

Spring 2018

Examining The Effectiveness Of An Esl Teacher Observation Tool

Hetal Ascher
Hamline University

Follow this and additional works at: https://digitalcommons.hamline.edu/hse_all



Part of the [Education Commons](#)

Recommended Citation

Ascher, Hetal, "Examining The Effectiveness Of An Esl Teacher Observation Tool" (2018). *School of Education Student Capstone Theses and Dissertations*. 4405.

https://digitalcommons.hamline.edu/hse_all/4405

This Thesis is brought to you for free and open access by the School of Education at DigitalCommons@Hamline. It has been accepted for inclusion in School of Education Student Capstone Theses and Dissertations by an authorized administrator of DigitalCommons@Hamline. For more information, please contact digitalcommons@hamline.edu, lterveer01@hamline.edu.

EXAMINING THE EFFECTIVENESS OF AN
ESL TEACHER OBSERVATION TOOL

by

Hetal Ascher

A capstone submitted in partial fulfillment of the requirements for the degree of a Master
of Arts in English as a Second Language

Hamline University
Saint Paul, Minnesota

May 2018

Primary Advisor: Ann Mabbott

Secondary Advisor: Kristina Robertson

Peer Reviewers: Laura Sepannen and Sarah O'Rourke

Table of Contents

Chapter One: Introduction.....	1
Context.....	2
Development of the ESL Observation Tool.....	3
Guiding Questions.....	5
Overview of Chapters.....	5
Chapter Two: Literature Review.....	7
Language Instruction vs. Mainstream Instruction.....	7
Teacher Evaluation.....	10
The Gap.....	20
Research Questions.....	21
Summary.....	22
Chapter Three: Methodology.....	23
Overview of Chapter.....	23
Mixed Methods Research Paradigm.....	24
Procedure.....	25
Data Analysis.....	29
Verification of Data.....	30
Ethics.....	30
Chapter Four: Results.....	32
Research Question One.....	32
Research Question Two.....	37
Chapter Five: Conclusion.....	42
Findings.....	42
Limitations.....	43
Implications.....	44
Suggestions for Further Research.....	45
Dissemination.....	45
Appendix ESL Teacher Observation Tool.....	47
References.....	50

List of Tables

Table 1 Minnesota Graduation Rates 2016.....20
 Table 2 Rubric Scores by Participant.....33
 Table 3 Average Rubric Score by Participant.....37

List of Figures

Figure 1 Minnesota Graduation Trend Comparison.....20
 Figure 2 Procedure.....29
 Figure 3 Rubric Score Averages.....37

Chapter One: Introduction

In my first three years at the school where I work as an English as a Second Language (ESL) teacher, I was observed 24 times. Nineteen of those times, I was observed by various administrators in my building. Out of those nineteen times, I never once received feedback about language instruction. As an ESL teacher, I knew that my primary role in the classroom was to teach language, more specifically, academic language (TESOL International Association, 2013). However, the expectations of my ESL supervisor did not always translate to the expectations of my building administrators. Because I received feedback often on other areas of teaching, I was able to grow and hone those specific skills. For example, I was struggling to make smooth transitions from activity to activity, and one of my observers recommended some resources. I bought a kitchen timer and my tasking problems were over. However, I was not able to make such feedback-based growth in language instruction because my building administrators simply did not notice the language instruction that was going on in my classroom.

Last year, my ESL supervisor sent out an invitation to all district ESL teachers to be on a team of teachers whose mission was to develop a tool for observing ESL teachers. Because I had personally felt the impact of receiving feedback on everything except my language instruction, I took the opportunity with enthusiasm and high hopes for impacting change in my district.

Our team began by discussing the the current system, the Charlotte Danielson Framework (2013). Danielson's (2013) rubric is organized into four parts: Planning and Preparation, Classroom Environment, Instruction, and Professional Responsibilities. As our team outlined the challenges of being observed using the Charlotte Danielson (2013) rubric, we realized that ESL teachers' jobs are significantly different than those of mainstream teachers. Our team consisted of teachers from various grade levels, various buildings, and teachers with a range of years of experience. Among other things, we taught in very different contents, and we worked in closets, corners, and even in other teachers' classrooms. The Danielson evaluation system was meant to evaluate mainstream classroom teachers and was simply not a good fit for ESL teachers. With the goal of developing a rubric aligned with the Danielson rubric, our team began by changing the title, instruction, to academic instruction, and then began distilling English language instruction best practices and integrating them into the ESL Teacher Observation Tool.

Context

The schools from this district are first-ring suburban schools. In the middle school, about 50% of the students qualify for free and reduced lunch, and 12% of the students are English learners. The top five language groups are Spanish, Hmong, Karen, Somali, and Nepali.

Currently at the secondary level, students are separated by World-class Instructional Design and Assessment Consortium (WIDA) levels. The levels are a gradient scale from 1 to 6; a level 1 speaker is able to use English in simple words or phrases with concrete ideas and by level 6, the student is nearly a native speaker

(Gottlieb, Cranley, Cammilleri, 2007). At the secondary level, ESL services are dependent on a students' WIDA level. Level 1, 2, and 3 students take sheltered content classes, which are smaller and specifically for English learners, in English, math, science, and social studies. Level 4 and 5 students take one or two co-taught classes, which are classes where a third of the class consists of English learners, and an ESL teacher and a mainstream teacher teach the content together.

Currently in my district, the Charlotte Danielson (2013) rubric is used to formally evaluate probationary teachers for a fifty-minute period of time three times a year, as well as three "drop-in" visits that can range from five to ten minutes three times a year. Tenured teachers are evaluated for one year out of every three years, and in the third year re-evaluation year, they are observed for a fifty-minute period three times a year, and three drop-in visits. Tenured teachers in their non-evaluation years are given five drop-in visits. During a formal evaluation year, typically, the building principals are responsible for two of the three visits, and the ESL supervisor is responsible for one of the visits. During non-evaluation years, the principal is responsible for the majority of the drop-in visits. The principals and teacher coaches use a program called ObserverTab, which contains the Charlotte Danielson rubric to quickly send an email to give feedback to the teachers.

Development of the ESL Observation Tool

The supervisor of the ESL program saw the need for an ESL specific observation tool because the Danielson rubric did not meet the diverse and specific needs of ESL teachers. She sent out an open request to all ESL teachers to participate on a development team for this evaluation tool. Her process was based on implementation science work

done by researchers at the University of North Carolina (Fixsen, Naoom, Blase, Friedman, Wallace, 2005). She selected teachers from all grade levels: lower elementary, upper elementary, middle school, and high school. The ESL teachers met after school several times throughout the year. The development team utilized various resources, including Charlotte Danielson (Danielson, 2013), SIOP (Vogt, & Short, 2012), WIDA Standards ("English Language Development (ESLD) Standards", 2012), and the WIDA *Essential Actions Handbook* (Gottlieb, 2013) to develop the categories and criteria for each element.

After the initial prototype was developed, the tool was shared with all the district ESL teachers at a district meeting. Each teacher from the development team worked with a group of teachers and collected feedback from the other ESL teachers in the district to gain perspective and buy-in. The development team then reworked the tool based on the feedback given from all of the teachers.

Subsequently, a few teachers, including me, volunteered to be observed during the pilot run using the tool. We worked with a building administrator with whom we had a long-standing relationship. We walked the administrator through the tool, and then the administrator practiced using the tool side-by-side with the ESL supervisor. We then met for a post-observation meeting discussing the evaluation. I found it to be a very enlightening experience because for the first time, my building principal understood the purposes behind what she was seeing during my observations. It felt gratifying to know that my work and specialist role was being acknowledged. Additionally, I was left feeling hopeful that next time I was observed, perhaps I would receive feedback about my language instruction and be able to grow as a language teacher.

After having used the tool and personally finding it to be useful, I want to see if it is effective in a more empirical sense. I would also like to know how evaluators perceive the usefulness of the tool. It is my hope to receive results that can have an impact on the professional growth and evaluation of ESL teachers.

Guiding Questions

If this tool is effective, it can be used in not only the district in which it was developed, but also in other districts that are looking for alternatives to their evaluation of ESL teachers, as well as providing a document that clearly outlines the expectations for ESL teachers.

There are multitudes of ways that a quality ESL instructional resource could be used; however, for the purpose of this study, the focus will be on the tool's use as an evaluation rubric. As more and more emphasis is placed on teacher evaluation, it is imperative that teachers are evaluated fairly. This study aims to begin to test the effectiveness of the tool, and uses the following two questions to guide the research: To what extent can the ESL Observation Tool assist evaluators in providing language instruction related feedback to ESL teachers? How does using an ESL specific observation tool compare to using Danielson?

