Title of Project:
Using the Danielson Framework to Evaluate the Teachers of English Language Learners in Washington State: A Mixed Methods Study

Researcher:
Randi Freeman
Anaheim University
randifreeman62@icloud.com

Research Supervisors:
Dr. Luke Plonsky

Randi Freeman

Final Report

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?

4. 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?
5. 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?
6. 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;

Bell et al., 2016; Bell, Jones, Qi, & Lewis, 2018; Qi et al., 2018; Roegman, Goodwin, Reed, & Scott-McLaughlin, 2016). 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

Alvarez, L., Catechis, N., Chu, H., Hamburger, L., Herpin, S. A., & Walqui, A. (2012, June). *Quality Teaching for English Learners (QTEL) impact study*. San Francisco, CA: WestEd.

Anderson, S. (2018, July). *Analysis of school employees evaluation survey (SEES). Final Report.* Olympia, WA: Office of Superintendent of Public Instruction (OSPI). Retrieved from https://www.k12.wa.us/sites/default/files/public/tpep/pubdocs/seeswrittenreport.pdf

Archibald, S. J. (2007). How well do standards-based teacher evaluation scores identify high-quality teachers? A multilevel, longitudinal analysis of one district. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 68*, 1235.

Arkansas Department of Education (ADE). (n.d.). English for speakers of other languages (ESOL) considerations for Teacher Excellence Support System (TESS). Retrieved from http://www.arkansased.gov/public/userfiles/HR_and_Educator_Effectiveness/TESS/TESS%20Supporting%20Documents/ELL_alignment_with_Danielson_rubric_template2.pdf

August, D. (2018). Educating English language learners: A review of the latest research. *American Educator, 42*(3), 4–9, 38.

August, D., & Blackburn, T. (2019a, April). *Promoting success for teachers of English learners through structured observations*. Washington, DC: Council of Chief State School Officers. Retrieved from https://ccsso.org/resource-library/promoting-success-teachers-english-learners-through-structured-observations-0

August, D., & Blackburn, T. (2019b, April). *Promoting success for teachers of English learners: Tool aligned with the Danielson Framework for Teaching.* Washington, DC: Council of Chief State School Officers. Retrieved from https://www.air.org/sites/default/files/ccsso-danielson-framework-tool-April-2019.pdf

Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing 13*(3), 257–279.

Bardovi-Harlig, K. (2013). Developing L2 pragmatics. *Language Learning, 1* (Suppl.), 68–86.

Bazeley, P. (2018). *Integrating analyses in mixed methods research.* Thousand Oaks, CA: Sage.

Bell, C. A., Jones, N. D., Qi, Y., & Lewis, J. L. (2018). Strategies for assessing classroom teaching: Examining administrators' thinking as validity evidence. *Educational Assessment, 23*(4), 229–249. doi: 10.1080/10627197.2018.1513788

Bell, C., Jones, N., Lewis, J., Qi, Y., Kirui, D., Stickler, L., & Liu, S. (2015). *Understanding consequential assessment systems of teaching: Year 2 final report to Los Angeles Unified School District* (Research Memorandum No. RM-15-12). Princeton, NJ: Educational Testing Service.

Bell, C., Jones, N., Lewis, J., Qi, Y., Stickler, L., Liu, S., & McLeod, M. (2016). *Understanding consequential assessment systems for teachers: Year 1 report to the Los Angeles Unified School District* (Research Memorandum No. RM-16-12). Princeton, NJ: Educational Testing Service.

Bloomberg, L. D., & Volpe, M. (2019). *Completing your qualitative dissertation: A road map from beginning to end* (4th ed.). Thousand Oaks, CA: Sage.

Borman, G. D., & Kimball, S. M. (2005, September). Teacher quality and educational equality: Do teachers with higher standards-based evaluation ratings close student achievement gaps? *The Elementary School Journal, 106*(1), 3–20. Retrieved from https://repository.upenn.edu/cpre_articles/2/

Bos, J., Sanchez, R., Tseng, F., Rayyes, N., Ortiz, L., & Sinicrope, C. (2012). *Evaluation of Quality Teaching for English Learners (QTEL) professional development*. (NCEE 2012-4005). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Boser, U. (2012, March). *Race to the Top: What have we learned from the states so far? A state-by-state evaluation of Race to the Top performance.* Washington, DC: Center for American Progress. Retrieved from https://files.eric.ed.gov/fulltext/ED535605.pdf.

