Improving Student Evaluations of TESOL Practitioners

K. James Hartshorn, Brigham Young University, Provo, Utah, USA
Judson M. Hart, Emmersion, Lehi, Utah, USA
Benjamin L. McMurry, Brigham Young University, Provo, Utah, USA

Abstract

Teacher evaluation in second language education continues to become more important to stakeholders and is increasingly associated with higher-stakes decisions that impact teachers in substantial ways. Nevertheless, many program administrators struggle to know what to include in teacher evaluations and how best to use the results. We know very little about the kinds of factors that lead toward more favorable teacher evaluations. Thus, in addition to identifying best practices for developing and utilizing teacher evaluation instruments and data, this study sought to identify factors that lead toward better teacher evaluations. This study analyzed 5461 student evaluations of their teachers using exploratory and confirmatory regression analysis to identify eight factors with the greatest influence on student recommendation of their TESOL practitioners. Results included pedagogical practices accounting for more than half of the variability associated with teacher recommendation (adjusted $R^2 > .55$, $p < .001$). These included exemplifying professionalism, relying on course outcomes, cultivating a positive learning environment, evaluating learning effectively, optimizing class time, planning lessons effectively, utilizing homework strategically, and providing meaningful and timely feedback. Implications and applications for these findings are discussed.

Keywords: Principled pedagogical practices for TESOL, student evaluation of teachers, TESOL practitioner improvement, net promoter scores

Introduction

The importance of teacher evaluation continues to increase in many global contexts including those associated with teaching English to speakers of other languages (TESOL). Boraie (2014) suggests “Teacher evaluation is here to stay, and

the question to be considered is not whether there should be a teacher evaluation system but how to evaluate teachers effectively." Moreover, evidence suggests that the stakes associated with teacher evaluation may be higher than ever before as more institutions use evaluations to shape institutional policy, practice, and employment decisions (e.g., Boraie, 2014; Fantini, 2018; Howard & Donaghue, 2015; Thomsen, 2014; Rucinski & Diersing, 2014). Efforts to improve the quality and efficacy of teacher evaluation has generated a great deal of scholarship over recent decades (e.g., Berk, 2005; Howard & Donaghue, 2015; Paulsen, 2002; Pennington & Young, 1989; Spencer & Schmelkin, 2002). What may be less clear, however, is an appropriate understanding of what teacher evaluations should include and what factors may influence student evaluations of teachers. As work continues to improve the value and utility of teacher evaluations, efforts to extract greater understanding about language learner needs evident in this medium should also increase. Therefore, this study examines thousands of recent teacher evaluations in a TESOL context to identify factors associated with teacher preparation and practice as well as the learning context that may help clarify important ESL learner perceptions. The resulting insights should be highly relevant for program administrators seeking to improve teacher evaluation, TESOL practitioners who hope to receive more favorable ratings from their students, and researchers interested in factors affecting classroom learning.

Literature Review

Relevance of Evaluation in TESOL

Teacher evaluation has evolved a great deal over the past century, impacting many TESOL contexts worldwide. Although some issues are specific to TESOL, others are applicable to much broader educational contexts. For example, early in the twentieth century in the United States, teacher evaluation generally emphasized traits such as the 83 attributes highlighted by Charters and Waples (1929). These included attractiveness, cheerfulness, dignity, health, intelligence, and so on. By mid-century, however, the focus had shifted toward better professional development as demonstrated through competency exams and certifications. Nevertheless, such credentials did not always translate to better student performance. Subsequently, reports in the United States such as A Nation at Risk (National

Commission on Excellence in Education, 1983) precipitated a shift in thinking toward learning outcomes. This was further solidified by the No Child Left Behind Act (NCLB, 2002), which marshalled in a new era of standardized testing and evaluation. The same influences that precipitated increased standardized testing in the United States impacted other countries as well, and by 2006, the number of nations incorporating standardized testing more than doubled compared to the decade before (Benavot & Tanner, 2007).

This new global culture of educational assessment and evaluation continues to impact teachers in many ways. For example, since 2009 nearly 65% of the states in the United States have dramatically overhauled their systems for teacher evaluation, which have had a substantial impact on institutional policy (e.g., Thomsen, 2014) as well as how institutions hire, fire, train, assess, and reward their teachers (Rucinski & Diersing, 2014). Nevertheless, the debate over whether these changes actually help students achieve learning outcomes remains contentious. There are numerous kinds of ESL/EFL programs around the world in a wide variety of settings ranging from government-sponsored schools for children or adults to community programs designed to help residents or refugees develop the English language skills they need to gain employment and flourish within in society. This vast diversity of contexts in which these programs operate may make the processes and results of this work particularly valuable. Regardless of whether the context may be ESL or EFL, whether working with children or adults, or whether educators are teachers or administrators, language educators are increasingly more likely to grapple with a variety of issues associated with teacher evaluations.

The growing importance and use of teacher evaluations in many language education contexts raise many crucial questions such as What should evaluations include? How can administrators ensure that they are fair? What should be done with evaluation data once it is collected? While many institutions continue to wrestle with these kinds of questions, we add an additional question of particular relevance to this study, What insights can evaluation-based research provide to the field of language teaching? We believe that a systematic study of carefully designed teacher evaluations may unlock valuable insights regarding student experience that could impact language development.

Effective Teacher Evaluation

In considering the benefits of evaluation-based research, it is important to remember that attempts to measure the contribution of a teacher on student learning will be complex and multidimensional. For example, some researchers estimate that the actual influence a teacher has on student performance ranges from 14% down to 1% depending on a wide array of factors (American Statistical Association, 2014). Therefore, we side with scholars such as Mathis (2012) and Boraie (2014) who suggest that effective teacher evaluation needs to be established on a multitude of factors. Although it may be appropriate to examine actual student achievement of learning outcomes as part of a broader evaluation, this should be done in view of the specific teaching and learning context, recognizing that even exceptional teachers may have little control over student performance. A broader body of evaluation might also include observations conducted by administrators and peers, citizenship measures associated with dependability, collegiality, productivity, mentoring, and cooperation as well as the learners' evaluations of their teachers. While these and many other factors may be important to examine as part of a broader teacher evaluation strategy, this study focuses exclusively on the tangible instrument used by students to evaluate their teachers.

We acknowledge that some teachers are skeptical regarding the validity and utility of evaluations generated by students. Concerns may range from assumptions that evaluation results are simply a function of teacher popularity or that easy classes will produce higher evaluation scores. Although students may lack expertise in a number of areas relative to overall teacher performance, the students themselves may be in the strongest possible position to evaluate the effectiveness of many aspects of their classroom experience. Moreover, many scholars such as Aleamoni (1999) have debunked numerous myths about student evaluations of teachers, providing strong evidence for the reliability and validity of "well-constructed instruments and procedures" (p. 155) and that such evaluation results can be used effectively to improve teaching.

With the intent of improving our own teacher evaluations and procedures associated with our intensive English program, we analyzed our instruments, carefully weighing relevant literature and the specific needs of our own institution. Our approach was consistent with recommendations of Spooren, Brockx,

and Mortelmans (2013) who suggested that when developing evaluation instruments, "institutions should be able to select the aspects that are most important, according to their educational vision and policy" and that many stakeholders such as "administrators, teachers, and students…should be involved in the definition of these characteristics" (p. 603). Through this process, we concluded that our teacher evaluations should include two elements that were not currently in use but that were needed to optimize the effectiveness and utility of the instrument. The first was an overall summary measure of performance that could be used across a variety of teachers and contexts. The second was to replace more generic indicators of performance with very focused areas of specific interest to the institution. The relevance and applications of these two components are described below.