Overview of Chapters

In Chapter One, I provided background, context, and rationale for this study. In Chapter One, I also provided the questions that will inform and guide the study. Chapter Two presents a literature review of the limited research that exists around ESL teacher evaluation and explores the complexities in teacher evaluation in a Midwestern state, as well as a research context for different models of teacher evaluation including Danielson,

SIOP, and independently created tools. The third chapter describes the mixed method study that was conducted. Chapter Four offers an analysis of the quantitative and qualitative data obtained from the study. The paper concludes with Chapter Five, a discussion of a plan for use of the outcomes of this study as well as recommendations for further study.

Chapter Two: Literature Review

The purpose of this study is to examine the effectiveness and practicality of the ESL Observation Tool, which was developed for use in a particular school district. In this district, almost all teachers are evaluated using the same Charlotte Danielson (2013) rubric. This literature review presents evidence that language instruction is distinctly different from mainstream instruction and as such requires its own evaluative tools and techniques. Then it offers a brief overview of the current state of teacher evaluation, including research on the Charlotte Danielson (2013) rubric. Since many schools around the world have noticed the need for ESL specific evaluation tools, a description of some of these evaluation tools will be presented, including the most popular ESL observation model, SIOP.

Language Instruction versus Mainstream Instruction

At first glance, one might see many similarities between mainstream instruction and language instruction; however, there are marked differences in content, pedagogy, and standards for teachers of English learners. For the purposes of this research, mainstream instruction will be defined as instruction delivered to a group of learners for whom English is their first language (Gandara, Maxwell-Jolly, Driscoll, 2005).

Teachers who teach ESL are prepared with specific language related coursework that is significantly different from mainstream teachers. For example, at Hamline University, ESL teachers are trained through courses such as Linguistics for Language Teachers, and Second Language Acquisition, whereas, a mainstream English teacher is

required to take courses about literature such as Literary and Cultural Theory (Additional License and Curriculum).

While examining any area of instruction, including reading, writing, vocabulary, and all core-contents, the research remains consistent. English Learners require different strategies and instruction than native-English speakers (Escamilla, 2009; Gottlieb, 2013). WIDA's *Essential Actions* handbook (2013) guides teachers to build stronger instruction through building academic language instruction into curriculum. Academic language instruction includes language production and instruction in three areas:

Linguistic Complexity—the organization, cohesion and relationship between ideas expressed in a variety of sentences that make up different registers, genres, and text types in oral or written language.

Language Forms and Conventions—the grammatical structures, patterns, syntax, and mechanics associated with sentence level meaning and use.

Vocabulary Usage—the specificity of words, phrases, or expressions, along with multiple meanings, cognates, and collocations, applied to given contexts.

(Gottlieb, 2013)

These three areas of academic language display the differences between typical mainstream instruction and language instruction. For example, it is common for content teachers to simply provide students with definitions and a sentence for content-specific vocabulary terms. However, in order to acquire new vocabulary, English learners require deeper context, oral practice, and non-content specific academic vocabulary instruction (Escamilla, 2009; Kinsella, 2005). As this example shows, the approach of an ESL teacher is significantly different from the approach of a mainstream teacher. It is possible

that these differences could result in frustration for mainstream teachers and administrators who assume that English learners can be taught in exactly the same way as mainstream students.

Sempek (2014) surveyed mainstream teachers to find if teachers had negative, neutral, or positive feelings toward educating English learners. He breaks down the data by demographic groups to see if there are certain groups that feel one way or another about aspects of teaching English learners. He finds that elementary teachers felt much more positively than secondary teachers about professional development for English learners. According to the survey, mainstream teachers report that they do not know how to keep their English learners from being overwhelmed. It seems these mainstream teachers are struggling with how to appropriately differentiate their instruction for their English Learners.

Taking the issue further, Turkan and Buzick (2016) have asserted that mainstream teachers should not be held responsible for the progress of English learners because mainstream teachers are not specifically trained to teach English Learners. They found that only 13% of teachers with English learners in their classes had received professional development on how to help their English learners succeed academically. Mainstream teachers around the United States are feeling the discomfort of being unable to meet the needs of the English Learners in the classroom. One conclusion that can be drawn from these studies is simply that language instruction is distinctly different from mainstream instruction. Another conclusion that will be explored in a later section is that mainstream teachers need tools to help them meet the needs of their English Learners.

In 2011, the Minnesota Department of Education approved WIDA (World-class Instructional Design and Assessment) as its English Language Development (ELD) standards (Minnesota Department of Education). The WIDA ELD standards are drastically different from the rest of the Minnesota content area standards. The Minnesota content area standards specifically identify what skills and concepts to teach, and the WIDA ELD standards are a framework for how to teach academic language. The WIDA ELD standards call for teachers to connect the WIDA language outcomes to the state standards of their state (WIDA 2012 Amplified Standards). In essence, it is a language standards framework that requires individualized analysis of academic language requirements by content standard and task. In Minnesota, ESL teachers teach a different set of standards, so now the question is, are they being evaluated based on those standards?

Teacher Evaluation

For a rich context in evaluation for ESL teachers, the purpose and role of teacher evaluation in general must be examined. The purpose of teacher evaluation has changed in recent history. Cohen and Goldhaber (2016) describe a time between the 1950s and 1980s when principals would observe classrooms, but the results would have little impact in staffing decisions. Recently, President Obama's Race to the Top program has renewed states' interest in quality teacher evaluation, and more specifically tying teacher evaluation to student growth scores (USDE, 2009). Now, there are two widely accepted purposes of teacher evaluation. Marzano (2012) describes one purpose of teacher evaluation as the process by which effective teachers are distinguished from ineffective teachers and the other purpose as a process which develops teachers' skills.

In recent years, this push to differentiate ineffective teachers from effective ones has become stronger in part due to studies which show correlation between student growth and teacher effectiveness. The three-year long Measures of Evaluation in Teaching (MET) Study conducted in 2013 confirms that student success is correlated with teacher effectiveness. The large-scale study with 3,000 math and English language arts teachers demonstrates that effective teachers yield student growth on standardized tests regardless of their abilities coming into the classroom (Bill and Melinda Gates Foundation, 2010; Bill and Melinda Gates Foundation, 2013). In light of these high stakes, valid and accurate measurement systems are essential for providing valid and accurate evaluations of teachers (Darling-Hammond, 2010).

Investigations like the MET Study have begun a trend of measuring teacher effectiveness through value-added measures, which is the process of using student test scores to measure teacher performance (Bill and Melinda Gates Foundation, 2010). However, this leads to an issue: not all subjects can be measured using the same type of tests (Cohen & Goldhaber, 2016). Additionally, certain subjects such as math are more predictive because typically a math teacher would be the only teacher teaching math, but subjects such as reading are more difficult to connect to the English teacher because English teachers are not the only ones that require reading to be done in class (Bill and Melinda Gates Foundation, 2010). Subsequently, ESL teachers face a similar issue as English teachers, because students should be learning language from their mainstream teachers and their ESL teachers. Classroom observations are more common than value-added measures in schools. Cohen and Goldhaber (2016) argue that classroom observations need more attention from researchers. Cohen and Goldhaber recount that

classroom observations have been happening since the 1950s, and maybe earlier, but there is limited research on the effectiveness of classroom observations.

Charlotte Danielson rubric. Charlotte Danielson's *Framework for Teaching* (2013) is the most popular rubric for teacher evaluation in the United States, with over 20 states having approved the model for use (www.danielsongroup.org). The framework is meant to be a generic instrument with content specific examples to guide evaluators (Danielson, 2013). The Danielson's *Framework for Teaching* (2013) is organized into four domains, which each contain several components, and are then described by specific elements. The following are the four domains and components:

Domain 1: Planning and Preparation

1a Demonstrating Knowledge of Content and Pedagogy

1b Demonstrating Knowledge of Students

1c Setting Instructional Outcomes

1d Demonstrating Knowledge of Resources

1e Designing Coherent Instruction

1f Designing Student Assessments

Domain 2: Classroom Environment

2a Creating an Environment of Respect and Rapport

2b Establishing a Culture for Learning

2c Managing Classroom Procedures

2d Managing Student Behavior

2e Organizing Physical Space

Domain Three: Instruction

- 3a Communicating with Students
- 3b Using Questioning and Discussion Techniques
- 3c Engaging Students in Learning
- 3d Using Assessment in Instruction
- 3e Demonstrating Flexibility and Responsiveness
- Domain Four: Professional Responsibilities
- 4a Reflecting on Teaching
- 4b Maintaining Accurate Records
- 4c Communicating with Families
- 4d Participating in the Professional Community
- 4e Growing and Developing Professionally
- 4f Showing Professionalism

(Danielson, 2013)

Overall, as the list suggests, the Danielson (2013) rubric is comprehensive, and it encompasses many elements of a teacher's job. However, Danielson recognizes that there is need for separate work to be done for specialist roles. Danielson has a separate rubric for assessing teachers of special education, but no other specialist role. Danielson's website provides a worksheet with guiding questions for working through the evaluation criteria of a specialist role.