Bowles, M. A. (2010). *The think-aloud controversy in second language research.* New York, NY: Routledge.

Brown, J. D. (2001). *Using surveys in language programs.* New York, NY: Cambridge University Press.

Brown, J. D. (2014). *Mixed methods research for TESOL.* Edinburgh: Edinburgh University Press.

Bunch, G. C., Schlaman, H., & Rutherford-Quach, S. (2019). Where the rubber meets the road: What teachers and students report when implementing high-quality learning designs for English learners. In A. Walqui & G. C. Bunch (Eds.), *Amplifying the curriculum: Designing quality learning opportunities for English learners* (pp. 187–206). San Francisco, CA: WestEd.

Cardenas-Lopez, E. (2015). *Teaching English learners in middle grades: A mixed method study of the dispositions, priorities, and instructional practices of urban bilingual, ESL, and content teachers* [Doctoral Dissertation, National Louis University]. Retrieved from https://digitalcommons.nl.edu/cgi/viewcontent.cgi?article=1116&context=diss

Castellón, M., Cheuk, T., Greene, R., Mercado-Garcia, M., Santos, M., Skarin, R., & Zerkel, L. (2015). Schools to learn from: How six high schools graduate English language learners college and career ready. Retrieved from https://ell.stanford.edu/sites/default/files/Schools%20to%20Learn%20From%20.pdf

Coady, M. R., Miller, D., Zeyjun, J., Heffington, D. V., Lopez, M. P. S., Olszewska, A., de Jong, E., Yilma, T, & Ankeny, R. (2020). Can effective teaching be observed?: Validation of an EL-modified *Framework for Teaching* observation protocol. *TESOL Quarterly, 54(1), 173–200.* doi:10.1002/tesq.544

Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher, 23*(7), 13–20.

Cowan, J., & Goldhaber, D. (2016). National Board Certification and teacher effectiveness: Evidence from Washington State. *Journal of Research on Educational Effectiveness, 9*(3), 233–258. http://dx.doi.org/10.1080/19345747.2015.1099768

Creswell, J. W. (2015). *A concise introduction to mixed methods research.* Thousand Oaks, CA: Sage.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd. ed.). Thousand Oaks, CA: Sage.

Danielson, C. (2007). *Enhancing professional practice: A Framework for Teaching* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Danielson, C. (2011). *The Framework for Teaching evaluation instrument.* Princeton, NJ: The Danielson Group.

Danielson, C. (2013). General questions about the Framework. Retrieved from https://www.danielsongroup.org/questions-about-the-framework-for-teaching

The Danielson Group. (2013). The Framework. Retrieved from https://www.danielsongroup.org/framework

The Danielson Group (2014). Special ed scenarios: Extended examples of levels of performance in special education. Retrieved from http://www.danielsongroup.org/special-education/

The Danielson Group. (2019). Frequently asked questions. Retrieved from https://danielsongroup.org/faq

de Jong, E. J., & Harper, C. A. (2005). Preparing mainstream teachers for English language learners: Is being a good teacher good enough? *Teacher Education Quarterly*, *32*(2), 101–124.

Doerr, S. E. (2012). *Charlotte Danielson's theory of teacher evaluations: A quantitative study of teachers' perceptions on the four domains* [Doctoral Dissertation, Capella University]. Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3499681).

Dwyer, C. A. (1994). *Development of the knowledge base for the Praxis III: Classroom performance assessment criteria.* Princeton, NJ: Educational Testing Service.

Dörnyei, Z. (2007). *Research methods in Applied Linguistics.* Oxford, UK: Oxford University Press.

Dörnyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research: Construction, administration, and processing* (2nd ed.)*.* New York, NY: Routledge.

Echevarria, J., Vogt, M., & Short, D. J. (2004). Making content comprehensible for English learners: the SIOP model (2nd ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Elfers, A., & Achberger, E. (2017, April). *Analysis of school employees' evaluation survey (SEES): Final report.* Seattle, WA: University of Washington, College of Education.

