti, Neff, & Gonzalez, 1992)	Reliability: Interrater estimates $r = 0.74$, $r = 0.84$, and $r = 0.60$, and $r = 0.81$from Luykx and Lee (2007) Validity: From "Science for All: Instruction intervention to promote science and literacy with linguistically diverse elementary students" project Training time: Not apparent Adapted from a previous instrument: No Assessment: 5-point Likert rating system on the frequency of the activity (i.e., science inquiry) and the number of students involved in the activity

Table 4 continued

Classroom observation instrument	Purpose of instrument	Pedagogical framework	Language framework	Additional psychometric properties
Classroom Observation Measure (COM; Ross & Smith, 1996)	The COM was designed to "systematically study a tracked intervention model for at-risk elementary school students in one urban school system" (Castellano & Datnow, 2004, p. 239)	Teaching methods deemed effective for at-risk learners (Castellano & Datnow, 2004)	In the context of cultural diversity in the school using comprehensive school reform models (Stringfield, Datnow, and Ross, 1998)	Reliability: Ross, Smith, Lohr, and McNelis (1994) report the high consistency ratings of reliability in percentage of interrater agreement, interrater correlations, and member checking as in qualitative reliability checks Validity: COM validation in literature can be found in Stringfield, Datnow, and Ross (1998); Ross, Alberg, and Wang (1998); as well as, in Ross et al. (1994) Training time: Not apparent Adapted from a previous instrument: Yes, the COM came from the Elementary Classroom Observation Measure (ECOM) Assessment: 5-point scale on nine classroom snapshot measures
Classroom Observation Schedule (COS; Padrón, Waxman, & Huang, 1999; Waxman & Padrón, 2004)	"Designed to focus on individual students in order to address potential inequities in the classroom" (Waxman & Padrón, 2004, p. 74).	"Student-mediating paradigm, which maintains that students actively process information and interpret classroom reality" (Waxman & Padrón, 2004, p. 73)	Language use and interaction in the classroom	Reliability: interrater reliability recorded as $r > 0.95$ Validity: reported in previous studies (Waxman & Huang, 1999) Training time: a few hours Adapted from a previous instrument: No Assessment: 30-second interval observations per student during a 60-minute session

Table 4 continued

Classroom observation instrument	Purpose of instrument	Pedagogical framework	Language framework	Additional psychometric properties
Dual Language Activity Setting Observation System (DLASOS; Rivera & Tharp, 2010)	The purpose of the DLASOS is to "seek to develop effective classroom environments for the teaching and learning of bilingual children" (Rivera & Tharp, 2010)	Sociocultural theory; Research-based teaching and learning while challenging students toward cognitive complexity	Sociocultural theory; Linguistic environment and language development observed during activity structure	Reliability: Underdevelopment Validity: Underdevelopment Training time: Underdevelopment Adapted from a previous instrument: Yes, the Activity Setting Observation System (ASOS), which is derived from the work of Tharp and Gallimore (1988), O'Donnell and Tharp (1990), and Rivera, Tharp, Youpa, Danton, Guardino, and Lasky (2005) Assessment: Activity setting, which is the basic unit of analysis in sociocultural theory

Table 4 continued

Classroom observation instrument	Purpose of instrument	Pedagogical framework	Language framework	Additional psychometric properties
Early Language and Literacy Classroom Observation: Addendum for English Language Learners (ELLCO–ELL or ELLCO–A; Castro, 2005)	Described by researcher: "This measure has been developed as an addendum to the Early Language and Literacy Classroom Observation Toolkit (ELLCO), to obtain information about specific classroom practices related to promoting language and literacy development among children who are ELL" (Castro, 2005, p. 2)	The use of specific classroom practices to foster language and literacy development	Based on the ELLCO, which observes "the extent to which classrooms provide children optimal support for their language and literacy development" (Halle et al., 2010)	Reliability: Training criterion = 90% agreement; Cohen's kappa = 0.46 mean for each item on the classroom observation scale; 94% agreement on Literacy Environment Checklist; 100% mean value for percent exact agreement for each item on the Literacy Activities Rating Scale (Halle, et al., 2010, p. 170) Validity: Not available Training time: Not apparent Adapted from a previous instrument: No Assessment: Authors recommends, the ELLCO be conducted with the ELLCO-ELL because while the ELLCO establishes a starting point for classroom observation the ELLCO-ELL then assess what is being done for ELL beyond ELLCO measures

Table 4 continued

Classroom observation instrument	Purpose of instrument	Pedagogical framework	Language framework	Additional psychometric properties
English Language Learner Classroom Observation Instrument (ELLCOI; Baker, Gersten, Goldenberg, Graves, & Haager, 1999; Gersten et al., 2005; Haager et al., 2003)	The ELLCOI observes the instruction practices of teachers during reading instruction for English language learners (Whitacre, Diaz, & Esquierdo, 2013)	Based on relevant research on teaching reading in a second language	Focusing on English language reading proficiency in order to outpace student's oral language development (Baker, Gersten, Haager, Dingle, & Goldenberg, 2005, p. 8)	Reliability: Inter-observer agreement median = 74%, ranging from 55% to 88% (Gersten, et al. (2005)–based on item-by-item agreement (Whitacre, Diaz, & Esquierdo, 2013). Validity: Validation study conducted (Baker, Gersten, Haager, and Dingle, 2006; Baker, Gersten, Haager, Dingle, & Goldenberg, 2005) Training time: Not apparent Adapted from a previous instrument: No Assessment: Developed as a moderate-inference Likert scale teaching observation tool (Gersten et al., 2005, p. 199)

Table 4 continued

Classroom observation instrument	Purpose of instrument	Pedagogical framework	Language framework	Additional psychometric properties
Early Language and Literacy Environment (ELLE; created by Mathematica –Atkins-Burnett et al., 2010)	"A measure of the support for language and literacy available in the environment's materials and activities" (Atkins-Burnett, Xue, Kopack, Induni, and Moiduddin, 2010, p.7)	Created by Mathematica by adapting scales from the ELLCO Research Edition, the ELLCO Addendum (Castro 2005), and the CHELLO (Nueman et al. 2007)	Measures the availability of literacy resources in English, Spanish, and other languages such as toys and puzzles, technology, books, and writing materials	Reliability: the Literacy Checklist in English was adequate (alpha = .74), as stated by Atkins-Burnett, Xue, Kopack, Induni, and Moiduddin (2010, p. 10); Lower reliability was found in the Family Child Care programs compared to Center-based programs; Overall internal consistency was acceptable for the ELLE Literacy Checklist in English Validity: Described in Atkins-Burnett, Xue, Kopack, Induni, and Moiduddin (2010, p. 1) Training time: Not apparent Adapted from a previous instrument: Yes, Early Language and Literacy Classroom Observation (ELLCO) Research Edition (Smith & Dickinson, 2002); the ELLCO Addendum (Castro, 2005); and the Child Home Early Language and Literacy Observation (CHELLO; Neuman et al., 2007) Assessment: Two sections that measure availability of literacy resources and the rating of book-reading activities
Language Interaction Snapshot (LISn; Sprachman, Caspe, & Atkins-Burnett, 2008)	"Designed to examine how the language environment differs for children, particularly in classrooms that include dual language learners" (Halle, et al., 2010, p. 212)	Teacher instruction using contextualized language	Language use in the classroom	Reliability: Video inter-rater reliability = 96%; Field inter-rater reliability = 89% Validity: criterion, concurrent, and construct reported (Halle et al., 2010, p. 215) Training time: 2 days Adapted from a previous instrument: No Assessment: 30-second cycles within 5 minutes of classroom snapshot measures

Table 4 continued

Classroom observation instrument	Purpose of instrument	Pedagogical framework	Language framework	Additional psychometric properties
Supports for English Language Learners Classroom Assessment (SELLCA; National Institute for Early Education Research, 2005)	SELLCA "assesses the degree to which the teacher incorporates the cultural backgrounds of the children in the classroom and encourages parent participation (National Research Council, 2008, p. 172)	Use of effective strategies to support English language development	Teachers' use of children's dominant language and the degree to which children's cultural backgrounds are incorporated in the classroom	Reliability: Underdevelopment–appeared in Halle and Vick (2007) compendium but not in Halle et al. (2010) compendium Validity: Underdevelopment Training time: Not apparent Adapted from a previous instrument: No Assessment: Assess the degree of teacher use of cultural background information in the classroom from 1 (minimal evidence) to 5 (strong evidence) was it pertains to instruction

Table 4 continued

Classroom observation instrument	Purpose of instrument	Pedagogical framework	Language framework	Additional psychometric properties
Sheltered Instruction Observation Protocal (SIOP; Echevarria, Vogt, & Short, 2012)	SIOP was developed to make content material comprehensible to English language learners (Haynes, n.d.; Echevarria, Vogt, & Short, 2012)	Sheltered instruction is an approach for teaching content to ELL in strategic ways that make the subject matter concepts comprehensible	The SIOP works while promoting the students' English language development	Reliability: Guarino, Echevarria, Short, Schick, Forbes, and Rueda (2001) established reliability and validity; "All but one subscale (Comprehensible Input; alphas = .873) achieved an a priori level of acceptance" (Echevarria & Short, 2004, p. 31); The other subscales had alpha ranges from 0.959 (Preparation) to 0.914(Lesson Delivery) Validity: Instrument's discriminate validity was tested (3 factors accounting for 98.4% of the variance); Instrument's concurrent validity was checked; Stability of classification was also checked (81.25% correct classification rate) Training time: Virtual training from 11 sessions Adapted from a previous instrument: No Assessment: 5-point Likert scale (0-4) on 8 components
School Observation Measure (SOM; Ross, Smith, & Alberg, 1999)	SOM was designed to "capture the frequency with which 24 instruction practices are implemented during direct observation of classrooms" (Ross, Smith, Lowther, Alberg, & Cheon, 2006, p. 3)	Based on 24 teaching strategies derived from national teaching standards and effective teaching methods (Ross, Smith, Albert, & Lowther, 2004)	No clear language framework identified	Reliability: Interrater reliability conducted by Lewis, Ross, and Alberg (1999)–SOM had sufficiently high reliability and validity as a research and evaluation instrument Validity: Content validity (Faris, 2006) Training time: Not apparent Adapted from a previous instrument: Yes, limitation in school program research of the COM led to the development of the SOM based on Session Teaching Behaviors and Method sections of the COM Assessment: 5-point rubric scoring in 7 areas

Table 4 continued

Classroom observation instrument	Purpose of instrument	Pedagogical framework	Language framework	Additional psychometric properties
Standards Performance Continuum (SPC; Tharp & Gallimore, 1988; Tharp et al., 2000)	SPC measures a teacher's performance based on the Standards for Effective Pedagogy during classroom instruction	Standards for Effective Pedagogy	Based on the sociocultural tenet that learning occurs best when novices collaborate and converse with more experienced and knowledgeable others on a shared task (Vygotsky, 1978)	Reliability: Rater reliability estimates are gained through raters bring trained properly (10-20 hours plus the SPC Manual) and continually assessed for accuracy (Hilberg, Doherty, Epaloose, & Tharp, 2004, p. 56) Validity: Illustrated in Hilberg, Doherty, Epaloose, and Tharp (2004) Training time:10-20 hours Adapted from a previous instrument: No Assessment: The SPC is a 5-point rubric based on the Standards for Effective Pedagogy

Table 4 continued

Classroom observation instrument	Purpose of instrument	Pedagogical framework	Language framework	Additional psychometric properties
Transitional Bilingual Observation Protocol (TBOP; Lara-Alecio & Parker, 1994)	Described by researchers, as driven by their pedagogical model which seeks to "identify classroom elements which teachers have the ability to adjust to enhance student learning (i.e., pedagogical utility)" and the potential for use in formative program evaluation "for formative judgments about the presence and absence of valued elements in the learning process" (Lara-Alecio & Parker, 1994, p. 121)	Four-dimensional pedagogical model for transitional-English-bilingual classrooms: Activity Structures, Language of Instruction, Language Content, and Communication Mode (Lara-Alecio & Parker, 1994)	Language works synergistically with pedagogy and represented by Language Content and Language of Instruction, which are defined in unison within the Four-dimensional pedagogical model	Reliability: 40 hours of observation from Parker, Tindal, and Hasbrouck (1994) found reliability at 0.82 - 0.98 (Cohen's Kappa); Further reliability coefficients are illustrated in Bruce, Lara-Alecio, Parker, Hasbrouck, Weaver, and Irby (1997): Language Content = 0.87 - 0.93(percent agreement) and 0.67 - 0.76 (Cohen's Kappa), Language of Instruction = 0.69 - 1.0 (percent agreement) and 0.47 - 0.60 (Cohen's Kappa), Communication Mode = 0.88 - 0.93 (percent agreement) and 0.80 - 0.87 (Cohen's Kappa) Validity: Utility of the pedagogical model and validation (Lara-Alecio & Parker, 1994); Operationalization of the pedagogical model into a protocol (Bruce, 1995) Training time: Not apparent Adapted from a previous instrument: No Assessment: time sampled 20-second observation measure with a 1-minute momentary time sample across 4 pedagogical domains

Table 4 continued

Classroom observation instrument	Purpose of instrument	Pedagogical framework	Language framework	Additional psychometric properties
Teaching For Meaning (TFM; Knight & Ackerman, 1997)	"developed to assess behaviors associated with teaching for meaning during the Connections project– a districtwide effort that focused on enabling elementary teachers to design and implement instruction that is meaningful to the diverse group of students they teach" (Knight & Smith, 2004, p. 100)	Meaningful instruction which "embeds skill learning in activities that feature conceptually challenging content and draw on the prior experiences and cultures of students to provide relevance" (Knight & Smith, 2004, p. 98)	Research on effective strategies for culturally and linguistically diverse classrooms	Reliability: Interrater reliability obtained through training using videos, comparisons with expert ratings, and paired observations in the field (Knight & Smith, 2004, p. 101); Cohen's kappa = 0.85 (9 observers) Validity: Content validity (Knight & Smith, 2004, p. 101) Training time: Not apparent Adapted from previous instrument: No Assessment: A three part measure on student engagement, results of 5 point ratings on teaching for meaning scales, and a qualitative focused observation
Timed Observations of Student Engagement /Language (TO/SEL; Foorman, Goldenberg, Carlson, Saunders, & Pollard-Durodola, 2004)	The purpose of the TO/SEL was to measure the language use during student engagement measures	Recording of observations in instruction content and grouping strategies	Language of instruction and student engagement assessment	Reliability: was not specifically found, but reliability of the Timed Observations of Student Engagement (TOSE) is reported at over 80% interrater reliability (Foorman & Schatschneider, 2003; Foorman et al., 2004) Validity: Drawn from the TOSE with modifications in a procedure described by Scanlon and Vellutino (1996) Training time: 2 days Adapted from a previous instrument: Yes, modified from the TOSE Assessment: Foorman and Schatschneider (2003) developed a time-sampling procedure and a language component was added

Note. * = Instrument is underdevelopment and will report psychometric properties.

The second research question called for exploring COI construction from theory to practice. Table 1 illustrated theory to practice for 37 COIs from systematic review results while Table 4 was used to add more information about 19 COIs which were closer to the range of serving the ELL classroom. Furthermore, in Table 1, the logic was to identify COI theory that started the instrument and created the model or framework. Next, the theory to practice transition was described by what the COI measured in the classroom. All three aspects (i.e., theory, model, and measure) were described in order to determine the approach of each COI as they ultimately were relate to serving the ELL U.S. classroom. By examining each theoretical foundation from COIs, the raters illustrated how author's described their COI's development as it pertained to serving the ELL classroom. As a result, the reader could decide the best COIs based on their needed purpose for capturing ELL information in the U.S. classroom. Also, an assumption grounded in this study was that in order to for a COI to serve the ELL classroom, theory had to focus on improving ELL academic achievement.

Three COIs were explicit adaptations of other COIs (i.e., B-TBRS, ELLE, and TO/SEL; Table 1). In fact, additional COIs have roots from other instruments. For instance, the SOM was created from the Session Teaching Behaviors and Methods section of the COM. Additionally, the COM came from the ECCOM, which focused on observation research with at-risk learners (Ross, Smith, Lohr, & McNelis, 1994). As a result, some COIs of the 19 chosen are questionable in their ability to serve the ELL classroom. However, the result of 19 COIs in this study is a general range of possible COIs that could serve the ELL classroom. By eliminating COIs produced from other

COIs, the number of specific COIs used in classrooms serving ELLs diminishes to the following: ELLCOI, TBOP, ASOS, and observation scales developed by Luykx and Lee (2007), SPC and TFM.

The TBRS (Teacher Behavior Rating Scale) served as the platform in developing the B-TBRS. While the ELLE came about under a collection of COI, which were the ELLCO toolkit, ELLCO Addendum, and the CHELLO (Neuman, Dwyer, & Koh, 2007). On the other hand, the TO/SEL changed from the TOSE because of an added language measure. COIs evolving into other COIs tend to emerge from modifications added to the original COI, which questions if the original theory was still feasible after the modification. In fact, some instruments begin to have a constructed history from other instruments.

For example, the LISn+EVR was derived from the LISn and the making of the LISn came from adaptations from the C-COS or Child-Caregiver Observation System (Boller, Sprachman, and the Early Head Start Research Consortium 1998; Love et al., 2009). While, the LISn is described by researchers as doing the following: "examines language interactions of an individual focus child with both adults and peers" (Atkins-Burnett et al., 2010, p. 77). Additionally, the LISn is described as "designed to examine how the language environment differs for children, particularly in classrooms that include dual language learners" (Halle et al., 2010, p. 212), thereby making this COI able to serve the ELL classroom.

Based on effective pedagogical standards, such as the SPC (Standards Performance Continuum) and TFM Classroom Observation Form, teachers are trained to

exhibit pedagogy considered beneficial to ELLs. Yet, effective strategies tend to change over time. Certainly, it can be argued that not all strategies are helpful to all ELLs and the concept of when to apply a strategy or instructional practice can also be difficult for teachers. However, an element of continued professional development can be helpful in the case of COIs based on effective pedagogies. For example, one use of the TBOP (Transitional Bilingual Observation Protocol) is in treatment and control settings where the treatment is receiving teacher professional development (e.g., Lara-Alecio et al., 2009).

The TBOP is based on the Four Dimensional Transitional Bilingual Pedagogical Model developed by Lara-Alecio and Parker (1994). The theoretical construction of the TBOP started with deeper intentions. For instance, the Transitional Bilingual Observation (TBO) model "is pedagogical in that it attempts to integrate important elements of bilingual education theory for the purposes of improving classroom instruction" (Lara-Alecio & Parker, 1994, p. 125). Operationalization of the TBO model was described in Bruce (1995; Bruce et al., 1997). The TBOP's explicit instrument implementation process recorded in literature is unique and not generally exhibited across literature for other COIs.

Different theories dictate each COI (e.g., Table 1). For instance, the ASOS is a classroom observation system that builds upon sociocultural theory; that is viewing the classroom as a social organization. From a general sociocultural lens, what is observed is the treatment of ELLs in a sociocultural context which has proven beneficial (e.g., Rivera & Tharp, 2004). Alternatively, the SIOP model is used to inform instrumentation

to measure sheltered instructional behaviors. The SIOP is used in the context of sheltered instruction and has proven its effectiveness as an instructional method.

However, the SIOP is described as an effective tool for measuring sheltered instruction, followed by sheltered instruction as a proven method to promoting ELL academic achievement (Echevarria, Vogt, & Short, 2012). In the case of the SIOP, the model only proves sheltered instruction as helpful toward ELL academic achievement, which is the intent. However, the SIOP model is not claiming sheltered instruction as exclusive to ELL academic achievement. Therefore, although certain instructional methods are effective in the context of ELL academic achievement, the research questions in this study sought to examine the COI theory and its exclusivity toward ELL assessment in the classroom.

Local Systematic Change Classroom Observation Protocol (LSC COP) is an observation instrument that has validated science education methods in certain instances as beneficial to all students (i.e., even ELL). Yet, the instrument has a general focus to measuring student academic achievement and learning. This instrument was included because of an included study which described ELL in the classroom during LSC COP measures. However, the LSC COP is limited in providing information specifically on ELL academic achievement (Horizon Research Inc., 2000). Additionally, the LSC COP was not meant for specific ELL estimates.

Moreover, the Teacher Roles Observation Schedule (TROS) is an instrument measuring the instruction behaviors of the teacher (Waxman, Wang, Lindvall, & Anderson, 1990b). Yet, the instrument does not inform the teacher's behavior in relation

to the students, like the TBOP. In fact, researchers of the TROS have suggested additional observation instruments be used in tangent with the TROS (Waxman, Tharp, & Hilberg, 2004). However, the researchers who created the TROS are aware of this limitation and usually partner TROS with the Classroom Observation Schedule (COS).

Teachers can be given more to inform their pedagogy in terms of curriculum and professional development in conjunction with COI use. For instance, the TBOP is associated with large-scale interventions (i.e., Project English Language and Literacy Acquisition [ELLA] & Project Middle School Science for English Language Learners [MSSELL]–see Lara-Alecio, Tong, Irby, Guerrero, Huerta, & Fan, 2012) that have provided teachers with scripted lessons and professional development. As a result, teachers become further aware and trained on how best to support ELL academic achievement. Therefore, if the TBOP is measuring the observed behavior of the teacher and interactions within the class in the context of Project MSSELL, then the TBOP is estimating the ability the teacher has gained to instruct in relation to supporting ELL academic achievement. Yet, supportive instruction for ELL academic achievement is not only for ELL classrooms in the context of bilingual/English as a second language programs, since studies have also confirmed that ELLs in mainstream classrooms can also be served from instructional methods focused on ELL academic achievement (e.g., COI used in mainstream classrooms–Echevarria, Vogt, & Short, 2012; Lara-Alecio, Tong, Irby, Guerrero, Huerta, & Fan, 2012)

The English Language Learner Classroom Observation Instrument (ELLCOI) is grounded in reading instruction and recent cognitive research on academic learning, this

approach has proved positive findings in capturing the success of reading interventions for ELL (Baker, Gersten, Goldenberg, Graves, & Haager, 1999). However, the ELLCOI's intention to evaluate ELLs on reading instruction only captures the benefits of positively influenced reading interventions. The ELLCOI is crafted to reading instruction and the ability of the ELLCOI functioning outside reading instruction is a concern during math and science accountability measures. The ELLCOI also had elements of sheltered instruction defined by Tikunoff, Ward, van Broekhuizen, Romero, Castaneda, Lucas, & Katz (1991), yet the researchers also expressed their target audience was in grade 1 reading instruction for ELLs. In contrast, the TBOP has been documented in use for every grade levels (e.g., Bruce et al., 1997; Kujawa, Cavazos, Meyer, Rodriguez, Lara-Alecio, Galloway, & Irby, 2001; Lara-Alecio et al., 2012; Rodriguez et al., 2002).

