Research Paper

Development and Validation of a Quantitative Research Literacy Questionnaire

Alireza Zaker[1]*, Mania Nosratinia[2]

[1]Assistant Professor, English Language Research Center, South Tehran Branch, Islamic Azad University, Tehran, Iran
a.zaker@srbiau.ac.ir
[2]Assistant Professor, English Department, Central Tehran Branch, Islamic Azad University, Tehran, Iran
mania_nosratinia@yahoo.com

Abstract
Simultaneous with the emergence of numerous new concerns and techniques in carrying out research in English language teaching (ELT) contexts, conducting ELT research has turned into a multi-dimensional process, and ELT researchers seem to require advanced research skills in order to address different pedagogical issues. Focusing on quantitative research, the present study was undertaken to develop and validate a quantitative research literacy (QRL) instrument for English as a foreign language (EFL) teacher training context. To this objective, a four-component model of QRL, encompassing developing research topic knowledge, research design knowledge, procedural knowledge, and data analysis knowledge was developed based on experts' opinion and an extensive review of the related literature. Testing of the tentative model through exploratory and confirmatory data analyses on a sample of 1180 EFL teachers across three education levels indicated that a 4-factor model of QRL with 9 sub-factors and 50 items could best explain QRL.

Keywords: *Factor analysis, Quantitative research, Questionnaire, Research literacy, Teacher training, Validation*

طراحی و روایی سنجی پرسشنامه سواد تحقیق کمی

همزمان با پیدایش دغدغه ها و روش های جدید و متعدد در انجام تحقیق در حوزه آموزش زبان انگلیسی، انجام تحقیق در این حوزه تبدیل به فرآیندی پیچیده و چند بعدی شده است، و به نظر می رسد که محققان در حوزه آموزش زبان انگلیسی نیازمند به برخورداری از مهارتهای پیشرفته تحقیق به منظور پرداختن به موضوعات مرتبط به آموزش می باشند. با تمرکز بروی تحقیق کمی، این تحقیق تلاشی بود در جهت طراحی و روایی سنجی یک ابزار برای سنجش سواد تحقیق کمی در محیط های تربیت معلم آموزش زبان انگلیسی. بدین منظور، یک مدل 4 عاملی سواد تحقیق کمی بر پایه نظرات دریافتی از خبرگان دانشگاهی در حوزه تحقیق در آموزش زبان انگلیسی و مطالعه عمیق تحقیقات در زمینه آموزش زبان انگلیسی طراحی گردید که شامل دانش طراحی عنوان تحقیق، دانش طرح ریزی تحقیق، دانش رویه تحقیق، و دانش تحلیل داده ها می گردید. آزمایش مدل اولیه بواسطه انجام تحلیل عاملی اکتشافی و تاییدی بروی داده های گردآوری شده از 1180 مدرس زبان انگلیسی با 3 سطح مختلف تحصیلات دانشگاهی آشکار نمود که یک مدل 4 عاملی سواد تحقیق کمی با 9 عامل فرعی و 50 سوال به خوبی سواد تحقیق کمی را توضیح می دهد.

واژگان کلیدی: پرسشنامه، تحقیق کمی، تحلیل عاملی، تربیت معلم، روایی سنجی، سواد تحقیق

Introduction

Prevalent trends in Teaching English as a Foreign Language (TEFL) are deeply rooted in the supposition that English as a Foreign Language (EFL) teachers and instructors function as reflective and transformative practitioners who investigate and explore classroom events and outcomes (Kumaravadivelu, 2012; Lightbown & Spada, 2013). As proposed by Richards and Lockhart (1994), teachers should "collect data about their teaching, examine their attitudes, beliefs, assumptions, and teaching practices, and use the information obtained as a basis for critical reflection about teaching" (p. 1). Furthermore, this inquiry which has also been addressed through *Action/Classroom Research* is believed to be "an appealing way to look more closely at puzzling classroom issues or to delve into teaching dilemmas" (Burns, 2010, p. 6). Hence the reflective appraising of the teaching practice is considered a key asset to EFL teachers (Farrell, 2012).

There is a general consent that undertaking a systematic, context-based, and well-designed applied English language teaching (ELT) research can make a significant contribution to ELT practitioners' existing understanding of ELT and the development of pedagogical techniques (Farrell, 2012; Jay & Johnson, 2002; Zaker, Nosratinia, Birjandi, & Yazdanimoghaddam, 2019). However, we have recently witnessed a surge of growth regarding the pedagogical practice, curriculum development, assessment, learner variables, and teacher education in recent years (Akbari, 2008; Bell, 2003; Ellis, 2010; Fahim & Zaker, 2014; Lightbown & Spada, 2013; Mitchell, Myles, & Marsden, 2013; Nation & Macalister, 2010; Nosratinia & Zaker, 2014). Accordingly, and quite reasonably, there has been a simultaneous emergence of new concerns and techniques in carrying out applied research in ELT contexts (Birjandi & Siyyari, 2010; Mackey & Gass, 2015), and both professional researchers and TEFL students seem to require advanced research skills in order to address different pedagogical issues (Blessinger, 2015).

Applied research is a category of research which "aims to find a solution to a specified practical problem under the conditions in which it appears in practice (Ary, Jacobs, Sorensen Irvine, & Walker, 2019, p. 15). Conducting applied research is primarily undertaken by dint of quantitative and qualitative routes (Best & Kahn, 2006; Creswell, 2014). Reflecting upon the implementation of applied research in the ELT domain, both qualitative and quantitative approaches are believed to have their own advantages and disadvantages; however, quantitative research seems to provide a better basis for answering pedagogical inquiries and inspecting the effectiveness of pedagogical techniques (Hadi & Closs, 2016). This special privilege is emanated from the fact that the modern approach to ELT research has been rooted in positivism through highlighting objectivity, generalizability, and being criterion-oriented (Breen & Darlaston-Jones, 2010). Furthermore, although qualitative measures are of inestimable value in research, they are not highly practical in terms of the required resources. In consequence, quantitative research is "privileged over other forms of enquiry, and other epistemologies, methodologies, and methods remain marginalised within the discipline" (Breen & Darlaston-Jones, 2010, p. 67).

In research to date, there has been no systematic attempt to develop a model and scale for estimating ELT practitioners' knowledge of systematic quantitative research, henceforth referred to as quantitative research literacy (QRL; P. Birjandi, personal communication, January 1, 2016). The absence of this instrument makes it definitely challenging to diagnose the problematic areas of QRL among ELT practitioners and researchers whose articles and studies exhibit numerous problems, ranging all the way from erroneous study designs (M. Alavi, personal communication, August 7, 2015) to plagiarism (Sabbaghan, 2010) and publication and presentation problems (L. Dale, personal communication, September 4, 2015). Accordingly, there were two motives behind this study; first to come up with a tangible, valid, and expert-supported model of QRL, calibrated

for EFL teachers, and second, to design a statistically validated instrument to allow for the quantification of the construct and consequently, its empirical investigation.