Elfers, A. M., & Plecki, M. L. (2017). *Washington's teacher and principal evaluation system: Examining the implementation of a complex policy*. Seattle, WA: Center of the Study of Teaching and Policy, University of Washington.

Elfers, A. M., & Plecki, M. L. (2019). *School leaders and teacher evaluation: Learning, leading, and balancing responsibilities.* Seattle, WA: Center for the Study of Teaching and Policy, University of Washington.

Ellis, R. (2006). Researching the effects of form-focused instruction on L2 acquisition. *AILA Review, 19*(1), 18–41.

Finnegan, P., & Choi, J. (n.d.). High graduation rates: English language learners. Retrieved from http://www.k12.wa.us/MigrantBilingual/pubdocs/ELgradstudy.pdf

Freeman, R. F. (2016, October). *Conversations with Charlotte: ESL teacher evaluations.* Paper presented at the WAESOL conference, Spokane, WA.

Freeman, R. F. (2017, March). *Conversations with Charlotte: ESL teacher evaluations using the Danielson Framework.* Paper presented at the TESOL convention, Seattle, WA.

Gallagher, H. A. (2004). Vaughn Elementary's innovative teacher evaluation system: Are   teacher evaluation scores related to growth in student achievement? *Peabody Journal of Education, 79*(4)*,* 79–107.

Gánem-Gutiérrez, G. A. (2013). Sociocultural theory and second language development: Theoretical foundations and insights from research. In M. Garciá Mayo, M. Gutierrez Mangado & M. Martínez-Adrián (Eds.), *Contemporary approaches to second language acquisition* (pp. 129–152). Philadelphia, PA: John Benjamins Publishing.

Garrett, R., & Steinberg, M. P. (2015). Examining teacher effectiveness using classroom observation scores: Evidence from the randomization of teachers to students. *Educational Evaluation and Policy Analysis, 37*(2), 224–242.

Gibbons, P. (2003). Mediating language learning: Teacher interactions with ESL students in a content-based classroom. *TESOL Quarterly, 37*(2), 247–273.

Gill, B., Shoji, M., Coen, T., & Place, K. (2016). *The content, predictive power, and potential bias in five widely used teacher observation instruments* (REL 2017–191). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic. Retrieved from https://ies.ed.gov/ncee/edlabs/regions/midatlantic/pdf/REL_2017191.pdf.

Goe, L. (2007, October). *The link between teacher quality and student outcomes: A research synthesis.* Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from https://eric.ed.gov/?id=ED521219

Goe, L., Bell, C., & Little, O. (2008, June). *Approaches to evaluating teacher effectiveness: A research synthesis.* Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from http://www.tqsource.org/publications/EvaluatingTeachEffectiveness.pdf.

Goertzen, H. T. (2015). *Examining teacher qualifications, characteristics, and practices related to English learners in the state of Alabama: An exploratory study on teacher quality* [Doctoral dissertation, Auburn University]. Retrieved from https://etd.auburn.edu/xmlui/bitstream/handle/10415/4710/Goertzen_FinalDissertation_7-22-15.pdf?sequence=2&isAllowed=y

Goldenberg, C. (2008). Teaching English language learners: What the research does and does not say. *American Educator, 32*(2), 8–44.

Graham, M., Milanowski, A., & Miller, J. (2012). *Measuring and promoting inter-rater agreement of teacher and principal performance ratings*. Washington, DC: Center for Educator Compensation Reform.

Greeno, J. G. (1997). On claims that answer the wrong questions. *Educational Researcher, 26*(1), 5–17.

Hammond, J., & Gibbons, P. (2005). Putting scaffolding to work: The contribution of scaffolding in articulating ESL education. *Prospect, 20*(1), 6–30.

Hartman, L., & Billings, E. (2019). What makes me who I am?: Engaging beginning-level English learners in quality learning. In A. Walqui & G. C. Bunch (Eds.), *Amplifying the curriculum: Designing quality learning opportunities for English learners* (pp. 165–186). San Francisco, CA: WestEd.