Two separate studies reported that teachers generally find the Charlotte Danielson rubric to be a helpful and useful tool for evaluation and receiving feedback (Stewart, 2013; Batchelor, 2008). Neither of these studies describe having a significant population

of English learners and neither of these studies specify if the teachers they interviewed included ESL teachers.

Along with ESL teachers, there are other specialists who feel that the current more generic model of teacher evaluation does not fit them. As mentioned above, Charlotte Danielson (2013) now provides a separate rubric for evaluating special education teachers. When the *Framework for Teaching* was originally published in 1996, special educators did not have a rubric specific to their needs. The Council for Exceptional Children's Position on Special Education Teacher Evaluation (2013) argued for differentiated teacher evaluation based on the specific role of that teacher, and an understanding on the part of the evaluator of that teacher's specific role. They pushed for differentiated evaluations, meaning that the evaluation is similar to that of their peers, but actually assesses the specific role of that particular special education instructor.

In a similar way, Hunt, Gurvitch, and Lund (2016) question the effectiveness of teacher evaluation systems that use the same evaluation criteria to assess every teacher. The article highlights concerns with evaluation of physical education teachers and suggests some solutions, including changing the observer's criteria to match the objectives of the specific teacher. This article is written from a physical education standpoint, but the rationale could be applied to any content that is not the mainstream core. Hunt, Gurvitch, and Lund (2016) also contend that generic rubrics often lead to generic feedback which may not lead to teacher growth.

In the frenzy to distinguish effective teachers from ineffective ones, the second half of Marzano's purpose of teacher evaluation must not be forgotten. Teacher evaluation must also support and develop teachers' skills. Danielson (2012) illustrates a

point that for many teachers, evaluation is something that is fraught with fear. Rather than being a conversation that guides teachers to improving their craft, it has become prescriptive and punitive. She argues that evaluators must have the skills to have a conversation with teachers about their observations, and evaluation must be something that is done with teachers not to teachers.

Researchers Cherasaro, Brodersen, Reale, and Yanoski, at Marzano Research Laboratory (2016) conducted a study which supports the idea of working with educators to provide useful feedback. The goal of the study was to support teacher evaluations by understanding what teachers value in feedback. The authors find that teachers' perception of their evaluators' credibility was the most important factor in how useful they found the feedback. If the teachers find the evaluator credible, they also find the feedback useful and accurate. This finding emphasizes the need for evaluators to be credible; credibility is defined as, "The perception that the evaluator has the knowledge and understanding to give valuable feedback," (Cherasaro et al., 2016, p. 2). Such knowledge and understanding comes from the professional experiences of the evaluator. So, how does an evaluator without an understanding of second language instruction gain the skills necessary to evaluate ESL teachers?

ESL evaluation tools. There are a limited number of quality ESL evaluation tools in current use. In the US, most ESL teachers and mainstream teachers are evaluated using the same evaluation tool. Researchers Holdheide, Goe, Croft, Reschly at the National Comprehensive Center for Teacher Quality (2010) assert that an observer without an understanding of ESL pedagogy is unlikely to be capable of judging the effectiveness of

the teaching they are observing. However, there is a lack of resources and observation tools for language teachers in the United States and around the world.

Researchers and teachers in other countries have begun to develop tools for their ESL context. In Saudi Arabia, Al-Thumali (2011), an ESL teacher, designed an observation rubric based on the standards that are used to evaluate mainstream Saudi Arabian teachers. Her research purpose was to discover if experienced EFL teachers perform better than less experienced EFL teachers and to create a tool which can help provide feedback to EFL teachers in Saudi Arabia. There was no evaluation tool in existence, so she created her own rubric in order to proceed with her work. Some domains in her work do not apply to typical teaching in an ESL environment; for example, Al-Thumali's first domain evaluates the English language proficiency of the teacher. Another unique standard derived from Saudi Arabian and Egyptian standards involves applying moral and social teaching to language learning.

Other Saudi Arabian researchers, Tawalbeh and Ismail (2014), developed a checklist similar to the SIOP checklist to evaluate the ESL instructors at the university level. Tawalbeh and Ismail discovered a deep need for professional development around quality language instruction and assessment. They found that many of their teachers did not meet criteria in these two categories and the researchers were able to attribute it to a lack of certification in the areas. In this case, a valid checklist was able to not only diagnose an issue related to student needs, but also it was able to provide instructors with the resources needed in order to improve instruction.

ESL Teacher Evaluation tools should not only provide support to ESL teachers, but also provide support to any teacher who teaches English learners. Circling back,

Turkan and Buzick (2016) found that teachers do not believe they should be evaluated on the success of their English learners because they do not have the skills to teach English learners. While on the surface this notion may seem reasonable, it exposes an underlying belief that an English learner is beyond the reach of a mainstream classroom teacher.

However, teachers do need tools and strategies to teach their English learners and an English Language evaluation model may make it easy and accessible for mainstream teachers to teach their English learners. As a result, Echevarria and Short (2000) have designed the well-known Sheltered Instruction Observation Protocol (SIOP), which provides teachers and schools with a simplistic checklist of observable ESL practices.

SIOP. Sheltered Instruction Observation Protocol is a lesson-planning model designed for mainstream content teachers that emphasizes best practices in teaching English learners. The SIOP model contains eight components:

Lesson Preparation

Building Background

Comprehensible Input

Strategies

Interaction

Practice/ Application

Lesson Delivery

Review & Assessment

(Center for Applied Linguistics, 2017)

SIOP began as a federally- funded project meant to fuse together research on best practices for sheltered content teaching for English learners. The SIOP model was built

on research and currently there is a large pool of research that supports the use of SIOP. For instance, Echevarria, Short, and CREATE (2011) created various science units to be implemented using SIOP. They found that the classes that implemented SIOP performed far better than the control classes who did not implement SIOP (Echevarria, Short, & Center for Research on the Educational Achievement and Teaching of English Language learners, 2011). In another study, Short and Echevarria worked with a school which was implementing SIOP. Since the teachers knew they were going to be evaluated using the SIOP checklist, they decided to start planning using SIOP. This began the process of growing their understanding of what it means to teach English learners, and it also began to expand the uses of the SIOP model (Short & Echevarria, 1999).

Unlike other evaluation models which typically have the evaluator score as the lesson is occurring, SIOP demonstrates flexibility by providing three different ways that a lesson can be scored:

- 1) during the observation itself, as individual features are recognized
- 2) after the observation, as the observer reflects on the entire lesson, referring to observational field notes or
- 3) after the lesson while watching a videotape of the lesson.

The third option could be performed by the observer alone, yet it is also a useful way to involve the teacher in the scoring (Echevarria & Short, 2000).

Although SIOP seems to be a step forward in the right direction, there are some major flaws in the model. In his report, Krashen (2013) questions the validity of SIOP's claim to be research-based. He points out that the limited empirical studies that SIOP has conducted not only have design flaws, but also lack statistically significant results.

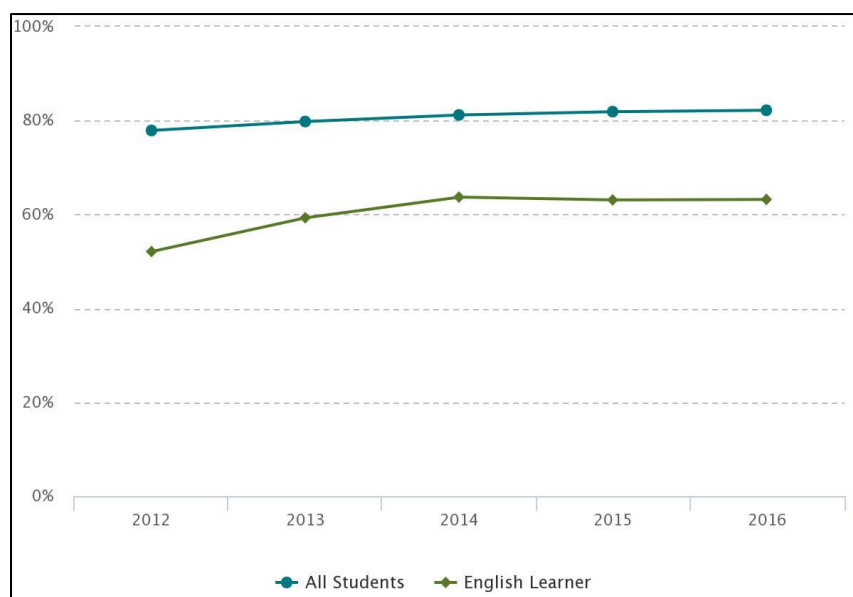
Additionally, according to Mabbott, a teacher educator, SIOP is weak in helping teachers understand how to teach language. For example, SIOP shows language objectives as modalities: Students can read (write, speak, listen) rather than any type of language structure (A. Mabbott, personal communication, April 23, 2018). Another major weakness of SIOP is that the authors of SIOP have conducted the majority of the research on SIOP, which is clearly not best practice in research.