Summary score of student sentiment

One challenge with many evaluations is the lack of an effective indication of overall student sentiment. While perhaps incomplete, a general sense of practitioner performance may help simplify evaluation and provide administrators with broad insight. Research suggests that a single overall score for teacher evaluation can be well correlated with other important measures and may be useful for summative purposes (Cashin & Downey, 1992; Fantini, 2018). An overall score facilitates broad comparisons across teachers and courses over time, allowing administrators to identify relevant trends and to make appropriate programmatic adjustments.

In this regard, the world of business may have a useful metric. Today numerous companies ranging from airlines and retail to healthcare and technology use what is called a net promoter score (NPS) to provide an overall evaluative indication of the products or services they provide (Reichheld, 2003, 2006). Customers are presented with a single question to answer using a scale of zero to ten: "How likely is it that you would recommend [company, product, or service] to a friend or colleague?" Proponents describe the NPS as providing essential insights (e.g., Martin, 2011; Reichheld & Markey, 2011), and that it has been used with great success across a wide array of businesses such as banking, cosmetology, telecommunications, healthcare, and so on (e.g., Hamilton, Lane, Gaston, Patton, MacDonald, Simpson, & Howie, 2014; Leisen Pollack & Alexandrov, 2013).

Traditionally, respondents providing a nine or ten were labeled promoters, those responding with a seven or eight were considered passives, and those responding with a one through six were identified as detractors. Typically, the NPS has been calculated by subtracting the proportion of promoters from the proportion of detractors while ignoring the passives (e.g., 70% promoters, 20% passives, and 10% detractors would yield an NPS of 60). Though its lack of sophistication has been seen by some as a considerable limitation (e.g., Krol, Boer, Delnoij, & Rademakers, 2015; Mandal, 2014), its simplicity has also been seen as its greatest strength by many proponents and critics alike (Bendle & Bagga, 2016). Many managers like the ease with which it can be elicited by clients and then analyzed and interpreted across a variety of contexts. Though widely popular in corporate America today, the NPS is not without controversy.

Because the 11-point scale is collapsed into just three categories and the passive scores are ignored, critics have noted that important numerical information is lost when the score is calculated (e.g., Bendle & Bagga, 2016; Schneider, Berent, Thomas, & Krosnick, 2008). They suggest that utilizing the complete scale may provide more evaluative insight. Thus, in an attempt to leverage the practical benefits of an overall score in our own teacher evaluation, we determined to calculate the average of the entire 0-10 scale, based on the same question regarding the respondent's likelihood of recommending the teacher to a friend or colleague. We also determined to strengthen the evaluation by connecting the score to additional quantitative and qualitative information provided by the respondents. We refer to our modified overall score based on the complete 0-10 scale as our teacher recommendation score.

Identifying Core Pedagogical Practices

In addition to the use of an overall summary score based on student recommendation of the TESOL practitioner, the second improvement to our student evaluation was the inclusion of specific pedagogical practices addressed in the literature and consistent with our values, culture, and beliefs about language learning. Before teachers can be evaluated effectively, an institution must have a clear idea of the criteria they would like to use. Though some programs are forced to use generic evaluations, this may not be the best approach (e.g., Hill & Grossman,

2013; Hunt, Gurvitch, & Lund, 2016; Pratt, 2002). Rather than using general standards that may be unrelated to core ideals for effective practice in specific contexts, standards must be clearly articulated and should grow out of the underlying purpose for the teaching and learning within a particular program setting. They should be consistent with the collective values and beliefs of the institution and be rooted in insights gleaned from theory and practice. There will also need to be an appropriate way to measure performance levels for each of these standards.

With these ideals in mind, we set out to identify key pedagogical practices that could be used for training and evaluation purposes specific to our TESOL context. We sought to identify the core pedagogical practices that best capture and reflect our institutional values, beliefs, and aspirations for teacher performance. The process drew on input from administrators, teachers, staff, and other stakeholders. These practices were refined in collaborative meetings over the course of many months and were based on the practical experience of the stakeholders as well as relevant literature. In working through this process, an attempt was made to balance the desire for a comprehensive list of the most consequential practices in second language teaching and learning with the need to keep the set of points simple enough to ensure it could be easily conceptualized and successfully implemented.

The final list included eight pedagogical practices designed to guide classroom preparation and teaching. Though the scientific evidence of the benefits of these practices may be stronger for some than for others, the consensus of stakeholders was that each practice that survived the winnowing process was very important to our context. An abundance of literature also made a strong case for the relevance of each. The eight pedagogical practices are listed here with an abbreviated sample of relevant literature (See Appendix A for more detail regarding each practice).

1. Rely on course outcomes (e.g., Basturkmen, 2010; Leung, 2012; Richards, 2013).
2. Plan lessons effectively (e.g., Baecher, Farnsworth, & Ediger, 2014; Liyanage, & Bartlett, 2010; Milkova, 2012; Pang, 2016).
3. Optimize class time (e.g., Calderón, Slavin, &, Sánchez, 2011; Murray & Christison, 2010; Rossiter, Derwing, Manimtim, &, Thomson, 2010; Tan, Nabb, Aagard, &, Kim, 2010; Walsh, 2006).

4. Cultivate a positive learning environment (e.g., Brown, 2006; Oxford, 1999; Tsiplakides, & Keramida, 2010; Tsui, 1996; Young, 1991).
5. Evaluate learning effectively (e.g., Abedi, 2010; Bailey & Heritage, 2014; Clark, 2012; Frey, Schmitt, &, Allen, 2012; Ketabi & Ketabi, 2014; McMillan, 2013).
6. Utilize homework strategically (e.g., Gershenson & Holt, 2015; Maltese, Tai, &, Fan, 2012; McReynolds, 2010; Walberg, Paschal, & Weinstein, 1985; Wallinger, 2008).
7. Provide meaningful and timely feedback (e.g., Hartshorn & Evans, 2015; Fordham, 2015; Su & Tian, 2016).
8. Exemplify professionalism (e.g., Alsalahi, 2015; Farrell, 2015; Lorimer, & Schulte, 2012; Orlich, Harder, Trevisan, Brown, &, Miller, 2016; Sawyer, Andzik, Kranak, Willke, Curiel, Hensley, & Neef, 2017; Vu, 2016).

Once these practices were established, they were distributed to teachers, posted in hallways, emphasized in teacher training and in-service meetings, and implemented as the focus of teacher observation and professional development activities. Thus, rather than utilizing a teacher evaluation that was detached from daily practice, our updated instrument was designed to reflect practices at the very core of what we expected of our teachers on a daily basis.

Although this list of pedagogical practices was specifically designed for our unique setting, these practices should be highly relevant for most language teaching and learning contexts. Nevertheless, we recognize that some institutions might benefit from placing greater emphasis on particular practices identified here or from focusing on other components that may not have been included in our list.