Hattie, J. A. C. (2003, October). Teachers make a difference: What is the research evidence? Paper presented at the Building Teacher Quality: What does the research tell us ACER Research Conference, Melbourne, Australia. Retrieved from http://research.acer.edu.au/research_conference_2003/4/

Hattie, J. (2009). *Visible learning: A synthesis of over 200 meta-analyses relating to achievement.* New York, NY: Routledge.

Heneman, H. G., III, & Milanowski, A. T. (2003, June). Continuing assessment of teacher reactions to a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education, 17*(2), *173–195.*

Heneman, H. G., III, Kimball, S., & Milanowski, A. (2006). *The teacher sense of efficacy scale: Validation evidence and behavioral prediction* (WCER Working Paper No. 2006-7). Madison, WI: Wisconsin Center for Education Research. Retrieved from http://www.wceruw.org/publications/workingPapers/Working_Paper_No_2006_07.pdf.

Heneman, H. G., III, Milanowski, A., Kimball, S. M., & Odden, A. (2006). *Standards-based teacher evaluation as a foundation for knowledge- and skill-based pay* (CPRE Policy Brief No. RB-45). Philadelphia: Consortium for Policy Research in Education. Retrieved from http://www.wcer.wisc.edu/cpre/publications/rb45.pdf

Heritage, M., Walqui, A., & Linquanti, R. (2015). *English language learners and the new standards: Developing language, content knowledge, and analytical practices in the classroom.* Cambridge, MA: Harvard University Press.

Holdheide, L. R., Goe, L., Croft, A., & Reschly, D. J. (2010, July). *Challenges in evaluating Special Education teachers and English language learner specialists.* Research and Policy Brief: National Comprehensive Center for Teacher Quality, 1–40. Retrieved from files.eric.ed.gov/fulltext/ED520726.pdf

Holtzapple, E. (2003). Criterion-related validity evidence for a standards-based teacher evaluation system. *Journal of Personnel Evaluation in Education, 17*(3), 207–219.

IBM Corporation (2017). IBM SPSS Statistics for MacIntosh, Version 25.0 [computer software]. Armonk, NY: IBM Corp.

Jiang, J. Y., Sporte, S. E., & Luppescu, S. (2015, March). Teacher perspectives on evaluation reform: Chicago's REACH students. *Educational Researcher, 44*(2), 105–116. doi: 10.3102/00131189X15575517

Johnston, J. A. (2013). *Elementary ELL interaction: Mainstream vs. sheltered instructional settings* [Unpublished doctoral dissertation]. Washington State University, Pullman, WA.

Jones, N. D., Buzick, H. M., & Turkan, S. (2013). Including students with disabilities and English learners in measures of educator effectiveness. *Educational Researcher, 42*(4), 234–241.

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle, WA: Bill & Melinda Gates Foundation. Retrieved from https://files.eric.ed.gov/fulltext/ED540960.pdf.

Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. *Journal of Human Resources*, *46*(3), 587–613.

Kimball, S. M., White, B., Milanowski, A. T., & Borman, G. (2004). Examining the relationship between teacher evaluation and student assessment results in Washoe County. *Peabody Journal of Education, 79*(4), 54–78.

Koch, C. (2014, August). *Guidance on building teacher evaluation systems for teachers of students with disabilities, English learners, and early childhood students.* Retrieved from www.isbe.net/peac/pdf/guidance/14-3-teacher-eval-sped-ell-preschool.pdf

LaFlair, G. T., Egbert, J., & Plonsky, L. (2015). A practical guide to bootstrapping descriptive statistics, correlations, *t*-tests*,* and ANOVAs*.* In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 46–77). New York, NY: Routledge.

Larson-Hall, J. (2016). *A guide to doing statistics in second language research using SPSS and R* (2nd ed.). New York, NY: Routledge.

Douglas, W. O. & Supreme Court Of The **United States**. (1973) **U.S.** Reports: **Lau v**. **Nichols**, **414 U.S. 563** . [Periodical] Retrieved from the Library of Congress, https://www.loc.gov/item/usrep414563/.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* New York, NY: Cambridge University Press.

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning, 60*(2), 309–365.