Two studies conducted by independent researchers both uncovered flaws in the SIOP model. One high school teacher and researcher, Vidot (2011), conducted a mixed-method study on the impact of SIOP implementation in a 90% Non-Native English speaking school. Vidot used test score data and triangulated it with interviews, and observations using the SIOP checklist. Although Vidot's work generally wrote positively about SIOP, the implementation of SIOP did not yield in student achievement based on test data. Another researcher, Rodriguez (2010), asserted that although SIOP is useful for middle- school and high- school learners, there has been very little research done on SIOP for elementary learners. Rodriguez conducted a qualitative study to examine the effectiveness of SIOP at the elementary level. She found that although SIOP was very helpful for intermediate and advanced learners, it was not useful for beginning learners because SIOP assumes a certain level of English proficiency. Hence, there is a compelling amount of evidence that SIOP is not highly effective as an evaluation model for mainstream or ESL teachers.

The Gap

Currently in the state of Minnesota, English learners are not graduating from high school at rates similar to white students (87 percent). Only 63 percent of English learners graduate from Minnesota schools in four years, which shows a gap greater than most racial gaps and greater than the socioeconomic gap. Not only was the graduation gap wider than ever in 2016, it has stayed close to the same in

the few preceding years. The boost in 2013 may be attributed to graduation requirements changing, i.e. doing away with the GRAD test as a requirement of graduating. While the



graduation rates for other students is slowly increasing, Minnesota graduation rates for English learners remain stagnant. The conclusion is logical;

(Figure 1. Longitudinal graduation rate comparison, Minnesota Report Card 2016)

Table 1 *Minnesota Graduation Rates, 2016*

Demographic	Percentage of Graduates
All Students	82%
Asian/Pacific Islander	83%
White	87%
Free/RP Lunch	68%
Hispanic	65%
Black	65%
English Learner	63%
Special Education	60%
Indigenous	52%

(Minnesota Report Card, 2016)

we must begin to examine the quality of education that English learners are provided. First, most English learners spend the majority of their time in mainstream classrooms; therefore, the issue of English learners being able to access mainstream content is of concern. However, the method in which ESL teachers are evaluated also requires attention. Much research has been conducted on teacher evaluation in general (Bill and Melinda Gates Foundation, 2013); however, little research has been done specifically on teacher evaluation for teachers of English learners. As the research suggests, mainstream evaluation tools have not yet evolved to distinguish effective ESL teachers from ineffective ones, nor do those tools require mainstream teachers to pay attention to the specific needs of their English learners (Holdheid, Goe, Croft, Reschly, 2010).

The current evaluation system of Charlotte Danielson (2013) is not specific or focused enough for teachers of English learners, and SIOP alone is too shallow to be of use as an evaluation tool. The ESL Observation Tool pulls together Charlotte Danielson's work (2013), SIOP, as well as other well-researched best practices, and it is intended to apply to teachers of all grade levels and proficiency levels. This study calls attention to the methods in which ESL teachers are evaluated by examining the effectiveness of a specific observation tool created by ESL teachers in a Midwestern school district.

Research Questions

This literature review provides a foundation on which to begin to answer the following research questions: To what extent can the ESL Teacher Observation Tool assist evaluators in providing language instruction related feedback to ESL teachers? How does using an ESL specific observation tool compare to using Danielson (2013)?

Summary

This chapter reviewed literature distinguishing language instruction from mainstream instruction, and provided context for this study in teacher evaluation including a description of the current model Charlotte Danielson (2013). This chapter also presented research on current models of teacher evaluation for teachers of English learners. Finally, this chapter presented the gap in research that this study aims to explore as well as the research questions that will guide the study. The next chapter describes the methodology of the study.

Chapter Three: Methodology

The purpose of this study is to gain insight into the effectiveness of the ESL Teacher Observation Tool by comparing the evaluations of ESL teachers using the Charlotte Danielson Rubric (2013) and evaluations of the same ESL teacher using the ESL Teacher Observation Tool. Additionally, the study will gather qualitative data on the feasibility of the ESL Teacher Observation Tool in a public, suburban school context.

The following research questions were explored through the methodology outlined in this chapter: To what extent can the ESL Teacher Observation Tool assist evaluators in providing language instruction related feedback to ESL teachers? How does using an ESL specific observation tool compare to using Danielson (2013)?

Mixed methods were used to investigate these two questions. Both an evaluation task, including written feedback and rubric scores, and a semi-structured interview provided insight into the research questions.

Because of a constraint in time, the evaluators only used the Instruction portion of both tools. However, using only the Instruction portion of the tool will not necessarily limit the scope of the feedback to instruction because all of the participants are professional evaluators and will be free to give feedback on whatever they notice.

Overview of Chapter

This chapter provides a justification of the use of a mixed methods paradigm, and then describes the two-pronged approach for data collection and the pilot study that was

completed. Additionally, the chapter presents the plan for data analysis, and concludes by outlining the precautions that were taken to safeguard the participants of this study.

Mixed Methods Research Paradigm

Tashakkori and Creswell (2007) define mixed methods as “research in which the investigator collects and analyzes data, integrates the findings, and draws inferences using both qualitative and quantitative approaches or methods in a single study or a program of inquiry” (p. 4). The focal point of this work is the effectiveness of the ESL Observation Tool. The effectiveness of a tool meant for use in a school cannot be examined without taking into account the setting in which it would be utilized. Therefore, this study calls for quantitative data in the area of comparison between the current tool and the ESL Tool, and qualitative data when considering the larger scope of needs of a school administration.

The qualitative data was collected through written feedback and interviews, which allow for a wide range of responses (Mackey & Gass, 2016, p.225). Most administrators in a school setting have responsibilities other than simply evaluating teachers. In order for this tool to be useful, it must fit the needs of evaluators in a framework that allows for the flexibility and open-endedness that qualitative research offers. On the other hand, quantitative data is research that involves numerical data and variable manipulation (Mackey & Gass, 2016, p. 189). Quantitative data are used to compare the scores from the Danielson rubric (2013) to the ESL Teacher Observation Tool.

Data Collection

Participants. Eight evaluators from the district volunteered to participate in the study. Three elementary principals, one middle school principal, two middle school

assistant principals, and two high school assistant principals participated. The range of experience as an evaluator varied from as many as 22 years to as few as 1 year and averaged at about 9 years. Additionally, the range of exposure to ESL practices also varied from participant to participant. One participant said that she had no experience with ESL strategies and instruction, whereas most participants said they had a little or some. None of the participants had extensive experience with ESL, and none of the participants held ESL teaching licenses or certificates.

Setting. The study took place in a suburban school district in the Midwest that serves about 8,000 students. Out of the students the district serves, about 17% of the students are English learners, and there are English learners at every site (Minnesota Report Card, 2016).

Procedure

Participants. All elementary, middle, and high school principals and assistant principals in the district were emailed with a brief description of the study and a request to participate. Eight out of fourteen candidates replied to the email and volunteered to participate. Sessions were conducted with participants one-on-one, in each participant's school office.

Pilot Study. A pilot study was conducted with a lead teacher the district. The role of a lead teacher is to be available to observe teachers using whatever tool or method that teacher chooses. Lead teachers are not formal evaluators, but serve as coaches to teachers. They often observe teachers and offer formative feedback.

The pilot study aimed to refine the parameters of the tasks and test the period in order to avoid making the time commitment for administrators excessive. One outcome

of the pilot study was a switch from handwritten feedback to electronic feedback. The use of an electronic form was closer to how an evaluator would actually give feedback.

Materials. A paper copy of the Instruction portion of the Charlotte Danielson (2013) rubric is required for the first set of data, and a copy of the Instruction portion of the ESL Observation Tool is required as well. Both copies of the complete tool were available to the evaluator.

The ESL Teacher Observation Tool. The ESL Teacher Observation Tool in its entirety contains five categories: environment, language instruction, assessment, engagement, and management. The guiding criteria for each of the components were the following:

- Align with the Charlotte Danielson (2013) Framework
- Aligned with research (SIOP, WIDA)
- Clarify what quality language instruction looks like
- Be applicable in diverse settings (small-group pull-out and classroom)
- Foster conversations about language instruction and student language progress
- Increase teacher attention to specific aspects of the language-rich classroom in order to guide their own professional development in meeting the needs of English learners
- Be used by administrators, team teachers, individual teachers for reflection and goal setting
- Work for a drop-in observation, perhaps with an ongoing focus in one area and a conversation about strategies for improvement

(Robertson, Grucelski, Hoehn, Turnbull, Ascher, 2016)

As mentioned earlier, school administrators have many responsibilities and in order to make best use of time, only the Instruction portion of both tools will be used. See Appendix A for the tool.

