Motivation for the Research
Currently, high-stakes decisions are being made regarding the quality of instruction provided to English language learners (ELLs) using teacher evaluations based on classroom observations. One of the most common frameworks being used to make these decisions in Washington State and other locations is the Danielson Framework for Teaching (FFT). Despite its wide-spread influence, there is a dearth of ELL-focused research on the FFT (Holdheide, Goe, Croft, & Reschly, 2010; Jones, Buzick, & Turkan, 2013; Staehr Fenner, Kozik, & Cooper, 2015). The purpose of this study was to investigate the appropriateness of using the FFT to evaluate the quality of teaching provided to ELLs with the overarching aim to advance the needs of ELLs by examining the evaluation of their teachers with the expressed goal of advocating (at the state-level) for any necessary changes to either the instrument (FFT rubric) or the evaluation process (the use of the FFT rubric).

Research Questions
A gap in the research on teacher quality in relationship to the achievement of K-12 ELLs (Holdheide et al., 2010; Jones et al., 2013; Loeb, Soland, & Fox, 2014; Master, Loeb, Whitney, & Wyckoff, 2016; Samson & Collins, 2012; Turkan & Buzick, 2016; Turkan, Croft, Bicknell, & Barnes, 2012) has contributed to a lack of consensus around what constitutes quality instruction for ELLs. In the absence of an agreed-upon definition of quality instruction for K-12 ELLs, the Quality Teaching for English Learners (QTEL) approach was selected as the theoretical framework for this study and used to operationalize the concept of quality instruction for ELLs.

The following six research questions reflect the integral nature of QTEL in this study:

1. To what extent are the critical attributes for quality teaching of English learners as defined by the QTEL approach reflected in the FFT rubric?
2. How do educators in Washington State define quality teaching for ELLs, and do their definitions differ based on the educators’ backgrounds?
3. To what extent do educators in Washington State perceive that the FFT measures quality teaching for ELLs, and do their perceptions differ based on their backgrounds?
To what extent do educators in Washington State report that using the FFT influences their planning of lessons and their classroom assessments relative to ELLs?

To what extent do educators in Washington State report that using the FFT influences the professional development they seek out or are offered regarding meeting the needs of their English learners?

To what extent do administrators evaluate the same QTEL lesson in the same way?

Research Methodology
The design of this mixed methods study was a complex application of a convergent design in which a theoretical lens intersects with an identified core design (Creswell & Plano Clark, 2018; Plano Clark & Ivankova, 2016). As the theoretical framework, QTEL’s principles, critical attributes, and definition of quality teaching for English learners were infused into all aspects of the design. Qualitative and quantitative data from three separate sources were collected and analyzed. These sources included the following: a textual analysis of the FFT and QTEL foundational documents; an online educator questionnaire in two versions (teacher and administrator); and a think-aloud interview conducted with five administrators as they used the FFT to individually rate the same digitally-recorded lesson specifically designed for ELLs.

The data collected in the questionnaire and think-aloud interview portions of this study came exclusively from Washington State educators. Fifty-two administrators (11.5% had ELL endorsements) and 133 teachers (29% had ELL endorsements) responded to the respective forms of the questionnaire for a total N of 185. The educators that responded to the questionnaire represented school districts from across the state. The five administrators that participated in the think-aloud interviews represented three elementary, one middle school, and one high school principals from two adjoining rural school districts with 6.4% to 7.5% of their student body comprised of ELLs. None of these administrators held an ELL endorsement.

A four-phase textual analysis was performed on QTEL and FFT foundational documents in order to answer Research Question 1. In the questionnaire strand, both forms of the researcher-designed questionnaire consisted of 51 items and included 13 biodata items; 25 statements designed to elicit the participants’ perceptions, beliefs, and experiences regarding the FFT using a six-point Likert scale, as well as one open-ended item; and 12 QTEL critical attribute items using a seven-point Likert scale. Both questionnaire versions allowed participants to leave comments on 17 of the items. The quantitative data were analyzed using descriptive statistics, t-tests, correlations, and effect sizes when possible to answer Research Questions 2-5. Qualitative data were analyzed thematically. In the think-aloud strand, a modified interview protocol was used (Bell et al., 2016). The five administrators independently watched the same digitally recorded QTEL lesson and scored it using the FFT rubric. Each interview was digitally recorded, and the collected data were analyzed both qualitatively (thematically) and quantitatively to answer Research Question 6.

Summary of Findings
The three strand results were integrated to answer the six research questions listed above.