The Study

The present study was theoretically rooted in the notion that teachers' framework of teaching is influenced by both research and teachers' reflection and observation of teaching and learning (Nation & Macalister, 2010; Nosratinia & Zaker, 2017). As it was stated, ELT practitioners are now expected to conduct systematic research (Best & Kahn, 2006) as one of the significant attempts to reflectively evaluate the pedagogical practice (Springer, 2010). As a result, conducting a systematic and well-designed ELT research is assumed to make a significant contribution to ELT practitioners' existing understanding of ELT and the developing pedagogical techniques (Zeichner, 1999). However, conducting ELT research has turned into a complex and multi-dimensional process, and TEFL researchers require advanced research skills in order to address different pedagogical issues (Blessinger, 2015).

As stated earlier, many TEFL majors exhibit poor QRL (M. Alavi, personal communication, August 7, 2015); therefore, the QRL instrument can be employed as a diagnostic tool for TEFL majors before conducting a research so that remedy can be provided for their problems through providing focused instruction and recommending references. QRL seems to be an asset to EFL practitioners; however, when it comes to assessing the degree of QRL, the need for a valid and reliable instrument becomes conspicuous. So far, there has been no systematic attempt to develop a model and instrument for estimating ELT practitioners' QRL. Such an instrument will allow for the quantification of the construct and, consequently, its empirical investigation. Furthermore, this instrument can function as a reliable and informative tool for TEFL graduate programs which mostly require that students carry out assisted or independent research.

A Tentative Model

Contemporaneous with the growing endorsement of the constructivist theory of cognitive and mental development (Ashton-Hay, 2006; Zaker, 2016), the ELT domain seems to spotlight ELT practitioners' critical mental engagement in pedagogical events (Farrell, 2012; Kumaravadivelu, 2012). In order to collect and use the abovementioned information in a reasonable way, ELT practitioners should be familiar with the principles of conducting a systematic applied research which can materialize the reflective evaluation of the pedagogical practice (Springer, 2010). Research has been simply defined as, "a process of steps used to collect and analyze information to increase our understanding of a topic or issue" (Creswell, 2010, p. 3). However, based on the integration of many new factors in the TEFL practice, conducting a research in this field is no longer synonymous with such a simple definition. In fact, we now witness the emergence of many new concerns and techniques in carrying out research in ELT contexts (Mackey & Gass, 2015). Quantitative research covers a vast array of aspects and peculiarities. In this study, the categories and the components of QRL were initially developed based on an extensive review of the existing literature and interview with content area experts (thoroughly explained in the following section. The four components/factors of the proposed model are briefly discussed hereunder.

Developing the Research Topic

According to Cohen, Manion, and Morrison (2011), the identification of an appropriate research area or topic is a crucial step in conducting research which should be based on being "original, significant, non-trivial, relevant, topical, (and) interesting to a wider audience and to advance the field" (p. 106). Furthermore, a formidable issue is to consider and access and

practicality (Cohen et al., 2011). The in-depth study of the existing body of research should be carried out "before the actual conduct of the study begins in order to provide a context and background that support the conduct of the study" (Ary, Jacobs, & Sorensen, 2010, p. 62). Besides, the focus of study should be narrowed to make it "potentially researchable" (Best & Kahn, 2006, p. 33). Also, the researcher should attempt to prove that the "proposed study is important" (Ary et al., 2010, p. 589).

Design Knowledge

Design knowledge seems to me the most inclusive component of QRL. To begin with, awareness of different quantitative research types is a key factor in designing a quantitative research. This includes being cognizant of the characteristics of experimental research, the qualities of quasi-experimental research, and the features of a descriptive quantitative research (correlational and ex post facto). It has been stated that research in a "hypothetico-deductive mode, and research that uses statistics, often commences with one or more hypotheses" (Cohen et al., 2011, p. 608). Having said that, the researcher needs to distinguish between different typical hypothesis types, i.e. null hypothesis and alternative hypothesis.

Researchers need to possess adequate mastery over variables and measurement scales. Furthermore, within the quantitative research framework, it is critical to be aware of research validity and its threats (see Ary et al., 2010, p. 645; Best & Kahn, 2006, p. 172; Cohen et al., 2011, p. 183). Moreover, as within this framework the researchers basically works on a sample defined as a "small proportion of the population that is selected for observation and analysis" (Best & Kahn, 2006, p. 13), sampling knowledge seems to be another factor in design knowledge.

Procedural Knowledge

The first factor involved in this area is data collection competence which highlights employing "the most appropriate instruments and procedures that provide for the collection and analysis of data on which hypotheses may be tested" (Best & Kahn, 2006, p. 346). Moreover, when conducting a research involving human beings as the participants, "it is important to consider the ethical guidelines designed to protect the participants" (Best & Kahn, 2006, p. 47). Indeed, "strict adherence to ethical standards in planning and conducting both qualitative and quantitative research is most important" (Ary et al., 2010, p. 590), and it is essential to appreciate and value "the fundamental rights, dignity, and worth of all people" (Best & Kahn, 2006, p. 50). The eleven aspects of researchers' obligations to participants are identified in the AERA standards (American Educational Research Association, 1992, p. 24).

Data Analysis Knowledge

The knowledge of data analysis (aka statistical knowledge) is believed to function as a key element of quantitative research. Data analysis knowledge deals with analyzing the quantitative data and checking the hypotheses (Best & Kahn, 2006). More specifically, it deals with a "body of mathematical techniques or processes for gathering, organizing, analyzing, and interpreting numerical data" (Best & Kahn, 2006, p. 354). Data analysis knowledge is "an indispensable tool for researchers that enables them to make inferences or generalizations about populations from their observations of the characteristics of samples" (Best & Kahn, 2006, p. 441).

Instrument Development and Validation
Participants
Following the preparation of the first draft of the QRL model and instrument, we reached a group of 16 esteemed TEFL scholars who specialized in Research Methodology; they came from Allameh Tabataba'i University, Islamic Azad University (Central Tehran, South Tehran, & Garmsar branches), and Shahid Beheshti University. These research experts were provided with the initial QRL model and instrument. In addition to evaluating the model and the items through a Likert-type evaluation framework, these experts shared their comments and suggestions in written and oral form with the researcher.

