

Research Report
ETS RR-16-24

Exploring the Relationships Among Teacher Confidence, Learning, and Test Performance Within the English-for-Teaching Course

Lin Gu

Spiros Papageorgiou

June 2016

ETS Research Report Series

EIGNOR EXECUTIVE EDITOR

James Carlson
Principal Psychometrician

ASSOCIATE EDITORS

Beata Beigman Klebanov
Senior Research Scientist

Heather Buzick
Research Scientist

Brent Bridgeman
Distinguished Presidential Appointee

Keelan Evanini
Research Director

Marna Golub-Smith
Principal Psychometrician

Shelby Haberman
Distinguished Presidential Appointee

Anastassia Loukina
Research Scientist

Donald Powers
Managing Principal Research Scientist

Gautam Puhon
Principal Psychometrician

John Sabatini
Managing Principal Research Scientist

Matthias von Davier
Senior Research Director

Rebecca Zwick
Distinguished Presidential Appointee

PRODUCTION EDITORS

Kim Fryer
Manager, Editing Services

Ayleen Stellhorn
Senior Editor

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

RESEARCH REPORT

Exploring the Relationships Among Teacher Confidence, Learning, and Test Performance Within the English-for-Teaching Course

Lin Gu & Spiros Papageorgiou

Educational Testing Service, Princeton, NJ

In responding to the increasingly expanded emphasis on English language teaching (ELT) around the globe, Educational Testing Service (ETS) collaborated with National Geographic Learning to develop the *ELTeach*[™] program (<http://www.elteach.com>), an online professional development program consisting of two courses, English-for-Teaching and Professional Knowledge for ELT, with the goal of improving ELT teachers' classroom English language proficiency. Within the English-for-Teaching course, there are three interrelated components: a precourse planner that asks teachers to rate their confidence in using English to teach English, a learning course that teaches the type of English needed to carry out essential classroom activities in English, and the *TEFT*[™] assessment, which is developed based on the same framework as the learning course. Using data collected during the pilot administration of the *ELTeach* program in 2012, this study examines the relationship between self-reported confidence level and engagement with the learning course, and the relationship between self-reported confidence level and performance on the *TEFT* assessment. The findings suggested how users with different confidence levels might make use of the learning materials and how they might score on the *TEFT* assessment.

Keywords *ELTeach*[™]; the *TEFT*[™] assessment; English for teaching; classroom English; teacher confidence

doi:10.1002/ets2.12110

With the emergence of English as a global language, there is a growing need for the use of English as a vehicle of communication in English as a foreign language (EFL) contexts. English has become a key, and in most cases, compulsory subject in the educational systems in many countries around the world. The propagation of English language education is highlighted by a notable trend in recent years, which is the introduction of English as a school subject earlier in national school curricula (Butler, 2004; Nunan, 2003). This, in turn, drives the rising demand for teachers who are qualified, capable, and confident to teach English.

In a typical EFL context, English language teaching (ELT) is usually delivered by teachers whose first language is not English. In fact, most of the world's English teachers do not speak English as their native language (Canagarajah, 1999). Although the use of the local language in EFL classrooms is common (Sešek, 2007), EFL teachers are often held to the expectation of being able to teach English in English. For example, Butler (2004) reported that a policy of enforcing a weekly 1 hour English-only class for Grades 3 to 6 was implemented in Korea starting from 2001 to encourage the use of English in English classes. Nunan (2003) also noted that the immersion-type curriculum implemented in schools where English is the medium of instruction in Hong Kong requires local teachers to teach school subjects, including English, in English.

As the demand to use English as a vehicle for teaching the language grows, the English language proficiency of EFL teachers has also received increasing attention in the ELT literature. Cullen (1994) asserted that language proficiency, either in a general sense or conceptualized as a kind of English for the specific purpose of teaching, refers to the ability to use the target language in real communication and should not be confused with linguistic knowledge, which emphasizes awareness of the system of the language. The language proficiency factor was included as one of the 10 core constructs of skill and expertise in language teaching by Richards (2010). Kamhi-Stein (2009) also identified teacher language proficiency as one of the two dominant themes in work on teacher preparation for nonnative English-speaking educators.