Loeb, S., Soland, J., & Fox, L. (2014). Is a good teacher a good teacher for all? Comparing value-added of teachers with their English learners and non-English learners. *Educational Evaluation and Policy Analysis*, *36*(4), 457–475.

Martinez, N. (2014). *Teacher and evaluator perspectives on the teacher evaluation process for teachers of English language learners: A change leadership plan* [Doctoral dissertation, National Louis University]. Retrieved from http://digitalcommons.nl.edu/diss/98/

Master, B., Loeb, S., Whitney, C., & Wyckoff, J. (2016). Different skills? Identifying differentially effective teachers of English language learners. *The Elementary School Journal*, *117*(2), 261–284.

McKay, S. L. (2005). Sociolinguistics and second language learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 281–299). Mahwah, NJ: Erlbaum.

Merritt, E. G., Palacios, N., Banse, H., Rimm-Kaufman, S.E., & Leis, M. (2017). Teaching practices in Grade 5 mathematics classroom with high-achieving English learner students. *The Journal of Educational Research, 110*(1), 17–31. http://dx.doi.org/10.1080/00220671.2015.1034352

Milanowski, A. (2004). The relationship between teacher performance evaluation scores and student achievement: Evidence from Cincinnati. *Peabody Journal of Education, 79*(4), 33–53.

Milanowski, A. T., & Heneman, H. G., III. (2001). Assessment of teacher reactions to a standards-based teacher evaluation system: A pilot study. *Journal of Personnel Evaluation in Education, 15*(3), 193–212.

Miller, T. D., & Hanna, R. (2014, March 24). *Four years later, are Race to the Top states on track?* Washington, DC: Center for American Progress. Retrieved from https://files.eric.ed.gov/fulltext/ED564595.pdf

Moss, J. T. (2015). *The Danielson model of teacher evaluation: Exploring teacher perceptions concerning its value in shaping and improving instructional practice* [Doctoral dissertation, Seton Hall University]. Retrieved from www.scholarship.shu.edu/cgi/viewcontent.cgi?article=3133&context=dissertations

National Academies of Sciences, Engineering, and Medicine (NASEM). (2017). *Promoting the educational success of children and youth learning English*. Washington, DC: The National Academies Press. https://doi.org/10.17226/24677.

National Academies of Sciences, Engineering, and Medicine (NASEM). (2018). *English learners in STEM subjects: Transforming classrooms, schools, and lives*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25182.

New Jersey Government (NJ.gov). (n.d.). Danielson/ELL Crosswalk. Retrieved from https://www.nj.gov/education/bilingual/resources/Danielson.pdf

Newton, R. R., & Rudestam, K. E. (2013). *Your statistical consultant* (2nd ed.). Thousand Oaks, CA: Sage.

New York City Department of Education (NYCDE). (2014, March). Specific considerations for teachers of English language learners. Retrieved from http://schools.nyc.gov/NR/rdonlyres/04D90E09-0E02-4EEE-9988-C7B7CD81D676/0/SpecificConsiderationsforTeachersofELLs.pdf

Nunan, D., & Bailey, K. M. (2009). *Exploring second language classroom research: A comprehensive guide.* Boston, MA: Heinle Cengage Learning.

Office of the Superintendent of Public Instruction (OSPI). (n.d.). Washington state criteria student growth rubrics with critical attributes (Version 1.2). Retrieved from https://www.k12.wa.us/sites/default/files/public/tpep/studentgrowth/WA%20Student%20Growth%20Rubric_with%20Critical%20Attributes.pdf

Office of the Superintendent of Public Instruction (OSPI). (n.d.). Student growth. Retrieved from https://www.k12.wa.us/educator-support/teacherprincipalevaluation-program/teachers-and-principals/student-growth

Office of the Superintendent of Public Instruction (OSPI). (n.d.). Teacher evaluation criteria and descriptors. Retrieved from https://www.k12.wa.us/sites/default/files/public/tpep/pubdocs/Teacher_Criteria-Descriptors.pdf

Office of the Superintendent of Public Instruction (OSPI). (2011). Washington state school districts: Maps and websites. Retrieved from http://www.k12.wa.us/maps/Maps.aspx