Cross-referencing COIs with other authors

The judgment to designate COIs was also determined by how researchers in the field observed COIs as serving the ELL classroom or not (i.e., COIs determined to be or not be explicitly created for serving the ELL classroom). Raters had to decide whether COIs could be categorized in the following groups: (a) COIs that assessed language and literacy with no mention of ELLs, (b) COIs that explicitly stating how they served the ELL classroom, and (c) COIs that were not explicit to stating how they served ELLs or used to assess language and literacy. The use of the following information sources were used as a cross-reference to see how other researchers were viewing COIs found in this study: (a) CECER–DLL (2011), (b) Halle, Whittaker, and Anderson (2010), (c)

Waxman, Tharp, and Hilberg (2004), and (d) discussions with accessible researchers concerning their COI presented in this study.

Cross-referencing with researchers and their COIs, the following COIs are coined culturally and linguistically diverse by Waxman, Tharp, and Hilberg (2004): SIOP, SPC, COS, TFM, ASOS, SOM, along with the SFA (Success For All) COI and the use of ethnographic and evaluative approaches to serving the ELL classroom. Descriptive information provided by Halle et al. (2010) alludes to the following COIs as related to serving the ELL classroom in one or more capacities: CASEBA, CIRCLE, ELLCO-A, LISn, OMLIT, QUEST, SELA, SSEGA, B-TBRS. While, CECER–DLL (2011) acknowledges the following COIs as related to language and literacy measures: CLEO, ELLCO, E-LOT, ELLE, SELA, and OMLIT.

Although, ethnographic and SFA observation measures were not included these potential systematic measures could be used to serve the ELL classroom as described by Waxman et al. (2004). A COI that should be noted is the DLASO (Rivera & Tharp, 2010), which was identified through accessible researcher contact and the DLASO is considered currently under development (i.e., within grey literature context). The DLASO is considered a dual language learner specific (DLL-specific) instrument with the theoretical framework in sociocultural theory and modeled after the ASOS. Also, the TBOP is worthy of mention, which has appeared in several quasi-experimental interventions focused on literacy and academic achievement among ELLs in content areas and in a variety of classroom settings—such as dual language, transitional bilingual, English as a second language, English as a foreign language, and structured

English immersion (e.g., Breunig, 1998; Gomez, Parker, Lara-Alecio, Ochoa, & Gomez, 1996; Irby et al., 2007; Kujawa et al., 2001; Lara-Alecio, Cmajdalka, Parker, Cuellar, & Irby, 1996; Lara-Alecio & Irby, 1996; Lara et al., 2007, 2009).

In summary, cross-referencing with other researchers whom complied COIs (i.e., CECER–DLL, 2011; Halle et al., 2010; Waxman et al., 2004) also confirmed agreement of COIs in the categories of serving the ELL classroom and language and literacy measures (i.e., the gradual filtering process in Figure 4). COIs considered to serve the ELL classroom were 11 with a possible 9 other COIs. Furthermore, the B-TBRS, CASEBA, DLASO, ELLCO-A, and SELLCA have been classified by other researchers as for Dual language learners (DLLs).

Discussion

In this study a small number of COIs were identified through systematic review protocols with a logic model in place to compile a sample of COIs. Although the constrains of allowing COIs to fall in the realm of serving the ELL classroom are relaxed, 13 COIs were illustrated through Figure 5. It's possible other COIs were neglected because of not having a strong presence in peer-reviewed journals. COIs were not as numerous as expected, however the researcher has illustrated that even less COIs are specifically designed for serving the ELL classroom. This systematic review did come across researcher created COIs, which questions the extensive validity and reliability that should be associated to a new COI. Similarly, methodological quality was evaluated with the MQQ (Methodological Quality Questionnaire), but was not heavily

mentioned due to the MQQ focusing on study quality (i.e., a systematic review protocol) and not an indication of COIs serving the ELL classroom.

Of 13 potential COIs the English Language Learner Classroom Observation Instrument (ELLCOI) appeared the most in literature search results. Reasons for such a pattern, point to the search command for this study. For instance, the search command purposefully sought for COIs that observe ELLs and the ELLCOI fits the search terms very well. Consider, another COI like the Transitional Bilingual Observation Protocol (TBOP), which does not have as many search terms contained in the COI name compared to the ELLCOI. As a result, COIs that go below observation and begin to take note of pedagogy (e.g., the TBOP) are not easily found through search terms presented in this systematic review. Therefore, the TBOP is more likely to not surface from search terms used in this systematic review, thereby resulting in lower accounts of the TBOP in literature. However, if this systematic review merely focused on identifying COIs which explicitly recorded pedagogically events in the ELL classroom, then this systematic review would have resulted in far less COIs and studies to examine.

In this study, one of the earliest COI is the Stallings Classroom Observation instrument with origins from 1970. On the other side of the spectrum, the latest study that captured a COI by the search command was Padrón et al. (2012). However, COIs are constantly developed and used to target specific aspects of implementation or intervention. In fact, attempts to compile listings of COIs that serve the ELL classroom are not always easy. A noble attempt from Waxman, Tharp, and Hilberg (2004) collected seven proficient COIs for serving the ELL classroom in a book.

The classroom can be described as saturated with information, where every action and behavior relay information to the observer. In fact, a tremendous amount of information is present in classroom observation. The task of an observation rater is to record a sample of the total observable classroom. As a result, since the observation pool of information is so great, the researcher needs to determine what is essential to record during the observation session. Considering COIs identified in this study, researchers will differ in what should be recorded and mentioned. After all from research question one, the identification of 13 COIs exhibits thinking of more than 24 researchers of what should be observed while serving the ELL classroom.

There are several measuring attributes listed from the 13 COIs. While every measure is uniquely different from the next, the similarity is contained by the sample (e.g., ELLs). One can only wonder if researchers can harness the observation measures of all COIs into one general service to the ELL classroom. In fact, some COIs in this study have evolved into systems or a battery of observation tools for classroom measure (e.g., the CLASS). The moving progress for COIs to continue toward relevancy in pedagogy for classroom teachers serving ELLs.

Conclusion

The systematic review identified 37 COIs which reported the presence of ELLs in the classroom. However, of the 37 COIs, 13 were determined explicitly to be for observing classrooms in which ELLs are served. The fact that this systematic review demonstrated a few viable COIs available in the context of serving the ELL classroom, continues to highlight the notion that not enough COIs have become apparent to explore

the ELL classroom. Although, the search started simply to identify what observation instruments were available, much more instrument information was provided (i.e., psychometric properties of 13 COIs). For instance, the 37 COIs were displayed to show theory to practice of COIs used in serving the ELL classroom. However, with rater examination the 37 COIs were reduced to 13 COIs fitting within the context of serving the ELL classroom. Further examination of the 13 COIs, would suggest even fewer COIs are centralized at serving the ELL classroom.

The range in observation instruments also illustrates the diversity in measurement oriented in targeting specific variables present in the ELL classroom. While observation instruments have served to bring the classroom behavior into a quantified measure, there are qualitative means of measure (e.g., ethnographic monitoring). In other words, researchers have expressed the use of both research paradigms to explain classroom behavior or even the use of mixed methodologies to create insightful instruments (e.g., Onwuegbuzie et al. 2011). As a result, future classroom observation could incorporate a mixed methods approach toward serving the ELL classroom.

Limitations

Although the search was comprehensive, one area of search left out was COIs used by educational entities outside of literature. There are many educational entities with the focus of teaching ELLs and an interesting follow up to this study would be a search for them. Also, the searching of conference papers was not directly included in the search methods. However, some conference papers did surface. The uncertainty

involved by not directly searching in conference databases (e.g., American Educational Research Association–AERA, Southwest Educational Research Association–SERA) precludes that more studies have a potential to surface. The limitation of grey literature is also a concern, since some instruments start at this level. Yet, observation instruments found in this study represent instruments that are further along in the development stage. Although, some instruments do surface in the literature as new instruments available for researchers.

Also, the MQQ is an attempt to estimate the methodological quality of studies that portrayed COIs. As a result, the MQQ did not evaluate the methodological quality of a COI but rather was part of the systematic review protocol to assess study eligibility. Nor was this study an attempt to demonstrate methodological quality among COIs. This systematic review was intended to identify COIs with the potential to serve the ELL classroom by examining how researchers described their COI. Another methodological limitation is the observance of teachers with ELLs at the pre-K level, because ELLs are not always described as part of the sample, because students at this grade level are usually tested with early literacy instruments rather than specifically created COIs for serving the ELL classroom. The lack of identified studies with ELL samples in the pre-K range may be due to the nature of establishing early literacy status, since young ELLs would not have mastery of English or Spanish.

Implications

Researchers have examined the rationale for using certain COIs while serving the ELL classroom. For example, researchers can observe the numerous classroom

characteristics that each COI can set out to record. In other words, each COI can touch upon different grade levels, content areas, state and regional demographics, ELL populations, instrument data type, social-economic status (SES), measurement, and the classroom setting (e.g., bilingual, dual language, main-stream). However, the most influential person present in the classroom that can play a role in ELL academic achievement is the teacher. Therefore, if the teacher is the main proponent for serving the ELL classroom well, then having access to COIs that observe the teacher's pedagogy and the student's response to such pedagogy would be vital to expanding research in serving ELLs. Researchers, practitioners and policy makers can build from this study by observing the COIs discussed that focused on capturing pedagogical behavior related to serving the ELL classroom.

CHAPTER III

PEDAGOGICAL DIFFERENCES OF EIGHT GRADE 5 TEACHERS BASED ON ARCHIVED DATA FROM THE TRANSITIONAL BILINGUAL OBSERVATION PROTOCOL DURING PROJECT MSSELL

The field of research in teaching effectiveness began with teachers as role models before shifting into teaching aimed at promoting learner outcomes (Creemers, Kyriakides, & Antoniou, 2013). The teacher effectiveness shift continued to what is recognized as value-added modeling as a means to assess teacher effectiveness (Lee, 2012). Additionally, researchers have used value-added models in randomized control trials in order to approximate teacher causal impact (Kane, McCaffrey, Miller, & Staiger, 2013). However, the purpose of this study is to investigate teachers' pedagogical differences during a grade 5 science intervention (i.e., Project Middle School Science for English Language Learners–MSSELL), and specifically to answer the research question–How does the science-literacy intervention (i.e., Project MSSELL) change the pedagogical difference of treatment teachers compared to control teachers.

In this study, the author conducted a frequency analysis on secondary data, which came from a grade 5 science intervention. The original study (i.e., Lara-Alecio, Tong, Irby, Guerrero, Huerta, & Fan, 2012) from where the archived data originated reported the student sample comprised of English language learners (ELLs) in a science classroom setting. Establishing the student sample as consisting of ELLs is vital to highlighting research that has contributed to moving ELL research onward. Similarly, classroom observation with ELLs has the potential to improve educational research

concerning ELLs' academic achievement (Snow, 2002). Since the archived data was based on recorded pedagogical behavior, a chi-squared test of independence was used to determine the pedagogical differences between treatment and control teachers. As a result, the secondary analysis (i.e., analysis of archived data) was used to describe the pedagogical differences of teachers in order to illustrate the beneficial aspects of Project MSSELL to serve the ELL classroom. Although Project MSSELL proved to be effective toward ELL academic achievement (Lara-Alecio, Tong, Irby, Guerrero, Huerta, & Fan, 2012), Transitional Bilingual Observation Protocol (TBOP) data results from Project MSSELL have not appeared in the literature in the form of peer-reviewed publications.

Project MSSELL was a two-year (2009-10) randomized trial study federally funded from the National Science Foundation (NSF Award No. DRL–0822343; Lara-Alecio & Tong, 2013) to improve science achievement and academic English proficiency for grade 5 and 6 students (Lara-Alecio et al., 2012). However, this study was focused on answering the following research questions: what were the pedagogical differences of treatment and control teachers during Project MSSELL; and, what pedagogical differences were displayed from treatment and control teachers as recorded from the TBOP during Project MSSELL? As a result, I analyzed archived data collected from the TBOP for both treatment and control teachers.

Literature Review

Cummins' (1986, 2008) noted the importance of social and academic registers for ELL in the continuum of language development. This continuum distinguished between Basic Interpersonal Communications Skills (BICS; i.e., social language) and

Cognitive-Academic Language Proficiency (CALP; i.e. academic language). With the distinction between BICS and CALP, researchers in the classroom could categorize classroom events and describe ELL serving classrooms in the context of social and academic registers. Additionally, Cummins expanded on primary language of the ELL as important to developing knowledge and skills. His theory in practical terms means "Conceptual knowledge developed in one language helps to make input in the other language comprehensible" (Cummins, 2000, p. 39).

Similarly, science education researchers have noted science has specific registers, that is, linguistic features such as academic language and syntactical structures (Gee, 2005; Lembke, 1990; Norris & Phillips, 2003; Wellington & Osborne, 2001). These science registers, vocabulary, and syntactic structures are especially challenging for English language learners (Ryoo, 2010). As a result, it's beneficial for researchers to acknowledge interventions that have been found to be effective (e.g., publications from What Works Clearinghouse and the National Science Foundation).

Researchers whom have acknowledged specific science registers as well as the social and academic registers related to serving ELLs, have gone to implement effective interventions for serving ELLs (e.g., Luykx, & Lee, 2007; Lara-Alecio et al., 2012). As a result, the importance of reexamining effective interventions like Project MSSELL are needed in order to describe further the inner workings that helped promote positive results (i.e., serving the ELL classroom well). One approach to reexamine for further benefit of the information from Project MSSELL was to gain permission to examine

classroom observation data. The classroom observation data was gathered by raters using the TBOP, a systematic COI used to quantify classroom pedagogical events.

Quantitative Observation

Observation studies with ELLs have gradually contributed to the development of instrumentation to quantify teacher pedagogy in the ELL classroom (Foorman et al., 2004; Foorman & Schatschneider, 2003; Gersten & Baker, 2000; Haager et al., 2003; Irby et al., 2007; Lara-Alecio et al., 2009; Saunders et al., 2006). Additionally, observation studies have brought instructional deficits to light (Ramírez et al., 1991), for instances, an early observation studies with English language learners concluded that instruction for ELLs was not cognitively demanding or interactive during reading instruction (Padrón 1994).

While COIs take different approaches to serving the English language learner classroom, quantitative observation studies have made an impact to ELL research. There are COIs which can be used to record classroom events pertaining to student or teacher behaviors or both teacher-student interaction behaviors (Haager, Gersten, Baker, & Graves, 2003; Lara-Alecio, et al., 2012; Waxman, Tharp, Hilberg, 2004; Waxman, Wang, Lindvall, & Anderson, 1983). Also, there are COIs tied to comprehensive school reform interventions and COIs focused on specific content area instruction that have also provided a means for documenting classroom pedagogical events benecifical to ELLs (Calderón, Slavin, & Sánchez, 2011; Foorman & Schatschneider, 2003, 2004; Irby et al., 2007; Lara-Alecio, et al., 2012; Saunders et al., 2006). In this study, the COI used was the Transition Bilingual Observation Protocol (TBOP), which is an instrument rooted in

pedagogy theory to serve ELLs. Trained raters using the TBOP can describe the pedagogical events occurring in the ELL classroom.

TBOP

The TBOP originally started from theory based on the Four Dimensional Transitional Bilingual Pedagogical Model, which incorporated pedagogical principles of bilingual education (Lara-Alecio et al., 1996). The Four Dimensional Transitional Bilingual Pedagogical Model first appeared in the *Bilingual Education Research Journal* and described four dimensional domains for pedagogy (Lara-Alecio & Parker, 1994). The Four Dimensional Transitional Bilingual Pedagogical Model then appeared in Bruce (1995) as operationalized through field testing and became known as the TBOP. The TBOP was used to measure transitional bilingual classrooms and has expanded to an array of classroom settings serving ELLs. Over time, trained raters using the TBOP have shown the usefulness in being able to describe pedagogical events in different grade levels and across a variety of bilingual education settings (Breunig 1998; Bruce et al., 1997; Gomez et al., 1996; Irby et al., 2007; Kujawa et al., 2001; Lara-Alecio et al., 1996, 1997, 2007, 2009, 2012; Meyer 2000; Rodriguez et al., 2002). In fact, the TBOP has also been used in mixed settings (i.e., classrooms with ELLs and non-ELLs; Lara-Alecio et al., 2012) and adapted for multicultural education (Lara-Alecio & Irby, 1996). The descriptions surrounding each TBOP pedagogical code are illustrated in Appendix B.

In order to determine how well the ELL classroom was served based on TBOP archived data, the following needed to be addressed: (a) The context of classroom instruction; (b) observation studies with ELLs; and (c) the classroom observation

instrument or TBOP (Transitional Bilingual Observation Protocol). Since raters use a classroom observation instrument (COI) to record what occurred during science instruction, an account of ELL academic learning during science is described. For specific academic achievement results during Project MSSELL, the reader is encouraged to see Lara-Alecio et al. (2012). Although, the TBOP is not a specific COI geared toward science instruction, the protocol is used by raters to record the teacher pedagogy oriented toward serving ELLs. In addition, the TBOP has a history of development and field testing which has lead to COI validity and reliability in different settings (Breunig, 1998; Gomez et al., 1996; Irby et al., 2007; Kujawa et al., 2001; Lara-Alecio & Irby, 1996; Lara-Alecio et al., 1996, 2007, 2009).

Method

In this study, I analyzed archived data from Project MSSELL 2009–2010 with the permission of the Project MSSELL research team. The archived data I received was from Project MSSELL raters who recorded observational events using the TBOP. As a result, I examined the pedagogical events raters recorded when they used the TBOP. Because raters recorded pedagogical events for both treatment and control teachers, I was able to examine the pedagogical differences between treatment and control teachers. Serving as the unit of measurement, 1,966 observations were collected during Project MSSELL with the TBOP across four middle schools and eight teachers. Data collected with the use of the TBOP, was from grade 5 teachers where four were treatment teachers and the other four were control teachers.

The study consisted of two parts focused on observing teachers' pedagogy during science instruction with the use of the TBOP. The first part was a descriptive account of archived data from the TBOP depicting the treatment and control teachers separately. The second part was a comparison between treatment and control teacher pedagogy. A chi-squared test of independence provided the numerical difference between treatment and control teachers.

Project MSSELL

Project MSSELL was a quasi-experimental study conducted during the 2009–2010 school year and incorporated the use of treatment and control teachers (Lara-Alecio et al., 2012). The teachers that participated in Project MSSELL were selected because of randomized selection by classrooms since Texas Education Code (1995) disallowed the random selection of individual students. As a result and with permission from principals, two schools were randomly assigned to the treatment condition and two schools to the control condition. Furthermore, with schools assigned to treatment or control conditions, teachers from those schools were then randomly assigned to their school's condition.

Described in Lara-Alecio et al. (2012), ELLs and low socioeconomics (SES) non-ELLs from the same school both received the same practice in order to diminish cross contamination between treatment and control teachers. Treatment teachers then taught the enhanced science practice, while control teachers taught the typical science practice. However, due to a low number of teacher participants needed for the Project MSSELL study additional teachers were recruited to complete the teacher sample. As a

result, the non-random method to complete the teacher sample led to a quasi-experimental research design (Lara-Alecio et al., 2012).

Archived Data Sample

The archived data from Project MSSELL was from the fall and spring 2009–2010 academic year and was collected from the use of the TBOP. The TBOP recorded pedagogical data from the classroom of eight teachers. Teachers were in two conditions, four teachers in the treatment condition and four teachers in the control condition. There were 1,966 recorded observations from using the TBOP, which came from the classroom observation of eight teachers across four schools. The 1,966 recorded observations was the total amount of observations conducted after three rounds of observation recorded with the use of the TBOP. Meaning, recorded observations are from the beginning, middle and end of the 2009–2010 school year of grade 5 teachers.

Additionally, the research design was a 2 x 2 design as follows: two schools with four teachers (n = 200 students) in a Science Enhanced Program (SEP) and two schools with four teachers (n = 200 students) in a Science Typical Program (STP). In Table 5, the school demographics for treatment and control conditions are illustrated. The similarity among the schools in their demographic characteristics justified the comparisons of the two conditions (i.e., treatment and control).

Table 5

School Demographics of Treatment and Control Conditions 2009-2010

Group	African Am. (%)	Hispanic (%)	White (%)	Native Am. (%)	Asian (%)	Low SES (%)	ELL (%)	Academic Rating
Treatment								
School 1	19.5	78.3	1.5	0.1	0.6	92.9	30.4	Recognized
School 2	11.9	84.1	2.4	0.0	1.6	92.1	31.3	Exemplary
Control								
School 1	26.8	68.4	2.4	0.0	2.5	88.0	31.5	Recognized
School 2	26.5	67.7	5.0	0.2	0.7	90.5	22.9	Recognized

Note. Am. = American. Data collected from Lara-Alecio et al. (2012).

Teachers taught science for 90 minutes and had classrooms where Hispanic English Language Learners (ELLs) and non-ELL minority students from low socioeconomic backgrounds (i.e., low-SES) were present. The students were from a large urban school district in Southeast Texas. Additionally, the school district served a diverse student population: 46% of the students in the district were native Spanish speakers and 85% of the students qualified for free or reduced lunch (TEA, 2010).

Observation Instrument

The Transitional Bilingual Observation Protocol, or TBOP, is a classroom observation instrument derived by the transitional bilingual model (Lara-Alecio & Parker, 1994). Raters using the TBOP can record four pedagogical practices which are coded (i.e., Activity Structure, Communication Mode, Language Content and Language

of Instruction) into observed classroom events. Validated and field tested in different settings, the TBOP has been used in multiple observation studies and research grants (Breunig, 1998; Gomez et al., 1996; Irby et al., 2007; Kujawa et al., 2001; Lara-Alecio & Irby, 1996; Lara-Alecio et al., 1996, 2007, 2009).