Empirical research has also been conducted to explore teachers' views of their English proficiency in relation to their professional development. An early questionnaire study conducted by Berry (1990) in Poland showed that language

Corresponding author: L. Gu, E-mail: lgu001@ets.org

improvement was rated above methodology and theory as the most needed component of training programs for English language teachers. A survey of Sri Lankan teachers by Murdoch (1994) showed that when asked to rank the relative importance of the various curriculum components, the participating teachers gave the highest rating to language improvement, and that “a teacher’s confidence was most dependent on his or her own degree of language competence” (p. 258). Seidlhofer (1999) investigated the status and role of nonnative English-speaking teachers perceived by Austrian teachers. The results of this questionnaire study indicated that, for the majority of the teacher participants, being a nonnative teacher of English is a source of insecurity. More recently, in a survey of Venezuelan teachers, Chacón (2005) found that teachers’ perceptions of their ability to teach were positively correlated with self-reported levels of English proficiency. In a multicountry survey study across Korea, Taiwan, and Japan, Butler (2004) found a discrepancy between teachers’ perception of their English proficiency and the level of proficiency that they felt they needed in order to teach at the elementary school level. Teachers rated their self-assessed proficiency consistently lower than the desired proficiency they felt was needed across all individual language skills as well as when overall English language proficiency was considered. The author surmised that this low perception of relative proficiency could undermine teachers’ confidence.

To summarize, these studies demonstrate that EFL teachers perceive a high level of English language proficiency to be an important asset. These studies have also identified language improvement as a high priority for teacher training in an EFL context. Despite the consensus that language proficiency is central to EFL teachers’ professional development, insufficient attention has been paid to teachers’ English proficiency in teacher training programs (Berry, 1990; Cullen, 1994; Richards, 2010). Freeman, Katz, Gomez, and Burns (2015) made a further distinction between general English language proficiency and the kind of English language teachers need to carry out their teaching tasks in English. They defined the English language teachers need as “English-for-teaching” and contended that it is the “particular language related to the specific classroom work” (p. 6) that merits special attention in teacher education.

In acknowledgement of the need to provide high-quality learning programs to support ELT teachers’ professional needs, ETS collaborated with National Geographic Learning to develop the *ELTeach*[™] program (<http://www.elteach.com>), an online professional development program, consisting of two courses, English-for-Teaching and Professional Knowledge for ELT. Each course includes a coordinated assessment leading to a score report and certificate for individual teachers. The coursework and assessment are offered as an integrated program; neither component is independent.

The material taught and assessed in the English-for-Teaching course is the essential English skills a teacher needs in order to be able to prepare and enact the lesson in a standardized (usually national) curriculum in English in a way that is recognizable and understandable to other speakers of the language (Young, Freeman, Hauck, Garcia Gomez, & Papageorgiou, 2014). Three functional areas are included in the construct, serving as crucial organizing principles for the learning material, the assessment, and the score reporting: managing the classroom, understanding and communicating lesson content, and providing feedback.

Within the English-for-Teaching course, there are three interrelated components: a precourse planner, a learning course, and the *TEFT*[™] assessment. At the onset of the English-for-Teaching course, information on teachers’ confidence in using English for teaching English is collected via the precourse planner. After having completed the precourse planner, the users are then presented with the opportunity to receive training by taking the online self-access learning course at their own pace. The confidence ratings obtained via the precourse planner are used to generate recommended learning areas for each individual test takers. Therefore, the lower the confidence rating is, the more learning material will be recommended, and vice versa. This allows teachers to focus on areas of need and to determine individually what they practice and how they progress. Following the completion of training, the users are scheduled to take the *TEFT* assessment, which is developed based on the same framework as the learning course.

Since the English-for-Teaching course is a newly developed professional training program that includes a precourse planner, a learning course, and an assessment, it is critical to examine the relationships among the three components within the English-for-Teaching course. Therefore, in this study we aim to answer the following two research questions:

1. To what extent does self-reported confidence level during precourse planning relate to engagement with the learning course?
2. To what extent does self-reported confidence level during precourse planning relate to performance on the *TEFT* assessment?

Design of the Study

The study analyzed data collected during the pilot administration of the *ELTeach* program in the fall of 2012 (for more details on the pilot testing, see Freeman, Katz, Le Dréan, Burns, & Hauck, 2013).

Study Sample

The study sample consisted of ELT teachers who participated in the pilot administration of the *ELTeach* program. The participants were first asked to complete the precourse planner. The amount of time they logged in the online curriculum was recorded automatically by the system. They were then presented with the opportunity to receive training by taking the online English-for-Teaching course at their own pace. Studying the provided learning materials was completely voluntary. The course was then followed by the administration of the *TEFT* assessment. A total of 729 participants participated in all three activities, among whom 683 had no missing values in their responses to the precourse planner. These 683 test takers constituted the sample for this study.