Office of the Superintendent of Public Instruction (OSPI). (2017, May 23). Teacher/principal evaluation program: State evaluation criteria, frameworks, and rubrics. Retrieved from http://www.k12.wa.us/TPEP/Frameworks/default.aspx

Office of the Superintendent of Public Instruction (OSPI). (2017, July). Comprehensive evaluation–Danielson–certificated classroom. Retrieved from https://www.k12.wa.us/sites/default/files/public/tpep/frameworks/danielson/dan-compvsfocus.pdf

Office of the Superintendent of Public Instruction (OSPI). (2017, September). Focused evaluation–Danielson–certificated classroom. Retrieved from https://www.k12.wa.us/sites/default/files/public/tpep/frameworks/danielson/danielson_focusedscoringdiagram.pdf

Office of Superintendent of Public Instruction (OSPI). (2018). Danielson Framework for Teaching (2011) rubrics by Washington State (Version 1.1). Retrieved from http://www.k12.wa.us/TPEP/Frameworks/Danielson/Danielson-rubrics-by-criteria-critical-attributes.pdf

Office of Superintendent of Public Instruction (OSPI). (2018). Washington State report card. Retrieved from http://reportcard.ospi.k12.wa.us/summary.aspx?groupLevel=District&schoolId=1&reportLevel=State&yrs=2017-18&year=2017-18

Packer, M. J., & Goicoechea, J. (2000). Sociocultural and constructivist theories of learning: Ontology, not just epistemology. *Educational Psychologist, 35*(4), 227–241.

Pennsylvania Department of Education (PDE). (2013). Possible examples of how the Framework for Teaching could apply to English as a second language teachers. Retrieved from http://dev.static.pdesas.org/content/documents/Examples%20for%20English%20as%20a%20Second%20Language%20Teachers.pdf

Plano Clark, V., & Ivankova, N. (2016). *Mixed methods research: A guide to the field*. Thousand Oaks, CA: SAGE Publications, Inc. doi: 10.4135/9781483398341

Plonsky, L. (Ed.). (2015a). *Advancing quantitative methods in second language research.* New York, NY: Routledge.

Plonsky, L. (2015b). Statistical power, *p* values, descriptive statistics, and effect sizes: A "back-to-basics" approach to advancing quantitative methods in L2 research. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 23–45). New York, NY: Routledge.

Plonsky, L., & Oswald, F. L. (2014). How big is 'big'? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912.

Pritchett, B. (2013, September 6). *States implement new teacher evaluations that include student performance.* [Web log comment]. Retrieved from http://news.heartland.org/newspaper-article/2013/09/06/teachers-most-states-encounter-new-evaluations

Qi, Y., Bell, C. A., Jones, N. D., Lewis, J. M., Witherspoon, M. W., & Redash, A. (2018). *Administrators' uses of teacher observation protocol in different rating contexts* (Research Report No. RR-18-18). Princeton, NJ: Educational Testing Service. https://doi.org/10.1002/ets2.12205

Quality Teaching for English Learners (QTEL). (2012). Using the principles of quality teaching for English learners to observe classroom instruction [Rubric]. Unpublished instrument. San Francisco, CA: WestEd.

Quality Teaching for English Learners (QTEL). (2019a). QTEL theoretical framework. Retrieved from https://www.qtel.wested.org/framework

Quality Teaching for English Learners (QTEL). (2019b). Who we are. Retrieved from https://www.qtel.wested.org/

Reykdal, C. (2019, May 31). Bulletin No. 031-19 Educator growth and development . Olympia, WA: Office of Superintendent of Public Instruction. Retrieved from https://www.k12.wa.us/sites/default/files/public/bulletinsmemos/bulletins2019/B031-19.pdf

Rimm-Kaufman, S. E., Larsen, R. A., Baroody, A. E., Curby, T. W., Ko, M., Thomas, J. B., & DeCoster, J. (2014). Efficacy of the responsive classroom approach results from a 3-year, longitudinal randomized controlled trial. *American Educational Research Journal, 51*(3), 567–603.