There are 48 pedagogical behaviors coded by the TBOP (e.g., see Appendix B)– 21 behavioral codes to capture an activity structure type, 18 behavioral codes to capture a communication mode type, 5 behavioral codes to capture a language content type, and 4 behavioral codes to capture a language of instruction behavior for the teacher and student. The combination of the 4 instructional domains becomes a means for correct implementation of the Project MSSELL intervention. Meaning, treatment teachers would display pedagogy conducive to English language learners achieving gains in English language proficiency and science literacy. Such intervention requirements would also play a role in fidelity or making sure treatment teachers were following Project MSSELL protocol. The TBOP viewed as an implementation instrument can show the pedagogical differences between Project MSSELL enhanced science instruction compared to mainstream classroom instruction in science (e.g. Science Typical Program).

Data Analysis

This study used archived data collected during Project MSSELL science instruction for grade 5 students. The archived data was in the context of serving English language learners (ELLs) and classroom observation with the intent to measure ELLs has the potential to improve education research concerning ELL academic achievement (Snow, 2002). Since the archived data could illustrate the frequency of pedagogical

events during the ELL classroom, a chi-squared test of independence was used to determine the pedagogical differences between treatment and control teachers. As a result, the frequency of coded pedagogical behavior in the ELL classroom was used to illustrate the pedagogical difference between treatment and control teachers.

The archived data in this study had a nominal scale, meaning each category of the TBOP is mutually exclusive and identical in dispersion from the mode. Such nominal level of scale represented in the data only allows for counting (Thompson 2006). The predictor variable was categorical or binary and assessed by either treatment or control condition of the Project MSSELL intervention. Project MSSELL was a comprehensive reform intervention demonstrating higher science academic achievement scores from English language learners (Lara-Alecio, Tong, Irby, Guerrero, Huerta, & Fan, 2012), and this study looked at teacher pedagogy during Project MSSELL.

Treatment and control teacher pedagogy recorded from the TBOP was collected by raters during Project MSSELL and analyzed with SPSS 21. The TBOP data is coded by four simultaneous instructional behaviors during science instruction–Activity Structure, Communication Mode, Language Content, Language of Instruction of the teacher and student. One classroom observation from the TBOP consisted of 60 recorded entries conducted by rater during science instruction, and each entry was a 20-second observation, suggesting that each observation measure of the TBOP occurred in short periods of time to ensure high levels of reliability (Rowley, 1978).

Results

The 1,966 observations equated to an amount of time exceeding 655 minutes or 10 hours of recorded instruction time across both treatment and control teachers. The average time per TBOP observation per teacher was 27.31 minutes. However, each teacher was observed three times during the school year which totaled the observation time to 81.92 minutes per teacher. During the average 81.92 minutes per teacher, Project MSSELL raters used the TBOP to record both treatment and control teachers' pedagogy to serve ELLs in the classroom. The average amount of time per teacher (i.e., 81.92 minutes) was also the amount of time used by classroom observation raters to systematically categorize what they were seeing in the classroom based on TBOP coded descriptions within each domain (i.e., Activity Structure, Communication Mode, Language Content and Language of Instruction). As a result, time used by classroom observation raters to code classroom events based on the TBOP was equal across all domains.

Although classroom observation raters used the same amount of time to code occurring events from each TBOP domain, differences became apparent when treatment and control teachers were compared. For instance, the central tendency of categorical variables were: pedagogical behavior pertaining to Activity Structure among treatment and control teachers were both observed at mode value equaled to 11, Communication Mode treatment teachers were observed with a mode of 4 and control teachers were observed with a mode of 16, for the Language Content domain treatment teachers were observed as having a mode of 4 while control teachers displayed a mode of 3. For the

pedagogical behavior of Language of Instruction both treatment and control teachers were observed as having a mode of 2. The most apparent difference between treatment and control teachers can be found between descriptive statistic values for Communication Mode and Language Content. For example, the mode between treatment and control teachers for Communication Mode was coded pedagogical behavior 4 versus 16. The difference in Communication Mode indicated that the treatment teachers had more verbal description in the classroom compared to verbal-aural communication from control teachers. Whereas treatment teachers were observed to have more dense cognitive occurrence compared to control teachers (i.e., that is mode of 4 in Language Content compared to 3).

Each TBOP domain has its own descriptions for the rater to choose the correct pedagogical behavior that best describes classroom events. For instance, the amount of coded TBOP descriptions for each domain is: Activity Structure is coded from 1 to 21, Communication Mode is coded from 1 to 18, Language Content is coded from 1 to 4 and Language of Instruction is coded from 1 to 5 (i.e., see Appendix B). Represented in Table 6 are the coded TBOP descriptions that showed the most pedagogical difference between treatment and control teachers.

In the domain of Activity Structure Table 6 shows that the control teachers (i.e., the science typical program) were observed to have more occurrences of the teacher lecturing while students listened as well as the teacher directing while students listened. Control teachers were also observed to provide direction while the student was to perform. However, in treatment teachers were observed to ask more questions from

which students responded. Additionally, the treatment teachers observed how well the students were performing. In Table 6, the last thing to mention of Activity Structure is the amount of interruption. Because control teachers were observed to have more interruptions compared to treatment teachers.

Table 6

Descriptive Statistics in Interaction and Selected Main Effects between Treatment and Control Groups

			Program Model	
Domain	Level		SEP	STP
Activity Structure	Lectures/listens	Count	75	224
		%	11.5%	17.1%
	Directs/listens	Count	67	163
		%	10.2%	12.4%
	Directs/performs	Count	69	147
		%	10.6%	11.2%
	Asks/answers	Count	88	134
		%	13.5%	10.2%
	Observes/performs	Count	143	266
		%	21.9%	20.3%
	Not applicable–interruption	Count	5	70
		%	0.8%	5.3%
Communication Mode	Writing	Count	158	246
		%	24.2%	18.8%
	Verbal	Count	221	338
		%	33.8%	25.8%
	Verbal-writing	Count	47	37
		%	7.2%	2.8%
	Verbal-reading	Count	15	3
		%	2.3%	0.2%
	Verbal-aural	Count	194	599
		%	29.7%	45.7%
	Not applicable	Count	1	44
		%	0.2%	3.4%
Language	Academic routines	Count	35	222

Table 6 continued

			Program Model	
Domain	Level		SEP	STP
Content		%	5.4%	16.9%
	Light cognitive	Count	169	789
		%	25.8%	60.1%
	Dense cognitive	Count	430	281
		%	65.7%	21.4%

Note. Illustrated are TBOP codes which displayed the largest difference between treatment and control teachers. SEP is science enhanced program and STP is science typical program.

From Table 6, coded pedagogical results continued to illustrate a difference between the treatment and control teachers as examination moved to the area of Communication Mode and Language Content domains. Writing as a mode of communication occurred more among treatment teachers. In fact, there was a lot of verbal activity recorded more so among treatment teachers. For instance, verbal, verbal-writing and verbal-reading were recorded higher pedagogical behaviors among treatment teachers compared to control teachers. Where verbal behavior was recorded higher among control teachers compared to treatment teachers was in verbal-aural, which is a combination of verbal and listening communication. As for the Language Content domain, control teachers were recorded as displaying more academic routines and light cognitive activities. Compared to control teachers, treatment teachers were observed to display more dense cognitive activities.

The frequency analysis of archived data from Project MSSELL was used to illustrate several differences between treatment and control teachers' pedagogy in

Activity Structure, Communication Mode, Language Content and Language of Instruction. For instance, there was a small cluster of five Activity Structure types, which occurred more often compared to the other 21 possible codes for Activity Structure for both treatment and control teachers. Generally, the Activity Structure cluster represented teacher-centered techniques. However, the teacher's consistent behavior to observe the student's performance at a given task is the highest observed Activity Structure event in the classroom for both treatment and control teachers. Percentage values for Activity Structure are different between treatment and control teachers. For instance, comparing treatment and control teacher pedagogical percentage results, the Activity Structure coded behavior that stood out the most were 3 (directs/listens), 4 (directs/performs), 6 (leads/performs), 8 (asks/answers), 11 (observes/performs), with codes 18 (Not Applicable–transition), and 19 (Not Applicable–interruption) not too far away. Among treatment teachers behavioral codes 4, 6, 8, and 11 were recorded as exhibiting higher percentages compared to control teachers. Behavioral codes 3, 18 and 19 among control teachers were also observed having higher percentage results.

Recorded pedagogical behavior in Communication Mode tended toward verbal–aural compared to 17 other possible categories (e.g., which provide an indication of student–teacher interaction during science instruction) among control teachers. Verbal–aural means that TBOP observations were recording a lot of speaking and listening among control teachers. In comparison, treatment teachers were observed to display high frequency events of verbal communication that was more positive in serving the ELL

classroom. Generally, pedagogical behavior pertaining to Communication Mode percentages clustered around codes 1 (writing), 4 (verbal) and 16 (verbal-aural). The control teachers had more code 16 communication during science instruction. Verbal-aural communication is heavily depended on lecture based instruction. However, pedagogical behavior pertaining to Communication Mode percentages for the treatment teachers were observed to concentrate more on behavioral code 1 and 4 compared to control teachers. The communication focus among treatment teachers of Project MSSELL indicates non-lecture based instruction.

Pedagogical behavior pertaining to Language Content was observed having a tendency toward light cognitive events (i.e., naming, eliciting information) among control teachers (e.g., typical science instruction). As intended, the treatment teachers were observed to display more frequency among cognitively challenging task in the classroom because of more pedagogical behavior recorded in Language Content code 4 (i.e., dense cognitive). The pedagogical behavior of Language Content was observed to show a majority on code 3 (i.e., Light cognitive) among control teachers. Light cognitive meant students were exposed to such academic tasks as repetitive drill or skill practice, reviewing content already introduced, and current events. In other words, Project MSSELL prepared treatment teachers to encourage and challenge students to use their cognitive abilities.

Pedagogical behavior pertaining to Language of Instruction was entirely observed in remain in the second language among control teachers, which indicates English-only instruction for science. The same pedagogical behavior was observed

among treatment teachers because teaching science in English is common instructional practice for both ELLs and non-ELLs. As a result, pedagogical behaviors pertaining to Language of Instruction is generally not mentioned because of the need to stay in the English language during science instruction. However, there are instances of language support worth noting that played a role in serving the ELL classroom among treatment teachers during Project MSSELL. For example, treatment teachers could provide first language clarification (i.e., Spanish) of science terms or concepts during instruction if they felt ELLs in the classroom could gain better understanding by providing language clarification. Generally, the archived data from Project MSSELL did not show large pedagogical differences pertaining to Language of Instruction compared to the other instructional domains–Activity Structure, Communication Mode and Language Content.

Further pedagogical comparison of treatment and control teachers were examined for chi-squared differences in Activity Structure, Communication Mode and Language Content domains. The reason for not including the pedagogical behavior of Language of Instruction was because there was no expectation for change in language during science instruction since both treatment and control teachers used English for instruction. The similarity of pedagogical behavior pertaining to Language of Instruction for both treatment and control teachers would statistically show no difference through chi-square analysis. In Table 7, the chi-square differences between treatment and control teachers' pedagogy as well as Phi and Cramer's V statistics demonstrate that there was a distinct difference between treatment and control teachers. The results in Table 7 show that treatment teachers during Project MSSELL were displaying pedagogy as coded by the

TBOP completely different from control teachers. The treatment teacher having a pedagogically different approach to serving ELLs in the classroom would have played a role in success pertaining to Project MSSELL (e.g., see Lara-Alecio et al., 2012).

Table 7

Chi-Squared Difference for Treatment and Control Groups

Domain	Chi-Square[1]	Phi	Cramer's V	*p*-values
Activity Structure	107.9	.234	.234	<.01
Lectures/listens	25.83	.885	.885	<.05
Directs/listens	19.98	.778	.778	<.05
Directs/performs	23.55	.845	.845	<.05
Asks/answers	20.82	.794	.794	<.05
Observes/performs	22.54	.826	.826	<.05
Communication Mode	122.0	.249	.249	<.01
Writing	19.92	.777	.777	
Verbal	28.42	.928	.928	<.05
Verbal-writing	17.08	.719	.719	<.05
Verbal-reading	6.49	.443	.443	
Verbal-aural	27.96	.920	.920	<.05
Language Content	392.3	.447	.447	<.01
Academic routines	26.95	.904	.904	<.05
Light cognitive	29.64	.948	.948	<.05
Dense cognitive	33.00	1.00	1.00	<.05

Note. [1] is the Pearson Chi-Square value. Language of Instruction was not included because English was the dominant language across both treatment and control groups. *p*-values were not reported for values greater than .05

Also in Table 7 are the chi-square differences between the pedagogical differences of coded behavior pertaining to each instructional domain. For instance, the pedagogical differences between treatment and control teachers in Activity Structure coded behaviors 1 (lectures/listens), 3 (directs/listens), 4 (directs/performs), 8

(asks/answers) and 11 (observes/performs) are statistically significant. As for pedagogical behavior pertaining to Communication Mode the pedagogical differences between treatment and control teachers has proven statistically significant in coded behaviors 4 (verbal), 14 (verbal-writing) and 16 (verbal-aural). Furthermore, pedagogical behavior pertaining to Language Content was observed to be statistically significant for behavior codes 2 (academic routines), 3 (light cognitive) and 4 (dense cognitive) as distinct differences between treatment and control teachers.

Discussion

Treatment teachers were observed to display more time to pedagogical behavior pertaining to Activity Structure codes 4 (directs/listens), 6 (leads/performs), 8 (asks/answers), and 11 (observes/performs) and may be due to their invested commitment to the Project MSSELL intervention. Additionally, the pedagogical behavior pertaining to Communication Mode during Project MSSELL was observed to demonstrate a dramatic change in communication in the classroom. In fact, frequency counts among treatment teachers were strongly observed to tend toward writing and verbal as the dominant forms of communication in the classroom. Adding to more pedagogical behavior, coded behavior in Language Content was observed to have percentages which implicated control teachers were spending more time in light cognitive activities in the classroom compared to treatment teachers who were observed spending more time in dense cognitive activities. The treatment teachers followed Project MSSELL guidelines and displayed more time to expose students to opportunities for cognitively challenging tasks during science instruction.

There was also a clear distinction between treatment and control teachers concerning each pedagogical domain (i.e., Activity Structure, Communication Mode, Language Content and Language of Instruction). In each pedagogical domain, treatment teachers were found to have a statistically significant difference between the pedagogy of control teachers. Only the pedagogical behavior of Language of Instruction was not considered since English played a dominant role in science instruction. Chi-square difference results further illustrated the discrepancy between treatment and control teachers to serve ELLs. As a result, treatment teachers displayed pedagogy in line with Project MSSELL's research design to serve ELLs.

During Project MSSELL, the TBOP was used in several roles such as to record teachers' pedagogy, as a fidelity check and implementation tool. Treatment teachers received professional development during the entire length of Project MSSELL, yet professional development was used to reinforce pedagogical behavior deemed conducive to serving ELLs in the classroom. As a result, it was essential to make the distinction between pedagogy from treatment and control teachers. Control teachers were to use anything they wanted to shape their pedagogy to ELLs in their classroom. On the other hand, treatment teachers were given professional development during Project MSSELL which they were later observed through the use of the TBOP in order to observe their pedagogical behavior. Therefore, the TBOP was used to make the clear distinction of pedagogical behavior that served the ELL classroom.

For this study, the reason for not using the TBOP as a repeated measure throughout the school year was mostly due to cost efficiency. Planning incorporated with

cost analysis could have led Project MSSELL researchers to use the TBOP at three strategic time periods during the school year in order to optimize cost and resources. In fact, Project MSSELL researchers have stated that the TBOP was used three times during the school year from verbal and email communications. Program fidelity checks still occurred during Project MSSELL and would have resulted in surface level evaluations to ensure all teachers were aligned to intervention objectives.

The pedagogical behavior of control teachers would have varied more so than treatment teachers, only in the sense that among control teachers there was no opportunity to unify their pedagogy among themselves. For instance, treatment teachers had the opportunity to align their pedagogy to Project MSSELL's research design and overall objectives to serving the ELL classroom. Because Project MSSELL worked closely with treatment teachers, there is reason to believe the intervention unified treatment teachers in a shared goal to serve ELLs. As a result, statistically significant results were observed between treatment and control teachers in determining the difference between them. Treatment teachers were observed to serve the ELL classroom better than control teachers based on pedagogical behaviors recorded with the TBOP. Therefore, further replication of the teacher professional development provided to treatment teachers during Project MSSELL is warranted because of the beneficial pedagogical difference in serving the ELL classroom observed among treatment teachers.

Conclusion

Project MSSELL's teacher professional development can be used to inform the pedagogy of teacher instruction by promoting further English proficiency and science literacy to ELLs and non-ELLs. Evidence is further constructed in this study toward the magnitude of what Project MSSELL did at the instructional level (e.g., Lara-Alecio et al., 2012). Researchers, practitioners and policy makers can observe the successful gains of ELLs in Lara-Alecio et al. (2012), but what this study contributed was further description of the Project MSSELL intervention at the instructional level. Evidence for Project MSSELL at the instructional level in this study, demonstrated that teacher pedagogy has the potential to align with top-down pedagogy and unify across teachers. The conducted frequency analysis from TBOP data helped to demonstrate treatment teachers as having different pedagogy compared to control teachers. Consider, control teachers' had the potential to display further variation in pedagogy because they lacked teacher professional development to develop pedagogy that can serve the ELL classroom. An intervention like Project MSSELL could unify teachers into shaping pedagogy locally in order to address the needs of ELLs in their school district.

Limitations

The limited archived data may have affected the adequate comparison of treatment and control teachers in order to gain insight into how much gain was expressed by teachers' pedagogy. There were three rounds of TBOP observations conducted during Project MSSELL and more observations conducted could have provided more information. However, this study reported count and percentage values in order to

illustrate differences between treatment and control teachers despite sample concerns. In combination with chi-square difference testing, this study made attempts to illustrate the distinct difference between treatment and control teachers.

This study examined archived data from Project MSSELL. Archived data also has limitations in that the examination of what happened during the intervention has already passed. As a researcher, it was necessary to ask questions and probe into the surrounding information of Project MSSELL and the information that was recorded from the TBOP. The result is an outsider looking in, which requires further learning the context of which the intervention occurred. Because this study reflects on the past, there is a disconnection with the actual time when the study occurred. However, attempts were made to gain further information from Project MSSELL researchers surrounding the use and coded behaviors of the TBOP.

CHAPTER IV

CAUSAL COMMENTARY OF TEACHERS' PEDAGOGY DURING PROJECT MIDDLE SCHOOL SCIENCE FOR ENGLISH LANGUAGE LEARNERS

The population of English language learners (ELLs) is increasing (Aud, Hussar, Johnson, Kena, Roth, Manning, Wang, & Zhang, 2012), yet education for ELLs has not reached comparable levels of academic achievement compared to mainstream students (e.g., native English students). However, researchers have demonstrated success in improving academic achievement among ELLs (e.g., Tong, Irby, Lara-Alecio, & Mathes, 2008) through the use of randomized control trials (RCT). Generally, RCT models are recognized as being the gold standard for evidence in determining the effectiveness of an intervention. Additionally, when making causal inferences a RCT model is preferable. In fact, RCT models use an experimental design, considered the only design to answer causal questions (Thompson, Diamond, McWilliam, Snyder, & Snyder, 2005). However, other research designs can suggest causal effects (Thompson, 2006).

The gold standard elevates specific methodology and design (e.g., randomized control trials) as having more evidence toward an intervention believed to work. Yet, such restrictions of research design or methods limit the researcher to quantitative research. However, casual inference as described by Pearl (2009) serves as another method to establish a bridge between quantitative and qualitative research in order to make causal inference from different research designs. In comparison, there are growing trends in quantitative and qualitative research that validity and by implication quality is

not a matter of designs or methods, but of inferences (Briggs, 2008; Brinberg & McGrath, 1985; Messick, 1994; Mishler, 1990; Seale, 1999; Shadish, Cook, & Campbell, 2002).

There are few RCT interventions in literature on improving the academic achievement of ELLs that follow the gold standard (Thordardottir, 2010). The gold standard is descriptive of RCT use and highlights the mark of methodological quality for evidence, which in turn is believed to indicate an intervention of most likely the reason for producing the outcome result. Yet, not all researchers are able to conduct RCT interventions in the field of bilingual education (i.e., the specific ELLs context of this study). However, there is a growing trend spanning across quantitative and qualitative research that designs or methods are not necessarily the root of inference (Briggs, 2008). Methods and design can aid inference but are not necessary for making inferences (e.g., experimental design).

This study sought to examine an intervention that illustrated positive results toward increasing ELL academic achievement, especially in science (i.e., Project Middle School Science for English Language Learners [MSSELL]; Lara-Alecio & Tong, 2013; Lara-Alecio, Tong, Irby, Guerrero, Huerta, & Fan, 2012). By reasoning for causal structure in relation to the Project MSSELL intervention and teachers' pedagogy, the benefit was insight into a causal structure for pedagogy. Therefore this study sought to answer the following research questions–Using Pearl's (2009) causality framework, how can a causal structure for teachers' pedagogy from Project MSSELL be conceived, and what would be the structure's appearance?

Conceptual Framework

Pearl (2009) described causality as beginning with three variables in a relationship as the initial construction for causal reasoning. Figure 6(a), illustrates three variables X, Y, and Z in a relationship with no directional arrows to indicate causality. The relationship of X to Y and Y to Z in Figure 6(a) is depicted by a black circle connected to another black circle by a black line. However, the black circles represent a node or the set of values for each X, Y, and Z. Therefore, the lines connecting one node to another node indicate the relationship. Efforts to establish causality from one variable to another is then represented by directional arrows. For instance, the directional arrows in Figure 6(b) mean there is a casual relationship from X to Z and Y to Z.

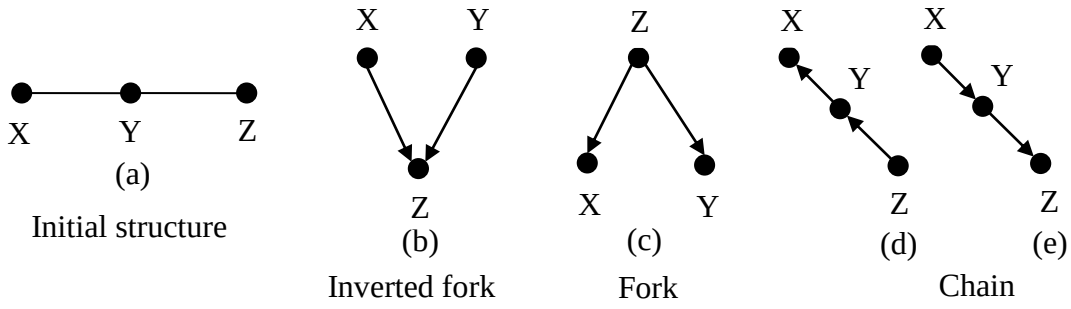

Figure 6. Initial causal structure as described by Pearl (2009).