Ten different countries were represented in the study sample. The majority of the participants were from three countries: China (39%), Italy (27%), and Mexico (13%). Close to 30% of the participants reported that they were not teaching at the time of data collection. We speculate that these individuals could have been students who were still enrolled in various teacher education programs. About 60% of the participants reported that they were teaching at either a primary or a junior high grade level. We also learned that about 60% of the participants were teaching in the public sector at the time of data collection.

It should be pointed out that the sample was from the pilot administration, which has implications for the motivation of the study participants compared to the motivation of the operational users of the *ELTeach* program. We return to this issue later, when we discuss the limitations of our study.

Precourse Planner

The participants were asked to rate their level of confidence in conducting classroom tasks in English by responding to the 22 questions in the precourse planner. Each question asked about one classroom teaching activity. Eight questions referred to teaching activities for managing the classroom: collecting student work, getting students' attention, organizing students into groups, disciplining students, checking students' understanding, motivating students, making announcements, and taking attendance. Nine questions focused on activities for understanding and communicating lesson content: giving activity instructions, understanding reading texts, understanding audio recordings, communicating lesson goals, explaining English grammar, understanding lesson goals, understanding instructions, explaining vocabulary, and giving examples. Five questions were based on activities for providing feedback: identifying students' written errors, correcting students' written errors, identifying students' spoken errors, correcting students' spoken errors, and giving feedback to students. Sample survey questions are provided in Appendix A. Confidence in performing these tasks in English was rated on a 3-point scale: not confident, somewhat confident, and very confident. Upon completing the precourse planner, the teachers received a personalized learning path that suggested which units to focus on in the learning materials. The participants, however, had access to the entire curriculum regardless of their confidence overall or on individual activities. As a self-access course, the amount of learning was completely under the individual participant's control.

The English-for-Teaching Course

The online self-access learning course teaches a bounded set of functional words, phrases, and language skills teachers use to carry out essential classroom activities in English. The course is organized by unit, and each unit focuses on one of the three functional areas.

A unit typically starts with a preview that introduces the unit goals, familiarizes learners with key words used in the unit, and provides learners with the opportunities to show what they already know. Learners are then introduced to the learning activities. Following the learning activities, learners complete a couple of practice exercises and reflect on what they have learned.

There are six units in the learning course. Units 1 and 2 focus on language use for the purpose of managing the classroom. Units 3 and 4 contain learning materials for understanding and communicating lesson content. The last two units

Table 1 Summary of Test Tasks by Skill and by Functional Area

Skill	Managing the classroom	Understanding/communicating lesson content	Providing feedback	Total
Reading	0	21	12	33
Writing	2	10	9	21
Listening	4	20	6	30
Speaking	9	9	3	21
Total	15	60	30	105

Table 2 Confidence Ratings by Functional Area

Functional area	Number of items	Possible range	Mean	SD	Alpha
Managing the classroom	8	8–24	16.54	5.20	0.94
Understanding/communicating lesson content	9	9–27	18.75	5.84	0.95
Providing feedback	5	5–15	10.50	3.48	0.94

Table 3 Hours Spent on Learning by Functional Area

Functional area	Mean	SD
Managing the classroom	19.52	16.01
Understanding/communicating lesson content	9.47	9.26
Providing feedback	4.74	4.48

teach the type of language needed for providing feedback. The focuses of the units and the activities covered within each unit are provided in Appendix B.

The *TEFT* Assessment

The assessment component of the English-for-Teaching course is the *TEFT* assessment. Based on the same framework as the learning materials, this computer-based test measures how learners perform on typical classroom tasks as represented in the learning materials through standardized procedures. Test tasks are organized into the three functional areas. Each task also targets one of the four language skills: reading, writing, listening, and speaking. The listening and reading tasks are of selected-response format (e.g., multiple choice) and are each scored dichotomously (0 or 1). The writing and speaking tasks require constructed responses and are each scored on a scale from 0 to 3. The test reports a total scaled score as well as a scaled score for each of the skill sections. Also reported is the information on a test taker's command of English in the three functional areas. For each functional area, the number of score points earned out of the total available and the percentile information are reported. The test form used in this study had 105 tasks in total. Table 1 summarizes the number of test tasks by skill and by functional area.