Roegman, R., Goodwin, A. L., Reed, R., & Scott-McLaughlin, R. M. (2016, May). Unpacking the data: An analysis of the use of Danielson's (2007) Framework for Professional Practice in a teaching residency program. *Educational Assessment, Evaluation and Accountability*, *28*(2), 111–137.

Rudestam, K. E., & Newton, R. R. (2015). *Surviving your dissertation: A comprehensive guide to content and process* (4th ed.). Thousand Oaks, CA: Sage.

Samson, J. F., & Collins, B. A. (2012, April). *Preparing all teachers to meet the needs of English language learners: Applying research to policy and practice for teacher effectiveness.* Washington, DC: Center for American Progress. Retrieved from files.eric.ed.gov/fulltext/ED535608.pdf

Sanchez, C. (2017, February 23). English language learners: How your state is doing. Retrieved from https://www.npr.org/sections/ed/2017/02/23/512451228/5-million-english-language-learners-a-vast-pool-of-talent-at-risk

Sartain, L., Stoelinga, S. R., & Krone, E. (2010, June). *Rethinking teacher evaluation: Findings from the first year of the excellence in teaching project in Chicago public schools. Research Report.* Consortium on Chicago School Research. Retrieved from https://consortium.uchicago.edu/sites/default/files/2018-10/Teacher%20Eval%20Final.pdf

Sartain, L., Stoelinga, S. R., & Brown, E. R. (2011). *Rethinking teacher evaluation in Chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation. Research Report*. Consortium on Chicago School Research. Retrieved from https://consortium.uchicago.edu/sites/default/files/201810/Teacher%20Eval%20Report20FINAL.pdf

Sato, M., & Ballinger, S. (Eds.). (2016). *Peer interaction and second language learning: Pedagogical potential and research agenda.* Philadelphia, PA: John Benjamins Publishing Company.

Schumacher, G. T. (2004). *Perceptions of the impact of a standards-based teacher evaluation system, based on the Danielson Framework for Teaching model, on teaching and student learning* [Doctoral dissertation, University of Wisconsin-Madison]. Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3128070).

Short, D. J., Becker, H., Cloud, N., Hellman, A. B., & Levine, L. N. (2018). *The six principles for exemplary teaching of English learners: Grades K-12.* Alexandria, VA: TESOL International Association.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*(1), 1–22.

Staehr Fenner, D. S., Kozik, P., & Cooper, A. (2014). Evaluating teachers of all learners. *Leadership, 43*(4), 8–12.

Staehr Fenner, D., Kozik, P., & Cooper, C. (2015). *Evaluating all teachers of English learners and students with disabilities.* Thousand Oaks, CA: Corwin.

Steinberg, M. P., & Donaldson, M. L. (2016, Summer). The new educational accountability: Understanding the landscape of teacher evaluation in the post-NCLB era. *Education Finance and Policy*, *11*(3), 340–359. doi:10.1162/EDFP_a_00186

Steinberg, M. P., & Garrett, R. (2016, June). Classroom composition and measured teacher performance: What do teacher observation scores really measure? *Educational Evaluation and Policy Analysis, 38*(2), 293–317. doi: 10.3102/0162373715616249

Steinberg, M. P., & Sartain, L. (2015a). Does teacher evaluation improve school performance? Experimental evidence from Chicago's excellence in teaching project. *Association for Education Finance and Policy, 10*(4), 535–572.

Steinberg, M. P., & Sartain, L. (2015b). Does better observation make better teachers? *Education Next, 15*(1), 71–76. Retrieved from http://educationnext.org/better-observation-make-better-teachers/

Storch, N. (2017). Sociocultural theory in the L2 classroom. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp.69–83). New York, NY: Routledge.

Swain, M. (2006). Languaging, agency and collaboration in advanced language proficiency. In I. H. Byrnes (Ed.), *Advanced language learning: The contribution of Halliday and Vygotsky* (pp. 95–108). London: Continuum.

Sweeley, T. M. (2004). *Teachers' attitudes towards Charlotte Danielson's four domains of teacher evaluation* [Doctoral dissertation, Widener University]. Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3120736).