Development of the Evaluation Instrument

The new teacher evaluation instrument made use of several item types to extract both quantitative and open-ended data from students. A question for each pedagogical practice was framed according to the frequency with which a student felt that the teacher effectively demonstrated that practice. Since the pedagogical practices were originally articulated for teachers and administrators rather than for students, some language was simplified in the instrument to help ensure student understanding (see Appendix B). For example, for optimizing class time, students

responded to a statement such as, "My Reading teacher, Ms. Jones, gets the most from each class hour." The student would then affirm the frequency by selecting from choices: always (5), almost always (4), usually (3), sometimes (2), or rarely (1). Depending on the student's selection of frequency a follow-up open response question appeared in the electronic form. Thus, if a student selected always, the survey then asked for the student to Provide an example of how Ms. Jones gets the most from each class hour. If students selected sometimes, they were asked to provide a suggestion how they feel Ms. Jones could get more from each class hour in the future. This process of discrete item followed by open-response item was repeated for each of the pedagogical practices. Students were required to respond to the discrete items. They were not required to supply evidence to support their rating of each practice but were given the opportunity to do so.

Following questions about teacher application of the pedagogical practices, the student provided an overall score of teacher performance, "On a scale of 0-10 how strongly would you recommend, the ELC use Ms. Jones to teach a class similar to this reading class again?" With the summary score and pedagogical practices incorporated into the evaluation instrument, we can pursue important questions that may benefit the broader field. Although the pedagogical practices identified by stakeholders were seen as important, it was unclear how the students perceived the relative importance of these practices within their own learning experience. Nor was it clear what additional factors beyond the pedagogical practices might impact teacher recommendation scores.

These additional factors included demographic variables related to the background of the teachers or students. For example, literature suggests that teacher practice might be influenced by the level of teacher preparation (e.g., Hartshorn, Evans, & Tuioti, 2014). Moreover, researchers have observed students who express concerns that their language teacher is not a native speaker of their target language (e.g., Braine, 2013). In addition, it is conceivable that other student factors that may impact language learning could also impact teacher recommendation such as student age, gender, proficiency level, and so on (e.g., DeKeyser, 2013; Dörnyei, 2014). Answering such questions could help address a significant gap in the literature and provide important insights about student perspectives of

pedagogical practices as well as general insights regarding student recommendations of their TESOL practitioners.

Research Question

With this review of literature in place, we now consider the specific research question articulated for this study: How well do student perceptions of teacher performance within the eight pedagogical practices and other demographic and programmatic factors account for teacher recommendation scores?

Methods

This section briefly addresses the data elicitation, the students who completed evaluations, and the analyses used in this study.

Data Elicitation

In order to address the research question, data from 5461 student evaluations were collected electronically by the institution examined in this study. This included sending the evaluations out to student email addresses through Qualtrics software one week prior to the end of the semester. Students in the Novice-High and Intermediate-Low Levels completed the evaluations in the computer lab to ensure additional language support if needed. Those who did not complete the evaluations right away were given reminders. The completion rate was over 98%. Any identifying student information was stripped prior to data being made available for analysis. The institution's Internal Review Board authorized the use of these preexisting data for study.

Students

Data used in this study was produced by students with a mean age of 25 (SD=7.31). Males made up 42% of the students while females made up 58%. Of the evaluations gathered for this study, 54% were from students continuing their study at the institution after at least one semester of previous study, and 46% were from new students who had just completed their first semester in the program. Evaluations came from students with various L1 backgrounds including Spanish (61%), Portuguese (10%), Chinese (9%), Korean (8%), Japanese (4%), Russian (2%), Mongolian (1%), French (1%), Thai (1%), and other languages represented

with less frequency (i.e., Albanian, Arabic, Chuvash, Farsi, Hungarian, Italian, Kazakh, Lithuanian, Loa, Malagasy, Tagalog, Tajik, Ukrainian). The language proficiency levels of the students, based on ACTFL guidelines (ACTFL, 2012), is presented in Table 1.

Table 1 Student Proficiency by Evaluation Percentages

Proficiency level	%
Novice High	1.53%
Intermediate Low	5.40%
Intermediate Mid	14.13%
Intermediate High	35.94%
Advanced Low	23.73%
Advanced Mid	16.72%
Advanced High	2.56%
Total	100.00%

Analyses

Teacher recommendation scores were based on means from the 0-10 scale described previously. Answering the research question involved multiple linear regression where the eight pedagogical practices and other relevant factors functioned as the explanatory variables for the response variable—the teacher recommendation score. Table 2 presents the list of variables that were included. They represent topics found in the literature and ideas of specific interest within the program. For convenience, these variables are organized into five different categories. The first category includes the eight pedagogical practices described previously.

Table 2 Explanatory Variables for Predicting Teacher Recommendation

Category	Explanatory Variables
Practices	1. Rely on course outcomes 2. Plan lessons effectively 3. Optimize class time 4. Cultivate a positive learning environment 5. Evaluate learning effectively 6. Utilize homework strategically 7. Provide meaningful and timely feedback 8. Exemplify professionalism
Teacher	1. Educational level of the teacher 2. Whether the teacher was full-time 3. Whether the teacher was a native speaker of English 4. Whether it was the teacher's first semester at the institution 5. Skill area taught (reading, writing, listening/speaking, grammar)
Student	1. Student age 2. Student sex 3. Student proficiency level 4. Total number of semesters at institution 5. Hours of homework completed per week 6. Whether the student was new or returning 7. Student confidence regarding skill improvement 8. Whether the student felt challenged in the course 9. Student's level of overall satisfaction with the course 10. Subsequent semester plans (i.e., stay, vacation, leave)
Exams	1. Final exam reading score 2. Final exam writing score 3. Final exam listening score 4. Final exam speaking score 5. Final exam vocabulary score 6. Combined final exam score 7. Class Proficiency grades 8. Class Citizenship grades 9. Teacher rating of student progress
Time	Class time (i.e., 8:15, 9:30, 12:15, 1:30)

The second category includes variables associated with the teacher. Level of education was based on a four-point scale according to the following: baccalaureate degree in progress (1), baccalaureate degree completed (2), master's degree in progress (3), master's degree completed (4). No additional distinctions were made for the few teachers who were pursuing or who had completed doctoral degrees.

The third category includes student demographic information such as student age, sex, language proficiency level (based on placement testing), the number of semesters the student had studied at the institution, and whether the student was new to the institution or continuing their study. It also included self-reported information such as the number of hours spent completing homework each week, how confident students were that they improved in the language skill taught in the course, the extent to which students felt challenged in the course, the students' level of satisfaction with the course, and their plans after the completion of the semester (i.e., whether they intended to continue studying at the institution, take a vacation for a semester and then return, or leave the institution altogether).

The final categories represent several measures of student performance and the time of day the respective classes were held. Measures of student performance includes institutional proficiency exams administered at the end of the semester in areas such as reading, writing, listing, speaking, and vocabulary. It also includes class proficiency grades based on student performance over the course of the semester and citizenship grades based on class participation and homework completion. The last measure of student performance was a single, overall rating provided by the teacher of student proficiency. The final category was based on the time of day the respective classes were held. Classes met Monday through Thursday for 65 minutes at 8:15 a.m., 9:30 a.m., 12:15 p.m., and 1:30 p.m.