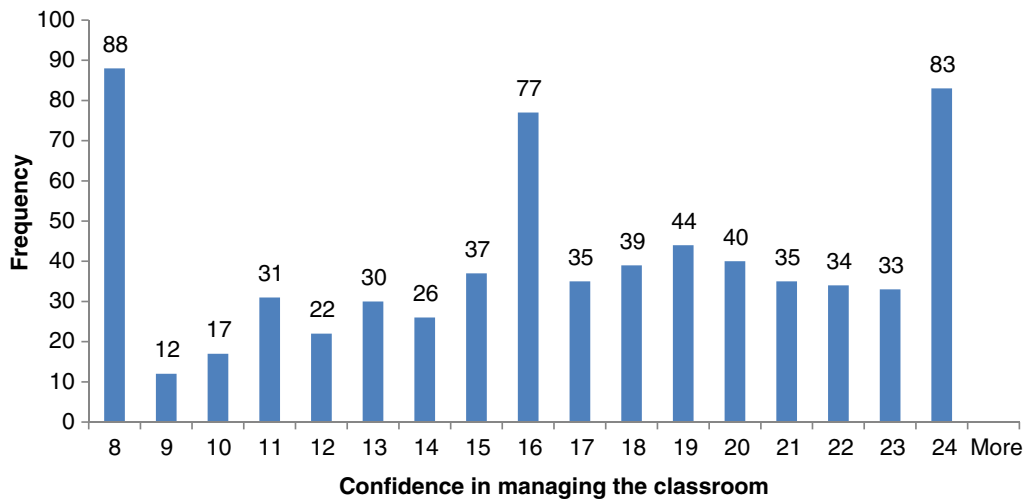
Data

Data were organized by functional area. Regarding confidence ratings, a candidate was assigned a score of 1 if a “not confident” response was given, 2 if a “somewhat confident” response was given, and 3 if a “very confident” response was given. Table 2 summarizes the descriptive statistics for the confidence ratings in each of the functional areas. A candidate's confidence level in a functional area was taken to be the sum of confidence ratings across all questions in that area.

Participants' engagement with the learning course was captured by the hours the participants logged in the curriculum for each functional area. It should be noted that this timing measure was only a proxy of the time the participants actually spent engaging with the curriculum. Table 3 summarizes the descriptive statistics for the hours spent on the learning materials by functional area. As expected, on average the participants spent the most amount of time on managing the classroom because the two units associated with this topic contained the most learning material (see Appendix B). They also spent the least amount of time on providing feedback because the two units on this topic were the shortest in length

Table 4 Test Scores by Functional Area

Functional area	Number of items	Possible range	Mean	SD	Kurtosis	Skewness	Alpha
Managing the classroom	15	0–37	29.71	4.98	1.02	–.95	0.79
Understanding/communicating lesson content	60	0–98	80.72	10.87	1.91	–1.23	0.91
Providing feedback	30	0–54	46.22	8.11	4.10	–1.83	0.88

**Figure 1** Distribution of confidence level in managing the classroom.

(see Appendix B). The standard deviations are high, suggesting that the self-access design of the course had direct implications for the variability of the amount of time spent as a measure of engagement, because participating teachers were free to devote as much or little time as they wished.

To obtain measures of participants' performance on the *TEFT* assessment, raw scores were summed across all items within each functional area. Table 4 summarizes the descriptive statistics for the scores by functional areas.

Analysis

We first explored the possibility of separating the entire sample into groups based on self-reported confidence level in a particular functional area. We examined the distribution of confidence level for each functional area. Figure 1 shows the distribution of confidence ratings summed across the eight questions in the functional area of managing the classroom. We observed three distinct groups. The first group ($N = 88$) had a total confidence rating of 8, having given a rating of 1 (not confident) to each of the 8 questions. The second group ($N = 77$) had a total rating of 16, meaning that on average these participants gave a rating of 2 (somewhat confident). The third group ($N = 83$) had a total rating of 24, having given a rating of 3 (very confident) to all eight questions. The same rating pattern held across all three functional areas, as the distribution in Figure 1 resembled those for the other two areas shown in Figure 2 and 3.

Within each functional area, we identified three distinct groups of different confidence levels. We therefore decided to separate the sample into three groups based on the highest and lowest confidence level scores, allowing for ties. Within each functional area, candidates with the top 25% of confidence level scores, including those who consistently gave the highest rating, constituted the high-confidence group. The candidates with the bottom 25% of confidence level scores, including the ones who consistently gave the lowest rating, represented the low-confidence group. Roughly 50% of the test takers in the middle were the medium-confidence group. Table 5 summarizes the number of candidates in each confidence group within each functional area.

For each functional area, we conducted a one-way analysis of variance (ANOVA) to compare the three groups on each of the two measures separately: time spent in hours and score on the *TEFT* assessment. All analyses were performed using SPSS.