Additionally, X and Y are said to be the parents of Z according to Pearl (2009) in Figure 6(b). Parent variables serve to provide information about the proceeding variable(s), such as Z being the parent to X and Y in Figure 6(c). The parent(s) of a

variable can be unknown if the parents of Z are in question from Figure 6(c); in other words, there is no information about how Z came into the picture in Figure 6(c). Furthermore, Pearl (2009) described the parent structure for variables X, Y, and Z in three distinct structures illustrated in Figure 6 as the inverted fork (b), the fork (c), and the chain (d). Although Figure 6(d) illustrates one possible structure of the chain, the casual direction of variables X, Y, and Z can flow in the opposite direction as in Figure 6(e).

Method

The Project MSSELL model can be described as a comprehensive school reform (CSR) intervention. Applying causal reasoning as described by Pearl (2009), the reasoning that followed converted the Project MSSELL model into the language needed for causal commentary. Observing the Project MSSELL model in Figure 7, treatment teachers are influenced by a multitude of variables (i.e., the overlapping circles each contributing to one another). In other words, in Figure 7, each variable overlaps with other variables indicating that each variable has an association to every other variable in the model. Therefore, the model demonstrates that all variables are vital for successful intervention outcomes. However, Pearl (2009) argues that making causal statements will come from local structure rather than globally. In other words, the causal implications of Project MSSELL are nestled within its global model depiction in Figure 7.

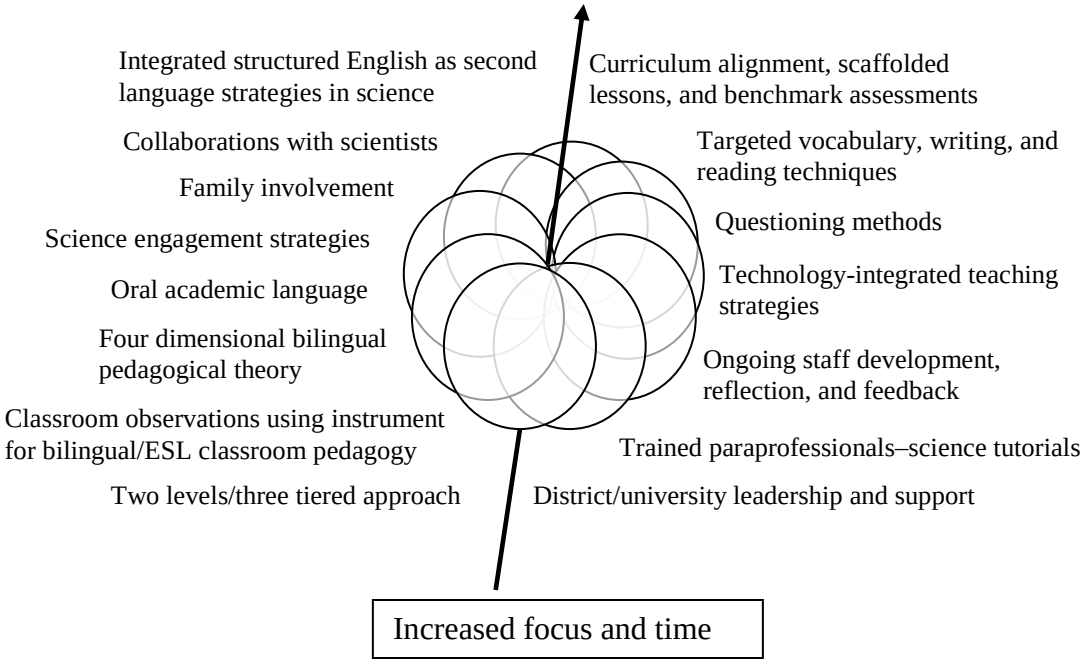

Figure 7. Project Middle School Science for English Language Learner Model, recreated from Lara-Alecio and Irby (2010)

Although Figure 7 illustrates all the variables involved in Project MSSELL, the intent is to focus on deconstructing the model from a three-dimensional figure to a two-dimensional figure in the context of Pearl (2009) casual reasoning. As a result, specific details about what each variable entailed in Figure 7 were referenced to Lara-Alecio et al. (2012). Continuing the deconstruction of Project MSSELL into a causal two-dimensional model, the researcher reasoned each variable as from a set of variables (i.e., a node). For example, in Figure 8, when each variable is converted into a node with relationships to every other node, the researcher translated the Project MSSELL model

into a complex causal model. In other words, numerous parent structures are represented in Figure 8 visually demonstrating the complexity of Project MSSELL.

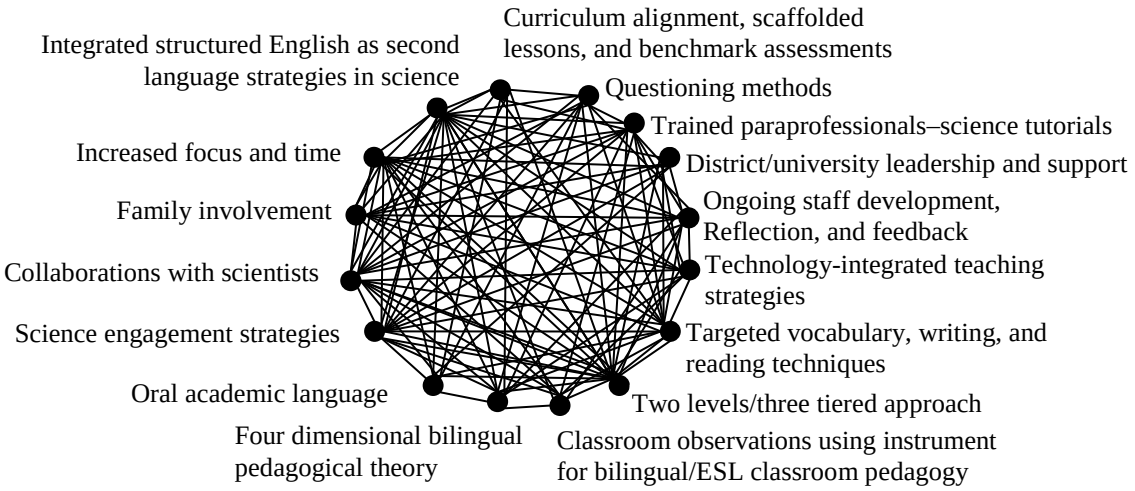

Figure 8. Project MSSELL model converted into numerous parent structures

The conversion from Figure 7 to 8 is reasonable, because Figure 8 still illustrates all variables overlapping each other by the depiction of numerous black lines connecting to other black circles. The difference between Figure 7 and 8 is that variables in Figure 8 were converted to nodes representing the set of all values associated to each variable from Figure 7. The black lines and circles in Figure 8 provide further information of the intricate and complex potential to discover casual relationships between nodes within the model. For the sake of visual clarity, the relationships between all variables (i.e., each

variable comes from a node) is not shown in Figure 8 because of the numerous relationships that exist in Project MSSELL. If Figure 8 did show all Project MSSELL variable relationships, then all 16 nodes would have connecting lines to every other node (e.g., 15 relationships per node). Figure 8 is the beginning to uncovering the causal relationships hidden in Project MSSELL.

Therefore, Figure 8 is an application derived from the Pearl (2009) causal framework. However, the focus was on teachers' pedagogy, the model displayed in Figure 8 needed further deconstruction to target causal structure related to teachers' pedagogy since no directional arrows are displayed. The process to derive a simple structure involved taking the 16 variables described in Project MSSELL and reassigning them into parent variables X and Y. The parent variables X and Y represent the pedagogy as designed by Project MSSELL teacher professional development and student instructional interventions. The reason for the reclassification of Project MSSELL variables was because a number of variables were reasoned to belong to the same node or set of values. For example, in Table 8, Project MSSELL variables were reclassified according to the two level approach described by Lara-Alecio et al. (2012). The two level approach grouped several Project MSSELL interventions at the teacher and student level. The reclassification of variables to either of the two levels (i.e., a node) helped to reduce the number of parent structures down to a specific area of interest, pedagogy.

Table 8

Project MSSELL Variables Reclassified

AV	Intervention component	Assigned relationship
1	Integrated structured English as second language strategies in science	TPD
2	Curriculum alignment, scaffolded lessons, and benchmark assessments	TPD, SII
3	Questioning methods	TPD
4	Trained paraprofessionals–science tutorials	SII
5	District/university leadership and support	TPD, SII
6	Ongoing staff development, reflection, and feedback	TPD
7	Technology-integrated teaching strategies	TPD, SII
8	Targeted vocabulary, writing, and reading techniques	TPD
9	Two levels/three tiered approach	TPD, SII
10	Classroom observations using instrument for bilingual/ESL classroom pedagogy	TPD
11	Four dimensional bilingual pedagogical theory	TPD
12	Oral academic language	TPD
13	Science engagement strategies	TPD
14	Collaborations with scientists	SII
15	Family involvement	SII
16	Increased focus and time	TPD, SII

Note. AV = an assigned value to Project MSSELL variables, which will be used in Figure 11 and 12. X = TPD = Teacher professional development; Y = SII = Student instructional interventions.

Project MSSELL was not based on exploratory means of shaping the teachers' pedagogy for the benefit of serving ELLs; rather Project MSSELL researchers used the Teacher Professional Development (TPD) and Student Instructional Interventions (SII) based on research and evidence of what works. The reassignment of variables to teacher and student level approaches help classify 8 variables to TPD, 3 variables to SII, and 5 variables related to both. As a result, Table 8 demonstrated the reasonable assignment of variables into two nodes (e.g., TPD, SII). The author's judgment to categorize all Project MSSELL variables into two nodes, relates to how the variables are described in the

literature (i.e., Lara-Alecio et al., 2012). In fact, several presentations on the Project MSSELL model illustrate that there are two levels (Lara-Alecio & Irby, 2010; Lara-Alecio, Irby, Tong, 2011; Lara-Alecio, Irby, Tong, Francis, Rodriguez, Guerrero, & Mansfield, 2010). The two levels are defined as Teacher Professional Development and Student Instructional Interventions. Therefore, for clarity X is defined as Teacher Professional Development (TPD) and represents the set of TPD values; and, Y is defined as Student Instructional Interventions (SII) and represents the set of SII values.

In returning to the conceptual framework described in Figure 6, the causal structure of interest (i.e., pedagogy) was extracted from Project MSSELL through reasonable model deconstruction. Figure 9 visually summarized the process of applying Pearl (2009) causal reasoning principles in order to discuss the parent structure related to pedagogy. The method first started with the Project MSSELL three-dimension figure, which was then converted to a two-dimension figure. Variables from the three-dimension figure were then depicted as nodes in the two-dimension figure. Then, in order to develop causal structure for pedagogy, all nodes from the two-dimension figure were viewed as variables and reclassified as the set of all values (i.e., nodes were redefined) related to the following parents: (a) teacher professional development and (b) student instructional interventions. As a result, the parent structure for pedagogy became a visible structure of commentary and displayed in Figure 9.

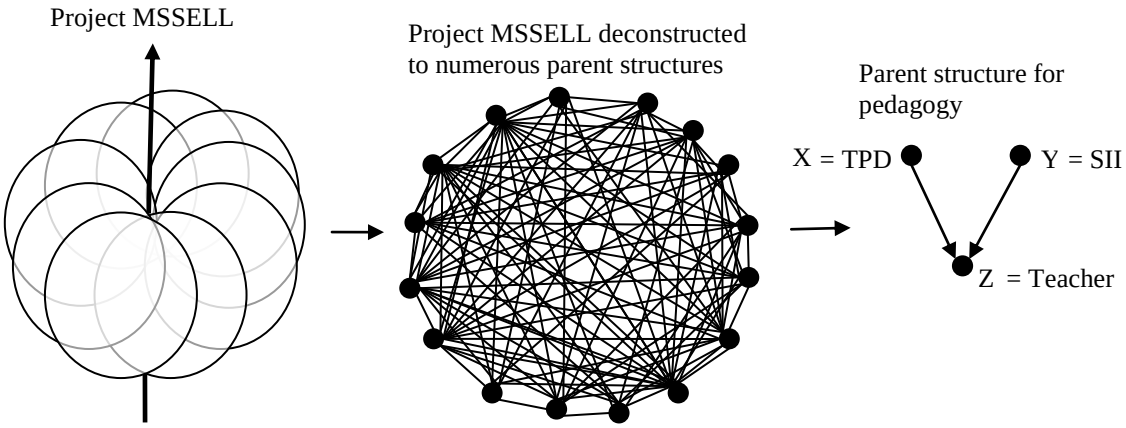

Figure 9. The process to reduce Project MSSELL to causal structure for pedagogy

The parent structure illustrated in Figure 9 is an inverted fork with TPD and SII as the parents of Z. In this case, Z was the teacher because Project MSSELL provided professional development and student interventions that would influence the pedagogy of the teacher. As the teacher (i.e., Z) was observed for pedagogy, then the pedagogy had unique characteristics that resembled Project MSSELL treatment and control conditions. Observation of teacher pedagogy was conducted through the use of the Transitional Bilingual Observation Protocol (TBOP). Having reasoned for the inverted fork formation in Figure 9, the commentary of why the inverted fork was the only structure possible in the context of Project MSSELL is explored.

Research Question Addressed

In this study, the researcher sought to comment on the development of causal structure for pedagogy during Project MSSELL. Through reason and causality described by Pearl (2009) Project MSSELL overlapping variables were converted into several

nodes. From the nodes, variables were reclassified into two specific nodes defined from Project MSSELL literature (i.e., the Level I and II approach described in Lara-Alecio et al., 2012). The literature on Project MSSELL is also supportive of a two node graphic representation to justify the illustrated change. Justification for reclassified variables in a two node context helped defined the causal directions between nodes.

Beginning with three variables and their initial relationship to each other not yet defined by causal direction as in Figure 10(a), was the conceptual start of this study. From the three variables in Figure 10(a) casual direction was then introduced and depicted in Figure 10(b), 10(c), and 10(d). Figure 10(d) is known as the chain, when describing causal structure, and can have the opposite causal direction as illustrated earlier in Figure 6(d). In fact, Figure 10 reiterates the same information from Figure 6 and extends two more causal structures relevant to describing Project MSSELL treatment and control teachers. Figure 10(e) represents the control teachers and Figure 10(f) represents the treatment teachers. By causal reasoning the parent structure derived to explain pedagogy during project MSSELL was illustrated in Figure 10 (i.e., an inverted fork structure).

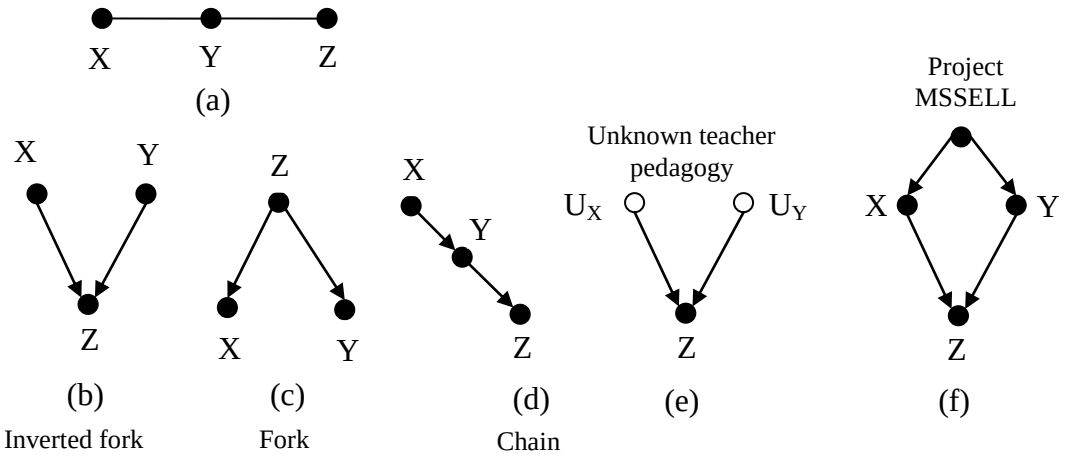

Figure 10. Basic causal structures

Without the context of Project MSSELL, initial variable relationship becomes unknown and expressed simply as X–Y–Z as in Figure 10(a). In the Figure 10(a) formation there is no concept of causal direction. As a result, initial review of different causal structures was an attempt to rule out causal structures contrary to causal intuition researchers can have in the context of Project MSSELL. The following causal structures are possible without knowing the context of Project MSSELL: Figure 10(b) X→Z←Y; Figure 10(c) X←Z→Y; Figure 10(d) X←Z←Y; and Figure 10(d) in the opposite direction X→Z→Y.

In addition, notice the white circles in Figure 10(e), which represent latent or unknown variables in the structure. For this study, the unknown variables refer to U_x and U_y or the unknown reason for control teachers to base their pedagogy (i.e., white circle

nodes) during Project MSSELL. In contrast, the black circles represent known values. For example, illustrated in Figure 10(f) are the known values for X and Y, which are derived from Project MSSELL since treatment teachers were given professional development for pedagogy. In this case, Project MSSELL was the parent to X (i.e., teacher professional development) and Y (i.e., student instructional interventions), and created a fork structure. In turn, X and Y became the parents to Z in another three variable relationship for determining causation. As a result, the causal structure displayed in Figure 10(f) became the initial structure to comment on causal relationships related to pedagogy in the context of Project MSSELL.

In the context of Project MSSELL, the values for X, Y, and Z become teacher professional development, student instructional interventions, and teacher. Although, Project MSSELL is a parent to X and Y as illustrated in Figure 10(f), the focus began with X and Y as parents to Z as depicted in Figure 9 and Figure 10(b). In fact, the structure receiving more attention is the inverted fork where X (i.e., teacher professional development [TPD]) and Y (i.e., student instructional interventions [SII]) had causal relationship to Z (i.e., the teacher or the internal thought process that lead to their pedagogy). However, Figure 10(b) as the causal structure best depicted to describe Project MSSELL in terms of pedagogy, was set aside briefly in order to comment on other causal structures.

Figure 10(c) is the fork structure and states that Z causes X and Y. Although Figure 10(c) is mathematically correct and could happen outside Project MSSELL, the implication of the teacher (Z) causing X (TPD) and Y (SII) did not follow the Project

MSSELL research design. Project MSSELL supported teachers in a positive way with the intention for teacher instruction to benefit ELL academic achievement. As a result, teachers were recruited to participate rather than dictate Project MSSELL objectives in professional development and student interventions. The notion of Z causing X and Y does not make sense in the context of Project MSSELL.

In Figure 10(d) the chain structure, Y (SII) causes Z (Teacher) and Z causes X (TPD). If Y (SII) caused Z and Z then caused X (TPD), why would X be so important to Project MSSELL when researchers already had the responsibility to lead teachers? The outcome as X or Y as opposed to Z, does not make sense in Project MSSELL because X and Y cannot deliver instruction like Z (i.e., the teacher). X within the context of Project MSSELL is most like feedback or additional professional development for Z, the teacher. Explained differently, if X (TPD) was the outcome result, then what was the purpose of having professional development (X) and not use it? The chain structure in Figure 10(d) illustrates X (TPD) as the end of the causal link. However, X (TPD) is best utilized to prepare teachers (Z) for pedagogy designed and monitored during Project MSSELL.

The same is true when Y is viewed as the end of a causal link. Taking Figure 10(d) and assessing the alternate chain structure or X→Z→Y to make sense in the context of Project MSSELL, does not work rationally. Y consisting of student instructional interventions does sound rational when teacher professional development is given to teachers in order to inform Y, student instructional interventions. However, Project MSSELL provided the student instructional interventions or Y. Y (SII) is

considered a given or constant since Project MSSELL provided curriculum and materials for teachers. Teachers volunteered to participate in Project MSSELL and the only logical conclusion from teacher professional development and student instructional interventions was the teacher would experience pedagogical change.

Therefore, the inverted fork of X→Z←Y is the only logical and rational choice in constructing causal structure for pedagogy in the context of Project MSSELL. Additionally, Figure 10(b), is a justified and reasonable description for how the Project MSSELL intervention was described in literature (Lara-Alecio & Irby, 2010; Lara-Alecio, Irby, & Tong, 2011; Lara-Alecio, et al., 2010). Other variables besides X and Y could influence Z (Teacher), yet Project MSSELL was described in enough detail for the researcher to structure the relationship of variables X, Y, and Z.

Control teachers in Project MSSELL are recognized as having no access to TPD or SII. As a result, attempts for causal structure reflect what was presented in Figure 10(e) and Figure 11(a). Treatment teachers in Project MSSELL are reflected in Figure 10(f) and Figure 11(b). Figure 11 illustrates the connection from Project MSSELL to teacher. Then, teachers exhibit pedagogy, which was a measurable outcome through the means of the TBOP during Project MSSELL. As a result, the variable ID is also introduced and represents pedagogy of the teacher (Z).

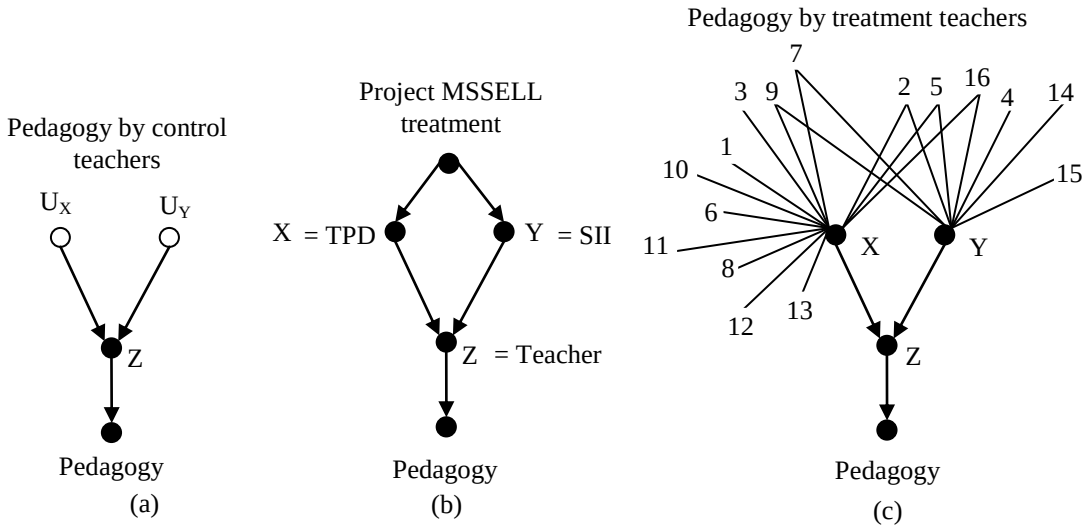

Figure 11. Deconstructing a causal model for pedagogy from Project MSSELL. Figure 11(a) represents the Project MSSELL control teachers. Figure 11(b) represents Project MSSELL as an intervention for treatment teachers. Figure 11(c) represents the Project MSSELL intervention categorized into Teacher Professional Development and Student Instructional Intervention components.