Teachers of English to Speakers of Other Languages (TESOL). (2010). *TESOL/NCATE standards for the recognition of initial TESOL programs in P-12 ESL teacher education.* Alexandria, VA: TESOL. Retrieved from https://www.tesol.org/advance-the-field/standards/tesol-caep-standards-for-p-12-teacher-education-programs

Teddlie, C., & Yu, F. (2007). Mixed methods sampling: A typology with examples. *Journal of Mixed Methods Research, 1*(1), 77–100.

Tellez, K., & Waxman, H. (2004). *Quality teachers for English Language learners: A research synthesis.* Philadelphia, PA: Mid-Atlantic Laboratory for Student Success, Publication Series No. 2. Retrieved from http://files.eric.ed.gov/fulltext/ED484732.pdf

Turkan, S. (2016). Considerations for evaluating instruction for English language learners. In R. E. Gabriel & R. L. Allington (Eds.), *Evaluating literacy instruction: Principles and promising practices* (pp. 44–57). New York, NY: Routledge.

Turkan, S., & Buzick, H. M. (2016). Complexities and issues to consider in the evaluation of content teachers of English language learners. *Urban Education, 5*(2), 221–248. doi: 10.1177/0042085914543111

Turkan, S., Croft, A., Bicknell, J., & Barnes, A. (2012). *Assessing quality in the teaching of content to English language learners.* (Research Report, ETS RR-12-10). Princeton: NJ. Retrieved from https://files.eric.ed.gov/fulltext/EJ1109845.pdf

Turner, J. L. (2014). *Using statistics in small-scale language education research: Focus on non-parametric data.* New York, NY: Routledge.

U.S. Department of Education. (2015). Race to the Top-Equity and Opportunity: Support to close achievement and opportunity gaps. Retrieved from https://www.ed.gov/documents/budget/2015/rtt-equity-opportunity-2015.pdf

van Lier, L. (1996). *Interaction in the language curriculum: Awareness, autonomy, and, authenticity.* London, UK: Longman.

van Lier, L. (2004). *The ecology and semiotics of language learning: A sociocultural approach.* Amsterdam, The Netherlands: Kluwer Academic.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Vygotsky, L., & Luria, A. (1994). Tool and symbol in child development. In R. van der Veer & J. Valsiner (Eds.), *The Vygotsky reader* (pp.99–174). Oxford, UK: Blackwell Publishers.

Walqui, A. (2019). Designing the amplified lesson. In A. Walqui & G. C. Bunch (Eds.), *Amplifying the curriculum: Designing quality learning opportunities for English learners* (pp. 43–69). San Francisco, CA: WestEd.

Walqui, A., & van Lier, L. (2010). *Scaffolding the academic success of adolescent English language learners.* San Francisco, CA: WestEd.

Walqui, A., & Heritage, M. (2018). Meaningful classroom talk: Supporting English learners' oral language development. *American Educator, 42*(3), 18–23, 39.

Walqui, A., & Bunch, G. C. (Eds.). (2019). *Amplifying the curriculum: Designing quality learning opportunities for English learners.* San Francisco, CA: WestEd.

Waxman, H. C., & Tellez, K. (2002). *Research synthesis on effective teaching practices for English language learners.* Publication series. Philadelphia: Mid-Atlantic Laboratory for Student Success. Retrieved from https://eric.ed.gov/?id=ED474821

Weber, G., & West, B. (2016, June 14). *Examination of the school employee evaluation survey*. American Institutes for Research (AIR)*.* Retrieved from http://www.k12.wa.us/TPEP/pubdocs/AIR_Analysis_2014-2015_SEES_Data.pdf

Whitehurst, G. J., Chingos, M. M., & Linquist, K. M. (2014, May). *Evaluating teachers with classroom observations: Lessons learned in four districts.* Brown Center on Education Policy at Brookings*,* 1–27. Retrieved from http://www.brookings.edu/~/media/research/files/reports/2014/05/13-teacherevaluation/evaluating-teachers-with-classroom-observations.pdf

Wittink, M. N., Barg, F. K., & Gallo, J. J. (2006). Unwritten rules of talking to doctors about depression: Integrating qualitative and quantitative methods. *Annals of Family Medicine, 4*(4), 302–209.