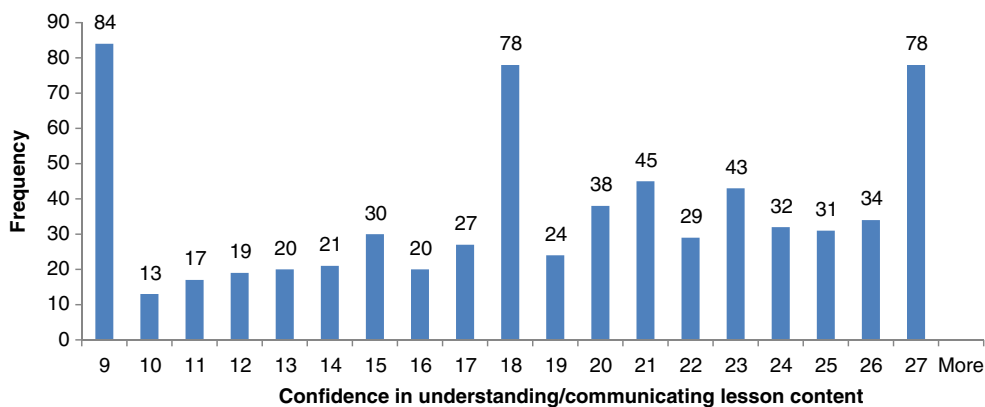


Figure 2 Distribution of confidence level in understanding/communicating lesson content.

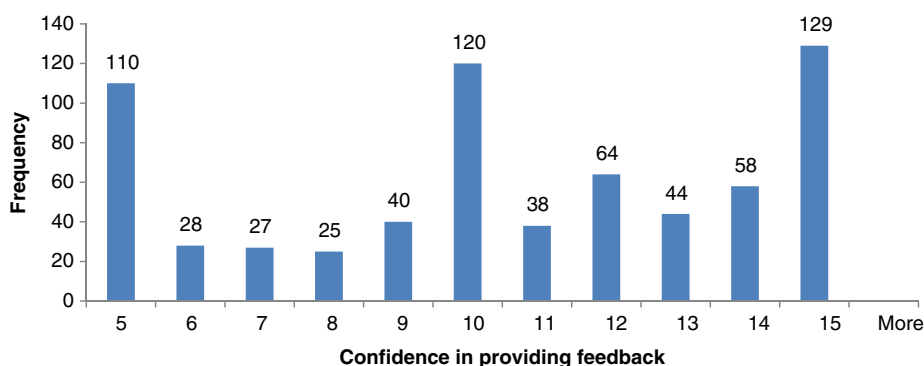


Figure 3 Distribution of confidence level in providing feedback.

Table 5 Number of Candidates in Each Confidence Group by Functional Area

Functional area	High-confidence	Medium-confidence	Low-confidence	Total
Managing the classroom	170	328	185	683
Understanding/communicating lesson content	174	334	175	683
Providing feedback	165	331	187	683

Results

Managing the Classroom

An ANOVA analysis was carried out to examine whether the three confidence groups differed in terms of time spent on the curriculum for managing the classroom. The result of the Levene’s test indicated that the assumption of homogeneity of variance could not be held. We therefore decided to report Welch’s *F*, which is a version of the *F*-ratio designed to be accurate when the assumption of homogeneity of variance has been violated. As summarized in Table 6, the ANOVA test for time spent was significant, although the effect size was small, $F(2, 390) = 9.84, p < 0.01, \eta^2 = 0.02$. According to Cohen (1988), the criteria for a small, medium, and large η^2 are 0.01, 0.06, and 0.14, respectively. Owing to uneven sample sizes in the comparison groups and violation of test assumption, we decided to use the Games-Howell procedure to perform the post-hoc between-group comparisons. The results showed that the high-confidence group spent significantly less time on learning than did the low-confidence group with a small effect size ($p < 0.01, d = 0.38$) and the medium-confidence group with a small effect size ($p < 0.01, d = 0.32$). Using Cohen (1988)’s guidelines, the criteria for a small, medium, and large *d* are 0.2, 0.5, and 0.8, respectively. The difference between the low- and medium-confidence groups was not significant ($p < 0.01$).