In the second part, the second version of the QRL instrument (including 78 items) was administered to 1180 individuals who were selected employing convenience sampling from Aale Taha University, Allameh Tabataba'i University, Islamic Azad University (Central Tehran and South Tehran branches), and Shahid Beheshti University. These participants were male and female (932 or 79% females; 248 or 21% males) B.A., M.A., and Ph.D. students who majored in English language teaching, English translation, and English literature. The age range of the participants was 18 to 58 (M_{age} = 25). The distribution of participants in each step was 9 for the initial piloting and 1171 for checking validity; however, the data collected from 717 participants demonstrated due care in answering.

Questionnaire Development
Step One: Deciding on the Content Area
The first step in developing a questionnaire is to identify what concepts should be included in it (Dörnyei, 2010). To facilitate this, researchers have at least two potential sources: (a) review of the existing literature, and (b) interview with content area experts. For this study, the related literature on QRL and research methodology was extensively investigated. Moreover, a number of research methodology experts were asked to share their opinions regarding the elements and areas of QRL. In addition, we studied more than 50 TEFL theses along with studying numerous published articles written by TEFL majors. Taking these steps resulted in proposing the initial model of QRL, including 4 factors and 12 sub-factors (Table 1). This model was the basis for the content validity of the QRL instrument. In other words, this phase addressed the adequacy of the developed test for sampling the universe of knowledge and skills which is being inspected, QRL, achieved through judgment of subject matter specialists and reviewing the literature (Best & Kahn, 2006; Cronbach & Meehl, 1955).

Table 1
The Initial Model of Quantitative Research Literacy Including the Factors and Sub-Factors

Factor 1: Developing the Research Topic

Embeddedness in the Existing Body of Research

Narrowing Down the Focus of the Study

Justifying the Significance of the Study

Practicality Concerns

Factor 2: Design Knowledge

Awareness of Different Quantitative Research Types

Knowledge of Variables and Measurement Scales

Research Validity Knowledge

Sampling Knowledge
Factor 3: Procedural Knowledge
Data Collection Competence
Implementing Ethics in Research
Factor 4: Data Analysis Knowledge
General Statistical Knowledge
Cognizance of Statistical Tests

Step Two: Drawing Up an Item Pool

Once the components to be included in the questionnaire were determined through a well-established model, it was time for the generation of an item pool. One of the main concerns in the development of a questionnaire is to strike a balance between the need to include all aspects of a given component through using multi-item scales and the necessity to keep the total number of items within a reasonable range (Dörnyei, 2010). Considering that the initial model of QRL included 12 components, the researchers aimed for including two to six items for each component depending on its complexity and multi-dimensionality. However, as it was unclear how the items would function in the real survey and it was predicted that some of the items would be eliminated in the expert review and piloting steps, more items were selected and generated (four to twelve for each sub-factor) for the pool.

The items of a new questionnaire could be either created by the researcher or borrowed from other questionnaires (Dörnyei, 2010). However, as no other research literacy or QRL scales were available, all of the items (*n* = 109) were written by the researchers and went through multiple revisions for the sake of brevity, conciseness, and avoiding double-barreled items. Some of the original items and their revised versions are presented below.

Original: I believe it is essential to make it clear what has given rise to the research, i.e. what is the reason for choosing the research topic.

Revised: I believe it is essential to determine what the reason for choosing the research topic is.

Original: When planning a research, it is important to consider whether I have the right personality, characteristics, experience, and interpersonal behavior to conduct the proposed piece of research.

Revised: I believe the researcher's personality and characteristics are key factors in planning a research.

Original: I believe the events and naturally occurring changes in participants between pretest and posttest might negatively affect the validity of an experimental research.

Revised: I believe events and changes in participants affect the validity of an experimental research.

When generating items, the researchers followed the guidelines suggested by Dörnyei (2010). During this process, it was attempted to use simple and natural language, avoid ambiguous words, negative constructions, and double-barreled questions (i.e. items that require a single answer to a statement that asks more than one question). In the end, a total of 109 items were incorporated into the questionnaire.

Step Three: Eliciting Expert Opinion

As was mentioned in the previous step, we included almost double the number of required items for each sub-factor. Therefore, in order to evaluate the QRL model and evaluate the 109 items, 16 TEFL scholars who specialized in Research Methodology were invited to evaluate the model and rate each item on a three-category Liker-type scale. Regarding the items, these experts were asked to read each statement carefully and tick Essential, Useful, or Not Relevant. These professionals were also asked to write their evaluation of items regarding their wording, unambiguity, and appropriateness in the form of a comment, if necessary; in some cases, they also shared their comments and evaluation in oral form.

The criterion of 70 percent of raters checking the *Essential* box was set as the minimum requirement for accepting an item (Dörnyei, 2010). Therefore, based on the obtained responses and comments, the number of items was reduced from 109 to 78. The items were deleted for low rating or comments on low content representativeness, relevance, and being double-barreled. Furthermore, some of the items were revised but kept in the instrument.

Step Four: Selecting Rating Scale

As for the rating scale, Likert scale, which is the most common scaling technique, was utilized in the present study. The researchers developed a scale which is a 6-point Likert scale anchored only at endpoints: 1 (strongly disagree) and 6 (strongly agree):

Strongly Disagree					Strongly Agree
1	2	3	4	5	6
-	-	-	+	+	+

The use of this rating scale is justified by three reasons. First, though the results are mixed, studies have generally shown that reliability and validity are optimum for scales with a moderate number of rating points than those with fewer or more points (Krosnick & Presser, 2010). However, reliability improvement becomes correspondingly smaller beyond 7 points. Second, the use of scales with endpoint labels enables researchers to calculate the linear relationship between variables using correlations or regression models (Weijters, Cabooter, & Schillewaert, 2010). Third, as suggested by Dörnyei (2010), incorporating an even number of choices with no neutral option would add to the validity of the collected data. The "I have no idea/undecided" option was not incorporated in the choices as many of the items have used the "I know" format. For the other items, it is a common belief that adding the neutral option will encourage the respondents to choose the easy option.

Step Five: Designing the Personal Information Section

The personal information section asked for respondent's name (optional to answer), gender, age, teaching experience, major, and education level. It is recommended not to put personal information near the beginning of the questionnaire as respondents initially expect to face interesting questions related to the topic than some factual items. Furthermore, some pieces of factual information might be sensitive and thus create resistance on the side of the respondents to express their real opinion about the items (Dörnyei, 2010). In line with these arguments, this section was placed at the end of the questionnaire.

Step Six: Writing Instructions

Two types of instructions were written for this study. A general instruction which most importantly explained what the study is about, for what purpose the data are collected, and that there is no right or wrong answer. Also, a specific instruction followed which explained how

respondents should answer each item and what the numerical values of the Likert scale stand for. To avoid mistake, the scale was repeated on the top of each page.