Data collection technique 1. The participants watched two video clips of an ESL teacher teaching a lesson. Then, they evaluated the videos of teaching using two different rubrics. The first rubric, which was familiar to them, was the Charlotte Danielson rubric (2013). All participants had taken training on how to use the Charlotte Danielson rubric and how to use it to inform their feedback after watching a classroom for five to ten minutes. First, participants watched a 5- minute video of an ESL teacher teaching a lesson. Then, they rated that teacher using the Danielson observation rubric and gave that teacher feedback both positive, meaning something that teacher did well, and constructive, meaning something the teacher could improve on. The video clips were alternated in order to counterbalance the video data (Mackey & Gass, 2016, p.204).

The following verbal instructions were given verbatim to each participant: *Please watch the following video clip and rate the teacher using the tool. If an element is not seen in a particular clip, just mark it as “not observed”. For this study, you will only rate the teacher using the Instruction portion of both tools. After you watch and rate this video, you will be asked to provide feedback to this teacher. Please feel free to give any kind of feedback that you would give this teacher if you were observing them. Please include both positive and constructive feedback.*

Participants used paper copies of the rubric to assess the videos. For the positive and constructive feedback, participants used Google forms in order to keep the format consistent with the district’s current procedures.

After completing the first task, participants were introduced to the ESL Teacher Observation Tool. Participants were presented with a brief overview of the categories and structure of the ESL Observation Tool. The participants had time to review the rubric and ask questions if needed.

Then the participants repeated the first step. They watched a different video clip of the same ESL teacher teaching a different part of the same lesson, and this time, they rated the teacher using the ESL Observation Tool. Then the participants provided feedback to this teacher.

Video clip selections. The first video clip is of an ESL teacher teaching a medium sized group of 2nd grade students. She focuses on teaching academic writing, specifically, teaching ways to name animals in order to make writing more interesting.

The second video clip is about 20 minutes into the same lesson as the previous video clip. Students are working in pairs on their writing and the teacher is offering assistance to one pair of students.

Data collection technique 2. Following the video evaluation and feedback simulation, participants were interviewed. The purpose of the semi-structured interview was to triangulate the rubric and feedback data. As mentioned earlier, in a school setting, perspectives of participants must be considered in order to implement a new system. The following questions were used to guide the conversation. Open-ended prompts, such as “anything else?” were used to encourage interviewees to share more of their thoughts (Mackey & Gass, 2016, p. 255). The interview questions help answer the research questions outlined in earlier chapters.

What insights did you gain from using this tool?

What improvements would you suggest be made to this tool?

Would you consider using this tool to observe your ESL teachers? Why or why not?

These open-ended questions ideally allowed administrators to use their knowledge of evaluation systems and experience with implementing change in schools to give insight on possibilities that have not yet been considered.



Figure 2, Procedure

Data Analysis

Grounded theory was used to analyze the qualitative data. Grounded theory allows themes to be derived from the data itself rather than from preconceived hypotheses (Mackey & Gass, 2016, p. 231). The feedback was analyzed by noting similarities in the feedback to see what commonalities exist.

In order to examine patterns in the data, first, the data was compiled to look for repeated feedback. Additionally, the amount of feedback given in particular categories using the two different tools was analyzed and also feedback given by individual

administrators was examined to see if each administrator was able to provide more language-related feedback using the ESL Teacher Observation tool. Language related feedback is any feedback given that would specifically improve the learning of English learners (Echevarría, Vogt, & Short, 2012). For example, when feedback was given about an objective, that counted as general instruction; however, when there was feedback given about a language objective, that was considered language instruction.

The open-ended interview occurred after the participants experienced evaluating video clips using both tools. The interview aimed to integrate the participants' experience with teacher evaluation. The interviews were recorded and transcribed. Grounded theory was employed again in order to build themes from the data and avoid researcher bias (Mackey & Gass, 2016, p. 231).

Verification of Data

The information gathered during the interview was used to triangulate the feedback and rubric scores. Within the observation task, the rotation of videos is based on the repeated measures design outlined by Mackey and Gass. (2016, p. 204). Evaluators observe and rate the videos in different orders to increase the internal reliability of the experiment.

Ethics

The study provided the following safeguards to protect participants: Volunteers were provided with the purpose of the study prior to participation. The participants were clearly informed that participation is completely voluntary and they could choose to opt-out. Additionally, written consent was obtained. The identity of the

district and participants was kept anonymous in this report. District level consent was also obtained in writing.

Conclusion

This chapter presented the research paradigm of the study, the data collection process, and the procedure of the study. It also included an overview of the ESL Teacher Observation Tool and a description of the pilot study as well as the plan for data analysis of both the quantitative data and qualitative data. Finally, the chapter presented the safeguards used for the protection of the participants. The next chapter provides an analysis of the collected data.

Chapter Four: Results

This study aimed to better understand the effectiveness of two evaluation tools, the ESL Teacher Observation Tool and the Danielson rubric (2013) in terms of ESL instruction. The data was collected in two ways. The first was a task where principals evaluated video clips of an ESL teacher using portions of both the ESL Teacher Observation Tool and the Danielson rubric. Following this task, the principals were interviewed and asked to provide their perceptions on the applicability and usefulness of the ESL Observation Tool compared to the tool they were already familiar with, the Danielson rubric.

Through the collection of this data, the following research questions were investigated: To what extent can the ESL Teacher Observation Tool assist evaluators in providing language instruction related feedback to ESL teachers? How does using an ESL specific observation tool compare to using Danielson? This chapter presents the rubric scores and feedback collected and a synthesis of the open-ended interviews conducted after the task in terms of each research question.

Research Question One

To what extent can the ESL Teacher Observation Tool assist evaluators in providing language instruction related feedback to ESL teachers?

Only one domain was tested, and as mentioned in the literature review, each domain contains several components. Within each component are three to six elements. Most participants chose to give a general score for each component, however, some participants preferred to score each element within the component separately. When that

occurred, scores from each element were averaged. Additionally, four of the participants chose to leave certain components as “not applicable” because they felt that component was not observed. The individual rubric scores cannot be validly compared due to the lack of correlation between the different elements.

Table 2	<i>Rubric Scores by Participant</i>							
<u>Danielson Rubric</u>								
Communicating with Students	3	3	3	3	4	3	3.3	2
Using Questioning and Discussion Techniques	3	n/a	3	2	3	2	2.6	2
Engaging students in Learning	3	n/a	3	3	4	n/a	n/a	3
Using Assessment in Instruction	3	n/a	n/a	3	3	3	3	2
Demonstrating Flexibility and Responsiveness	3	n/a	n/a	3	3	3	3	3
<u>ESL Teacher Observation Tool</u>								
Meaningful opportunities to engage	3	2	4	2	4	2	3	2
Background knowledge	3	3	4	3	4	2	3	2
Academic language is explicitly taught	3	3	4	3	4	3	2	3

Using the Danielson rubric (2013), participants provided some general instruction-related feedback and some language instruction related feedback. The most repeated general instruction themes were about the teacher’s use of cuing, hand signals, to show understanding and her clear statement of the learning target. Cuing and using hand signals and pacing relate to classroom management, while clear learning targets relate to generic instruction.

Using the Danielson rubric (2013), participants provided more general instruction related feedback than language related feedback. The language feedback themes that were prevalent included praise of the teacher's use of visuals with vocabulary and the teacher's use of verbal practice.

While the praise from the ESL Observation Tool held some similarities to the Danielson rubric (2013), the constructive feedback was notably different. All of the constructive feedback was language related. Three of the participants asked the teacher about including more verbal practice, and many participants commented about the connection to background knowledge and native language. The follow-up interview allows a glimpse of understanding as to why this difference occurred.

During the interview, several participants noted that this tool is particularly useful for them because they do not have a full and deep understanding of ESL practices. One high school administrator explained that the tool, "would give me a better foundation on what I should be seeing in a classroom that has ESL students in it." Another participant's comment inadvertently exemplified Holdheide. et. al.'s (2010, p. 16) work by paralleling ESL instruction with music instruction. The participant shared that he has a music background so when he observes a music classroom he knows what he is looking for. However, for evaluators who do not have a music background, he questioned, "if you don't have a music background do you know what you're really looking for? How do you know that students are really learning?" Finally, the participant shared that walking into an ESL classroom he would need a tool like the ESL Teacher Observation Tool to help him know what to look for.