Table 6 ANOVA Test Results for Managing the Classroom

Group	Hours Mean	Hours SD	ANOVA	TEFT Mean	TEFT SD	ANOVA
Low ($N = 170$)	21.49	17.66	$F(2, 390) = 9.84$ ($p < 0.01$)	29.64	5.52	$F(2, 680) = 0.33$ ($p = 0.72$)
Medium ($N = 328$)	20.54	17.16		29.86	4.78	
High ($N = 185$)	15.91	11.02		29.50	4.85	

Table 7 ANOVA Test Results for Understanding and Communicating Lesson Content

Group	Hours Mean	Hours SD	ANOVA	TEFT Mean	TEFT SD	ANOVA
Low ($N = 174$)	12.29	11.59	$F(2, 370) = 26.50$ ($p < 0.01$)	78.08	12.15	$F(2, 680) = 7.08$ ($p < 0.01$)
Medium ($N = 334$)	9.81	8.61		81.70	10.16	
High ($N = 175$)	6.01	6.38		81.49	10.46	

Table 8 ANOVA Test Results for Providing Feedback

Group	Hours Mean	Hours SD	ANOVA	TEFT Mean	TEFT SD	ANOVA
Low ($N = 165$)	5.76	5.24	$F(2, 373) = 34.42$ ($p < 0.01$)	46.46	7.90	$F(2, 680) = 0.98$ ($p = 0.38$)
Medium ($N = 331$)	5.27	4.46		45.78	8.38	
High ($N = 187$)	2.89	3.01		46.77	7.79	

The confidence groups were also compared regarding *TEFT* performance. The Levene's test indicated that the assumption of homogeneity of variance was not violated. The outcome of the ANOVA test was not significant, $F(2, 680) = 0.33$, $p = 0.72$, indicating that, regardless of confidence level, all groups performed similarly on the test.

Understanding/Communicating Lesson Content

The ANOVA test results in the functional area of understanding and communicating lesson content are summarized in Table 7. When comparing the time spent on the course, the assumption of homogeneity of variance could not be held. Results based on Welch's F indicated that the overall difference was significant with a medium effect size, $F(2, 370) = 26.50$, $p < 0.01$, $\eta^2 = 0.06$. Using the Games-Howell procedure, the results of the follow-up between-group comparisons showed that the high-confidence group spent significantly less time on the curriculum than did the low-confidence group with a medium effect size ($p < 0.01$, $d = 0.67$) and the medium-confidence group with a medium effect size ($p < 0.01$, $d = 0.50$), and that the difference between the low- and medium-confidence group was not significant ($p < 0.01$).

The confidence groups were also compared regarding *TEFT* performance. The Levene's test indicated that the assumption of homogeneity of variance was not violated. The outcome of the ANOVA test was significant with a small effect size, $F(2, 680) = 7.08$, $p < 0.01$, $\eta^2 = 0.02$, indicating an overall performance difference across the groups. The follow-up between-group comparisons by using the Games-Howell procedure indicated that the low-confidence group scored significantly lower than the medium-confidence group with a small effect size ($p < 0.01$, $d = 0.32$) and the high-confidence group with a small effect size ($p < 0.01$, $d = 0.30$) and that the medium- and high-confidence groups did not perform differently on the test.

Providing Feedback

Table 8 summarizes the ANOVA test results in the functional area of providing feedback. When comparing the time spent on the curriculum, the assumption of homogeneity of variance could not be held. Results based on Welch's F indicated that the overall difference was significant with a medium effect size, $F(2, 373) = 34.42$, $p < 0.01$, $\eta^2 = 0.07$. The follow-up between-group comparisons by using the Games-Howell procedure showed that the high-confidence group spent

significantly less time on learning than did the low-confidence group with a medium effect size ($p < 0.01$, $d = 0.67$) and medium-confidence group with a medium effect size ($p < 0.01$, $d = 0.63$) and that the difference between the low- and medium-confidence groups was not significant.

The confidence groups were also compared regarding *TEFT* performance. The Levene's test indicated that the assumption of homogeneity of variance was not violated. The outcome of the ANOVA test was not significant, $F(2, 680) = 0.98$, $p = 0.38$, indicating that, regardless of confidence level, all groups performed similarly on the test.

Discussion

In this study we explored the relationships among the three components within the English-for-Teaching course: confidence reported by using the precourse planner, time spent on the learning course, and performance on the *TEFT* assessment. We carried out our investigation by functional area because categorization by functional area served as a crucial organizing principle for the content presented in the course.

The findings from our study suggested that there was an association between the confidence level reported by participants during the precourse planning and the time spent on learning course. We found that the high-confidence learners spent significantly less time on the curriculum than did the medium- and low-confidence groups in all three functional areas. This finding was largely consistent with the design of the *ELTeach* program. Based on the results of precourse planning, low- and medium-confidence learners were advised to work on more learning activities, and we found that they indeed spent more time on the curriculum. High-confidence learners were recommended fewer learning activities, and we found that they did spend less time. However, we did not find significant differences in time between the low- and the medium-confidence groups.