Step Seven: Initial Piloting

After adding the instructions section, the researchers set out to conduct the first phase of the field testing. At this stage, the items were administered to a sample of respondents who were characteristically similar to the target sample for whom the questionnaire was designed. A panel of 9 students from Islamic Azad University, South Tehran Branch volunteered to take part in this stage of the study. Participants took part in a round-table meeting where the researchers were present to discuss with them each item of the questionnaire. Volunteers were required to go through a certain number of items in a given period of time and give feedback on the items regarding their appropriateness of wording, clarity of meaning, and naturalness. Based on the comments received from this panel, the wording of some items was changed.

Step Eight: Ensuring Face Validity and Construct Validity

Validity of an instrument deals with the adequacy of the developed test for sampling the universe of knowledge and skills which is being inspected (Best & Kahn, 2006; Cronbach & Meehl, 1955). Therefore, numerous measures were taken in order to provide evidence for the validity of the QRL model and instrument. Basically, in this study three main strategies were implemented in order to investigate the validity of the QRL model and instrument, content, face, and, most importantly, construct validity. Although some of these measures were already discussed (i.e. content validity in steps 1, 2, & 3 of this section), in this step, the other two aspects of instrument validity, face validity and construct validity, are discussed.

Face validity. In the course of developing a questionnaire face validity should not be neglected. As Dörnyei (2010, p. 13) argues, "producing an attractive and professional design is half the battle in eliciting reliable and valid data." A properly designed questionnaire which looks nice to the eyes convinces respondents to spend time and effort on responding a questionnaire. Therefore, to make the questionnaire attractive, great care was taken on such graphic issues as the layout, typesetting, and margin.

Construct validity. Construct validity of an instrument is believed to deal with the correspondence of the patterns of correlation and covariance among the items in the instrument with the proposed construct model (J. D. Brown, 2001; Dörnyei, 2010; Kline, 2015). The commonest framework for gathering information in support of the construct validity of a survey is factor analysis which is employed to help researchers ascertain the underlying structure of an instrument. In the present study, the data collected employing the 78-item version of the instrument from 1171 participants were analyzed by the two categories of factor analysis, i.e. exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). For the construction of a measuring device, researchers start with an EFA and then move to the CFA.

From the 1020 returned instruments, a number of 717 instruments qualified for the validation analyses as the items were answered carefully and completely. The following sections report the validation analyses.

Uncovering the QRL's factor structure. Several exploratory factor analyses were conducted to shed light on the factorial structure of the questionnaire. All analyses in this part were conducted through the SPSS software. Prior to the main EFA analyses, the suitability of the data for EFA was checked. In order for the correlation (or covariance) matrix to be factorable, the Bartlett's test should be significant and the KMO index should be at least 0.60 (Pallant, 2016). After ensuring that the data matrix was appropriate for EFA analysis, several Principal Axis Factoring (PAF) were performed for the extraction of the factors (see T. A. Brown, 2015). The

decision as to the number of factors to be retained was based on theoretical justification, inspection of the scree plot, and justifiability of the factor loadings. The Promax technique was also used in order to rotate the extracted factors. Promax (which is one of the oblique rotation techniques) was selected because the factors measured by the questionnaire are assumed to be correlated (see T. A. Brown, 2015). Since EFA is not capable of handling nested factor solutions, separate EFA analyses were run for each of the four main factors.

Developing the research topic. The EFA analyses for the *Developing the Research Topic* factor revealed that four factors could explain 33 percent of the variance. The factor loadings are displayed in Table 2. It appears that each item loads on only one factor and each factor is measured by a number of highly loading items.

Table 2
Factor Loadings for Developing the Research Topic

	Factor			
	1	2	3	4
A11	.830			
A12	.447			
A10	.293			
A5		.657		
A4		.624		
A6			.532	
A7			.415	
A2				.543
A1				.293

Design knowledge. In order to extract the factors, PAF was applied several times and the factors were rotated through the Promax technique. The results indicated that the initial four-factor model is not logical and the first two factors must be merged. Hence, a three-factor model emerged that could explain 48 percent of the variance in the data. The factor loadings are displayed in Table 3. All loadings are high and all items load on the expected factors.

Table 3
Factor Loadings for Design Knowledge

	Factor		
	1	2	3
A21	.889		
A22	.774		
A14	.760		
A19	.757		
A17	.671		
A18	.623		
A13	.544		
A15	.535		
A20	.511		
A23	.456		
A16	.429		
A40		.706	
A41		.697	
A39		.543	

A42	.475	
A44	.421	
A32		.695
A25		.450
A26		.424
A34		.379
A31		.373

Procedural knowledge. PAF along with Promax rotation revealed that two factors could best explain the factorial structure of the data. The two factors explained over 37 percent of the variance in the data. The factor loadings are displayed in Table 4; as reported, it is clear that items load on their relevant factors.

Table 4

Factor Loadings for Procedural Knowledge

	Factor	
	1	2
A57	.793	
A59	.771	
A60	.654	
A58	.636	
A63	.493	
A56	.468	
A55	.460	
A54	.447	
A62	.341	
A46		.683
A47		.672
A45		.650
A48		.585
A49		.479
A50		.401
A51		.322

Data Analysis Knowledge. The EFA analysis revealed that a unidimensional (i.e. one-factor) solution best captured the factorial structure of the data, and the single factor explained about 61 percent of the variance in the data. The factor loadings are displayed in Table 5. Apparently, all loadings are acceptable.

Table 5

Factor Loadings for Data Analysis Knowledge

	Factor
	1
A70	.836
A73	.818
A71	.814

A72	.807
A74	.800
A69	.783
A67	.780
A75	.763
A66	.720
A68	.704
A76	.659
A64	.654
A65	.647
A78	.566
A77	.561

Testing the hypothesized models. After the initial inspection of the factorial structure of the different sections of the questionnaire, Confirmatory Factor Analyses (CFA) were performed. All analyses were done in Mplus 7.11 (Muthén & Muthén, 1998–2010). Since there were missing data, the Robust Maximum Likelihood (MLR) estimation technique was employed for parameter estimation. This technique is robust against violations of normality (Kline, 2015).

The evaluation of the fit of each model was based on an inspection of the plausibility of parameter estimates and the overall fit indices provided in Mplus. In each case, it was made sure that there are no out-of-bound or implausible estimates. Having ensured the plausibility of individual parameter estimates, overall fit indices were examined. The inspection of model fit in Mplus is based on the following indices:

●Model chi-square
●Root Mean Square Error of Approximation (RMSEA)
●Comparative Fit Index (CFI)
●Standardized Root Mean Square Residual (SRMR)

The model chi-square tests the "exact-fit" hypothesis for the model. Since almost all models are expected to diverge from the data to some extent, the chi-square indicates poor fit in almost all cases. T. A. Brown (2015) recommends that RMSEA values should be less than 0.6 to indicate good fit. Usually, a 90% interval is also reported around the RMSEA. The upper bound of this interval should not be higher than 0.8. Moreover, the CFI must be higher than 0.90 and the SRMR should not exceed 0.80.