Because of the exploratory nature of the research question, data were randomly divided into two halves. The first half (2731 evaluations) was used to run an exploratory stepwise regression analysis, and then the second half (2730 evaluations) was used to run a confirmatory regression analysis to test the model identified through the first stepwise analysis (see Mark & Goldberg, 2001). Because stepwise regression can be prone to overfitting and distorted p-values, the default criteria in the SPSS software was adjusted to help minimize distortion and ensure

that any variables associated with teacher recommendation would be truly meaningful (the typical variable entry, p=.05, and removal, p=.10, was replaced with p=.001 for entry and p=.002 for removal based on a Bonferroni adjustment for the number of explanatory variables used in the study, see Wilkinson & Dallal, 1981).

Results and Discussion

Exploratory and Confirmatory Analyses

The research question addressed the extent to which the eight pedagogical practices and other demographic or programmatic factors influenced teacher recommendation scores. Table 3 presents the results of the exploratory and confirmatory regression analyses. Of the 41 variables included, the analysis generated a model of nine variables accounting for teacher recommendation. These are listed on the left side of the Table 3. Statistics associated with the exploratory analysis are included in the middle section of the table and account for more than 55% of the variability associated with teacher recommendation (adjusted R^2 of .558). The exploratory model included all eight of the pedagogical practices discussed previously along with one additional variable associated with the combination of the student final exam scores. In general, the exploratory analysis suggests that the greater the presence of these model factors, the greater the recommendation of a given teacher.

Table 3 Factors Impacting Teacher Recommendation

	Exploratory				Confirmatory			
Model	*B*	*SE*	*ß*	*p*	*B*	*SE*	*ß*	*p*
(Constant)	-1.398	0.223		<.001	-1.713	0.173		<.001
Exemplify professionalism	0.920	0.079	0.231	<.001	0.699	0.064	0.181	<.001
Rely on course outcomes	0.572	0.079	0.149	<.001	0.563	0.066	0.147	<.001
Cultivate positive environment	0.512	0.074	0.133	<.001	0.499	0.061	0.131	<.001
Evaluate learning effectively	0.326	0.074	0.094	<.001	0.536	0.057	0.160	<.001
Optimize class time	0.308	0.068	0.090	<.001	0.421	0.059	0.119	<.001

	Exploratory				Confirmatory			
Model	*B*	*SE*	*ß*	*p*	*B*	*SE*	*ß*	*p*
Plan lessons effectively	0.272	0.078	0.078	<.001	0.460	0.063	0.130	<.001
Utilize homework strategically	0.288	0.059	0.088	<.001	0.317	0.050	0.095	<.001
Provide meaningful feedback	0.264	0.055	0.087	<.001	0.162	0.047	0.052	.001
Combined final exam scores	0.138	0.028	0.068	<.001	0.100	0.023	0.048	<.001
	Adjusted R^2=.558				Adjusted R^2=.662			

Note that the standardized betas (β) in the table show the relative importance of each of the variables to the overall model. For example, exemplifying professionalism is the most important part of the model produced by the exploratory analysis followed by relying on course outcomes and cultivating a positive learning environment. Although all of the factors provide a meaningful contribution to the model, exemplifying professionalism carries about one and a half times the weight of the next most important factor, relying on course outcomes, and more than three times the weight of the combined final exam scores.

The results of the confirmatory analysis are included in the right portion of Table 3. Though based on the same variables identified in the exploratory analysis, the confirmatory analysis accounted for just over 66% of the variability associated with teacher recommendation (adjusted R2 of .662). Though the confirmatory analysis accounted for a little more variation, than the exploratory analysis, these results seem comparable and tend to underscore the importance of the model factors in teacher recommendation.

Additional insights may be gleaned by examining just a few of the responses to the follow-up questions included in the evaluation that focus on the pedagogical practices valued the most highly by the students. For example, when asked to provide evidence of professionalism, students often commented on the teachers' overall attitude toward teaching and concern for student success. One student described his teacher with representative comments such as "very professional"

and "punctual," and, in addition to planning classes well, she had "a very good attitude," and was "interested in the improvement of the students." Many other learners provided evidence of professionalism by referring to other pedagogical practices included in our list. Comments like, "she's well prepared and [uses] time wisely," were frequent throughout the data. Many students also referred to the demeanor of their teachers. They generally connected professionalism with kindness, patience, and dedication among other valued attributes.

Students who provided evidence of the teacher relying on course outcomes generally emphasized one or more of the following—introducing course outcomes at the beginning of the semester, writing or stating course outcomes at the beginning of a class period, or using activities or resources that were clearly connected to the outcomes. One representative response illustrates this with the description, "At the beginning of the course, she explained the outcomes of the class and how we were supposed to reach them." The student continued by indicating that in class they always were engaged in "activities to help us to [reach] those outcomes." Another student response echoed that of many others, "She wrote [on] the board the outcomes for the day." As with professionalism, many comments were connected to other pedagogical practices such as, "She always prepares and really focuses on the course outcomes."

Other explanatory variables had no apparent influence on teacher recommendation. Clarifying that these variables did not impact recommendation may be as important to understanding teacher recommendation as the variables that did affect it. Variables not impacting teacher recommendation include level of education, full-time versus part-time status, whether it was the teacher's first semester at the institution, and the language skill taught by the teacher (i.e., reading, writing, listening/speaking, or grammar). These findings suggest that levels of education and experience were not valued by students as much as effective practice. These findings also indicate that non-native speakers were recommended just as much as native speakers, a conclusion consistent with other scholars who have recognized potential strengths in native and nonnative teachers alike, but no disadvantage for non-native teachers (e.g., Braine, 2013; Shin & Manochphinyo, 2017; Walkinshaw & Oanh, 2014).

There were many additional factors which similarly had no observable impact on teacher recommendation. Some of these were associated with the students themselves such as student age, sex, language proficiency, time spent on homework, how long the student had studied at the institution, how challenged the student felt by the course, or the students' plans for the subsequent semester. Additional areas with no impact on teacher recommendation suggest student ability to separate the evaluation of their teachers from other conspicuous elements of their classroom experience. For example, teacher recommendations were unaffected by the extent to which students felt they improved in the specific skill area taught or their overall level of satisfaction with the course. Other factors with no apparent impact on teacher recommendation included when classes were held and eight of the nine measures of student assessment. These metrics included various classroom-based evaluations as well as student performance on institutional exams for each language skill.

Although performance on individual exams assessing the discrete language skills of reading, writing, listening, speaking, and vocabulary were not associated with teacher recommendation, the combined scores for these exams were meaningful enough to be included in the model. It should be noted, however, that the contribution of the combined exam scores provided the weakest impact to the model in both the exploratory and confirmatory analyses. Since exam performance in this study was tied to specific proficiency levels, this factor of combined exam scores is essentially an overall measure of proficiency.

An important question is why there might be a relationship between higher proficiency levels and teacher recommendations, especially when anecdotal evidence suggests that the highest proficiency learners may be more demanding and have higher expectations than their lower-proficiency counterparts. Although the answer is not obvious, here are two possibilities. It could be that higher proficiency students are able to have more meaningful interactions with their teachers in ways that foster greater language development that is recognized by the students. It could also be that if higher-proficiency learners maintain higher expectations of their teachers, the institution may place some of their best teachers in those positions where they garner positive evaluation data while teachers with less de-

veloped skills may opt out of those more challenging assignments. Additional research may be needed to better understand these possibilities.