When exploring the relationship between precourse confidence and performance on the *TEFT* assessment, we found that, in the two functional areas of managing the classroom and providing feedback, all groups performed very similarly on the test regardless of confidence level. One potential way to interpret this finding is that after receiving the training through the learning course, the groups seemed to have equal chances to succeed on the test, although without a precourse measure of *TEFT* competence the impact of curriculum engagement on test performance cannot be statistically evaluated in the study. In the functional area of understanding and communicating lesson content, however, we found that the low-confidence group performed significantly worse than the medium- and high-confidence groups, indicating that even after having been exposed to the curriculum the low-confidence group still lagged behind the other two groups in handling tasks related to understanding and communicating lesson content. This finding suggested that the low-confidence group could benefit from receiving more training in the area of understanding and communicating lesson content than what was offered in the learning course.

A few study limitations need to be pointed out. First, the data came from a pilot administration instead of an operational administration and hence the test takers may not have been as motivated as operational users of the *ELTeach* program who have the possibility of receiving or not receiving a certificate based on the quality of their participation. Second, the majority of the participants came from three countries, which limited the extent to which the study results can be generalized to the entire target test-taking population. Third, the grouping of the participants relied on their self-reported confidence level, which might not have been a true measure of their confidence in using English for teaching English. Fourth, in this study engagement was defined by the hours spent on the course. The richness of these participants' engagement with the learning course was not fully captured by this indicator. Variables other than time-on-task were not investigated because they were not available. It is recommended that future researchers collect detailed information on the nature and intensity of learners' interaction with the learning course. Finally, because the data used for this study were purely observational, we encourage future researchers to design controlled experimental studies to investigate the effect of the learning curriculum offered by the *ELTeach* program in a more rigorous way.

Despite the aforementioned study limitations, we believe that our study constitutes an initial step in providing evidence of the effect of construct-relevant training provided by the English-for-Teaching course on teachers' classroom English language proficiency.

References

Berry, R. (1990). The role of language improvement in in-service teacher training: Killing two birds with one stone. *System*, 18, 97–105. [http://dx.doi.org/10.1016/0346-251X\(90\)90032-Z](http://dx.doi.org/10.1016/0346-251X(90)90032-Z)

- Butler, Y. G. (2004). What level of English proficiency do elementary school teachers need to attain to teach EFL? Case studies from Korea, Taiwan, and Japan. *TESOL Quarterly*, 38, 245–278. <http://dx.doi.org/10.2307/3588380>
- Canagarajah, A. S. (1999). Interrogating the native speaker fallacy; non-linguistic roots, non-pedagogical results. In G. Braine (Ed.), *Non-native educators in English language teaching* (pp. 77–92). Mahwah, NJ: Lawrence Erlbaum.
- Chacón, C. T. (2005). Teachers' perceived efficacy among English as a foreign language teachers in middle school in Venezuela. *Teaching and Teacher Education*, 21, 257–272. <http://dx.doi.org/10.1016/j.tate.2005.01.001>
- Cohen, J. (1988). *Statistical power analysis for the behavioral science* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cullen, R. (1994). Incorporating a language improvement component in teacher training programmes. *ELT Journal*, 48, 162–172. <http://dx.doi.org/10.1093/elt/48.2.162>
- Freeman, D., Katz, A., Gomez, P. G., & Burns, A. (2015). English-for-teaching: Rethinking teacher proficiency in the classroom. *ELT Journal*, 2, 129–139. <http://dx.doi.org/10.1093/elt/ccu074>
- Freeman, D., Katz, A., Le Dréan, L., Burns, A., & Hauck, M. (2013). *ELTeach global pilot report 2012*. Retrieved from http://elteach.com/ELTeach/media/Documents/ELTeach_GPR_9-20-13.pdf
- Kamhi-Stein, L. D. (2009). Teacher preparation and nonnative English speaking educators. In A. Burns & J. C. Richards (Eds.), *The Cambridge guide to second language teacher education* (pp. 91–101). Cambridge, England: Cambridge University Press.
- Murdoch, G. (1994). Language development provision in teacher training curricula. *ELT Journal*, 48, 253–265. <http://dx.doi.org/10.1093/elt/48.3.253>
- Nunan, D. (2003). The impact of English as a global language on educational policies and practices in the Asia-Pacific region. *TESOL Quarterly*, 37, 589–613. [10.2307/3588214](http://dx.doi.org/10.2307/3588214)
- Richards, J. C. (2010). Competence and performance in language teaching. *RELC Journal*, 41, 101–122. <http://dx.doi.org/10.1177/0033688210372953>
- Seidlhofer, B. (1999). Double standards: Teacher education in the expanding circle. *World Englishes*, 18, 233–245. <http://dx.doi.org/10.1111/1467-971X.00136>
- Sešek, U. (2007). English for teachers of EFL—Toward a holistic description. *English for Specific Purposes*, 26, 411–425. <http://dx.doi.org/10.1016/j.esp.2006.11.001>
- Young, J. W., Freeman, D., Hauck, M. C., Garcia Gomez, P., & Papageorgiou, S. (2014). *A design framework for the ELTeach program assessments* (Research Report No. RR-14-36). Princeton, NJ: Educational Testing Service. <http://dx.doi.org/10.1002/ets2.12036>