The fit indices for the four first-order CFA models are presented in Table 6. Based on the overall fit indices presented earlier, it is clear that all four models show adequate fit. Further inspection of the individual parameter estimates also showed that all of them were logical and there were no out of bound estimates. Hence, it may be concluded that the models adequately fit the data.

Table 6
Fit Indices for the First-Order CFA Models

	Chi-square	RMSEA		CFI	SRMR

	Index		90 interval		%	
Developing the Research Topic	25.623	0.018	0.000 - 0.038	0.988	0.025	
Design Knowledge	444.188	0.053	0.047 - 0.058	0.925	0.043	
Procedural Knowledge	205.131	0.055	0.047 - 0.064	0.921	0.052	
Data Analysis Knowledge	104.299	0.065	0.052 0.078	0.937	0.039	

The path diagrams along with the relevant parameter estimates for each CFA model are reported in Figures 1 through 4.

Figure 1
Path diagram for Developing the Research Topic
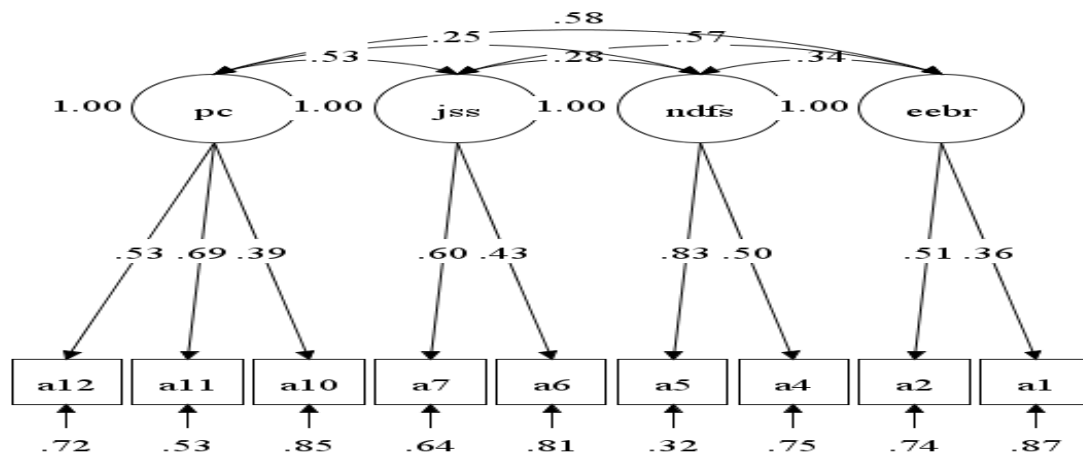
Note. eebr = embeddedness in the existing body of research; jss = justifying the significance of the study; ndfs = narrowing down the focus of the study; pc = practicality concerns.

Figure 2
Path diagram for Design Knowledge
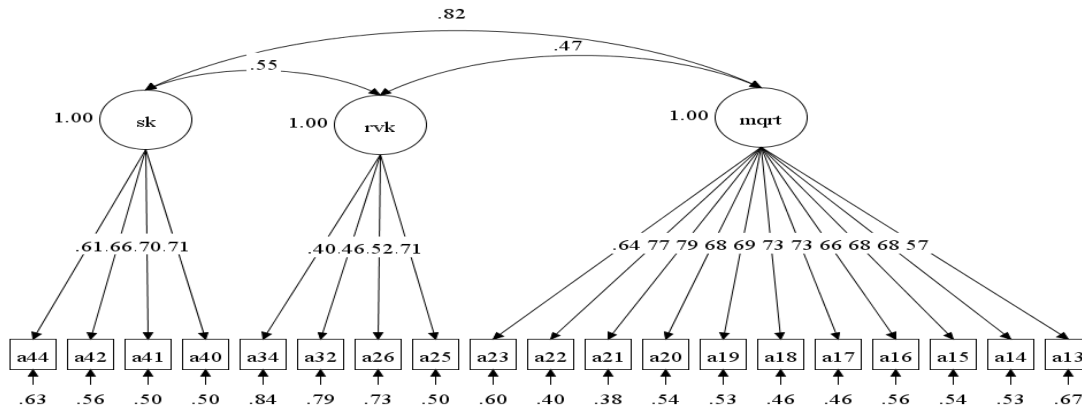
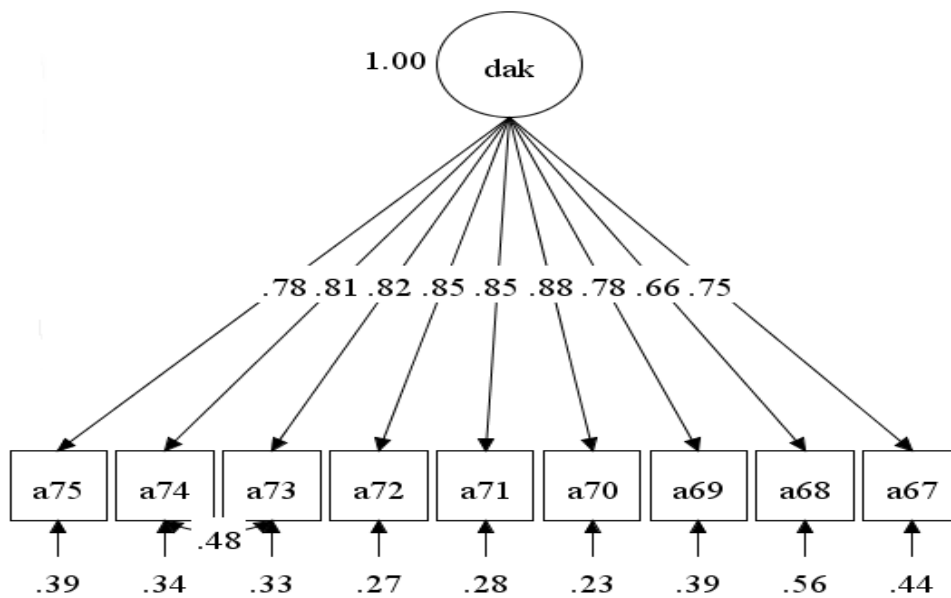*Note.* mqrt = mastery of quantitative research types, variables, and measurement scales; rvk = research validity knowledge; sk = sampling knowledge.