Implications

Of the numerous variables examined in this study, the eight pedagogical practices established by the observed institution accounted for more than half of the variability associated with teacher recommendations. The relative impact of these practices within the regression model is summarized in Table 4, which averages the contributions of each factor from the exploratory and confirmatory analyses into an overall set of percentages. Although some of these variables are more important to the model than others, collectively they provide meaningful insight into those teacher practices highly regarded by the ESL learners in this study. These findings should be valuable to program administrators, teachers who may be considering ways to better meet student expectations, and researchers interested in factors affecting classroom learning.

Table 4 Contribution Percentages for Factors Explaining Teacher Recommendation

Teacher recommendation factor	Percent
Exemplify professionalism	19.80%
Rely on course outcomes	14.22%
Cultivate positive environment	12.69%
Evaluate learning effectively	12.21%
Optimize class time	10.04%
Plan lessons effectively	10.00%
Utilize homework strategically	8.79%
Provide meaningful feedback	6.68%
Combined final exam scores	5.57%
Total	100.00%

Using Evaluation Data

This study presents specific findings that should provide practitioners and administrators with useful insights as they seek to identify teaching strengths

and areas for improvement. Focusing on these practices may also be beneficial in teacher preparation, teacher in-service training, or other professional development activities. Although application of these specific findings may improve the efficacy of language teaching in certain contexts, other well-designed teacher evaluations are also likely to benefit teachers who seek to improve their practice. Nevertheless, the ultimate efficacy of such instruments may depend on how the results are applied. Thus, we briefly address considerations necessary for effectively utilizing results from this study as well as other kinds of teacher evaluations used in other contexts.

Practitioners and administrators should keep in mind that the appeal for students to participate in evaluation of their teachers is largely to provide feedback to improve teaching (Chen & Hoshower, 2003). Yet, the research shows that evaluation by itself does not produce more effective teaching (e.g., Ballantyne, Borthwick, & Packer 2000; Kember, Leung, & Kwan 2002; Spooren, Brockx, & Mortelmans, 2013). There may be a number of reasons why this is true. Some practitioners do not seek to improve their teaching based on evaluation results because they may not know how to do so (e.g., Arthur 2009; Bamber & Anderson 2012; Dresel & Rindermann 2011). Some teachers may simply disregard evaluation results because of concerns regarding their validity or utility (e.g., Simpson & Siguaw, 2000). Still others may see evaluations strictly as summative. They may think that conspicuous use of evaluations for improvement may highlight their weaknesses (Baber & Anderson, 2012). Finally, other teachers simply may not have the motivation to improve the quality of their teaching (Edström 2008; Hendry, Lyon, & Henderson-Smart, 2007).

If improved teaching is the ultimate purpose of the evaluation, practitioners and administrators need to understand the conditions that are necessary for evaluation data to lead toward improved teaching. We address just three of these conditions here. First, evaluation instruments must be designed and used in ways that allow them to effectively capture and present meaningful information to the stakeholders. If evaluation instruments are perceived as ineffective by stakeholders, they should be reevaluated. Where appropriate, faulty instruments should be improved or replaced. Nevertheless, if institutions have exercised due diligence in successfully creating instruments that give the students a valued voice and

effectively provide teachers with relevant, constructive, and informative data, students will be more inclined to participate and teachers are likely to glean many useful insights.

Second, practitioners must have a desire to continue to develop the quality of their teaching. Nasser and Fresko (2002) observe that utilizing evaluation results to improve practice requires a certain disposition and willingness of the teacher. Golding and Adam (2016) refer to this as an "improvement attitude" (p. 5). This is in contrast to a perception where practitioners consider their teaching to be good enough. This characteristic of successful teachers who want to improve is illustrated by practitioners who are student-centered and who look to evaluation data for insights regarding adjustments they could make to better help students meet learning outcomes (McGowan & Graham, 2009; Hendry, Lyon, & Henderson-Smart, 2007). Golding and Adam (2016) noted that the highly effective teachers they observed utilized "a reflective approach to their teaching" (p. 5) where they consistently asked themselves how they could improve, and they conscientiously looked to their evaluation data as formative feedback to help them adjust their practice in ways that benefited their students.

The final notion we address here is that practitioners are more likely to improve their teaching when they meet regularly with program administrators to review their evaluation results and to develop and act on a plan to implement appropriate adjustments. Spooren, Brockx, and Mortelmans (2013) suggest that rather than being left to analyze, interpret, and apply evaluation data on their own, "teachers should be able to rely on expert consultation" (p. 628) regarding the content of their evaluations, which might include discussing results with "colleagues or educational experts" (p. 623). Lang and Kersting (2007) observed that providing teachers with evaluation data without consultation was not effective over the course of multiple semesters. On the other hand, Dresel and Rindermann (2011) noted that consultations with teachers regarding their evaluation results had a positive effect on the quality of teaching. Similarly, Penny and Coe (2004) observed that teaching efficacy improved as practitioners were given opportunities to reflect and discuss their teaching. Thus, we encourage administrators to ensure that their teacher evaluation instruments are designed well and used appropriately, that the teaching and learning environment encourages a focus on

improvement, and that teachers are given ample opportunities to process and apply insights from evaluations with input and support from administrators.

Limitations and Future Research

Though this was a fairly large study which included thousands of evaluations, it took place at only one institution where more than half of the student participants were native speakers of Spanish. Thus, additional studies at other institutions and in other contexts should be pursued in an effort to test the pervasiveness of these findings. Moreover, we acknowledge the potential effects of simplifying the language of the data collection instrument and of representing each of the practices with a single survey item. Using multiple items for each of the practices would likely increase the validity and reliability of these constructs. Further study could also examine the possible effects of L1 or national background on teacher recommendations. In addition, future research could incorporate other factors that may impact teacher recommendation that were not part of this study. Finally, it could also seek to provide additional insight regarding the possible relationship between student proficiency level and teacher recommendation observed in this study.

Conclusion

This study highlights the growing importance of teacher evaluation in TESOL contexts. It suggests that evaluation instrument development should be anchored to relevant literature and practices that best capture an institution's values, beliefs, and aspirations for teacher performance, drawing on input from administrators, teachers, staff, and other stakeholders. It suggests that teacher evaluations should elicit both qualitative and quantitative data that is informative to both administrators and the teachers themselves.

This study also examined research questions about desirable pedagogical practices and their relationship to teacher recommendations. The results suggest that higher teacher recommendations may have less to do with specific elements of the program or the demographic background of students or teachers, and more to do with the extent to which teachers successfully implement specific pedagogical practices within the classroom. These include exemplifying professionalism,

relying on course outcomes, cultivating a positive learning environment, evaluating learning effectively, optimizing class time, planning lessons effectively, utilizing homework strategically, and providing meaningful feedback. These insights should be leveraged by administrators and practitioners to better understand and meet the expectations and learning needs of their students.

References

Abedi, J. A. M. A. L. (2010). Research and recommendations for formative assessment with English language learners. Handbook of formative assessment, 181-197.

ACTFL. (2012). ACTFL Proficiency Guidelines 2012. White Plains, NY: American Council on the Teaching of Foreign Languages.