Appendix A

Sample Survey Questions

Screen capture of a question that addresses the functional area of managing the classroom.

Pre-Course Planner

For each topic choose your level of confidence (*very confident, somewhat confident, or not confident*). You must respond to all topics. After responding to the last topic, click **Submit** to see your course plan.

Topic: Taking attendance

How confident do you feel using English to take attendance?

Very confident
 Somewhat confident
 Not confident

Screen capture of a question that addresses the functional area of understanding and communicating lesson content.

Pre-Course Planner

For each topic choose your level of confidence (*very confident, somewhat confident, or not confident*). You must respond to all topics. After responding to the last topic, click **Submit** to see your course plan.

Topic: Understanding reading texts

How confident do you feel understanding reading texts in English in student textbooks?

- Very confident**
- Somewhat confident**
- Not confident**

Screen capture of a question that addresses the functional area of providing feedback.

Pre-Course Planner

For each topic choose your level of confidence (*very confident, somewhat confident, or not confident*). You must respond to all topics. After responding to the last topic, click **Submit** to see your course plan.

Topic: Identifying students' spoken errors

How confident do you feel identifying students' spoken errors?

- Very confident**
- Somewhat confident**
- Not confident**

Appendix B

Unit and Activity Focuses

Units		Unit Structure
1. Managing the Classroom A	1	Greeting students
	2	Discussing the date and weather
	3	Taking attendance
	4	Reviewing and collecting students work
	5	Review A
	6	Making announcements A
	7	Making announcements B
	8	Assigning homework
	9	Dismissing the class
	10	Review B
2. Managing the Classroom B	1	Using classroom materials
	2	Teaching classroom language
	3	Directing students
	4	Giving test and quiz instructions
	5	Review A
	6	Changing activities (getting students' attention)
	7	Disciplining
	8	Checking student understanding
	9	Encouraging participating
	10	Motivating students
	11	Review B
3. Understanding Lesson Content	1	Understanding instructions in student materials
	2	Understanding listening texts in student materials
	3	Understanding reading texts in student materials
	4	Writing content from student materials
	5	Review
4. Communicating Lesson Content	1	Understanding and communicating lesson goals
	2	Engaging students in the topic
	3	Explaining lesson content
	4	Modeling and giving examples
	5	Review A
	6	Asking for and writing student examples
	7	Giving activity instructions A
	8	Giving activity instructions B
	9	Organizing students
	10	Review B
5. Assessing Student Work	1	Identifying and correcting written errors
	2	Identifying and correcting spoken errors
	3	Assessing student comprehension
	4	Review
6. Giving Feedback	1	Giving positive feedback
	2	Encouraging self-correction A
	3	Encouraging self-correction B
	4	Review

Suggested citation:

Gu, L., & Papageorgiou, S. (2016). *Exploring the relationships among teacher confidence, learning, and test performance within the English-for-Teaching Course* (Research Report No. RR-16-24). Princeton, NJ: Educational Testing Service. <http://dx.doi.org/10.1002/ets2.12110>

Action Editor: Heather Buzick

Reviewers: Pablo Garcia Gomez and Sultan Turkan

ETS and the ETS logo are registered trademarks of Educational Testing Service (ETS). MEASURING THE POWER OF LEARNING and TEFT are trademarks of ETS. ELTEACH is a trademark of Cengage Learning, Inc. All other trademarks are property of their respective owners.

Find other ETS-published reports by searching the ETS RESEARCHER database at <http://search.ets.org/researcher/>