The Impact of Implementing Critical Appraisal on EFL Teachers' Data Analysis Knowledge

Alireza Zaker, Ph.D. Candidate, English Department, Science and Research Branch, Islamic Azad University, Tehran, Iran
a.zaker@srbiau.ac.ir
Mania Nosratinia*, Assistant Professor, English Department, Central Tehran Branch, Islamic Azad University, Tehran, Iran
mania_nosratinia@yahoo.com
Parviz Birjandi, Professor, English Department, Science and Research Branch, Islamic Azad University, Tehran, Iran
pbirjand@srbiau.ac.ir
Massood Yazdani Moghaddam, Assistant Professor, English Department, Garmsar Branch, Islamic Azad University, Garmsar, Iran
mym1300@gmail.com

Abstract
Condemning a laissez faire approach to English Language Teaching (ELT), English as a Foreign Language (EFL) teacher trainers unanimously agree that the building blocks of teachers' teaching framework are profoundly influenced by conducting research. Focusing on quantitative research, this study endeavored to scrutinize the impact of the Critical Appraisal of Published Research (CAPR) in undergraduate teacher training programs on EFL teachers' Data Analysis Knowledge (DAK). To this objective, 30 male and female EFL teachers were non-randomly selected and randomly assigned to two groups. In two Research classes, the experimental group received the CAPR whereas the control group received traditional teacher-centered instruction with summative assessment. The DAK section of the Quantitative Research Literacy (QRL) questionnaire was employed as the pretest and posttest. Subsequent to corroborating participants' pre-treatment homogeneity in terms of DAK, analyzing the post-treatment data through running an independent-samples *t*-test, *eta squared* = .338 (representing a large effect size), indicated the existence of a significant difference in the post-treatment DAK scores between the two groups. The obtained results confirmed that the CAPR has a significantly better impact on EFL teachers' DAK which is a key area of QRL. Therefore, it seems accurate to argue that ELT teacher training programs should endeavor to involve the students in a mentally engaging process, e.g. CAPR, where the content of the course is put into practice by the students, something which is required for balancing the concrete and the abstract.

Keywords: Critical appraisal, data analysis knowledge, quantitative research, research literacy, teacher training

Introduction

Prevailing Teaching English as a Foreign Language (TEFL) trends are deeply rooted in the premise that English as a Foreign Language (EFL) instructors and teachers perform as reflective and transformative practitioners who explore and investigate classroom events and outcomes (Kumaravadivelu, 2012; Lightbown & Spada, 2013). As postulated by Richards and Lockhart (1994), teachers ought to "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). Consequently, the reflective inspection of the teaching practice is regarded as a key asset to EFL teachers (Farrell, 2012).Happening through the systematic analysis of the classroom events and pedagogical practices, this careful inspection enables teachers to behave more sensibly, reasonably, and justifiably (Farrell, 2012; Lightbown & Spada, 2013).

The capacity to engage in the critical and systematic analysis of the teaching practice, which is mainly developed through receiving training, gaining experience, and self-discovery, can substantially influence EFL teachers' perception of the teaching techniques and practices (Borg, 2008; Freeman, 2002). Furthermore, this critical and self-involved process would make it possible for ELT practitioners to detach themselves from the limits of subjectivity and ''discover meaning [they] might otherwise miss'' (Jay & Johnson, 2002, p. 78). According to Springer (2010), conducting research signifies acquiring knowledge "through deduction, induction, and the application of the scientific method… informed by key assumptions, including empiricism, conditionality, precision, parsimony, objectivity, and theoretical motivation" (p. 26). In actual fact, conducting systematic "Applied Research" (Best & Kahn, 2006, p. 21), which is by definition practical, problem-oriented, objective, and contextual (Creswell, 2014), can be considered one of the significant attempts to reflectively evaluate the pedagogical practice (Springer, 2010).

Applied research is primarily undertaken by dint of quantitative and qualitative routes (Best & Kahn, 2006; Creswell, 2014); however, the former, i.e. quantitative research, is believed to be more practical and appropriate when it comes to addressing pedagogical concerns and issues (Hadi & Closs, 2016). The knowledge of data analysis is the fourth factor of literacy in quantitative research (Zaker, Nosratinia, Birjandi, & Yazdanimoghaddam, 2019), functioning as the key element of quantitative research. This knowledge deals with analyzing the quantitative data and checking the hypotheses (Best & Kahn, 2006). However, previous research has reported that knowledge of data analysis is the weakest area of literacy in quantitative research among EFL teachers across different educational levels (Zaker, 2019). Hoping to address this shortcoming and considering the peculiarities of ELT teacher training programs, more specifically Research courses, this study endeavored to scrutinize the way implementing the Critical Appraisal of Published Research (CAPR) affects EFL teachers' knowledge of data analysis. In order to meet the objective of the present study, the subsequent research question was formulated:

Q: Is there any statistically significant difference between the impact of the critical appraisal of published research and traditional teacher-centered instruction on EFL teachers' data analysis knowledge?