All administrators said they would use this tool to observe their ESL teachers. Moreover, many administrators requested additional training on the tool for a variety of reasons. One participant wanted to know more about the context of use for the tool. Another two participants pointed out that it requires training to know what some of the criteria on the tool look like. One of those participants exemplified this by asking, “What are activities in an ESL classroom? What do those look like differently when you’re teaching language objectives?” Overall, participants displayed an eagerness to use the tool and learn more about it.

An additional theme that emerged from the interviews is relates to the element, “explicit connection with students’ home language and cultures enhances understanding.” This topic sparked personal narratives for two of the participants. One participant described how she asked one of her ESL teachers in a post-observation meetings how they feel about home language use in their classroom. Another participant said that although she definitely wants her ESL teachers to make this type of connection, she is not sure what exactly that looks like other than a literal translation from English to home language. Overall, several areas of the ESL Teacher Observation Tool require deeper training and perhaps clearer description.

The feedback evaluators gave after watching the video clips provides valuable insight into the effectiveness of the ESL Teacher Observation Tool. Overall, evaluators did give more language-related feedback in terms of quantity. All except one evaluator gave feedback related to language, and every piece of constructive feedback is language related. However, the depth of these responses is lacking. When comparing feedback given with Danielson (2013), the ESL Teacher Observation Tool produced several pieces

of feedback that were directly copied from the ESL Teacher Observation Tool. This was not the case for the feedback given using Danielson rubric. However, a few evaluators were able to expand on the criteria from the ESL Teacher Observation Tool. One principal recommended that the teacher ask students to practice saying the sentences before they write them. This may indicate that the evaluator understood the bullet point from the ESL Teacher Observation Tool which says, “activities meaningfully integrate multiple modalities and maximize output of language”, (Robertson, Grucelski, Hoehn, Turnbull, Ascher, 2016). Another evaluator suggested that, the teacher provide more opportunities to produce language and interact, and suggested using whiteboards or turn and talk to create “more structured engagement”. Again, this feedback is showing more depth or coaching for the ESL teacher.

Overall, the results show that feedback related to ESL teaching increased when administrators used the ESL Teacher Observation Tool. Although the feedback given using the ESL Teacher Observation Tool does not contain the depth that the feedback of Charlotte Danielson (2013) does, interviews show that this tool is still causing forward progress in helping evaluators to recognize academic language. One evaluator, talking about her own experiences evaluating ESL teachers said, “I feel like I will say things like ‘I didn’t see enough academic language’ but if I were pressed I wouldn’t be able to really articulate what that looks like.” Then, she explained that this tool is specific enough for her to be able to make suggestions on how to improve the language instruction.

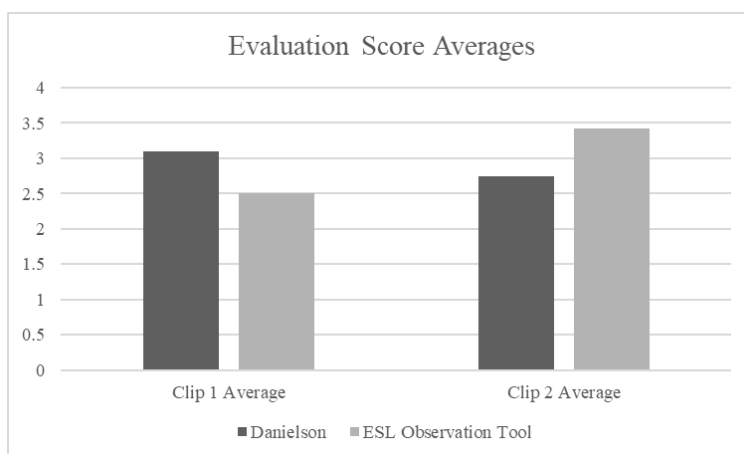
Research Question Two

How does using an ESL specific observation tool compare to using Danielson?

When desegregated by video clip, table 3, the data shows that video clip 1 was scored higher in 3 of 4 cases using the Danielson rubric (2013). Video clip 2 showed the opposite with all of the ESL Teacher Observation Tool scores being higher than the Danielson scores.

		<u>Clip 1</u>				<u>Clip 2</u>			
Danielson Rubric	3.0	4.0	4.0	2.7	2.7	2.7	2.3	2.3	
ESL Observation Tool	3.0	2.8	2.8	2.4	3.0	3.0	3.4	3.0	

In video clip 1, the average score for the Danielson rubric (2013) was higher than the ESL Teacher Observation Tool with a difference of 0.6. Video clip



2 showed the opposite change with a difference of 0.7. It is possible that this difference means that proficient teaching according to the Danielson rubric looks different that proficient teaching according to the ESL Observation Tool. In an interview, one administrator explained the same saying that, “using the Danielson teacher model on EL teachers does not really [work]: the rigor is just not there.” She explained that in her experience it is possible to earn the highest marks in Danielson (2013), distinguished, and

yet, be only an average ESL teacher. This observation is supported by Holdheide et. al. (2010) who explains that distinguishing between an effective and ineffective ESL teacher requires an understanding of the unique instructional methods and skills related to ESL teacher. With a rubric that does not require an understanding of these unique skills and methods, it is not surprising that a discrepancy between evaluations based on the Danielson rubric and reality.

As mentioned in the literature review, WIDA's Essential Actions handbook explains that practice and instruction need to take place at three levels: linguistic complexity, language forms and conventions, and vocabulary usage (Gottlieb, 2013). Using the Danielson tool (2013), evaluators only commented on vocabulary usage. However, using the ESL Teacher Observation Tool, evaluators were able to make comments on all three levels. For example, after using the ESL Teacher Observation Tool one evaluator expressed satisfaction about the teacher's instruction by commenting, "The structure of showing the use of bumble bee over and over again and then changing the describing words, " flying creature" etc. Seemed to work well for the students." Although this evaluator does not seem to recognize the teacher's use of repetition and replacing words to teach syntax, using the ESL Observation Tool, he was able to accurately identify an example of explicit language instruction. This example is the only example where the evaluator achieved depth in language feedback. However, this shows that it is possible for a principal without an ESL background to notice language instruction. Perhaps with further training and professional development, a greater depth could be achieved.

Additionally, the quantity of language related feedback increased when evaluators used the ESL Teacher Observation Tool. Using Danielson (2013), two comments about adding a speaking element to the lesson appeared, whereas using the ESL Teacher Observation Tool, four evaluators commented on speaking. Using Danielson, background knowledge was not mentioned; however, four evaluators recommended making a connection to background knowledge when using the ESL Teacher Observation Tool.

During the interviews, many participants noted some differences between the two tools. Several mentioned that the Danielson (2013) is far more detailed and comprehensive than the ESL Observation Tool. Some explained this difference as positive, saying, “one of the things that we often struggle with Danielson is that it’s designed to catch everything and so when you’ve got 23 components within 4 domains...it’s just sometimes too big.” Yet some have the concern that because of its simplicity compared to the Danielson (2013) rubric, it may not be enough to meet the needs of administrators. Although he prefaced by saying that he cannot give concrete feedback until knowing how the tool would be used, one principal suggested, “if it’s a replacement for domain 3, then it needs to contain a little bit more. It is focused, but it can also be too focused.” This suggests that the purpose of the ESL Teacher Observation Tool needs further clarification and perhaps changes need to be made based on the defined purpose.

Additionally, one administrator discovered that the scores on the ESL Teacher Observation Tool do not correlate well with the scores on the Danielson (2013) rubric. She described how challenging it is to achieve a 4, or distinguished, on the Danielson (2013) rubric and usually, a great lesson lands at a 3, or proficient. She criticized that the

ESL Observation Tool contains language similar to the Danielson band 2, or basic, in the 3 band. The data from the rubric scores substantiates this inconsistency. Although the difference of the two rubric score averages was only 0.1, on Danielson, a score of 4 was only awarded twice, whereas using the ESL Teacher Observation Tool a 4 was awarded 6 times. There are two possible explanations for this occurrence. On one hand, it is possible that the ESL Teacher Observation Tool measures aspects that are different enough to create a significant inconsistency in the scores. On the other hand, it is possible that this inconsistency coincides with the aforementioned issue that the Danielson (2013) is far more detailed and comprehensive than this ESL Observation Tool and is therefore impossible to compare. Further refinement of the tool and further research must be done to understand the cause of these differences.

Overall, there are some major differences between the Danielson (2013) and ESL Teacher Observation Tool. Some of these differences include the perceived complexity of the Danielson (2013) versus the perceived simplicity of the ESL Observation Tool. Additionally, the rubric score data showed a difference in score based on the video clip, one positive and one negative, which may suggest that the observation tool has an impact on the outcome of the teacher's performance evaluation.