Teacher pedagogy (ID) could have several influential variables that affect Z (Teacher). For example, Figure 11(a) illustrated the unknowns that can influence the pedagogy of a teacher. However, in the context of Project MSSELL U_x and U_y are defined and the causal structure became represented by Figure 11(b). In greater detail, Figure 11(c), summarized the original 16 overlapping Project MSSELL variables (i.e., the three-dimensional figure of Project MSSELL in Figure 7) are reclassified into two nodes representing X and Y. Therefore, the set of all values of X (i.e., TPD) are Project MSSELL components 1-3, 5-13 and 16. The same conclusion can be drawn for Y

values, namely that they were Project MSSELL components 2, 4-5, and 14-16. Of the 16 variables reclassified into two nodes Figure 11(c) also showed that five variables were classified as values contributing to both X and Y.

By observing how researchers described their Project MSSELL model and variables the researcher was able to convert the MSSELL model into casual structure using Pearl's (2009) framework. Specifically, the Project MSSELL three-dimensional model was converted into a two-dimensional model and then restructured into a three variable causal structure. The transformation of the three-dimensional model also included the use of causal language to describe causal structure, such as language to describe the parents of variables. As a result, the parents of Z (i.e., teacher professional development and student instructional interventions) were connected to Project MSSELL and had a causal relationship toward Z (i.e., Teacher), as seen in Figure 11(b). In other words, the parent variables that have a causal reference to Z are X (TDP) and Y (SII).

Each variable described in the Project MSSELL model was assessed for reclassification toward teacher professional development (TPD) or student instructional interventions (SII). Researcher decision making to determine variable reclassification was based on Project MSSELL literature describing Level I (i.e., teacher professional development) and Level II (i.e., student instructional interventions) design structures (Lara-Alecio & Irby, 2010; Lara-Alecio et al., 2010). Furthermore, Figure 11(c) and Figure 12(a) illustrated how Project MSSELL variables were still represented in the causal structure.

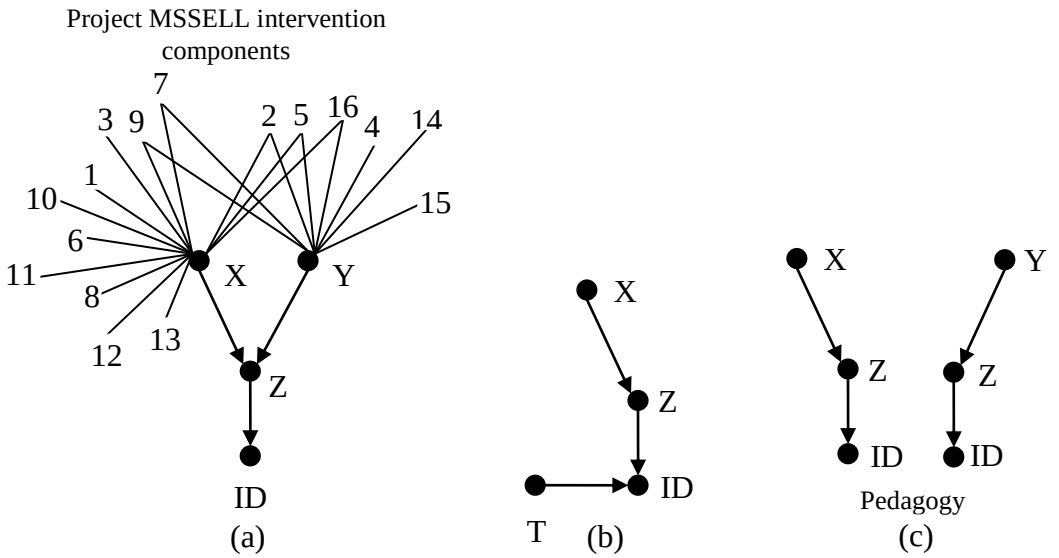

Figure 12. Derived causal structure for pedagogy

In Pearl (2009) the process to separate causal direction from one variable to the next variable is through the process called *d*-separation, where *d* means directional. Through the *d*-separation criterion researchers can observe the independence of Z (Teacher) from T (TBOP) given ID (i.e., pedagogy; Pearl 2009). In reference, the classroom observation measure was the TBOP and represented by T in Figure 12(b). In this study, the *d*-separation of Z and T was beneficial because it allowed researchers to assess ID as the outcome from Z and T. Results from the use of the TBOP (i.e., T) provided a measure for pedagogy (ID) during Project MSSELL. Although, there was no access to Project MSSELL teachers for further information the use of T becomes the main proponent for describing ID captured from Z.

The reason for T was to record pedagogy during Project MSSELL. Because the directional path between T to Z is not present as demonstrated in Figure 12(b), the causal structure of Z → ID ← T (i.e., inverted fork) indicated that Z and T are independent of each other and have *d*-separation. The absence of a directional arrow between Z and T conforms to the context of Project MSSELL because T represented the classroom observation measure, which had to remain separate from the teacher (i.e., Z). Additionally, due to the nature of an inverted fork (Pearl, 2009) parents such as T and Z to ID in Figure 12(b) can have *d*-separation for the purpose of assessing specific causal structures.

In Figure 12(c) the splitting of structure X → Z ← Y into single parent structures illustrated *d*-separation further. As a result, X → Z → ID and Y → Z → ID represent two alternative structures. Yet, as the probability of X becomes more apparent as causal toward ID, Y reduces in causal direction toward ID. However, in Figure 12(a) X and Y are considered independent in relation to Z and void any directional arrows from X to Y because of the inverted fork structure (Pearl 2009). X and Y became dependent if the rest of the structure was considered. For example in Figure 12(a), X and Y are dependent on Project MSSELL to exist. The structure X ← Project MSSELL → Y is bound by characteristics defined by the fork structure, as illustrated by the node representing Project MSSELL and described earlier in Figure 11(a) (i.e., Rule 1 condition for *d*-separation; Pearl 2009).

In other words, X → Z ← Y is blocked from forming direction arrows between X and Y because of the inverted fork structure. Because the structure described as X → Z

← Y has a middle node defined as Z with a descendent described as ID outside the set of nodes descriptive of MSSELL, the formation of directional arrows from MSSELL to ID are not in context to the reality of the study. X and Y in Figure 12(a) illustrate *d*-separation due to the existence of ID, and making X and Y necessary in producing ID during Project MSSELL. As a result, new direction arrows passed the inverted fork represented by X → Z ← Y are possible and witnessed in Figure 12(b) with the introduction of T.

By including classroom observation instruments (COIs) to measure pedagogy, the causal structure illustrated in Figure 12(a) can introduce another node represented by T as in Figure 12(b). The use of a COI was another aspect of Project MSSELL to measure pedagogy from treatment and control teachers. Project MSSELL used the Transitional Bilingual Observation Protocol (TBOP) as the COI for observing pedagogical behavior or pedagogy. For this study, T specifically referred to scores given by raters using the TBOP to record measures of pedagogy. The rationale for introducing T and how T relates to pedagogy is summed by the following probability function:

P (classroom measure | teacher, pedagogy) or P (T | Z, ID).

Meaning, the probable outcome of T given the teacher and pedagogy became the probability estimate for deciphering control teachers from treatment teachers during Project MSSELL. Distinction from treatment and control teachers is based on the assumption that Project MSSELL was developing pedagogical behaviors among treatment teachers for the benefit of serving the ELL classroom.

The rationale of the probability function is derived from the notion that teachers would display pedagogy that would either be conducive to academic achievement for ELLs or not. In other words, the conditional probability from the set of T classroom measures, given the teacher and their pedagogy, would describe the pedagogy. T was substantiated from raters using the TBOP instrument, which recorded pedagogy during Project MSSELL. Additionally, the difference between pedagogy from treatment and control teachers became an assessment of ID by T and demonstrating the positive results produced by Project MSSELL (i.e., as described in Lara-Alecio et al., 2012).

Outside of Project MSSELL

External from the intervention (i.e., Project MSSELL), we can rationalize the teacher's (i.e., Z) pedagogy would result from the teacher's judgment to adhere to an instructional method. The idea for a teacher making sense of guiding instruction from somewhere whether it was from internal beliefs on teaching or years of experience can be represented as variable U_X in Figure 11(a). In turn, the next reasonable variable to influence the teacher or Z would be the student element or Y. Teachers need students in order to teach, indicating the teacher-student relationship as not an unreasonable connection. Therefore, the variable Y represented the student component, which can also point to curriculum or what the teacher was expected to teach in the classroom. As a result, two assumptions concerning pedagogy are the likelihood of X and Y present in effecting the teacher. Furthermore, apart from Project MSSELL, a general causal structure representing pedagogy is shown in Figure 11(a), also known as the Project MSSELL control teachers.

Researchers can use counterfactuals in a deterministic approach. For example, consider two student types such as English language learner (ELL) and non-ELL, where A, B, C, D, and U stand for the following:

- U = Teacher,
- C = Pedagogy of teacher,
- A = English language learner,
- B = Non-English language learner, and
- D = Academic achievement.

The assumptions are teachers in the Project MSSELL intervention are cooperating and not displaying behavior outside of the intervention or outside the realm of standard teaching practices. Also, researchers are not acting in an unethical manner when providing professional development to treatment teachers. The counterfactual sentence can be expressed in the following way: If academic achievement gains were more concerned about ELLs in the classroom, then the ELL would be affected by pedagogy even if the non-ELL had not received pedagogy serving the ELL classroom.

Table 9 expressed how variables (e.g., Teacher) can go from Step 1 in Model M to Step 2 in Model M not A (i.e., non-ELL), which displays the outcome Project MSSELL can have on ELLs. Meaning, two possible outcomes for ELLs exist in what actually happened (i.e., Lara-Alecio et al., 2012) and what could hypothetically happen as in not A. Therefore, Project MSSELL can be described as the source of orientation from making practical applications to developing further hypothetical influence to ELLs.

In fact, Pearl (2009) expressed the notion of background variables as main carriers of information from the actual world to the hypothetical world.

Table 9

Evaluating Counterfactuals (Pearl, 2009)

Step 1 Model M	(U) Teacher
C = U	(C) Pedagogy of teacher
A = C	(A) English language learner
B = C	(B) Non-English language learner
D = A or B	(D) Academic achievement
Facts: D	
Conclusions: U, A, B, C, D	
Step 2 Model M not A	(U) Teacher
C = U	(C) Pedagogy of teacher
Not A	(A) English language learner
B = C	(B) Non-English language learner
D = A or B	(D) Academic achievement
Facts: U	
Conclusions: U, not A, C, B, D	

In reference to Figure 13, if an ELL was affected by treatment teachers from Project MSSELL, then the agreement between treatment teachers to teaching according to Project MSSELL design would result in pedagogy for ELLs. Alternatively, if ELLs were affected by pedagogy not serving the ELL classroom, then ELL's academic results were speculated to be below mainstream academic achievement (i.e., contributing to the academic achievement gap between ELLs and non-ELLs). As a result, if a teacher

displayed pedagogy that served the ELL classroom well, then ELLs would experience positive gains toward academic achievement.

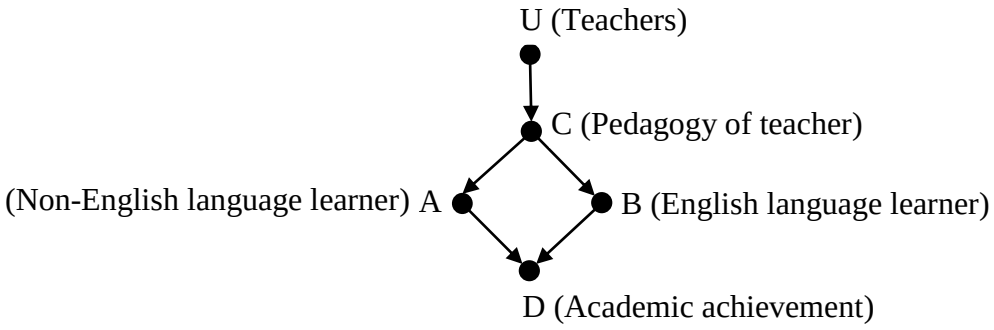

Figure 13. Causal relationships in the two student groups

Project MSSELL literature provided in detail a description of the model, which allowed the researcher to reason for causal structure related to pedagogy. Furthermore, counterfactuals can illustrate the causal relationship between teacher and student. For example, if student A and B in Figure 13 receive the same pedagogy, then the expected academic achievement can only benefit either student A or B but not both. In other words, if pedagogy is beneficial for non-ELLs, then the expected academic achievement is not the same since the pedagogy would only benefit the non-ELL. For an intervention, such as Project MSSELL the expected pedagogy for student A and B was expected to be different because Project MSSELL was training the teacher to exhibit pedagogy to serve the ELL classroom. However, the goal of this study was to illustrate causal structure in relation to teacher pedagogy in light of a global model like Project MSSELL. As a result,

ELL academic achievement can be derived from pedagogy by observing final results from Project MSSELL (i.e., Lara-Alecio et al., 2012).

Conclusion

Project MSSELL does adhere to the gold standard for providing evidence of what works. For instance, the intervention is quasi-experimental, the sample has been randomized at the classroom level, there is a control group, and was conducted over time (Lara-Alecio, Tong, Irby, Guerrero, Huerta, & Fan, 2012). Certainly these design characteristics have validated Project MSSELL as having a general connection between the intervention and the effect (i.e., academic achievement of English language learners' science literacy and English proficiency in middle school). However, there is no specific direction of how an intervention such as Project MSSELL would produce an effect on student outcomes in the sense of making causal inferences (e.g., Briggs, 2008 in making casual inferences).

Therefore, the goal was to observe the pedagogy of teachers during Project MSSELL with the lens of causal structure as explained by Pearl (2009). By using reasoned judgment of Project MSSELL and Pearl's causation framework, the impact was to illustrate viable causal structure for pedagogy with observable directions toward English language learner academic achievement. In fact, Lara-Alecio et al. (2012) illustrated actual academic gains by ELLs. Supporting the notion, pedagogy conducive to academic achievement among ELLs would have causal relationship to ELL academic achievement gains. In this study I speculated by causal reasoning how teacher's

pedagogy from Project MSSELL is related to ELL academic achievement. In addition, I speculated by what means we can recognize a causal link.

Project MSSELL had both treatment and control teachers, which resulted in causal structure for the treatment teachers having identified or defined variables (i.e., represented by black circles). The control teachers had latent variables represented by white circles outlined in black. For instance, earlier in Figure 11(a) the control teachers were represented with causal structure having two unknown variables U_X, U_Y. Figure 11(b) then portrayed causal structure for the treatment teachers with defined variables (i.e., black circles), since Project MSSELL had defined all variables during intervention. However, the control teachers were left to decide how they would determine their X and Y apart from Project MSSELL. As a result, control teachers would gather their X and Y from whatever source available to dictate their pedagogy to serve the ELL classroom.

In general, observing the pedagogical differences between causal structure for treatment and control teachers, would illustrate the lack of guidance control teachers had in displaying pedagogy to serve the ELL classroom. Meaning, teachers made decisions about pedagogy, which in turn had causal implications toward ELL academic achievement. However, decision making on pedagogy conducted by control teachers would most likely surround the notion of simplicity (e.g., whatever is easier to do). Most control teachers would hypothetically base their pedagogy on whatever was easy to follow, such as the principal's expectations or the teacher's own belief of how academic achievement is gained. The danger involved resided in what would teachers naturally

gravitate to concerning their pedagogy in the absence of structure (e.g., Project MSSELL).

What Lara-Alecio et al. (2012) illustrated with Project MSSELL and in conjunction with this dissertation study was that structure for pedagogy could be followed by teachers. Furthermore, if teachers followed the Project MSSELL intervention design, then academic achievement gains could be observed among English language learners. Also, among treatment teachers there was a shared purpose, which unified the tangible short goals and vision for what the school needed in order to empower its English language learners.

CHAPTER V

CONCLUSION: A SYNTHESIS

This dissertation study consisted of three chapters oriented toward discussion on Classroom Observation Instruments (COIs) and English Language Learners (ELLs). While Chapter 1 provided an overview of the following chapters, Chapters 2, 3 and 4 shared a general similarity in describing the Transitional Bilingual Observation Protocol (TBOP) in different settings. Chapter 5 presents a synthesis from preceding chapters in order to make sense of how chapters relate to one another. In fact, Figure 14 illustrates the number of relationships connecting chapters together.

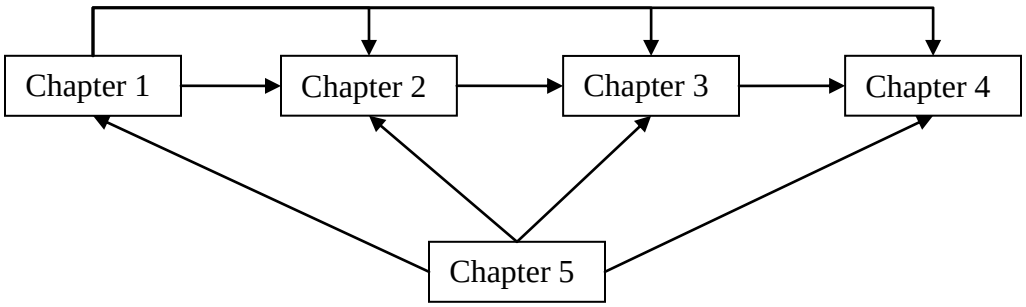

Figure 14. How chapters relate to one another

In Figure 14, arrows from Chapter 1 connect to Chapters 2, 3 and 4 because the function of Chapter 1 was to provide a descriptive overview. The overview in Chapter 1 includes research questions and potential sites for publication for Chapters 2, 3 and 4. Additionally, the arrow directed from Chapter 1 to Chapter 2 illustrated the transition from dissertation overview to systematic literature. The relationship between Chapter 1

and Chapter 5 began the process of synthesis to draw from Chapter 1 relevant information for Chapter 5.

In Chapter 2 a systematic review was conducted which sought to identify COIs used in classrooms serving ELLs. Once COIs were identified generally, Chapter 2 methodology guided the process to derive from 37 COIs which COIs were more specifically tailored to serving the ELL classroom. Figure 15 depicts a directed arrow toward Chapter 3 because data from a COI (i.e., the Transitional Bilingual Observational Protocol or TBOP) identified in Chapter 2 then became the focus in Chapter 3. Chapter 5 then drew from literature described in Chapter 2 in order to synthesize information.

Chapter 3 examined data from Project Middle School Science for English Language Learners (MSSELL). The TBOP was the COI used during Project MSSELL, which observed the pedagogy of treatment and control teachers. In Figure 15, a directed arrow points from Chapter 3 to Chapter 4 indicating results from the TBOP are further discussed in Chapter 4. Chapter 5 also drew information from Chapter 3 in order to add further to the collective contribution from multiple chapters.

Chapter 4 continued the discussion from Chapter 3 in the context of a casual commentary for pedagogy during Project MSSELL. The relationship of Chapter 3 to Chapter 4 is displayed by a directed arrow from Chapter 3 pointing to Chapter 4. In Chapter 4, pedagogy was discussed in terms of causal reasoning and commentary in order to rationalize for a causal structure pertaining to pedagogy during Project MSSELL. Figure 14 illustrates Chapter 4 information was added to the synthesis in Chapter 5 in order to complete the general meaning of all findings from this dissertation.

Chapter 5 drew from all chapters in order to synthesis this dissertation study. As depicted in Figure 14, Chapter 5 had directional arrows pointing to Chapters 1, 2, 3 and 4 (i.e., all chapters) because Chapter 5 was designed to take information from all chapters. The directional arrows portrayed in Figure 14 however are not causal but rather the flow of information that represented the relationships between chapters made by the author. As a result, Figure 14 illustrated the synthesis process and logic involved by the author to bring information together in a meaningful way. Additionally, the process of synthesis sought to summarize chapter findings and describe recommendations.

Summary of Study Significance

Through the use of systematic review protocols 37 COIs were generally found to have been used in classrooms with ELLs. However, through further COI examination 19 of the 37 COIs were considered generally to have potential to serve the ELL classroom. While there were several COIs, the frequency of the instruments in the literature also expressed how widely used they were, yet no attempt was made in this study to measure the breadth and scope of each COI. The English language Learner Observation Instrument (ELLCOI) was the most frequent COI in the literature, yet this may be due to the search terms used and therefore inconclusive. Although several researchers have constructed COIs to observe the English language learner classroom, researchers differ in what they choose to observe.

In fact, through the use of systematic review the author observed COIs had diversity and depended on different theories. For instance, the ASOS and DLASOS depended on socio-cultural theory as a means to make sense of the classroom to serve

ELLs. The TBOP was constructed from pedagogical theory with the rationale that the teacher is the main proponent to serving the ELL classroom. Along the standards of effective pedagogy, the TFM and SPC were COIs that used teaching standards in order to serve the ELL classroom. Still other COIs were used to record teacher and student interactions as a means to observe teacher–student interactions that should be observed in the classroom in order to confirm pedagogy was serving the ELL classroom (e.g., COM).