1. Not all of QTEL’s critical attributes are reflected in the FFT. The textual analysis using the QTEL rubric revealed that the FFT rubric does not address how to approach language learning in any way. Thus, the FFT rubric does not include the most central of the critical attributes for quality teaching of ELLs, as defined by the QTEL approach. These critical attributes are specifically focused on how teachers combine the language-learning needs with the content learning needs of ELLs.
2. Background factors influence educators’ definitions of quality teaching for ELLs. The educators’ responses to questionnaire items appear to indicate that educators believed that quality instruction for ELLs required different instructional practices from those used with either Special Education students or mainstream students. However, while educators appeared to believe that a quality lesson for ELLs would benefit other students who struggle with academic English, the converse was not true.

3. The FFT does not measure quality instruction for ELLs. Descriptive statistics from one scale (FFT Measures Quality Instruction) consisting of two subscales (FFT Measures General Quality Instruction; FFT Measures Quality Instruction for ELLs) revealed that educators in Washington State slightly disagreed (M=3.42, SD=1.07, Median=3.4) that the FFT Measures Quality Instruction (whole scale). However, their responses to the idea that the FFT measures quality instruction in general were closer to slightly agree (M=3.60, SD=1.26, Median=3.7) while their responses to the idea that the FFT measures quality instruction for ELLs were closer to slightly disagree (M=3.33, SD=1.11, Median=3.3). The educators’ job category moderately influenced their responses to the ELL Quality Scale, and the effect was stronger ($d=0.87$) than for the General Quality Scale ($d=0.69$). Teachers slightly disagreed with the idea that the FFT does a good job of measuring quality instruction for English language learners while administrators slightly agreed. There was a strong positive relationship between the educators’ responses to the ability of the FFT to measure general quality instruction and the ability of the FFT to measure ELL quality instruction ($r[185]=.69$), and the relationship was stronger for teachers than for administrators.

4. The FFT does not strongly influence the planning and assessments for ELLs.

5. The FFT does not influence the availability of ELL specific PD.

6. Five independent evaluations of the same QTEL lesson lack inter-rater reliability. The five administrators assigned the QTEL lesson the same overall (proficient) rating but varied in the number of components they rated, the scores assigned to the components that they did rate, the approaches they used to arrive at the ratings at both the component and summative levels, and the evidence they chose to justify their ratings.

After the initial integration of findings from the three strands was used to answer the research questions, a subsequent analysis was performed by integrating the findings from the textual analysis with the findings of the think-aloud interviews in order to deepen the understanding of the interplay between the QTEL approach and the FFT rubric.

1. Administrators responding to the strengths of the QTEL lesson were inconsistent in choosing FFT components to align their evidence to because QTEL attributes are missing from the FFT.

2. The FFT rubric’s inherent structural inability to capture ELL-oriented evidence may contribute to the administrators’ lack of inter-rater reliability in RQ6.

**Implications**

Because no previous study about the FFT had started with ELLs at the center, in other words, used an ELL focused framework to evaluate the FFT rubric from, and because the FFT is missing critical indicators of quality instruction for ELLs (as independently confirmed by Gill, Shoji, Coen, and Place, 2016), this study provided evidence missing from previous studies about the FFT. Unlike the FFT rubric, which focuses on what the teacher is doing, not on who the teacher is teaching (Danielson, 2007), the QTEL approach focuses specifically on the needs of ELLs. In addition, previous studies have raised serious concerns about inter-rater reliability when using the FFT for classroom observations (Bell et al., 2015;
The findings from this dissertation study demonstrate how a lack of inter-rater reliability is compounded when we evaluate teachers of ELLs. The policy and planning implications of this study are that states should not use the FFT to make high-stakes decisions regarding teacher evaluations, and most especially for the teachers of ELLs because there is not yet any empirical evidence that supports the use of the FFT for this purpose. Furthermore, without further research, it is advised that EL-modified versions of the FFT not be used to evaluate teachers of ELLs. The reasoning behind this recommendation is that the current EL-modified versions of the FFT (August & Blackburn, 2019; Coady et al., 2020) are built using the original framework of the FFT, which has been found to be missing some of the essential attributes of quality instruction for ELLs. Therefore, it is possible to leave out ELL-critical components in a modified version of the FFT.

In conclusion, the findings of this study indicate that the FFT rubric is not an appropriate tool to use in high-stakes evaluations of teachers of ELLs. Furthermore, it has been demonstrated that the evaluation process using the FFT lacks both validity and reliability because the FFT does not completely, accurately, and consistently measure quality instruction for ELLs. Given these findings, meeting the burden of evaluation and accountability may be better served at the systems level (school, district, state) rather than at the teacher level because a single teacher in isolation is rarely solely responsible for ELL student achievement or opportunity gaps.
References