Conclusion

The amount of language related feedback was greater in quantity and quality when using the ESL Observation Tool. Additionally, the perceptions generally seem positive with all interviewees stating that they would use the ESL Teacher Observation Tool to observe their ESL teachers. Generally, administrators felt that it was helpful in clarifying what an ESL teacher's classes should look like. The key concerns that arose

from the data were the ESL Teacher Observation Tool's lack of consistency with Danielson (2013). Participants also noted that although the tool provides helpful guidance about what to look for in an ESL classroom, more training may still be needed in accompaniment with the tool in order for evaluators to know what each component looks like. The next chapter will discuss major findings, implications, and suggestions for further research.

Chapter Five: Conclusion

In this study, I attempted to answer the following research questions: To what extent can the ESL Teacher Observation Tool assist evaluators in providing language instruction related feedback to ESL teachers? How does using an ESL specific observation tool compare to using Danielson (2013)? Prior to this investigation, there had already existed an identified need for a specifically ESL-oriented observation tool. A team of teachers led by a program supervisor created this tool and began piloting it among themselves. This research aimed to gain deeper insight by drawing comparisons with the existing tool and gathering feedback from administrators who have used the tool. The purpose of this research was to refine the ESL Teacher Observation Tool based on administrator feedback and to provide a point of reference to administrators who are deciding whether to implement this tool in their schools or districts.

Findings

One of the most important findings was that administrators perceive the ESL Teacher Observation Tool to be useful in evaluating and giving instruction-related feedback to ESL teachers. As mentioned in the literature review, Marzano (2012) described two purposes of teacher evaluation: differentiating effective teachers from ineffective teachers and developing skills of teachers. Based on the feedback forms and the follow-up interviews, this tool may be able to assist administrators with the second purpose. For differentiating effective teachers from ineffective ones, I believe the data presents compelling motivation to investigate further if the current method of evaluating ESL teachers is truly effective. As mentioned in the literature review, due to new initiatives, state governments have begun placing an emphasis on teacher evaluation

(USDE, 2009). However, the data from this study suggests that principals need more tools and training to effectively evaluate ESL teachers.

Limitations

There are several limitations of this study. The issue of participant bias is a major concern in two ways. Eight administrators volunteered to participate out of fourteen administrators. The reasons as to why the other six administrators chose not to participate is unknown. It is possible that only the administrators who felt positively about ESL teacher evaluation made time for the study, and that the remaining six administrators would have felt more unfavorably toward the ESL Teacher Observation Tool. The second layer of bias in this study resulted from the fact that I personally knew many of the administrators who volunteered in the study. It is possible that they responded to the study more favorably because of our relationship.

This study worked under the assumption that a professional without linguistic knowledge is capable of giving high quality language-related feedback. The reason for this assumption is simply because in many schools, whether they are capable or not, they still must give feedback. However, this assumption is still a limiting factor, and therefore is worth exploring further.

The last limitation is concerning the rubric data. Originally, the plan was to use the rubric scores to compare and contrast the tools. However, the interviewer feedback made it clear that the tools are impossible to compare component for component, rendering the rubric scores inadequate.

Implications

While this research was underway, the ESL supervisor uploaded the tool onto the drop-in observation system, and it is now available to any administrator from the district. However, an important finding from the interviews is that administrators need further training to understand the tool. Additionally, a few participants noticed issues about the consistency of the ESL Observation Tool rating system compared with the Danielson rubric (2013) in some of the elements. Thus, refining the tool and providing additional training become two essential follow-up actions.

Based on the results of this study, there are several areas of refinement to consider. The team that created this tool will likely be the team to refine the tool as well. The following are suggestions for the team to consider.

- Change the language and gradient of the ESL Teacher Observation Tool to match Danielson (2013). For example, scoring a four should be similar to Danielson (2013) in that a teacher who scores a four in that category did an exemplar performance of that element. This will not only made the rubric more fair when compared to Danielson (2013), but also will require less training for administrators as they are already familiar with the way the Danielson (2013) rubric is oriented.
- Offer coaching sessions or training sessions where administrators can learn to use the tool. Several administrators have mentioned that they do not always know what to look for in certain categories. For example, in the element “explicit connection with students’ home language and cultures enhances understanding,” many administrators knew that the teacher in the sample evaluation did not achieve this

target, but were not prepared to coach the teacher with suggestions on how this target could have been achieved.

- Because this study was limited to the administrators who volunteered, it is exceedingly important to work with the administrators who did not participate and note their suggestions for improving the tool. Their perspectives may be significantly different from the ones in this research.

Further Research

Although some specialist fields such as Special Education have ample research on teacher evaluation and even their own version of the Danielson (2013) Rubric, there is still a large gap in ESL teacher evaluation. The most concerning issue for further research came directly from a follow-up interview. An administrator felt that an ESL teacher could achieve a mark of distinguished on the Charlotte Danielson (2013) and still be an average ESL teacher. As the United States' most popular framework (Danielson, 2012), it is critical that the use of this tool on ESL teachers is examined closely.

Dissemination

I hope to share this research in two main ways. First, I hope to continue working with the original design team for this tool to refine it based on the suggestions above. Furthermore, I am aiming to present the process for making the tool, along with the tool itself at Korea's International TESOL conference in October 2019.

Conclusion

The goal of this study was to begin to answer two questions: To what extent can the ESL Teacher Observation Tool assist evaluators in providing language instruction related feedback to ESL teachers? How does using an ESL specific observation tool

compare to using Danielson (2013)? The overarching outcome of this study is that although it still needs to be refined, the ESL Teacher Observation Tool is useful for providing language related feedback. I hope that this tool will not only allow administrators to give useful language related feedback, but also result in a greater value for ESL teachers' unique expertise.

Appendix: ESL Observation Tool: Language Instruction

Environment is conducive to language learning. 2A, 2B, 2E*	
4 Well Done clear and effective	<ul style="list-style-type: none"> • Interactions reflect respect, caring and trust between individuals with varying cultural backgrounds and languages. • Activities are culturally appropriate and convey high expectations for ALL students. • Classroom organization ensures access to learning for ALL students.
3 Satisfactory	<ul style="list-style-type: none"> • Most interactions reflect respect, caring and trust between individuals with varying cultural backgrounds and languages. • Activities are occasionally culturally appropriate and convey high expectations for students. • Classroom organization ensures access to learning for students.
2 Attempted needs improvement	<ul style="list-style-type: none"> • Interactions inconsistently reflect respect and or there is a lack of trust between individuals with varying cultural backgrounds and languages. • Activities are not culturally appropriate and convey low expectations for students. • Classroom is disorganized and access to learning is inconsistent.
1 Missing	<ul style="list-style-type: none"> • Interactions are disrespectful or disruptive. • Activities are culturally offensive and convey low expectations for students. • Classroom is disorganized and access to learning is extremely limited.
N/A	

Academic language is explicitly taught. 3A*	
4 Well Done clear and effective	<ul style="list-style-type: none"> • Target academic language is clearly explained and modeled. • Teacher draws on students' native languages. • Scaffolded support is provided for students to use target academic language. • Opportunities for meaningful, authentic use of target academic language. • Visuals and graphic organizers are used for language reinforcement.
3 Satisfactory	<ul style="list-style-type: none"> • Target academic language is explained and modeled. • Scaffolded support is provided for students to use target academic language. • Opportunities for meaningful, authentic use of target academic language is inconsistent. • Visuals and supports are offered for language reinforcement.
2 Attempted needs improvement	<ul style="list-style-type: none"> • Target academic language is inconsistently explained. • Scaffolded support is inconsistently or ineffectively provided for students. • Limited opportunities for use of target academic language. • Visuals and supports may be present, but not utilized.

1 Missing	<ul style="list-style-type: none"> • Target academic language is not explained and modeled. • Teacher is unaware of students' native languages. • Scaffolded support is not provided for students. • No expectation for use of target academic language. • Visuals and graphic organizers are not used for language reinforcement.
N/A	

Background knowledge is built upon. 1B*	
4 Well Done clear and effective	<ul style="list-style-type: none"> • Explicit opportunities to connect prior knowledge and experiences to learning. • Explicit connection with students' home language and cultures enhances understanding. • Vocabulary is pre-taught and makes connection with first language if possible (cognates.)
3 Satisfactory	<ul style="list-style-type: none"> • Inconsistent opportunities to connect prior knowledge and experiences to learning. • Interest in students' home language and cultures but inconsistent connection to increasing understanding. • Vocabulary is pre-taught.
2 Attempted needs improvement	<ul style="list-style-type: none"> • Inconsistent opportunities to connect prior knowledge to learning. • New vocabulary is introduced but not explicitly connected with the learning.
1 Missing	<ul style="list-style-type: none"> • Lack of connection to prior knowledge and experiences related to learning. • Lack of knowledge of students' home language and cultures. • Very limited vocabulary instruction.
N/A	

Meaningful opportunities to engage in all four modalities with an emphasis on productive language skills. 3C*	
4 Well Done clear and effective	<ul style="list-style-type: none"> • Wide variety of activities that require students to listen, read, speak and write during the lesson. • Activities meaningfully integrate multiple modalities and maximize output of language.
3 Satisfactory	<ul style="list-style-type: none"> • Activities allow students to engage in multiple domains - to listen, read, speak and write during the lesson. • Activities integrate multiple modalities and expect basic output of language.