However, whether the use of teaching standards, pedagogical theory or socio-cultural theory were used in constructing COIs, this study did not attempt to order which pedagogical or theoretical basis for a COI was better but rather reasoned for COIs that could serve the ELL classroom, more so than the original 37 COIs found through systematic review protocols. Although it can be argued that the 19 of 37 COIs labeled as serving the ELL classroom can really be reduced to a smaller amount, this study allowed for less restrictive conditions for COIs to enter the boundaries of serving the ELL classroom. For instance, of the 19 COIs, if pedagogical theory was stressed as vital for deciding which COIs should be labeled as serving the ELL classroom, then this study would conclude the following COIs as having pedagogical theory strengthening the COIs to serve the ELL classroom: TBOP, SPC, ASOS, DLASOS and the classroom observation guidelines based on observation scales developed by Luykx and Lee (2007). Meaning, other COIs are relying on pedagogy standards and effective strategies or methods for serving the ELL classroom.

Yet does literature match with the instruments? In short, yes, but instruments also measure more than what literature calls for. Therefore, perhaps an observation instrument needs to be directed to one measure since the variety in observation instruments would suggest that there are multidimensional aspects of the classroom that could be measured. By focusing an observation instrument to a particular task, then conventional wisdom is that the instrument can measure exact criteria. For example, the Sheltered Immersion Observation Protocol (SIOP) is used to examine the sheltered immersion model as an observational start. Therefore, the start of classroom observation begins with what is the reason for observation? An observation instrument can be directed to a content area (e.g., science, math, etc) or to the teacher as was the topic of this study.

Teacher effectiveness (i.e., pedagogy) would need to complement the literature of research on how ELLs learn (or based on standards). If we align teacher effectiveness to standards then our observation instruments become a measure of the standards, which are present in some COIs (e.g., SPC). In fact, large-scale interventions can utilize researchers to develop COIs based on intervention standards and guidelines (e.g., Success For All COIs, Marzano observation protocols). Therefore, if an observation instrument needs a directive, then what would comprehensive observation instrument look like? The comprehensive observation instrument would have to be a system (e.g., Activity Structure Observation System), which could include multidimensional measures of the ELL classroom. As a result, a researcher could use parts or all of the observation system to observe for their targeted research objective. Such a system would need

general research community agreement if researchers are to agree on the use of a standard or "go to" form of observation measurement for serving the ELL classroom. Systems of observation can observe different educational levels such as state, district, school (e.g. School Observation Measure), classroom, teacher, teacher-student interaction-level, and student-level.

Through examination of archived data (i.e., referring to Chapter 3) the implications for Project MSSELL as an intervention for empowering teachers and increasing ELL academic achievement becomes evident from treatment and control teacher results. Additionally, in Figure 15 the process of the dissertation to gradually go from the broad topic of COIs used among ELLs to causal structure in pedagogy provide further description of the effectiveness of Project MSSELL. In Figure 15, Chapter 3 demonstrates how findings from one chapter make a bridge to the proceeding chapter in order to illustrate the gradual synthesis of information presented in this dissertation.

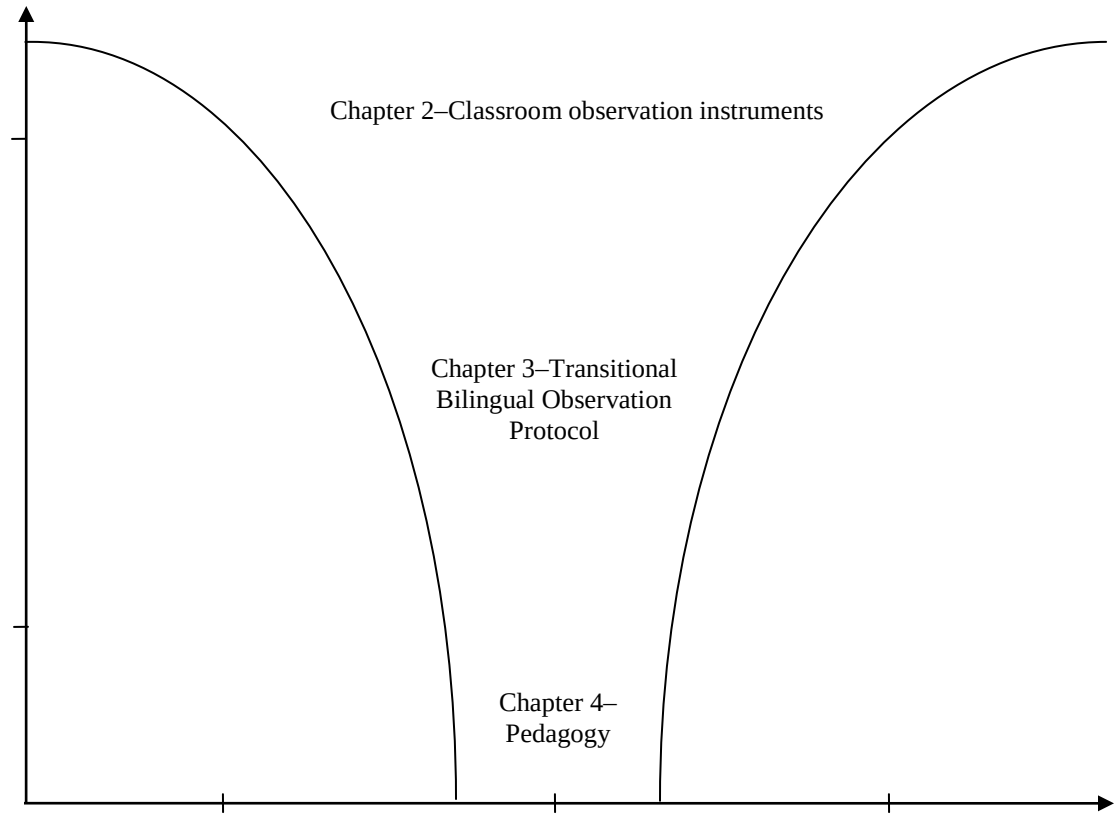

Figure 15. How chapters begin to narrow in focus

The most visible difference between treatment and control teachers as recorded from the use of the TBOP was in the coded pedagogical behavior pertaining to Communication Mode. The other pedagogical domains (i.e., Activity Structure, Language Content and Language of Instruction) of the TBOP were still observed and pedagogical differences between treatment and control teachers exhibited positive gains for ELLs during Project MSSELL. In my opinion, Lara-Alecio, Tong, Irby, Guerrero, Huerta, and Fan (2012) provided evidence to support Project MSSELL as having a positive impact on increasing the achievement of English language learners in English proficiency and Science literacy. In fact, Project MSSELL is related to Project English

Language and Literacy Acquisition (ELLA), which has also demonstrated the potential to improve English language learners in literacy acquisition. Project ELLA has gone into a scale-up innovation grant to be conducted across the state of Texas, another indicator of intervention effectiveness to serve ELLs by Project MSSELL researchers. The results from this study reinforce and validate the positive academic impact Project MSSELL had in promoting academic achievement for ELLs.

Commentary on causal inference implications in Project MSSELL was constrained to teacher-level variables. In addition, the causal reasoning for potential causal structures within Project MSSELL illustrated components of causality related to teacher pedagogy. The rationale for constructing causal structures in Chapter 4 were to serve an alternative means for causal implications for future interventions that are not able to be experimental or quasi-experimental in research design (e.g., ethnographic study). By examining further the pedagogical behaviors recorded by the TBOP during Project MSSELL the author brought forth additional discussion that contributed to validating the effectiveness of Project MSSELL to serve ELLs (i.e., Lara-Alecio et al., 2012).

Summary of Key Findings

When searching through the literature, COIs designed to observe teacher behavior conducive to serving the ELL classroom were few in number. Yet, research educational entities and researchers interested in capturing classroom behaviors proven to serve the ELL classroom can also be found outside of the research literature as in grey literature (e.g., Marzano observation protocols; Marzano, Pickering, & Pollock, 2001).

The concern for COIs outside of the literature is a call for evidence to prove COIs not found in the literature to have psychometrically valid and reliable characteristics. In this dissertation study 19 COIs were identified and few COIs take the approach of assessing the teacher (e.g., TBOP and Teacher Record Observation Schedule). For the purpose of this study, COIs capturing teacher pedagogy was the focus and COIs expressing different aspects of observation in the ELL classroom were included to form a bigger picture.

What was illustrated from Project MSSELL treatment teachers was their ability to align to Project MSSELL goals for student outcomes (Lara-Alecio et al., 2012). This is evident in the pedagogical differences illustrated in Chapter 3 between the treatment and control teachers. Treatment teachers displayed a focus toward writing activities and challenging students at higher cognitive levels by means of the TBOP. Such pedagogical differences from treatment teachers compared to control teachers only exemplified Project MSSELL's use of teacher professional development and student instructional interventions to empower treatment teachers to serve the ELL classroom. Pedagogical behaviors observed among control teachers displayed a lack in serving the ELL classroom and ultimately had an effect on ELLs that may have not been helpful to their academic achievement.

Project MSSELL can relate to comprehensive school reform interventions, because of the all-inclusiveness taken by researchers to establish the appropriate school environment necessary to increase the academic achievement of ELLs. In response, this dissertation study examined possible causal structure from Project MSSELL in order to

infer causal relationship between teachers and their pedagogy to serve the ELL classroom. Taking global properties of Project MSSELL (e.g., quasi-experimental design) aside and observing the local properties, such as student instructional interventions and teacher professional development, allowed for careful causal reasoning.

Project MSSELL was already considered to have causal inference as an effective intervention because of the nature of quasi-experimental studies, yet the causal commentary in this study sought to inform the specific causal structure for teacher pedagogy during Project MSSELL. The goal was to provide a causal structure for others to utilize when unable to conduct an experimental research design. Furthermore, the research design (i.e., randomized control trial) was appropriate in deducing causal inferences from the Project MSSELL intervention. As a result, this dissertation study extended the casual inference implications of a quasi-experimental study. For example, the Pearl (2009) framework for causal reasoning was used to construct causal structure in order to identify the inner mechanics of causation related to pedagogy to serve ELLs during Project MSSELL.

Limitations

The strengths of this dissertation are that rigorous methods were used to find out more information concerning pedagogy during Project MSSELL through a secondary analysis. The systematic review followed established protocols in order to increase the rigor of this study. However, search terms involved in the systematic review protocol may have played a role in not finding other COIs. The notion of other COIs not found in

the literature are minimal because systematic review procedures called for comprehensively searching through numerous databases. In fact, five search engines were used to access over a hundred databases. Such exhaustive searching would minimize COIs not found in the literature. There is still the possibility that a COI could have escaped the sight of raters used during the systematic review process.

Secondary analysis involved the use of archived data from Project MSSELL which was recorded through the use of the TBOP. The author thoughtfully examined the archived data along with access to Project MSSELL researcher for feedback on any additional information that could lend itself to further describing teacher pedagogy during Project MSSELL. The analysis involved frequency analysis and chi-square testing and led to interesting findings from nominal data. However, weakness related to conducting this dissertation study was considered minimal to none because data collection and analysis were not complex. Not much can result from nominal data, yet the researcher carefully observed the research questions and sought to use the data to find answers.

Another limitation was the author's ability to comment of causal structure, since causal reasoning and inference were a newly developed concept. However, the author spent time reading and researching how Pearl (2009) described and used causal reasoning to make inferences. Pearl (2009) served as a model to follow when commenting on teacher pedagogy in a causal context. Also, new trends in causality research continue to develop quickly and causality described here might change in the future or the future can provide more clarity in making causal inferences. As a result, as

causality research continues to expand there is a danger in causality expressed in this dissertation to have gone out of date or expressed differently.

Implications

Systematic review results imply that not many COIs are available to researchers, practitioners and policy makers to assess teacher pedagogy in the form of serving the ELL classroom. Therefore, COIs found in this systematic review can improve or strengthen their relationship to serving the ELL classroom since 19 COIs were loosely included as serving the ELL classroom. Pedagogical theory should play an important role in COIs serving the ELL classroom because the teacher can be viewed as the initiator of learning in the classroom. Also, researchers or practitioners interested in developing new COIs can observe findings in this study to begin planning and developing COIs that can serve the ELL classroom.

The implications of this dissertation study suggest that Project MSSELL was able to get treatment teachers exhibiting pedagogy to serve the ELL classroom. Treatment teachers had a tendency toward Project MSSELL outcome goals for pedagogy. The teacher behaviors illustrated here indicate that if unified teacher pedagogy is of value then Project MSSELL demonstrates one method of accomplishing such a task. Additionally, Project MSSELL displays what more can result from having teachers unified in pedagogy (e.g., science literacy gains; Lara-Alecio et al., 2012).

The causal commentary in Chapter 4 can be used to reason for causal inference when experimental research design is not possible. In Chapter 4, the process to deconstruct the Project MSSELL model lead to a three-node causal structure that can be

used to help researchers or practitioners begin to observe and describe pedagogy to serve ELL in a casual inference context. Making causal inferences needs to always be taken carefully and thoughtfully because of the much needed justification to utilize causality, but with attempts taken here future studies can enforce the causal structure proposed for teacher pedagogy.

Recommendations

Although conducting a systematic review has a specific process (Torgerson, 2003), the systematic review resulted in compiling COIs used in ELL settings. Identifying COIs for the field of bilingual education is helpful in recognizing COIs that have under gone testing for validity and reliability. The systematic review furthers practice in bilingual education by informing researchers and practitioners of potential COI use in bilingual education settings. Therefore, illustrated in Figure 16 is how recommendations are connected to improving practice, especially among researchers wanting to know what is currently available in classroom observation tools. A listing of available COIs for the mainstream classroom is also a struggle for educational researchers, because generally there is no source that keeps up or take note of all observations instruments being created.

Figure 16 also displays the connections to other areas, such as practice, theory, policy, and future research. Although theory was not heavy discussed, the theoretical implications would be the expansion of critical race theory. By describing the gap between ELLs and non-ELLs, the notion of critical race theory was subtly introduced in this dissertation. Critical race theory in its broadest terms is about equality issues

between people groups. In this dissertation, I worked to illustrate the deficit of COIs to serving ELLs, provide findings from Project MSSELL to illustrate what worked with ELLs, and in Chapter 4 demonstrated causal structure that would aid researchers to determine causal relationships to empowering the pedagogy to serve the ELL classroom. In sum, the efforts in Chapters 2, 3, and 4 speak to what can be done about the equality issues among ELLs and non-ELLs (i.e., critical race theory).

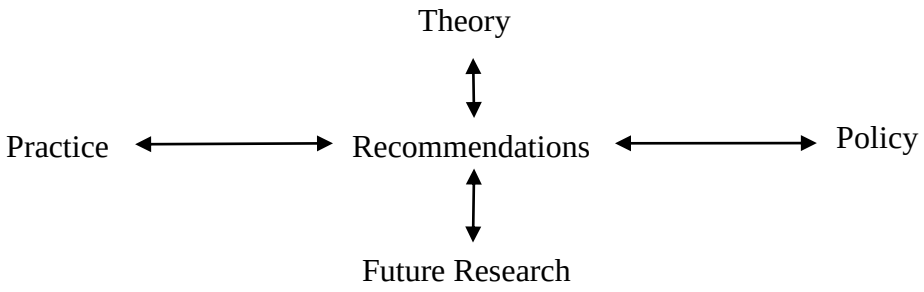

Figure 16. How recommendations inform other areas

Practice

In conducting a secondary analysis from Project MSSELL data, which looked at grade 5 pedagogy in science, the researcher had access to archived data collected by others. Rarely, are secondary studies conducted to review the quality of previous findings reported by researchers. In other words, a secondary analysis contributes validity back to the original study, since in this dissertation further information was presented to contribute to Project MSSELL's overall effectiveness among ELLs. Much

of research does not do replication studies, and though this study was not a replication it did reanalyze original data from Project MSSELL. Also, secondary analysis, when conducted can demonstrate another level of trustworthiness in the original study findings when outcomes arrive at the same conclusion. The practice of secondary analysis is a recommended need in the field of bilingual education.

Policy

Policy recommendations should include additional scale-up research to observe and support original findings in Project MSSELL. After all, Project MSSELL is a quasi-experimental study, which suggests there is a level of causal inference indicating a relational influence. The creation of the MET (Measure of Effective Teaching) database has arisen from the need to establish pedagogy vital for student academic learning. Yet, further study is needed to distinguish bilingual measures for effective teaching. Although, effective bilingual teaching is a call for future research, in this context the necessity of policy to provide opportunity for such a database to be created is what is recommended.

As for the commentary on causal reasoning, much of bilingual education research has taken a step back from making causal inferences. There are guidelines as to making causal inferences, yet a growing trend to make causal inferences seems to be increasing, but void in the field of bilingual education. Further attempts are necessary in bilingual education to make causal claims so that policy makers can formulate policy that supports best practices in the classroom. Causality is a different field, yet to venture into causality is necessary to provide applications and solidify what works.

Future Research

Future research can benefit by determining and coming to consensus among researchers of what measures or indicators are necessary for effective pedagogy in the bilingual classroom. If indictors are established, then a measurement device (e.g., a classroom observation instrument) to standardize what is necessary to observe can dictate the estimated teacher pedagogy to serve the ELL classroom. Current findings seems to indicate a diversity as to what during teacher pedagogy is measured by a COI; thereby leaving confusion concerning what COI to use and what measures are necessary from the ELL classroom. Future research seems to follow classroom observation systems, which include an arsenal of COIs for specific purposes. However, COIs need to serve the ELL classroom in order for research for ELLs academic gains to be actualized. What may be the test of time is to observe the popularity across time to see what COIs researchers and practitioners use.

REFERENCES

Acosta, S., Garza, T., Hsu, H.-Y., & Goodson, P. (*under review*). Validating the methodological quality questionnaire (MQQ): A rating scale for systematic reviews. *International Journal of Social Research Methodology.*

Acosta, S., & Garza, T. (2011). The podcasting playbook: A typology of evidence-based podagogy for prek-12 classrooms with English language learners. *Research in the Schools*, *18*(2), 40-57.

Acosta, S., Goltz, H. H., Goodson, P., Padrón, Y., Garza, T., & Johnston, A. (*under review*). English language learners and large-scale, high-stakes testing: A systematic review. *Hispanic Journal of Behavioral Sciences*.

Anderson, L. W. (1987). The classroom environment study: Teaching for learning. *Comparative Education Review*, 31(1), 69-87.

Atkins-Burnett, S., Sprachman, S., & Caspe, M. (2010). *"Language Interaction Snapshot with End of Visit Ratings (LISn+EVR)"* [Measurement instrument]. Princeton, NJ: Mathematica Policy Research.

Atkins-Burnett, S., Xue, Y., Kopack, A., Induni, M., & Moiduddin, E. (2010). *Instructional practices in Los Angeles universal preschool. Final report.* Princeton, NJ: Mathematica Policy Research.

Atwater, J., Lee, Y., Montagna, D., Reynolds, L., & Tapia, Y. (2009). *Classroom CIRCLE: Classroom code for interactive recording of children's learning environments.* Kansas City, KS: Juniper Gardens Children's Project.

Aud, S., Hussar, W., Johnson, F., Kena, G., Roth, E., Manning, E., . . . Zhang, J. (2012). *The condition of education 2012* (NCES 2012-045). Washington, DC. Retrieved December 10, 2012, from http://nces.ed.gov/pubsearch

August, D., Branum-Martin, L., Cardenas-Hagan, E., & Francis, D. (2009). The impact of an instructional intervention on the science and language learning of middle grade English language learners. *Journal of Research on Educational Effectiveness, 2*(4), 345-376.

Bailey, A. L., Butler, F. A., Stevens, R., & Lord, C. (2007). Further specifying the language demands of school. In A. L. Bailey (Ed.), *Language demands of school: Putting academic language to the test*. New Haven, CT: Yale University Press.

Baker, S. K., Gersten, R., Haager, D., & Dingle, M. (2006). Teaching practice and the reading growth of first-grade English learners: Validation of an observation instrument. *Elementary School Journal, 107*(2), 199-221.

Baker, S. K., Gersten, R., Haager, D., Dingle, M., & Goldenberg, C. (2005). *Assessing the relationship between observed teaching practice and reading growth in first grade English learners: A validation study.* Los Alamitos, CA: Instructional Research Group. Retrieved December 10, 2012, from ERIC database. (ED486282)

Benavides, A. H. (2004). Editor's introduction. *Bilingual Research Journal, 28*(1), iii-vi. doi:10.1080/15235882.2004.10162608

Blank, R. K., & Pechman, E. M. (1995). *State curriculum frameworks in mathematics and science: How are they changing across the states?* Washington, DC: Council

of Chief State School Officers.

Boller, K., Sprachman, S., & the Early Head Start Research Consortium. (1998). *Child-caregiver observation system*. Princeton, NJ: Mathematic Policy Research, Inc.

Boscardin, C. K., Aguirre-Muñoz, Z., Chinen, M., Leon, S., & Shin, H. S. (2004). *Consequences and validity of performance assessment for English language learners: Assessing OTL in grade 6 language arts* (CSE Tech. Rep. No. 635). Los Angeles: University of California, National Center for Research on Evaluation, Standards, Student Testing (CRESST).

Borko, H., Stecher, B. M., Alonzo, A., Moncure, S., & McClam, S. (2005). Artifact packages for measuring instructional practice: A pilot study. *Educational Assessment, 10*(2), 73-104.

Braun, H. I. (2005). *Policy information perspective: Using student progress to evaluate teachers: A primer on value-added models*. Princeton, NJ: Educational Testing Service.

Breunig, N. A. (1998). *Measuring the instructional use of Spanish and English in elementary transitional bilingual classrooms*. (Doctoral dissertation). Retrieved December 12, 2012, from Linguistics and Language Behavior Abstracts (LLBA). (9904616)

Brewer, D. J., & Stasz, C. (1996). Enhancing opportunity to learn measures in NCES data. In G. Hoachlander, J. E. Griffith, & J. H. Ralph (Eds.). *From data to information: New directions for the national center for education statistics*

(NCES 96-901, pp. 3:1-28). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Briggs, D. (2008). Comments on Slavin: Synthesizing causal inferences. *Educational Researcher, 37*(1), 15-22.

Brinberg, D., & McGrath, J. E. (1985). *Validity and the research process*. Newbury Park, CA: Sage.

Bruce, K., Lara-Alecio, R., Parker, R., Hasbrouck, J., Weaver, L., & Irby, B. J. (1997). Inside transitional bilingual classrooms: Accurately describing the language learning process. *Bilingual Research Journal 21*(2&3), 123-142.

Bruce, K. (August, 1995). *Measuring bilingual classroom activities according to a four-dimensional model.* Unpublished master's thesis, Texas A&M University.

Calderón, M., Slavin, R., & Sánchez, M. (2011). Effective instruction for English learners. *The Future of Children, 21*(1), 103-127.