Figure 3
Path diagram for Procedural Knowledge
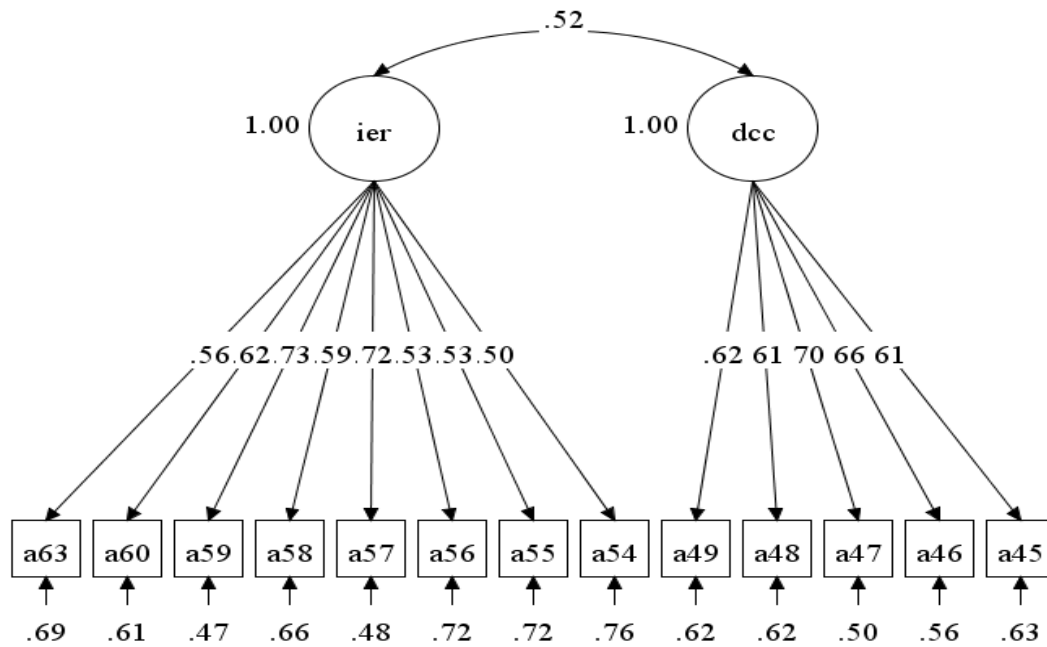
Note. dcc = data collection competence; ier = implementing ethics in research.

Figure 4
Path diagram for Data Analysis Knowledge

Note. dak = data analysis knowledge.

After ensuring that first-order CFA models show adequate fit, they were merged together to get an overall picture of the underlying structure of the entire set of items included in the questionnaire. To this end, two other CFA models were hypothesized. The first model was

correlated-factors model which assumed that the four general factors could be just assumed to be correlated with each other. The second model was a hierarchical CFA model which assumed that a single higher-order factor could explain the correlations among the factors. The fit indices for the two models are presented in Table 7. It appears that the correlated-factors model shows adequate fit. The only problematic index for the hierarchical factor model, however, is the CFI value which is slightly lower than the 0.90 level. All other indices show very good fit for both models. The inspection of the individual parameters and the relevant standard errors also revealed no out-of-bound estimates. This provides further support for the fit of the models.

Table 7
Fit Indices for the Correlated-Factors and Hierarchical CFA Models

|	Chi-square	RMSEA		CFI	SRMR
		Index	90 % interval		
Correlated-factors	2406.771	0.039	0.037 - 0.041	0.90	0.064
Hierarchical	2451.811	0.040	0.047 - 0.058	0.893	0.066

In order to select the model with best fit, the Akaike Information Criterion (AIC), Byesian Information Criterion (BIC), and Sample-Size Adjusted BIC (ABIC) were used for model comparison. Based on these criteria, models with smaller information values/indices show better fit (Wang & Wang, 2012). These indices are displayed in Table 8.

Table 8
Information Indices for the Correlated-Factors and Hierarchical CFA Models

|	AIC	BIC	ABIC
Correlated-factors	118510.840	119288.603	118748.808
Hierarchical	118562.913	119331.526	118798.081

Based on the values presented in Table 8, it is evident that all indices are smaller for the correlated-factors model. Hence, this model seems to show better fit and is the model of choice in this study.

Figure 5
Path diagram for the correlated-factors confirmatory factor analysis model
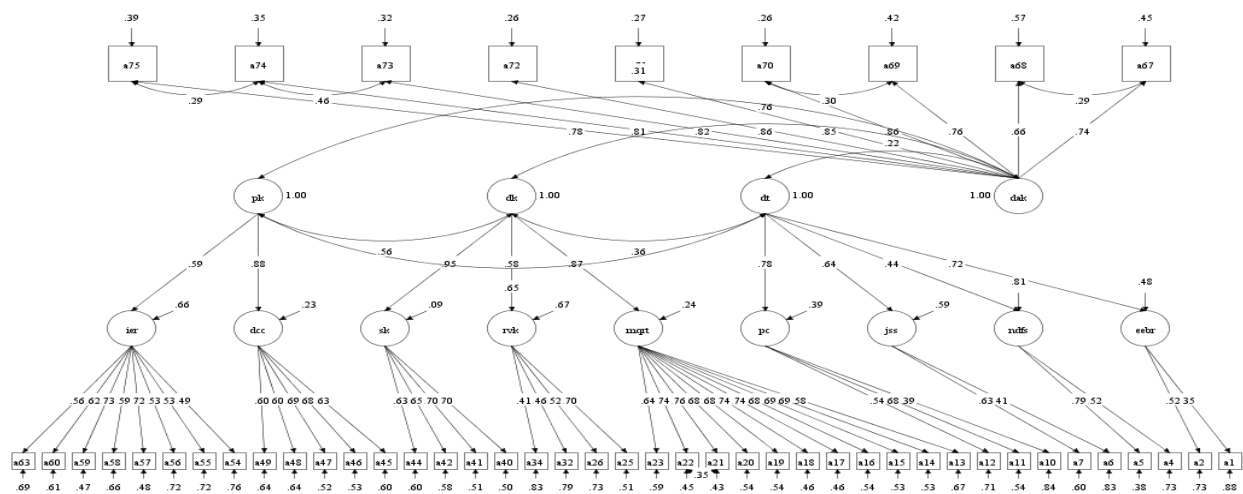
Note. pk = procedural knowledge; dk = design knowledge; dt = developing the research topic; dak = data analysis knowledge; ier = implementing ethics in research; dcc = data collection competence; sk = sampling knowledge; rvk = research validity knowledge; mqrt = mastery of quantitative research types, variables, and measurement scales; pc = practicality concerns; jss = justifying the significance of the study; ndfs = narrowing down the focus of the study; eebr = embeddedness in the existing body of research.