2 Attempted needs improvement	<ul style="list-style-type: none">• Activities focus on receptive rather than productive language.• Activities integrate some modalities but miss opportunities to increase output of language.
1 Missing	<ul style="list-style-type: none">• Teacher lecture and no expectation for student language production.
N/A	

References

- Additional Licenses and Curriculums. Retrieved May 06, 2018, from <https://www.hamline.edu/education/additional-licenses/licenses-curriculums/>
- Al-Thumali, A. A. (2011, January 1). Evaluating EFL Intermediate Teachers' Performance in the Light of Quality Standards in Saudi Arabia. Online Submission,
- Bill and Melinda Gates Foundation. (2010). Learning about Teaching: Initial Findings from the Measures of Effective Teaching Project. Research Paper. MET Project. Retrieved from Educational Information Resources Center
- Bill and Melinda Gates Foundation. (2013). Ensuring Fair and Reliable Measures of Effective Teaching: Culminating Findings from the MET Project's Three-Year Study. Policy and Practice Brief. MET Project. Retrieved from Educational Information Resources Center
- Cherasaro, T. L., Brodersen, R. M., Reale, M. L., Yanoski, D. C., Regional Educational Laboratory Central, National Center for Education Evaluation and Regional Assistance, & Marzano Research, L. (2016). Teachers' Responses to Feedback from Evaluators: What Feedback Characteristics Matter? REL 2017-190. Retrieved from Educational Information Resources Center
- Cohen, J., & Goldhaber, D. (2016). Building a More Complete Understanding of Teacher Evaluation Using Classroom Observations. *Educational Researcher*, 45(6), 378-387. doi:10.3102/0013189x16659442
- Collins, J. L. (2009, January 1). What's Good for the Goose Is Good for the Gander: Implementing the SIOP Model into an Urban Elementary Classroom of African American Students. ProQuest LLC,

Center for Applied Linguistics. (n.d.). SIOP - Learn about SIOP. Retrieved May 05, 2017, from

<http://www.cal.org/siop/about/>

Danielson, C. (2013). *The framework for teaching: evaluation instrument*. Princeton, NJ:

Danielson Group.

Danielson, C. (2012). *Observing Classroom Practice*. *Educational Leadership*, 70(3), 32-37.

Retrieved from Educational Information Resources Center

Danielson Group » Charlotte Danielson. (n.d.). Retrieved May 06, 2017, from

<https://www.danielsongroup.org/charlotte-danielson/>

Echevarría, J., & Short, D. J. (2000). *Using Multiple Perspectives in Observations of Diverse*

Classrooms: The Sheltered Instruction Observation Protocol (SIOP). Retrieved from

Educational Information Resources Center

Echevarría, J., Short, D. J., & Center for Research on the Educational Achievement and

Teaching of English Language Learners, (2011). *The SIOP Model: A Professional*

Development Framework for a Comprehensive School-Wide Intervention. CREATE

Brief. Retrieved from Educational Information Resources Center

Echevarría, J., Vogt, M., & Short, D. (2012). *Making Content Comprehensible for English*

Learners: the SIOP® model. Boston: Pearson.

Escamilla, K. (2009). *English Language Learners: Developing Literacy in Second-Language*

Learners—Report of the National Literacy Panel on Language-Minority Children and

Youth. *Journal Of Literacy Research*, 41(4), 432-452. Retrieved from Educational

Information Resources Center

- Fixsen, D., Naoom, S., Blase, K., Friedman, R., Wallace, F. (2005). *Implementation Research: A Synthesis of the Literature*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, National Implementation Research Network
- English Language Development (ELD) Standards. (n.d.). Retrieved May 30, 2017, from <https://www.wida.us/standards/eld.aspx>
- Gandara, P., Maxwell-Jolly, J., Driscoll, A., Policy Analysis for California Education, B. C., Center for the Future of Teaching and Learning, S. C., & California Univ., S. I. (2005). Listening to Teachers of English Language Learners: A Survey of California Teachers' Challenges, Experiences, and Professional Development Needs. Policy Analysis For California Education, PACE. Retrieved from Educational Information Resources Center
- Gottlieb, M. (2013). Essential actions: a handbook for implementing WIDA's framework for English language development standards. Madison, WI: Published for the WiDA Consortium by the Board of Regents of the University of Wisconsin System. Retrieved from <https://www.wida.us/standards/eld.aspx>
- Gottlieb, M., Cranley, E., Camilleri, A. (2007). Understanding the WIDA English Language Proficiency Standards. Madison, WI: Published for the WiDA Consortium by the Board of Regents of the University of Wisconsin System. Retrieved from <https://www.wida.us/standards>
- Holdheide, L. R., Goe, L., Croft, A., Reschly, D. J., & National Comprehensive Center for Teacher, Q. (2010). Challenges in Evaluating Special Education Teachers and English Language Learner Specialists. Research & Policy Brief. Retrieved from Education Resources Information Center

- Hunt, K., Gurvitch, R., & Lund, J. L. (2016). Teacher Evaluation: Done to You or with You?.
JOPERD: The Journal Of Physical Education, Recreation & Dance, Vol 87, Issue 9, 21-27.
- Kinsella, K. (2005). Teaching Academic Vocabulary. Aiming High. Retrieved May 5, 2017, from <https://www.scoe.org/>
- Krashen, S. (2013). Does SIOP Research Support SIOP Claims?. International Journal of Foreign Language Teaching, Vol 8, Issue 1, 19-24.
- Marzano, R (2012). Teacher Evaluation: What's Fair? What's Effective?. *Educational Leadership*, Vol 70, Issue 3.
- Minnesota Department of Education. "English Language Development Standards." *English Language Development Standards*, education.mn.gov/MDE/dse/el/Std/.
- Minnesota Report Card. (2016). Retrieved May 07, 2017, from <http://rc.education.state.mn.us/>
- Roberston, K., Froemming, A., Grucelski, K., Hoehn, S., Turnbull, K., Ascher, H., (2016). *EL Observation Tool. Unpublished Manuscript.*
- Rodriguez Moux, S. (2010, January 1). Teacher's Perceptions of Sheltered Instruction Observation Protocol for Teaching Young English Language Learners: A Qualitative Case Study. Retrieved from ProQuest LLC.
- Sempek, A. N. (2014). Elementary and secondary teachers' attitudes towards the education of english language learners in the mainstream classroom in rural schools (Order No. 3618069). Available from ProQuest Dissertations & Theses Global. (1530298390). Retrieved from <http://ezproxy.hamline.edu:2048/login?url=http://search.proquest.com.ezproxy.hamline.edu:2048/docview/1530298390?accountid=2810>

- Short, D. J., Echevarria, J., & Center for Research on Education, D. C. (1999). The Sheltered Instruction Observation Protocol: A Tool for Teacher-Research Collaboration and Professional Development. Educational Practice Report 3.
- Tashakkori, A., & Creswell, J.W. (2007). Editorial: The new era of mixed methods. *Journal of Mixed Methods Research*, Vol 1, Issue 1. Retrieved from <http://mmr.sagepub.com/content/1/1/3>.
- Tawalbeh, T. I., & Ismail, N. M. (2014). Investigation of Teaching Competencies to Enhance Students' EFL Learning at Taif University. *International Education Studies*, 7(11), 84-96. Retrieved from Educational Information Resources Center
- TESOL International Association. (2013). Implementing the Common Core State Standards for English Learners: The Changing Role of the ESL Teacher. A Summary of the TESOL International Association Convening. April 2013
- The Council for Exceptional Children's Position on Special Education Teacher Evaluation. (2013). *Teaching Exceptional Children*, 45(3), 73-76. Retrieved from Education Full Text
- Turkan, S. s., & Buzick, H. M. (2016). Complexities and Issues to Consider in the Evaluation of Content Teachers of English Language Learners. *Urban Education*, 51(2), 221-248. Retrieved from Education Full Text
- U.S. Department of Education. (2009). Race to the top program executive summary. Retrieved from <http://www2.ed.gov/programs/racetothetop/executive-summary.pdf>
- Vidot, J. L. (2011, January 1). The Efficacy of Sheltered Instruction Observation Protocol (SIOP) in Mathematics Instruction on English Language Learner Students. Retrieved from ProQuest LLC