Aleamoni, L. M. (1999). Student rating myths versus research facts from 1924 to 1998. *Journal of Personnel Evaluation in Education, 13*, 153-166. doi.org/10.1023/A:1008168421283

Alsalahi, S. M. (2015). Stages of teacher's professionalism: how are English language teachers engaged? *Theory and Practice in language studies, 5*(4), 671. doi.org/10.17507/tpls.0504.01

American Statistical Association (2014). ASA Statement on Using Value-Added Models for Educational Assessment. April 8. www.amstat.org/policy/pdfs/ASA_VAM_Statement.pdf

Arthur, L. (2009). From performativity to professionalism: lecturers' responses to student feedback. *Teaching in Higher Education, 14*, 441-454. doi.org/10.1080/13562510903050228

Baecher, L., Farnsworth, T., & Ediger, A. (2014). The challenges of planning language objectives in content-based ESL instruction. *Language Teaching Research, 18*(1), 118-136. doi.org/10.1177/1362168813505381

Bailey, A. L., & Heritage, M. (2014). The role of language learning progressions in improved instruction and assessment of English language learners. *TESOL Quarterly, 48*(3), 480-506. doi.org/10.1002/tesq.176

Ballantyne, R., Borthwick, J., & Packer, J. (2000). Beyond student evaluation of teaching. *Assessment & Evaluation in Higher Education, 25*, 221-236. doi.org/10.1080/713611430

Bamber, V., & Anderson, S. (2012). Evaluating learning and teaching: institutional needs and individual practices. *International Journal for Academic Development, 17*, 5-18. doi.org/10.1080/1360144X.2011.586459

Basturkmen, H. (2010). *Designing Courses in English for Specific Purposes*. New York: Palgrave Macmillan. doi.org/10.1057/9780230290518

Benavot, A., & Tanner, E. (2007). The growth of national learning assessments in the world, 1995-2006. Background paper for the Education for All Global Monitoring Report 2008: Education for All by 2015: Will we make it? UNESCO.

Bendle, N. T., & Bagga, C. K. (2016). The metrics that marketers muddle. *MIT Sloan Management Review, 57*(3), 73.

Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International journal of teaching and learning in higher education, 17*(1), 48-62.

Boraie, D. (2014). Nine Features of an Effective Teacher Evaluation System [Blog post]. Retrieved from http://blog.tesol.org/nine-features-of-an-effective-teacher-evaluation-system/

Braine, G. (Ed.). (2013). *Non-native educators in English language teaching*. Routledge. doi.org/10.4324/9781315045368

Brown, H. D. (2006). *Principles of Language Learning and Teaching* (5th ed.). New York: Pearson ESL.

Calderón, M., Slavin, R., & Sánchez, M. (2011). Effective instruction for English learners. *The Future of Children, 21*(1), 103-127. doi.org/10.1353/foc.2011.0007

Cashin, W.E., & Downey, R.G. (1992). Using global student rating items for summative evaluation. *Journal of Educational Psychology, 84*, 563-572. doi.org/10.1037/0022-0663.84.4.563

Charters, W. W. & Waples, D. (1929). The commonwealth Teacher-Training Study. Chicago: University of Chicago.

Chen, Y., & Hoshower, L. (2003). Student evaluation of teaching effectiveness: An assessment of student perception and motivation. *Assessment and Evaluation in Higher Education, 28*, 71-88. doi.org/10.1080/02602930301683

Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review, 24*(2), 205-249. doi.org/10.1007/s10648-011-9191-6

DeKeyser, R. M. (2013). Age effects in second language learning: Stepping stones toward better understanding. *Language Learning, 63*(s1), 52-67. doi.org/10.1111/j.1467-9922.2012.00737.x

Dörnyei, Z. (2014). *The psychology of the language learner: Individual differences in second language acquisition*. Routledge. doi.org/10.4324/9781410613349

Dresel, M., & Rindermann, H. (2011). Counseling university instructors based on student evaluations of their teaching effectiveness: A multilevel test of its effectiveness under consideration of bias and unfairness variables. *Research in Higher Education, 52*, 717-737. doi.org/10.1007/s11162-011-9214-7

Edström, K. (2008). Doing course evaluation as if learning matters most. *Higher education research & development, 27*(2), 95-106. doi.org/10.1080/07294360701805234

Fantini, A. E. (2018). Teacher assessment and evaluation. In The TESOL Encyclopedia of English Language Teaching (pp. 1-11). New York: John Wiley & Sons. doi.org/10.1002/9781118784235.eelt0652

Farrell, T. S. (2015). It's not who you are! It's how you teach! Critical competencies associated with effective teaching. *RELC Journal, 46*(1), 79-88. doi.org/10.1177/0033688214568096

Fordham, S. K. (2015). *Teacher and peer written feedback in the ESL composition classroom: Appropriation, stance, and authorship*. The University of Arizona.

Frey, B. B., Schmitt, V. L., & Allen, J. P. (2012). Defining authentic classroom assessment. *Practical assessment, research & evaluation, 17*(2).

Gershenson, S., & Holt, S. B. (2015). Gender gaps in high school students' homework time. *Educational Researcher, 44*(8), 432-441. doi.org/10.3102/0013189X15616123

Golding, C., & Adam, L. (2016). Evaluate to improve: useful approaches to student evaluation. *Assessment & Evaluation in Higher Education, 41*, 1-14. doi.org/10.1080/02602938.2014.976810

Hamilton, D. F., Lane, J. V., Gaston, P., Patton, J. T., MacDonald, D. J., Simpson, A. H. R. W., & Howie, C. R. (2014). Assessing treatment outcomes using a single question. *Bone Joint J, 96*(5), 622-628. doi.org/10.1302/0301-620X.96B5.32434

Hartshorn, K. J. & Evans, N. W. (2015). The Effects of Dynamic Written Corrective Feedback: A 30-Week Study. *The Journal of Response to Writing, 2*, 6-34.

Hartshorn, K. J., Evans, N. W., & Tuioti, E. A. (2014). Influences on teachers' corrective feedback choices in second language writing. *Writing and Pedagogy, 6*, 251-282.

Hendry, G. D., Lyon, P. M., & Henderson–Smart, C. (2007). Teachers' approaches to teaching and responses to student evaluation in a problem–based medical program. *Assessment & Evaluation in Higher Education, 32*, 143-157. doi.org/10.1080/02602930600801894

Hill, H., & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review, 83*(2), 371-384. doi.org/10.17763/haer.83.2.d11511403715u376

Howard, A., & Donaghue, H. (2105). *Teacher evaluation in second language education*. New York: Bloomsbury.