Figure 6

Path diagram for the hierarchical confirmatory factor analysis model
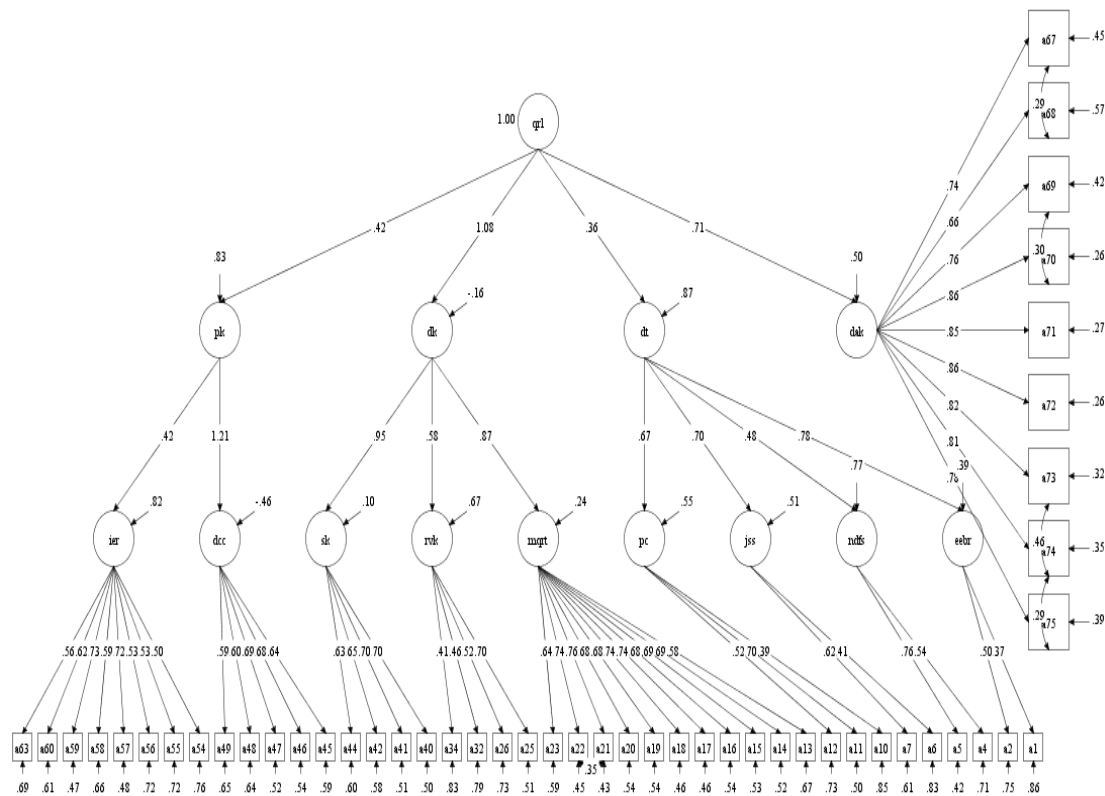
Note. pk = procedural knowledge; dk = design knowledge; dt = developing the research topic; dak = data analysis knowledge; ier = implementing ethics in research; dcc = data collection competence; sk = sampling knowledge; rvk = research validity knowledge; mqrt = mastery of quantitative research types, variables, and measurement scales; pc = practicality concerns; jss = justifying the significance of the study; ndfs = narrowing down the focus of the study; eebr = embeddedness in the existing body of research.

Discussion

The model and instrument development commenced with an extensive review of the pertinent literature, followed by developing the initial model, drafting the items, seeking expert reviews, and revising the items. Taking these steps resulted in proposing the initial model of QRL and an instrument with 78 items. However, conducting factor analysis was needed to provide the requisite support for the validity of the newly-developed instrument through inspecting the internal structure of the instrument and the underlying construct affecting the variance in the scores (Kline, 2015; Tabachnick & Fidell, 2014).

The obtained results of the EFA analyses seemed, to a large extent, to confirm the organization of the item and the appropriateness of the proposed QRL model. This provided a

reasonable support for the validity of the model along with establishing the basis for conducting CFA (Dörnyei, 2010; Kline, 2015). However, a number of items from the initial 78 items did not demonstrate acceptable factor loadings, and 61 items qualified as acceptable items of QRL model. Following conducting EFA and the initial inspection of the factorial structure of the different sections of the questionnaire, CFAs were performed.

The fit indices for the four first-order CFA models indicated that all four models, representing the four factors of QRL, show adequate fit. Hence, it was concluded that the models adequately fit the data. However, from the 61 items, 50 items seemed to demonstrate acceptable qualities and remain in the model. Following this, after ensuring that first-order CFA models show adequate fit, they were merged together to get an overall picture of the underlying structure of the entire set of items included in the questionnaire. To this end, two other CFA models were hypothesized, i.e. correlated-factors model and hierarchical CFA model. In order to select the model with best fit, the AIC, BIC, and ABIC were used for model comparison. This resulted in the correlated-factors model qualifying as the model of choice in this study.

As observed in the results of the EFA and CFA procedures, the final and confirmed model of QRL (Table 9) closely corresponded with the initial QRL models, prior and after expert reviews. This goes to show that the item and model development phase at the early stages of the study has appropriately covered and categorized the elements of QRL available in the literature (Dörnyei, 2010). The obtained results indicated that QRL, as a multi-dimensional construct (Breen & Darlaston-Jones, 2010), can be legitimately estimated through a four-factor model, including *Developing the Research Topic, Design Knowledge, Procedural Knowledge, and Data Analysis Knowledge.*

Table 9
The Validated Model of Quantitative Research Literacy Including the Factors and Sub-Factors

Factor 1: Developing the Research Topic
Embeddedness in the Existing Body of Research
Narrowing Down the Focus of the Study
Justifying the Significance of the Study
Practicality Concerns
Factor 2: Design Knowledge
Mastery of Quantitative Research Types, Variables, and Measurement Scales
Research Validity Knowledge
Sampling Knowledge
Factor 3: Procedural Knowledge
Data Collection Competence
Implementing Ethics in Research
Factor 4: Data Analysis Knowledge (unidimensional)

Conclusion

Being rooted in positivism, valuing objectivity, possessing better generalizability, and being criterion-oriented (Breen & Darlaston-Jones, 2010) has resulted in the prioritization of "quantitative research … over other forms of enquiry" (Breen & Darlaston-Jones, 2010, p. 67). However, prior to this study, there were no systematic attempts to develop a model and instrument for estimating ELT practitioners' QRL. Consequently, the main and major objective

of this study was the development of QRL model and instrument. Furthermore, the researchers attempted to determine the underlying constructs of QRL and the factorial structure of QRL.