Castellano, M., & Datnow, A. (2004). The influence of school reform on classroom instruction in diverse schools: Findings from an observational study of success for all. In H. C. Waxman, R. G. Tharp & R. S. Hilberg (Eds.), *Observational research in U.S. classrooms: New approaches for understanding cultural and linguistic diversity* (pp. 231-265). New York, NY: Cambridge University Press.

Castro, D. C. (2005). *Early language and literacy classroom observation. Addendum for English language learners*. Chapel Hill, NC: University of North Carolina, FPG Child Development Institute.

Castro, D. C., Espinosa, L., & Páez, M. (2011). Defining and measuring quality early childhood practices that promote dual language learners' development and learning. In Zaslow, M., Martinez-Beck, I., Tout, K., & Halle, T. (eds.). *Quality Measurement in Early Childhood Settings* (pp. 257-280). Baltimore: Brookes Publishing.

Center for Advanced Research on Language Acquisition (CARLA). (n.d.). *Culture and language learning initiatives*. Retrieved April, 6, 2013, from http://www.carla.umn.edu/culture/initiatives.html

Center for Early Care and Education Research–Dual Language Learners (CECER–DLL). (2011). *Research brief #5: Early care and education quality measures: A critical review of the research related to dual language learners.* Chapel Hill: The University of North Carolina, FPG Child Development Institute, Author.

Creemers, B., Kyriakides, L., & Antoniou, P. (2013). Establishing the field of teacher effectiveness research: Moving from investigating personal characteristics of teachers to understanding effective teaching practices. In *Teacher Professional Development for Improving Quality of Teaching*, (pp. 65-79). Netherlands: Springer Science+Business Media Dordrecht. doi: 10.1007/978-94-007-5207-8_5

Cummins, J. (1986). Empowering minority students: A framework for intervention. *Harvard Educational Review, 56* (1), 18-35.

Cummins, J. (2000). Language, power and pedagogy: Bilingual children in the crossfire. Clevedon, UK: Multilingual Matters Ltd.

Datnow, A., & Yonezawa, S. (2004). Observing school restricting in multilingual, multicultural classrooms: Balancing ethnographic and evaluative approaches. In H. C. Waxman, R. G. Tharp, & R. S. Hilberg (Eds.), *Observational research in U.S. classrooms: New approaches for understanding cultural and linguistic diversity* (pp. 174-204). New York, NY: Cambridge University Press.

Doherty, R. W., Hilberg, R. S., Epaloose, G., & Tharp, R. G. (2002). Standards performance continuum: Development and validation of a measure of effective pedagogy. *Journal of Educational Research, 96*(2), 78-89.

Doyle, W. (1977). Paradigms for research on teacher effectiveness. In L. S. Shulman (Ed.), *Review of research in education* (Vol. 5, pp. 163-198). Itasca, IL: Peacock.

Drews, T. S. (2009). *Application of the instructional congruence framework: Developing supplemental materials for English language learners.* (Doctoral dissertation). Retrieved December 12, 2012, from ProQuest Dissertations & Theses Full Text database. (304846138)

Echevarria, J., & Short, D. J. (2004). Using multiple perspectives in observations of diverse classrooms: The sheltered instruction observation protocol (SIOP). In H. C. Waxman, R. G. Tharp & R. S. Hilberg (Eds.), *Observational research in U.S. classrooms: New approaches for understanding cultural and linguistic diversity* (pp. 21-47). New York, NY: Cambridge University Press.

Echevarria, J., Vogt, M., & Short, D. (2012). *Making content comprehensible for English language learners: The SIOP model* (4th ed.). Boston, MA: Pearson.

Faris, J. A. (2006). *Prominent practices of effective teachers.* (Doctoral dissertation). Retrieved August 22, 2014, from ProQuest Dissertations & Theses Full Text. (304912401)

Foorman, B. R., Goldenberg, C., Carlson, C. D., Saunders, W. M., & Pollard-Durodola, S. D. (2004). How teachers allocate time during literacy instruction in primary-grade English language learner classrooms. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research*. Baltimore, MD: Paul H. Brookes Publishing Co.

Foorman, B. R., & Schatschneider, C. (2003). Measurement of teaching practices during reading/language arts instruction and its relationship to student achievement. In S. Vaughn & K. L. Briggs (Eds.), *Reading in the classroom: Systems for the observation of teaching and learning* (pp. 1-30). Baltimore, MD: Brookes Publishing Co.

Freedson, M., Figueras-Daniel, A., & Frede, E. (2009). *Classroom assessment of supports for emergent bilingual acquisition.* New Brunswick, NJ: National Institute for Early Education Research.

Garza, T., & Acosta, S. (2013, April). *The psychometric characteristics of the methodological quality questionnaire: Assessing educational research evidence in systematic reviews.* Poster presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Gee, J. P. (2005). Language in the science classroom: Academic social languages as the heart of school-based literacy. In R. Yerrick & W. M. Roth (Eds.), *Establishing*

scientific classroom discourse communities: Multiple voices of teaching and learning research (pp.19-37). Mahwah, NJ: Lawrence Erlbaum Associates.

Gersten, R., Keating, T. J., & Irvin, L. K. (1995). The burden of proof: Validity as improvement of instructional practice. *Exceptional Children, 61*(6), 510-519.

Gersten, R., & Baker, S. (2000). What we know about effective instructional practices for English-language learners. *Exceptional Children, 66*(4), 454-470.

Gersten, R., Baker, S. K., Haager, D., & Graves, A. W. (2005). Exploring the role of teacher quality in predicting reading outcomes for first-grade English learners. an observational study. *Remedial and Special Education, 26*(4), 197-206. doi: 10.1177/07419325050260040201

Gomez, L., Parker, R., Lara-Alecio, R., Ochoa, S. H., & Gomez, R. (1996). Naturalistic language assessment of LEP students in classroom interactions. *Bilingual Research Journal, 20*(1), 69-92.

Goodson, B. D., Layzer, J. I., & Layzer, C. J. (2005). *Quality of early childhood care settings: Caregiver rating scale (QUEST)* [Measurement instrument]. Abt Associates Inc.: Cambridge, MA.

Goodson, B. D., Layzer, C. J., Smith, W. C., & Rimdzius, T. (2006). *Observation Measures of Language and Literacy Instruction in early childhood (OMLIT)* [Measurement instrument]. Cambridge, MA: Abt Associates, Inc.

Grehan, A. W., Smith, L. J., & Ross, S. M. (2004). *The early reading first literacy observation tool*. Memphis, TN: The University of Memphis, Center for Research in Educational Policy.

Guarino, A. J., Echevarria, J., Short, D., Schick, J. E., Forbes, S., & Rueda, R. (2001). The sheltered instruction observation: Reliability and validity assessment. *Journal of Research in Education*, 11(1), 138-140.

Haager, D., Gersten, R., Baker, S., & Graves, A. W. (2003). The English-language learner classroom observation instrument for beginning readers. In S. Vaughn & K. L. Briggs (Eds.), *Reading in the classroom: Systems for the observation of teaching and learning* (pp. 111-144). Baltimore: Brookes Publishing Co.

Halle, T., & Vick, J. E. (2007). *Quality in early childhood care and education settings: A compendium of measures*. Washington, DC: Prepared by Child Trends for the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Halle, T., Whittaker, J. V., & Anderson, R. (2010). *Quality in early childhood care and education settings: A compendium of measures* (2nd ed.). Washington, DC: Child Trends. Prepared by Child Trends for the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Hammer, M. R., & Bennett, M. J. (1998). *The Intercultural Development Inventory (IDI) manual* [Measurement instrument]. Portland, OR: The Intercultural Communication Institute.

Haynes, J. (n.d.). SIOP: Making content comprehensible for ELLs. *EverythingESL.net*. Retrieved May 26, 2014, from http://www.everythingesl.net/inservices/using_siop_model_08621.php.php

Hilberg, R., S., Doherty, R. W., Epaloose, G., & Tharp, R. G. (2004). The standards performance continuum: A performance-based measure of the standards for effective pedagogy. In H. C. Waxman, R. G. Tharp, & R. S. Hilberg (Eds.), *Observational research in U.S. classrooms: New approaches for understanding cultural and linguistic diversity* (pp. 48-71). New York, NY: Cambridge University Press.

Hitchcock, J., Dimino, J., Kurki, A., Wilkins, C., & Gersten, R. (2011). *The impact of collaborative strategic reading on the reading comprehension of grade 5 students in linguistically diverse schools*. (NCEE 2011-4001). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Holland-Coviello, R. (2005). *Language and literacy environment quality in early childhood classrooms: Exploration of measurement strategies and relations with children's development.* (Doctoral dissertation). Retrieved December 12, 2012, from ProQuest Dissertations & Theses Full Text. (305448761)

Horizon Research Inc. (2000). Local systemic change through teacher enhancement classroom observation protocol. Retrieved March 23, 2014, from http://www.horizon-research.com/local-systemic-change-through-teacher-enhancement-classroom-observation-protocol/

Intercultural Development Inventory (IDI). (n.d.). Intercultural development inventory. *The roadmap to intercultural competence using the IDI*. Retrieved April 10, 2014, from https://idiinventory.com/products/

Irby, B. (2013). Collaborative research: A longitudinal randomized trial study of middle school science for English language learners (project MSSELL). Retrieved September 27, 2014, from http://www.nsf.gov/awardsearch/showAward?AWD_ID=0822153&HistoricalAwards=false

Irby, B., Tong, F., Lara-Alecio, R., Meyer, D., & Rodríguez, L. (2007). The critical nature of language instruction compared to observed practices and high stakes tests in transitional bilingual classrooms. *Research in the Schools, 14*(2), 27-36.

James-Burdumy, S., Mansfield, W., Deke, J., Carey, N., Lugo-Gil, J., Hershey, A., . . . Faddis, B. (2009). *Effectiveness of selected supplemental reading comprehension interventions: Impacts on a first cohort of fifth-grade students–Executive summary* (NCEE 2009-4058). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment*. Seattle, WA: Bill & Melinda Gates Foundation. Retrieved March 10, 2014, from the ERIC database. (ED540959)

Kieffer, M.J., Lesaux, N., Rivera, M., & Francis, D. J. (2009). Accommodations for English language learners taking large-scale assessments: A meta-analysis on effectiveness and validity. *Review of Educational Research, 29*(3), 1168-1201.

Klingner, J. K., & Vaughn, S. (1996). Reciprocal teaching of reading comprehension strategies for students with learning disabilities who use English as a second language. *The Elementary School Journal, 96*(3), 275-293.

Knapp, M. S., & Adelman, N. E. (1995). *Teaching for meaning in high-poverty classrooms*. New York, NY: Teachers College Press.

Knapp, M. S., Shields, P. M., & Turnbull, B. J. (1995). Academic challenge in high-poverty classrooms. *Phi Delta Kappan, 76*(10), 770-776.

Knight, S. L., & Ackerman, C. (1997). *Teaching For Meaning (TFM) classroom observation manual* [Measurement instrument]. College Station, TX: Texas A&M University.

Knight, S. L., & Smith, R. G. (2004). Development and use of a classroom observation instrument to investigate teaching for meaning in diverse classrooms. In H. C. Waxman, R. G. Tharp, & R. S. Hilberg (Eds.), *Observational research in U.S. classrooms: New approaches for understanding cultural and linguistic diversity* (pp. 97-121). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511616419.005

Kujawa, N., Cavazos, G., Meyer, D., Rodriguez, L., Lara-Alecio, R., Galloway, M., & Irby, B.J. (2001). Field-initiated research report August, 2001: A study addressing three components of the national bilingual research agenda for English language learners on high stakes assessment. Retrieved February 7, 2013, from http://repository.tamu.edu/handle/1969.1/92873

Landry, S. H., Crawford, A., Gunnewig, S., & Swank, P. R. (2001). *Teacher Behavior Rating Scale* [Measurement instrument]. Unpublished research instrument. Houston, TX: Center for Improving the Readiness of Children for Learning and Education, University of Texas Health Science Center at Houston.

Lara-Alecio, R., Cmajdalka, S. L., Parker, R. I., Cuellar, R. Y., & Irby, B. J. (1996). *A three year study of a new pedagogical theory/model in a bilingual education program using mathematics as a vehicle of instruction.* Paper presented at the Annual Meeting of the American Educational Research Association. New York, NY.

Lara-Alecio, R., & Irby, B. J. (1996). Bilingual education and multicultural education: An inclusively oriented delivery system. *Journal of Educational Issues of Language Minority Students, 17*, 11-24.

Lara-Alecio, R., & Irby, B. J. (2010). Project MSSELL: A randomized longitudinal study, 5th grade report. Retrieved March 20, 2013 from http://mssell.tamu.edu/research.html

Lara-Alecio, R., Irby, B. J., & Tong, F. (2011, April). *Middle school science for English language and literacy acquisition (MSSELL): A randomized trial longitudinal discovery research k-12 (DR-K12) study.* Paper presented with CADRE at the Annual Meeting of the American Educational Research Association, New Orleans, LA. Retrieved May 5, 2013, from http://www.aera.net/repository

Lara-Alecio, R., Irby, B. J., Tong, F., Francis, D., Rodriguez, L., Guerrero, C., & Mansfield, T. (2010). Project MSSELL: Advisory board meeting, 6th grade

implementation. Retrieved March 20, 2013 from http://mssell.tamu.edu/research.html

Lara-Alecio, R., & Parker, R. I. (1994). A pedagogical model for transitional English bilingual classrooms. *Bilingual Research Journal, 18* (3&4), 119-133.

Lara-Alecio, R., & Tong, F. (2013). Collaborative research: R&D: A longitudinal trail study of middle school science for English language learners (project MSSELL). Retrieved September 27, 2014 from http://www.nsf.gov/awardsearch/showAward?AWD_ID=0822343&HistoricalAwards=false

Lara-Alecio, R., Tong, F., Irby, B. J., Guerrero, C., Huerta, M., & Fan, Y. (2012). The effect of an instructional intervention on middle school English learners' science and English reading achievement. *Journal of Research in Science Teaching, 49* (8): 987–1011. doi: 10.1002/tea.21031

Lara-Alecio, R., Tong, F., Irby, B. J., & Mathes, P. (2009) Teachers' pedagogical differences during ESL block among bilingual and English-immersion kindergarten classrooms in a randomized trial study. *Bilingual Research Journal, 32*(1), 77-100.

Lara-Alecio, R., Tong, F., Irby, B. J., Mathes, P., Guerrero, C., Quiroz, A., & Cox, K. (2007). *Using an observation protocol in bilingual and ESL classrooms in institute for education science-funded research.* Paper presented at the Annual Meeting of the American Educational Research Association. Chicago, Illinois.

Lee, L. (2012). *Voices less heard: Teachers' perspectives on the evaluation of teacher effectiveness through the use of value-added modeling.* (Doctoral dissertation). Retrieved February 8, 2014, from ProQuest Dissertations & Theses Full Text database. (3550569)

Lee, O., Deaktor, R. A., Hart, J. E., Cuevas, P., & Enders, C. (2005). An instructional intervention's impact on the science and literacy achievement of culturally and linguistically diverse elementary students. *Journal of Research in Science Teaching, 42*(8), 857-887.

Lee, O., & Fradd, S. H. (1998). Science for all, including students from non-English-language backgrounds. *Educational Researcher*, *27*(4), 12-21. doi:10.3102/0013189X027004012

Lee, O., Maerten-Rivera, J., Penfield, R., LeRoy, K, & Secada, W.G. (2008). Science achievement of English language learners in urban elementary schools: Results of a first-year professional development intervention. *Journal of Research in Science Teaching, 45(1),* 31-52.

Lembke, J. (1990). *Talking science: Language, learning and values*. Norwood, NJ: Ablex Publishing Corporation.

Lewis, E. M., Ross, S. M., & Alberg, M. (1999, July). *School observation measure: Reliability analysis*. Memphis, TN: Center for Research in Educational Policy. The University of Memphis.

Littell, J., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. New York, NY: Oxford University Press.

Love, J. M., Atkins-Burnett, S., Vogel, C., Aikens, N., Xue, Y., Mabutas, M., . . . Sonnenfeld, K. (2009). *Los Angeles universal preschool programs, children served, and children's progress in the preschool year: Final report of the first 5 LA universal preschool child outcomes study*. Princeton, NJ: Mathematica Policy Research Inc. Retrieved February 8, 2013, from the ERIC database. (ED533140)

Luykx, A., & Lee, O. (2007). Measuring instructional congruence in elementary science classrooms: Pedagogical and methodological components of a theoretical framework. *Journal of Research in Science Teaching, 44*(3), 424-447.

Martinez, J. F., Bailey, A. L., Kerr, D., Huang, B. H., & Beauregard, S. (2010). *Measuring opportunity to learn and academic language exposure for English language learners in elementary science classrooms.* Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST). (CRESST report 767)

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.

Mashburn, A., & Pianta, R. (2006). Social relationships and school readiness. *Early Education and Development, 17*(1) 151-176. doi:10.1207/s15566935eed1701_7

Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis, 21*(1), 29-45.

Medley, D. M. (1992). Structured observation. In M. C. Alkin (Ed.), *Encyclopedia of educational research* (6th ed., pp. 1310-1315). New York, NY: Macmillan.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Education measurement* (3rd ed., pp.13-103). New York, NY: Macmillan.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13-23.

Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice, 14*(4), 5-8.

Meyer, D. J. (2000). *Evaluating urban elementary bilingual classrooms through the four-dimensional transitional bilingual pedagogical theory*. (Doctoral dissertation). Retrieved February 12, 2013, from ProQuest Dissertations & Theses Full Text. (304647007)

Mishler, E. G. (1990). Validation in inquiry-guided research: The role of exemplars in narrative studies. *Harvard Educational Review, 60*, 415–442.

Moll, L. C., Amanti, C., Neff, D., & Gonzalez, N. (1992). Funds of knowledge for teaching: Using qualitative approach to connect homes and classrooms. *Theory into Practice*, *31*(2), 132-141.

National Center for Education Statistics (NCES). (2006). *The condition of education 2006* (NCES 2006-071). Washington, DC: U.S. Government Printing Office.

National Center for Education Statistics (NCES). (2011). *The condition of education 2011* (NCES 2011-033). Retrieved August 1, 2013, from http://nces.ed.gov/pubs2011/2011033.pdf.

National Research Council. (2008). Early childhood assessment: Why, what, and how. Committee on Development Outcomes and Assessments for Young Children, C.

E. Snow and S. B, Van Hemel (Eds.). Board on Children, Youth, and Families, Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Neuman, S. B., Dwyer, J., & Koh, S. (2007). *Child/Home Early Language & Literacy Observation tool (CHELLO)* [Measurement instrument]. Baltimore, MD: Brookes Publishing.

No Child Left Behind (NCLB) Act of 2001. (2002, January 8). *Public Law 107-110, 115 Stat. 1425.* Retrieved July 19, 2014, from http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf

Norris, S. P., & Phillips, L. M. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education*, *87*, 224-240.

O'Donnell, C. R., & Tharp, R. G. (1990). Community intervention guided by theoretical development. In A. S. Bellack, M. Hersen, & A. E. Kazdin (Eds.), *International handbook of behavior modification and therapy* (2nd ed., pp. 251-266). New York, NY: Plenum Press.

Padrón, Y. (1994). Comparing reading instruction in Hispanic/limited English proficient schools and other inner-city schools. *Bilingual Research Journal, 18*, 49-66.

Padrón, Y. N., Waxman, H. C., & Huang, S. L. (1999). Classroom and instructional learning environment differences between resilient and non-resilient elementary school students. *Journal of Education for Students Placed at Risk of Failure*, 4(1), 63-81.

Parker, R., Tindal, G., & Hasbrouck, J. (1994). *Observing activity structures in classrooms: The activity structure scale.* Unpublished manuscript, Department of Special Education and Clinical Sciences, University of Oregon, Eugene, OR.

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). New York, NY: Cambridge University Press.

Pianta, R. C., LaParo, K. M., & Hamre, B. K. (2011). *Quality counts: Striving for five star early learning, CLASS* [PDF document]. Retrieved September 1, 2013, from http://www.elcmdm.org/QualityCounts/SPMMay23.pdf

Porta, M. (2008). *A Dictionary of Epidemiology* (5th ed.). New York, NY: Oxford University Press, Inc.

Ramírez, J. D., Yuen, S.D., Ramey, D.R., & Pasta, D.J. (1991). *Final report: Longitudinal study of structured English immersion strategy, early-exit and late-exit transitional bilingual education programs for language-minority children. Vol. 1* (Report No. 300-87-0156). San Mateo, CA: Aguirre International. Retrieved September 1, 2013, from http://eric.ed.gov/PDFS/ED330216.pdf

Ramos, E. (2010). Let us in: Latino underrepresentation in gifted and talented programs. *Journal of Cultural Diversity, 17*(4), 151-153.

Rivera, H. H., & Tharp, R. G. (2004). Sociocultural activity settings in the classroom: A study of a classroom observation system. In H. C. Waxman, R. G. Tharp, & R. S. Hilberg (Eds.), *Observational research in U.S. classrooms: New approaches for understanding cultural and linguistic diversity* (pp. 205-230). New York, NY: Cambridge University Press.

Rivera, H., & Tharp, G. (2010). *Dual Language Activity Setting Observation System (DLASOS)* [Measurement instrument]. Unpublished instrument. Retrieved April 1, 2014, from authors.

Rivera, H., Tharp, R. G., Youpa, D., Dalton, S., Guardino, G., & Lasky, S. (1999). *ASOS: Activity Setting Observation System coding and rulebook*. Center for Research on Education Diversity and Excellence, Santa Cruz, University of California.

Rivera, H., Tharp, R., Youpa, D., Dalton, S., Guardino, G., & Lasky, S. (2005). *Activity Setting Observation System (ASOS)* [Measurement instrument]. Center for Research on Education Diversity and Excellence, UC Berkeley: Publisher: eScholarship University of California http://escholarship.org/uc/item/5010t7g4

Rodriguez, L., Lara-Alecio, R., Galloway, M., & Irby, B. J. (2002). Field-initiated research report AERA, April 2002: A study addressing three components of the national bilingual research agenda for English language learners on high stakes assessment. Retrieved September 1, 2013, from http://repository.tamu.edu/handle/1969.1/92874

Ross, S. M., & Smith, L. J. (1996). *Classroom observation measure observer's manual*. Memphis, TN: University of Memphis/Memphis City Schools.