Hunt, K., Gurvitch, R., & Lund, J. L. (2016). Teacher Evaluation: Done to You or with You? *Journal of Physical Education, Recreation & Dance, 87*, 21-27. doi.org/10.1080/07303084.2016.1226215

Kember, D., Leung, D., & Kwan, K. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching? *Assessment and Evaluation in Higher Education, 27*, 411-425. doi.org/10.1080/0260293022000009294

Ketabi, S., & Ketabi, S. (2014). Classroom and formative assessment in second/foreign language teaching and learning. *Theory and Practice in Language Studies, 4*, 435. doi.org/10.4304/tpls.4.2.435-440

Krol, M. W., Boer, D., Delnoij, D. M., & Rademakers, J. J. (2015). The Net Promoter Score-an asset to patient experience surveys? *Health Expectations, 18*(6), 3099-3109. doi.org/10.1111/hex.12297

Lang, J. W. B., & Kersting, M. (2007). Regular feedback from student ratings of instruction: Do college teachers improve their ratings in the long run? *Instructional Science, 35*, 187-205. doi.org/10.1007/s11251-006-9006-1

Leisen Pollack, B., & Alexandrov, A. (2013). Nomological validity of the Net Promoter Index question. *Journal of Services Marketing, 27*(2), 118-129. doi.org/10.1108/08876041311309243

Leung, C. (2012). Outcomes-based language teaching. In A. Burns & J. C. Richards (Eds.), *The Cambridge Guide to Pedagogy and Practice in Language Teaching* (161-179). New York: Cambridge University.

Liyanage, I., & Bartlett, B. J. (2010). From autopsy to biopsy: A metacognitive view of lesson planning and teacher trainees in ELT. *Teaching and Teacher Education, 26*, 1362-1371. doi.org/10.1016/j.tate.2010.03.006

Maltese, A. V., Tai, R. H., & Fan, X. (2012). When is homework worth the time? Evaluating the association between homework and achievement in high school science and math. *The High School Journal, 96*, 52-72. doi.org/10.1353/hsj.2012.0015

Mandal, P. C. (2014). Net promoter score: a conceptual analysis. *International Journal of Management Concepts and Philosophy, 8*(4), 209-219. doi.org/10.1504/IJMCP.2014.066899

Mark, J. & Goldberg, M. A. (2001). Multiple regression analysis and mass assessment: A review of the issues. *The Appraisal Journal*, 89-109.

Martin, R. L. (2011). The innovation catalysts. *Harvard Business Review, 89*(6), 82-87.

Mathis, W. (2012). Research-Based Options for Education Policy Making: Teacher Evaluation. National Education Policy Center. http://nepc.colorado.edu/files/pb-options-1-teval_0.pdf

McGowan, W. R., & Graham, C. R. (2009). Factors contributing to improved teaching performance. *Innovative Higher Education, 34*, 161-171. doi.org/10.1007/s10755-009-9103-6

McMillan, J. H. (2013). *Classroom Assessment: Pearson New International Edition: Principles and Practice for Effective Standards-Based Instruction*. Pearson Higher Ed.

McReynolds, J. C. (2010). Do Try this at Home: Motivating Second-grade ESL Students to Understand the Importance of Homework Completion to Improve Their Learning (Doctoral dissertation, Gratz College).

Milkova, S. (2012). Strategies for effective lesson planning. Center for Research on learning and Teaching, 1-4.

Murray, D. E., & Christison, M. (2010). *What English language teachers need to know II: facilitating learning*. Routledge. doi.org/10.4324/9780203846292

Nasser, F., & Fresko, B. (2001). Interpreting student ratings: Consultation, instructional modification, and attitudes towards course evaluation. *Studies in Educational Evaluation, 27*, 291-305.doi.org/10.1016/S0191-491X(01)00031-1

National Commission on Excellence in Education, (1983). A nation at risk: The imperative for educational reform. Washington, DC: US Government Printing Office

No Child Left Behind Act, 20 § U.S. C. 6301 (2002).

Orlich, D. C., Harder, R. J., Trevisan, M. S., Brown, A. H., & Miller, D. E. (2016). *Teaching strategies: A guide to effective instruction*. Cengage Learning.

Oxford, R.L. (1999). Anxiety and the language learner: new insights. In J. Arnold (Ed.), *Affect in Language Learning* (pp. 58-67). Cambridge: Cambridge University Press.

Pang, M. (2016). Pedagogical reasoning in EFL/ESL teaching: revisiting the importance of teaching lesson planning in second language teacher education. *TESOL Quarterly, 50*, 246-263. doi.org/10.1002/tesq.283

Paulsen, M. B. (2002). Evaluating teaching performance. New Directions for Institutional Research, 2002(114), 5-18. doi.org/10.1002/ir.42

Pennington, M. C., & Young, A. L. (1989). Approaches to faculty evaluation for ESL. *TESOL Quarterly, 23*(4), 619-646. doi.org/10.2307/3587535

Penny, A. R., & Coe, R. (2004). Effectiveness of consultation on student ratings feedback: A meta-analysis. *Review of Educational Research, 74*, 215-253. doi.org/10.3102/00346543074002215

Pratt, D. D. (2002). Good teaching: One size fits all? New directions for adult and continuing education, 2002(93), 5-16. doi.org/10.1002/ace.45

Reichheld, F. (2006). *The Ultimate Question: driving good profits and true growth.* Harvard Business Press: Boston, MA.

Reichheld, F. F. (2003). The one number you need to grow. *Harvard business review, 81*(12), 46-55.

Reichheld, F. F., & Markey, R. (2011). *The Ultimate Question 2.0: How net promoter companies thrive in a customer-driven world*. Harvard Business Press.

Richards, J. C. (2013). Curriculum approaches in language teaching: Forward, central, and backward design. *Relc Journal, 44*(1), 5-33. doi.org/10.1177/0033688212473293

Rossiter, M. J., Derwing, T. M., Manimtim, L. G., & Thomson, R. I. (2010). Oral fluency: The neglected component in the communicative language classroom. *Canadian Modern Language Review, 66*(4), 583-606. doi.org/10.3138/cmlr.66.4.583

Rucinski, M., & Diersling, C. (2014). America's teacher evaluation system revolution. Harvard Political Review. http://harvardpolitics.com/united-states/americas-teacher-evaluation-system-revolution/

Sawchuk, S. (2015). Teacher evaluation: An issue overview. Education Week, 35(3).

Sawyer, M. R., Andzik, N. R., Kranak, M. P., Willke, C. P., Curiel, E. S., Hensley, L. E., & Neef, N. A. (2017). Improving Pre-Service Teachers' Performance Skills Through Behavioral Skills Training. *Behavior Analysis in Practice*, 1-5. doi.org/10.1007/s40617-017-0198-4

Schneider, D., Berent, M., Thomas, R., & Krosnick, J. (2008, June). Measuring customer satisfaction and loyalty: Improving the 'Net-Promoter' score. Presentation at the Annual Meeting of the American Association for Public Opinion Research, New Orleans, Louisiana.

Shin, T. W., & Manochphinyo, A. (2017). The Perceptions of Malaysian Students toward Native and Non-native ESL Teachers. *The English Teacher, 46*(1), 17.

Simpson, P., & Siguaw, J. (2000). Student evaluations of teaching: An exploratory study of the faculty response. Journal of Marketing Education, 22, 199-213. doi.org/10.1177/0273475300223004

Spencer, K. J., & Schmelkin, L. P. (2002). Student perspectives on teaching and its evaluation. Assessment & Evaluation in Higher Education, 27(5), 397-409. doi.org/10.1080/0260293022000009285

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research, 83*, 598-642. doi.org/10.3102/0034654313496870

Su, T., & Tian, J. (2016). Research on corrective feedback in ESL/EFL classrooms. *Theory and Practice in Language Studies, 6*(2), 439. doi.org/10.17507/tpls.0602.29

Tan, F., Nabb, L., Aagard, S., & Kim, K. (2010). International ESL graduate student perceptions of online learning in the context of second language acquisition and culturally responsive facilitation. *Adult Learning, 21*(1-2), 9-14. doi.org/10.1177/104515951002100102

Thomsen, J. (2014). A Closer Look: Teacher Evaluations and Reduction-in-Force Policies. Education Commission of the States.