Results obtained from EFA and CFA provide the requisite support for the validity of the newly-developed instrument. As a result, considering all of the steps taken in developing the final 50 items, it seems sensible to argue that the QRL instrument with 4 factors, 9 sub-factors, and 50 items can function as a valid instrument for estimating the degree of QRL (Dörnyei, 2010; Kline, 2015). One important point to be made is in the absence of any other QRL model and instrument, it was impossible to check the concurrent validity of the instrument with any similar measure.

It is not unbeknownst to ELT researchers that participants' internal factors, which are highly diverse and influential (Zaker, 2016), along with other features of the context and participants can influence the qualities of the data and findings of studies of this nature (Best & Kahn, 2006; Tabachnick & Fidell, 2014); this suggests that the aforementioned results should be checked and confirmed in other ELT contexts. Therefore, it is sensible to argue that the real test of the QRL model and instrument validity and relevance would be the way it functions in the empirical studies which attempt to validate its structure further, a measure we strongly recommend.

Note
You are welcome to request the questionnaire through correspondence with the corresponding author.

References
Akbari, R. (2008). Postmethod discourse and practice. *TESOL Quarterly, 42*(4), 641-652. doi:10.1002/j.1545-7249.2008.tb00152.x

American Educational Research Association. (1992). Ethical standards of the AERA. *Educational Researcher, 21*(7), 23–26.

Ary, D., Jacobs, L. C., & Sorensen, C. (2010). *Introduction to research in education* (8th ed.). Boston: Cengage Learning.

Ary, D., Jacobs, L. C., Sorensen Irvine, C. K., & Walker, D. (2019). *Introduction to research in education* (10th ed.). Boston: Cengage Learning.

Ashton-Hay, S. (2006). *Constructivism and powerful learning environments: Create your own!* Paper presented at the the 9th International English Language Teaching Convention: The Fusion of Theory and Practice, Middle Eastern Technical University, Ankara, Turkey. http://eprints.qut.edu.au/17285/1/17285.pdf

Bell, D. M. (2003). Method and Postmethod: Are They Really So Incompatible? *TESOL Quarterly, 37*(2), 325-336. doi:10.2307/3588507.

Best, J. W., & Kahn, J. V. (2006). *Research in education* (10th ed.). Boston, MA: Pearson Education, Inc.

Birjandi, P., & Siyyari, M. (2010). Self-assessment and peer-assessment: A comparative study of their effect on writing performance and rating accuracy. *Iranian Journal of Applied Linguistics, 13*(1), 23-45.

Blessinger, P. (2015). Educational research methodology framework. Retrieved from Patrick Blessinger website: http://www.patrickblessinger.com/ermf/.

Breen, L., & Darlaston-Jones, D. (2010). Moving beyond the enduring dominance of positivism in psychological research: Implications for psychology in Australia. *Australian Psychologist, 45*(1), 67-76. doi:10.1080/00050060903127481

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

Brown, T. A. (2015). *Confirmatory factor analysis for applied research, 2nd ed*. New York, NY, US: Guilford Press.

Burns, A. (2010). *Doing action research in English language teaching: A guide for practitioners*. New York: Routledge.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). New York: Routledge.

Creswell, J. W. (2010). *Educational research* (4th ed.). Boston: Pearson.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. Los Angeles: SAGE Publication.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests *Psychological Bulletin, 52*(4), 281-302. doi:10.1037/h0040957.

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

Ellis, R. (2010). A principled approach to incorporating second language acquisition research into a teacher education programme. *Reflections in Language Teaching, 9*(1), 11-23.

Fahim, M., & Zaker, A. (2014). EFL learners' creativity and critical thinking: Are they associated? *Humanising Language Teaching, 16*(3).

Farrell, T. S. C. (2012). Reflective practice as professional development. In C. Coombe, L. England, & J. Schmidt (Eds.), *Reigniting, retooling and retiring in English language teaching* (pp. 23-32). East Lancing: University of Michigan Press.

Jay, J. K., & Johnson, K. L. (2002). Capturing complexity: a typology of reflective practice for teacher education. *Teaching and Teacher Education, 18*(1), 73-85. doi:10.1016/s0742-051x(01)00051-8

Kline, R. (2015). *Principles and practice of structural equation modeling* (4th ed.). London: Guildford.

Krosnick, J. A., & Presser, S. (2010). Question and questionnaire design. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of survey research* (2nd ed., pp. 263-313). Bingley: Emerald.

Kumaravadivelu, B. (2012). *Language teacher education for a global society: A modular model for knowing, analyzing, recognizing, doing, and seeing*. New York: Routledge.

Lightbown, P. M., & Spada, N. (2013). *How languages are learned* (4th ed.). Oxford: Oxford University Press.

Mackey, A., & Gass, S. M. (2015). *Second language research: Methodology and design* (2nd ed.): Routledge.

Mitchell, R., Myles, F., & Marsden, E. (2013). *Second language learning theories*. New York: Routledge.

Nation, I. S. P., & Macalister, J. (2010). *Language curriculum design*. New York: Routledge.

Nosratinia, M., & Zaker, A. (2014). Metacognitive attributes and liberated progress:The association among second language learners' critical thinking, creativity, and autonomy. *SAGE Open, 4*(3). doi:10.1177/2158244014547178

Nosratinia, M., & Zaker, A. (2017). Scrutinizing the impact of teachers' critical thinking and teaching autonomy on their teaching success and learners' use of language learning strategies. *Journal of Language Teaching and Research, 8*(1), 122-132. doi:10.17507/jltr.0801.15

Pallant, J. (2016). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (6th ed.). London: Open University Press.

Richards, J. C., & Lockhart, C. (1994). *Reflective teaching in second language classrooms*. Cambridge: Cambridge University Press.

Sabbaghan, S. (2010). Iranian TEFL students' perceptions of plagiarism. *Journal of English Language Studies, 1*(4), 23-40.

Springer, K. (2010). *Educational research: A contextual approach*. NJ: Wiley.

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* (6th ed.). Boston: Pearson Education.

Wang, J., & Wang, X. (2012). *Structural equation modeling: applications using Mplus*. New York: John Wiley.

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing, 27*(3), 236-247. doi:10.1016/j.ijresmar.2010.02.004

Zaker, A. (2016). Social constructivism and metacognition in an EFL context: Inspecting the contribution of critical thinking to EFL learners' social intelligence. *Humanising Language Teaching, 18*(6). Retrieved from www.hltmag.co.uk/dec16/index.htm

Zaker, A., Nosratinia, M., Birjandi, P., & Yazdanimoghaddam, M. (2019). The impact of implementing critical appraisal on EFL teachers' quantitative research literacy. *Journal of Language and Translation, 9*(2), 141-155.

Zeichner, K. (1999). The new scholarship in teacher education. *Educational Researcher, 28*(9), 4-15. doi:10.3102/0013189X028009004.

Acknowledgements