Ross, S. M., Alberg, M., & Wang, L. W. (1998). *The impacts of alternative school restructuring designs on at-risk learners: A longitudinal study* (Progress Report to U. S. Department of Education of At-Risk Students). Memphis, TN: Center for Research in Educational Policy, University of Memphis.

Ross, S. M., Smith, L. J., Lohr, L. L., & McNelis, M. J. (1994). Math and Reading instruction in tracked first-grade clases. *The Elementary School Journal, 95*(1), 105-109.

Ross, S. M., Smith, L. J., & Alberg, M. (1999). *The School Observation Measure (SOM)* [Measurement instrument]. Memphis, TN: Center for Research in Educational Policy, The University of Memphis.

Ross, S. M., Smith, L. J., Lowther, D. L., Alberg, M. J., & Cheon, J. (2006). *The school observation measure: Observer's manual.* Memphis, TN: Center for Research in Educational Policy, The University of Memphis. Retrieved June 14, 2014, from http://edit.educ.ttu.edu/site/jcheon/manual/SOM_manual.pdf

Ross, S. M., Smith, L. J., Alberg, M., & Lowther, D. (2004). Using classroom observation as a research and formative evaluation tool in education reform: The school observation measure. In H. C. Waxman, R. G. Tharp, & R. S. Hilberg (Eds.), *Observational research in U.S. classrooms: New approaches for understanding cultural and linguistic diversity* (pp. 144-173). New York, NY: Cambridge University Press.

Rowley, G. L. (1978). The relationship of reliability in the classroom research to the amount of observation: An extension of the Spearman-Brown formula. *Journal of Educational Measurement, 15*, 165-180.

Ryoo, K. (2010, May). *Learning science, talking science: The impact of technology-enhanced curriculum on students' science learning in linguistically diverse*

classrooms. A paper presented at the annual meeting of the American Educational Research Association, Denver, CO.

Saini, M., & Shlonsky, A. (2012). *Systematic synthesis of qualitative research*. New York, NY: Oxford University Press.

Saunders, W. M., Foorman, B. R. & Carlson, C. D. (2006). Is a separate block of time for oral English language development in programs for English learners needed? *Elementary School Journal, 107*(2), 181-198.

Scanlon, D. M., & Vellutino, F. R. (1996). Prerequisite skills, early instruction, and success in first-grade reading: Selected results from a longitudinal study. *Mental Retardation and Development Disabilities Research Reviews, 2*(1), 54-63. doi: 10.1002/(SICI)1098-2779(1996)2:1<54::AID-MRDD9>3.0.CO;2-X

Schunk, D. H. (1992). Theory and research on student perceptions in the classroom. In D. H. Schunk & J. L. Meece (Eds.), *Student perceptions in the classroom* (pp. 3-23). Hillsdale, NJ: Erlbaum.

Seale, C. F. (1999). *The quality of qualitative research*. London, UK: Sage.

Servin, B. (1983). *Spanish-English bilingual pupils' reactions toward three varieties of Spanish along various attitude dimensions*. (Doctoral dissertation). Retrieved December 10, 2012, from ProQuest Dissertations & Theses Full Text. (303119144)

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Slavin, R. E., & Cheung, A. (2005). A synthesis of research on language of reading instruction for English language learners. *Review of Educational Research, 75*(2), 247-284.

Smith, M. W., Brady, J. P., & Anastasopoulos, L. (2008). *User's Guide to the Early Language & Literacy Classroom Observation: Pre-K Tool*. Baltimore, MD: Paul H. Brookes Publishing Company.

Smith, M. W., & Dickinson, D. K. (2002). *Early Language and Literacy Classroom Observation (ELLCO) toolkit, research edition* [Measurement instrument]. Baltimore, MD: Brookes Publishing Co.

Smith, M. W., & Dickinson, D. K., with Sangeorge, A., & Anastasopoulos, L. (2002). *Early Language & Literacy Classroom Observation (ELLCO) toolkit, research edition* [Measurement instrument]. Newton, MA: Education Development Center.

Smith, S. (2004). *Pilot of a new classroom assessment instrument: Supports for social-emotional growth* [unpublished study]. New York: Child and Family Policy Center, Steinhardt School of Education, Culture, and Human Development, New York University.

Smith, S., Davidson, S., Weisenfeld, G., & Katsaros, S. (2001). *Supports for Early Literacy Assessment (SELA)* [Measurement instrument]. New York, NY: New York University.

Snow, C.E. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND. Retrieved November 13, 2012, from http://www.rand.org/pubs/monograph_reports/2005/MR1465.pdf

Sprachman, S., Caspe, M., & Atkins-Burnett, S. (2008). *Language Interaction Snapshot (LISn)* [Measurement instrument]. Princeton, NJ: Mathematica Policy Research.

Stallings, J. A. (1973). *Follow through program classroom observation evaluation 1971-72.* Menlo Park, CA: Stanford Research Institute.

Stallings, J., & Freiberg, J. H. (1991). *Observation for the improvement of teaching. Publication series 91-3.* Philadelphia, PA: Temple University, Center for Research in Human Development and Education. Retrieved December 10, 2012, from ERIC database. (ED360412)

Stallings, J. A., & Mohlman, G. G. (1988). Classroom observation techniques. In J. P. Keeves (Ed.), *Educational research, methodology and measurement: An international handbook* (pp. 469-471). Oxford, UK: Pergamon Press.

Stevens, F. I. (1993). Applying an opportunity-to-learn conceptual framework to the investigation of the effects of teaching practices via secondary analyses of multiple-case-study summary data. *Journal of Negro Education, 62*(3), 232-248.

Stipek, D., & Byler, P. (2004). The early childhood classroom observation measure. *Early Childhood Research Quarterly, 19*(3), 375-397. http://dx.doi.org/10.1016/j.ecresq.2004.07.007

Stringfield, S., Datnow, A., & Ross, S. (1998). *Scaling up school restructuring in multicultural, multilingual contexts: Early observations from Sunland County*

(Research Report No. 2). Santa Cruz: Center for Research on Education, Diversity & Excellence, University of California.

Texas Education Code. (1995). *Texas education code–section 29.056. Enrollment of students in program.* Retrieved August 1, 2013, from http://law.onecle.com/texas/education/29.056.00.html

Texas Education Code. (1996). *Texas education code–section 89.1201. Adaptations for special populations.* Retrieved March 9, 2015, from http://ritter.tea.state.tx.us/rules/tac/chapter089/ch089bb.html

Tharp, R. G. (2005). *Activity setting observation system rule book* [Measurement instrument]. Santa Cruz, CA: Center for Research on Education, Diversity & Excellence, UC Berkeley.

Tharp, R. G., Estrada, P., Dalton, S. S., & Yamauchi, L. (2000). *Teaching transformed: Achieving excellence, fairness, inclusion, and harmony*. Boulder, CO: Westview Press.

Tharp, R. G., & Gallimore, R. (1988). *Rousing minds to life. Teaching, learning, and schooling in social context.* Cambridge: Cambridge University Press.

Thomas, W. P., & Collier, V. P. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement*. Santa Cruz, CA, and Washington, DC: Center for Research on Education, Diversity & Excellence. Retrieved December 10, 2012, from http://www.usc.edu/dept/education/CMMR/CollierThomasComplete.pdf

Thomas, W. P., & Collier, V. P. (2003). The multiple benefits of dual language. *Educational Leadership, 61*(2), 61-64.

Thompson, B. (2006). *Foundations of behavioral statistics: An insight-based approach*. New York, NY: The Gilford Press.

Thompson, B., Diamond, K. E., McWilliam, R., Snyder, P., & Snyder, S. W. (2005). Evaluating the quality of evidence from correlational research for evidence-based practice. *Exceptional Children, 71*(2), 181-194.

Thordardottir, E. (2010). Towards evidence-based practice in language intervention for bilingual children. *Journal of Communication Disorders*, 43(6), 523-537. doi: 10.1016/j.jcomdis.2010.06.001

Tikunoff, W. J., Ward, B. J., van Broekhuizen, L. D., Romero, M., Castaneda, L. V., Lucas, T., & Katz, A. (1991). *A descriptive study of significant features of exemplary special alternative instructional programs*. Los Alamitos, CA: The Southwest Regional Educational Laboratory.

Tong, F., Irby, B., Lara-Alecio, R., & Mathes, P. (2008). The effectiveness of a three-year longitudinal randomized trial study on Hispanic bilingual students' English and Spanish acquisition. *Hispanic Journal of Behavioral Sciences, 30*(4), 500-529.

Torgerson, C. J. (2003). *Systematic Reviews*. London, UK: Continuum Books.

Vaughn, S., Hughes M. T., Schumm, J. S., & Klingner, J. K. (1998). A collaborative effort to enhance reading and writing instruction in inclusion classrooms. *Learning Disability Quarterly, 21*(1), 57–74.

Vaughn, S., Mathes, P., Linan-Thompson, S., Cirino, P., Carlson, C., Pollard-Durodola, S., . . . Francis, D. (2006). Effectiveness of an English intervention for first-grade English language learners at risk for reading problems. *The Elementary School Journal, 107*(2), 153-180. doi:10.1086/510653

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. In M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Trans.), *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Waxman, H. C., & Huang, S. L. (1999). Classroom observation research and the improvement of teaching. In H. C. Waxman & H. J. Walberg (Eds.), *New directions for teaching practice and research* (pp. 107-129). Berkeley, CA: McCutchan.

Waxman, H. C., & Padrón, Y. N. (2004). The uses of the classroom observation schedule to improve classroom instruction. In H. C. Waxman, R. G. Tharp & R. S. Hilberg (Eds.), *Observational research in U.S. classrooms: New approaches for understanding cultural and linguistic diversity* (pp. 72-96). New York, NY: Cambridge University Press.

Waxman, H.C., Tharp, R. G., & Hilberg, R.S. (2004). *Observational research in U.S. classrooms: New approaches for understanding cultural and linguistic diversity*. New York, NY: Cambridge University Press.

Waxman, H. C., Wang, M. C., Lindvall, C. M., & Anderson, K. A. (1983). *Teachers roles observation schedule technical manual*. Pittsburgh, PA: University of Pittsburgh, Learning Research and Development Center.

Waxman, H. C.,Wang, M. C., Lindvall, C. M., & Anderson, K. A. (1990a). *Classroom Observation Schedule technical manual* (revised ed.) [Measurement instrument]. Philadelphia, PA: Temple University, Center for Research in Human Development and Education.

Waxman, H. C.,Wang, M. C., Lindvall, C. M., & Anderson, K. A. (1990b). *Teacher Roles Observation Schedule technical manual* (revised ed.) [Measurement instrument]. Philadelphia, PA: Temple University, Center for Research in Human Development and Education.

Weinstein, R. S. (1989). Perceptions of classroom processes and student motivation: Children's views of self-fulfilling prophecies, In C. Ames & R. Ames (Eds.), *Research on motivation in education: Goals and cognition* (Vol. 3, pp. 187-221). San Diego, CA: Academic Press.

Wellington, J. & Osborne, J. (2001). *Language and literacy in science education*. Buckingham, UK: Open University Press.

Whitacre, M., Díaz, Z, & Esquierdo, J. (2013). Pre-service teachers: An analysis of reading instruction in high needs districts dual language classrooms. *International Journal of Instruction, 1*(6), 5-10.

Wright, W. E. (2005). *Evolution of federal policy and implications of No Child Left Behind for language minority students: Policy brief.* Retrieved July 19, 20014, from http://files.eric.ed.gov/fulltext/ED508474.pdf

APPENDIX A

Criterion-Based Assessment: Methodological Quality Questionnaire

Criterion 1 Theoretical or conceptual definition Page no.	Rating No = 0 pts. Yes = 1 pt.	Extension – Statement (2 pts.)
1. Was the construct or phenomenon of interest theoretically or conceptually defined? *(If not, at minimum the theory must be named or standards cited.)* Rationale:	__No __Yes	1. The characteristics of the construct of interest or the relationship between the parameters were clearly defined. Rationale: _Disagree __Agree
Criterion 2 Operational definition Page no.	Rating No = 0 pts. Yes = 1 pt.	Extension – Statement (2 pts.)
2. Was the construct defined operationally? *(A statement of how the variable(s) corresponding to the construct was measured must be provided.)* Rationale:	__No __Yes	2a. Qn: The process of how the construct was measured was described (the corresponding variables). 2b. Ql: the measurement process comprised describing how the research was conducted, what the sampling methods were employed, and how and where the data was collected was described. Rationale: __Disagree __Agree
Criterion 3 Research design Page no.	Rating No = 0 pts. Yes = 1 pt.	Extension – Statement (2 pts.)
3. Was the research design described? *(If not, at minimum, the research design must be named.)* Rationale:	__No __Yes	3. The research design must be grounded in or linked to the research question or hypothesis. Rationale: _Disagree __Agree
Criterion 4 Sampling Design Page no.	Rating No = 0 pts. Yes = 1 pt.	Extension – Statement (2 pts.)
4. Was a sampling method/strategy named or described? Rationale:	__No __Yes	4. The sampling method was described in enough detail to be replicable. Rationale: __Disagree __Agree
Criterion 5 Sample Page no.	Rating No = 0 pts. Yes = 1 pt.	Extension – Statement (2 pts.)
5. Was the sample well-characterized? *(Description must include all of the following: age/grade, race/ethnicity, language groups and/or English proficiency classification [ELL/LEP])* Rationale:	__No __Yes	5. The socio-demographic and other characteristics of the sample and context (e.g., school, year) that might influence the constructs of interest were listed and described in detail. Rationale: __Disagree __Agree

Criterion-Based Assessment Continued

Criterion 6 Evidence of reliability and validity (Qn); trustworthiness, credibility, and dependability (Ql) Page no.	Rating No = 0 pts. Yes = 1 pt.	Extension – Statement (2 pts.)
6a. Qn: Was evidence of reliability and validity provided for data collected? Rationale: 6b. Ql: Were trustworthiness, credibility, and/or dependability addressed? *(Researcher[s] must address either trustworthiness or credibility.)* Rationale:	__No __Yes	6a. Qn: Information about instrument development and adaptations for specialized populations (e.g., language translations) or short versions were reported. 6b. Ql: Triangulation, data saturation, and/or member checking were discussed. Rationale: __Disagree __Agree
Criterion 7 Data analysis Page no.	Rating No = 0 pts. Yes = 1 pt.	Extension – Statement (2 pts.)
7. Was the data analysis plan consonant with the research question and design? *(The data analysis techniques must be appropriate for the research design.)* Rationale:	__No __Yes	7. Data analysis rendered usable data for interpretation and application in educational practices. Rationale: __Disagree __Agree
Criterion 8[a] High-Stakes Test Page no.	Rating No = 0 pts. Yes = 1 pt.	Extension – Statement (2 pts.)
8. Was the COI described in terms of implications for students or teaching/learning practices? Rationale:	__No __Yes	8. Evidence was provided that identified impacts on target population (e.g., ELL), evidence-based practices, or curricula. Rationale: __Disagree __Agree
Criterion 9[a] Policy Implications Page no.	Rating No = 0 pts. Yes = 1 pt.	Extension – Statement (2 pts.)
9. Did the policy implications on ELL flow from the findings? Rationale:	__No __Yes	9. Policy implications for ELL and issues resulting from high-stakes testing were discussed and clarified with specific suggestions for policymakers and school districts. Rationale: _Disagree __Agree
[a] Criteria 8 and 9. These are criteria whose wording would change according to the research question; however both evaluated how successfully researchers' linked their findings to practitioners (Criterion 8) and policy (Criterion 9). Note: These criteria do not address whether limitations were included or not in the report. Limitations were addressed separately in the findings. They were not included in the methodological quality score (MQS).		

Criterion-Based Assessment Continued

Methodological Quality Questionnaire Score Summary		
Criteria		*Score*
C1	Theoretical or conceptual definition	
C2	Operational definition	
C3	Research design	
C4	Sampling Design	
C5	Sample	
C6	Validity and reliability evidence	
C7	Data analysis	
C8	High-Stakes Test	
C9	Policy Implications	
Total Score (Maximum score = 27)		

APPENDIX B

Transitional Bilingual Observation (TBO) Model Pedagogy Codes
(Lara-Alecio & Parker, 1994)

Teacher Behaviors	Activity Structure Descriptions
Lectures (Lec)	teacher lectures instructing students about content/subject matter/skills, presents info verbally or on chart, overhead, or AV materials, explains how something works
Directs (Dir)	teacher gives directions, orders, directives, procedures to follow for academic assignments
Demonstrates (Dem)	teacher demonstrates or models desired student academic performance, demonstration/modeling something students will later perform themselves
Leads (Led)	teacher leads students through a desired performance while students perform the task with or slightly behind the teacher
Asks (Ask)	teacher verbally asks questions related to content/subject matter/skills; asks/directs students to perform a content/subject matter/skills related task. Teacher's behavior during a teacher-led/controlled discussion.
Evaluates (Ev)	any overt teacher behavior which is part of a judgment of correctness or quality of a content/subject matter/skills response or performance, including teacher giving academic feedback to students and making verbal corrections
Answers (Ans)	verbally answering content/subject matter/skills area questions from students; making clarifications. Teacher's behavior during a student led/controlled discussion
Observes (Obs)	observing or supervising students during academic activities including informal socializing with students, including those times when a teacher may be physically in the room but is not actively engaged in overt observation or supervision
Student Behaviors	**Activity Structure Descriptions**
Listens (Lis)	student is passively listening, watching
Asks (Ask)	student asking questions related to content/subject matter/skills. Student behavior during student-led/controlled discussion
Performs (Per)	student performs an academic task; a response to a directive; note-taking; paraphrasing
Answers (Ans)	fairly brief verbal response to a content/subject matter/skills area question. Student answers questions related to skill/subject area; student behavior during a teacher-led/controlled discussion

TBO Model Instructional Practice Codes Continued

Discovers (Dis)	discovering an answer to a content/subject matter/skills question or problem/ involves trial and error, exploratory learning. Students work individually
Cooperates (Cop)	cooperatively learning or helping each other, students work in groups of 2 or more
Non-Academic Activities	
Feedback (NA feed)	giving positive or negative verbal feedback to students about their non-academic behavior, includes activities related to discipline of students
Free Time (NA free)	free time or play
Transition (NA tran)	housekeeping-beginning and end-of-day activities including managerial routines such as taking attendance, collecting money, lunch count, cleaning desks, etc.: setting up or preparing for an activity, putting materials away. Also includes non-academic discussion, demonstration, directives for social behaviors which occur within the classroom
Interruption (NA int)	any interruption to the classroom instruction activity including fire drills, intercom messages, unplanned visitors, child becoming ill, etc.
Outside (NA out)	of the classroom-activity on the playground, hallway, bus area, cafeteria, in assemblies, etc.
Interactive Instruction (Interact)	teaching with active student responding, typical of direct instruction lessons. Teacher models, leads, tests students and students perform and orally respond to questions as an integral part of instruction

Activity Structure		
	Code	Teacher Behavior / Student Behavior
	1	lectures / listens
	2	lectures / performs
	3	directs / listens
	4	directs / performs
	5	demonstrates / listens
	6	leads / performs
	7	asks / performs
	8	asks / answers
	9	answers / asks
	10	evaluates / performs
	11	observes / performs
	12	evaluates / discovers
	13	evaluates / cooperates
	14	observes / discovers

TBO Model Instructional Practice Codes Continued

	15	observes / cooperates
	16	Not Applicable–feedback
	17	Not Applicable–free time
	18	Not Applicable–transition
	19	Not Applicable–interruption
	20	Not Applicable–outside
	21	interactive instruction
Mode		
	Code	Description
	1	Writing
	2	Reading
	3	Aural
	4	Verbal
	5	writing–reading
	6	writing–aural
	7	writing–verbal
	8	reading–writing
	9	reading–aural
	10	reading–verbal
	11	aural–writing
	12	aural–reading
	13	aural–verbal
	14	Verbal–writing
	15	verbal–reading
	16	verbal–aural
	17	Aural–reading–verbal
	18	Not Applicable (NA)
Language Content		
	Code	Description
Social Routines (Social)	1	social exchanges and conversation
Academic Routines (Academic)	2	preparing for recess, returning books, learning strategies, handing in assignments, structuring homework
Light Cognitive (Light Cog)	3	current events, discussion of the school fiesta, multicultural education issues, repetitive drill or skills practice, reviewing content already introduced
Dense Cognitive (Dens Cog)	4	new content-area information, conceptually loaded communication with specialized vocabulary and procedures
Language of	Code	Description

TBO Model Instructional Practice Codes Continued

Instruction		
Content Presented in L1 (L1)	1	(native language)-indicates Spanish-only introduction, a beginning point for students with very low English-proficiency
Content Presented in L2 (L2)	2	(second language)-indicates English-only instruction
L1 Introduces L2 (L1-2)	3	indicates instruction primarily in L1, but additionally, English vocabulary is taught for key ideas, concepts, and procedures
L2 Clarified by L1 (L2-1)	4	indicates instruction primarily in English, but with L1 used as "back-up" as needed to ensure understanding
	5	Not Applicable (NA)

Source. Lara-Alecio, R., and Irby, B. J. (2010). Project MSSELL: A randomized longitudinal study, 5th grade report. Retrieved March 20, 2013 from http://mssell.tamu.edu/research.html

APPENDIX C

Systematic Review Coding sheet

	Initial Study Characteristics	
1	Is the study conducted in the U.S.?	Yes or No
2	Is a classroom observation instrument described or referenced?	Yes or No
	Sample Target	
3	Are there English Language Learners (ELL) in the study?	Yes or No
4	Is the ELL sample in the pre-K through grade 12 range?	Yes or No
5	Are the ELL Spanish speaking?	Yes or No
6	Do the ELL make up more than 50% of the study sample?	Yes or No
	The Classroom Observation Instrument (COI)	
5	Does the study describe a COI used in conjunction with ELL?	Yes or No
		Fill in the Blank Response
6	What is the name of the COI?	
7	Provide the COI citation reference.	
8	What is the purpose of the COI as described by researchers?	
9	What are the COI measures and constructs?	
10	How does the COI address ELL in the classroom?	
11	What COI theoretical framework or model depiction is described?	
12	What COI language theory is described?	
13	What COI pedagogical theory is described?	
14	What COI psychometric properties are described in the study (e.g., reliability and validity)?	
15	What other COI are described in the study (i.e., name additional COI mentioned in the study)?	