Tsiplakides, I. and Keramida, A. (2010). Promoting positive attitudes in ESL/EFL classes. *The Internet TESL Journal, XVI*(1).

Tsui, A.B.M. (1996). Reticence and anxiety in second language learning. In K.M. Bailey and D. Nunan (Eds.), Voices from the Language Classroom: Qualitative Research in Second Language Education (pp. 145-167). New York: Cambridge University Press.

Vu, M. T. (2016). The kaleidoscope of English language teacher professionalism: A review analysis of traits, values, and political dimensions. *Critical Inquiry in Language Studies, 13*(2), 132-156. doi.org/10.1080/15427587.2016.1146890

Walberg, H.J., Paschal, R.A., & Weinstein, T. 1985, 'Homework's Powerful Effects on Learning', *Educational Leadership, 42*(7), pp. 76-79.

Walkinshaw, I., & Oanh, D. H. (2014). Native and non-native English language teachers: Student perceptions in Vietnam and Japan. *SAGE Open, 4*(2), 2158244014534451. doi.org/10.1177/2158244014534451

Wallinger, L.M. 2008, 'The Role of Homework in Foreign Language Learning', *Foreign Language Annals, 33*(5), pp. 483-496. Abstract doi.org/10.1111/j.1944-9720.2000.tb01993.x

Walsh, S. (2006). Talking the talk of the TESOL classroom. *ELT Journal, 60*(2), 133-141. doi.org/10.1093/elt/cci100

Wilkinson, L., & Dallal, G.E. (1981). Tests of significance in forward selection regression with an F-to enter stopping rule. *Technometrics, 23*, 377-380. doi.org/10.1080/00401706.1981.10487682

Young, D.J. (1991). Creating a low-anxiety classroom environment: What does language anxiety research suggest? *The Modern Language Journal, 75*(iv), 426-439. doi.org/10.2307/329492

About the Authors

K. James Hartshorn has been involved in second language education in the United States and Asia for three decades. James currently serves as a faculty member in the Linguistics Department at Brigham Young University where he mentors new TESOL professionals by helping them optimize the efficacy of their ESL teaching and research.

Judson M. Hart holds an MA in TESOL and currently serves as Director of Language Assessment at Emmersion.

Benjamin L. McMurry has nearly 20 years of experience working with non-native English speakers. He currently serves as the Program Coordinator at the English Language Center at Brigham Young University. His research interests include materials development and evaluation.

Appendix A

Principled Pedagogical Practices

1. Rely on course outcomes

Teachers understand the course outcomes for the skill and proficiency level in which they teach and effectively communicate them to students. They can describe student behaviors that demonstrate these outcomes, and they successfully design classroom-learning activities that help students progress toward achieving them. Teachers engage in ongoing informal and formal assessment activities and provide personalized feedback based on the course outcomes.

2. Plan lessons effectively

Teachers carefully plan lessons so language development will be optimized during the class period. Teachers plan to incorporate an appropriate number and variety of learning activities that are meaningful and engaging. These activities build incrementally from more simple uses of language to more complex uses that are authentic and communicative. Teachers consider the best ways to ensure that communication of explanations and expectations are clear and concise in order to maximize student language practice. This includes preparing the board or other materials well ahead of class time. Teachers also prepare contingency plans in order to adjust for a variety of unforeseen circumstances and changing student needs.

3. Optimize class time

Teachers feel a sense of urgency about using as much of the classroom time as possible for meaningful language practice. They convey this sense of urgency to their students by starting class on time and by carefully managing activities and transitions in order to maximize communicative language practice. However, rather than rushing through their lessons, teachers skillfully connect activities and ensure that students achieve the needed level of mastery before moving on. They anticipate potential threats to effective use of class time such as problems with technology, excessive student questions, inappropriate student behaviors and so

on. Their responses to such challenges are principled and appropriately bring the class back on course. Teachers also end class on time.

4. Cultivate a positive learning environment

Teachers understand the necessity of a positive learning environment in order to optimize learning. They recognize that positive teacher-student interaction is at the heart of the environment they seek to cultivate. They foster genuine concern for their students and their learning based on principles of respect and trust. They leave personal concerns behind as they plan and teach their classes. They are consistent and equitable in their classroom practices and help students to see how classroom policies and activities facilitate language development. They create a non-threatening learning environment that is cheerful, upbeat, and optimistic. They inspire students to do their best, and they help them experience the joy of effectively applying what they learn. They sincerely praise students and regularly express confidence in their abilities.

5. Evaluate learning effectively

Teachers are committed to the ongoing evaluation of student learning. They skillfully use diagnostic tests, classroom instruction, language practice, and formal and informal assessments to clarify individual learner needs in relation to established course outcomes. They also regularly solicit qualitative input from their students regarding learning materials and methods. This information is then used to make appropriate adjustments in lesson planning and the selection of materials and methods used in the classroom. Teachers help students to understand the rationale for adjustments that are made as well as areas where continuity may be necessary.

6. Utilize homework strategically

Teachers understand the potential for effective homework to help students achieve course outcomes. Rather than assigning busy work, they carefully consider the quantity and specific kinds of learning activities that are needed by their students in order to foster language development or to help them better understand and diagnose learner needs. They are able to effectively communicate the rationale

for various types of homework to their students. They demonstrate the value of the homework in the way they follow up and process the homework. They know when it may be appropriate to review certain types of homework in class and when the class time should be used for other activities. They utilize student performance on homework to inform their ongoing instruction in the classroom.

7. Provide meaningful and timely feedback

Teachers know that feedback is essential to effective learning. They regularly provide students with feedback that is meaningful—it focuses on the most important language elements for each learner; students understand the feedback, why it was given, and how to apply it. Though teachers ensure that ongoing feedback is timely, they are careful not to overload the students' cognitive ability to process and apply the feedback. Along with feedback, teachers provide students with abundant opportunities to practice and apply the feedback in a variety of learning contexts.

8. Exemplify professionalism

Teachers value and participate in orientations, training, and workshops. They are well prepared, punctual, and complete all administrative tasks on time. They act and look the part of a professional in the classroom including adhering to the dress and grooming standards and maintaining appropriate teacher-student boundaries. They are respectful and courteous with their students and other teachers with whom they share resources such as classrooms, offices, technologies, and learning materials. They consistently evaluate their own teaching and seek to improve through feedback from students, administrators, and peers. They appropriately apply the relevant feedback they receive.

Appendix B

Survey Components Presented to Students

My [Course] teacher, [Teacher]

	Rarely	Sometines	Usually	Almost Always	Always
teaches toward course outcomes.	○	○	○	○	○
plans lessons effectively.	○	○	○	○	○
gets the most from each class hour.	○	○	○	○	○
creates and keeps a positive learning environment.	○	○	○	○	○
evaluates learning effectively.	○	○	○	○	○
gives useful homework.	○	○	○	○	○
provides helpful feedback quickly and frequently.	○	○	○	○	○
is a good example of a professional teacher.	○	○	○	○	○