Literature Review

Quantitative vs. Qualitative

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 basically made possible through taking two major routes, quantitative and qualitative (Best & Kahn, 2006; Creswell, 2014). Reflecting upon the implementation of applied research in the ELT domain, both quantitative and qualitative approaches seem to have their own pros and cons; 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).Consequently, 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).

As stated earlier, there is a general consent that undertaking a systematic, context-based, and well-designed applied ELT research, either quantitative or qualitative, can make a significant contribution to ELT practitioners' existing understanding of ELT and the development of pedagogical techniques (Farrell, 2012; Jay & Johnson, 2002).However, we have recently witnessed a surge of growth regarding the pedagogical practice, curriculum development, assessment, learner variables, and teacher education (Akbari, 2008; Bell, 2003; Ellis, 2010; Lightbown & Spada, 2013; Mitchell, Myles, & Marsden, 2013; Nation & Macalister, 2010). Accordingly, and quite reasonably, there has been a simultaneous emergence of new concerns and techniques in carrying out applied research in ELT contexts (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).

Data Analysis Knowledge

Coined by Birjandi (P. Birjandi, personal communication, January 1, 2016), the term Quantitative Research Literacy (QRL) is legitimately used to address researchers' (including ELT practitioners') knowledge of conducting systematic quantitative research. In fact, QRL can determine the validity of research findings and the accuracy of the path taken for achieving ELT pedagogical goals. The knowledge of data analysis (aka statistical knowledge) is the fourth factor of QRL (Zaker et al., 2019) which is believed to function as the key element of quantitative research; other QRL factors are: Developing Research Topic Knowledge, Research Design Knowledge, and Procedural Knowledge. Data Analysis Knowledge (DAK) deals with analyzing the quantitative data and checking the hypotheses (Best & Kahn, 2006). More specifically, DAK deals with a "body of mathematical techniques or processes for gathering, organizing, analyzing, and interpreting numerical data" (Best & Kahn, 2006, p. 354). Best and Kahn further add that DAK "is a basic tool of measurement, evaluation, and research" (p. 354).

It is believed that DAK 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). However, previous research has reported that DAK is the weakest area of QRL among EFL teachers across different educational levels (Zaker, 2019). In a similar vein, Creswell (2012) postulated that DAK is where researchers in the realm of education and language teaching are faced with a considerable challenge. Apropos of these findings, and knowing that DAK is a key factor of QRL (Cohen, Manion, & Morrison, 2011), it sounds reasonable to argue that EFL teacher trainers and curriculum developers should pay due attention to working on this neglected area of QRL which can function as an enabling tool for EFL teachers in conducting and understanding ELT research (Ary et al., 2019).

Critical Appraisal

As DAK is mainly gained in academic contexts, e.g. undergraduate and graduate courses, exploring the effectiveness of different instructional procedures for developing EFL teachers' QRL and its four main aspects/factors (Zaker et al., 2019) seems to be quite essential and legitimate. Therefore, considering the peculiarities and possibilities of ELT teacher training

programs, more specifically Research courses, this study endeavored to scrutinize the way implementing the CAPR affects EFL teachers' DAK. The term *Critical Appraisal* has been defined as "a systematic process used to identify the strengths and weaknesses of a research article in order to assess the usefulness and validity of research findings" (Young & Solomon, 2009, p. 82). In this process, it is of paramount importance to assess the suitability of the study design, assess the methodological features, and consider the appropriateness of the statistical methods and their subsequent interpretation (Young & Solomon, 2009).

Implementing the multidimensional CAPR seems highly prone to bias and subjectivity without employing a framework or criterion. In fact "a structured approach to critical appraisal could potentially improve the quality of this process" (Young & Solomon, 2009, p. 83). Among the instruments and guidelines for implementing the CAPR, the guideline provided by Best and Kahn (2006) seems to be one of the highly practical and concise frameworks for guiding the CAPR. This CAPR guideline is presented in Table 1.

Table 1. *Guideline for Implementing the Critical Appraisal of Published Research*

Element of Articles	Expected Features of the Element
Title	▪ clear and concise
Problem and Hypotheses	▪ clearly stated ▪ significance of problem ▪ specific question raised ▪ clear statement of hypothesis or ▪ research question ▪ testable hypothesis ▪ assumptions stated ▪ important terms defined
Review of Literature	▪ adequate coverage ▪ well organized ▪ important findings noted ▪ studies critically examined ▪ related to problem and hypothesis
Procedures	▪ subjects and methodology described in detail ▪ adequate sample

	▪ appropriate design ▪ variables controlled ▪ appropriate data gathering instruments
Data Analysis/Results	▪ effective use of tables ▪ effective use of figures ▪ concise but complete report of findings ▪ appropriate statistical or other treatment of data ▪ logical analysis
Discussion/Conclusions	▪ problem restated or addressed ▪ hypotheses restated or addressed ▪ clear and concise ▪ conclusions based on results ▪ statement of practical or theoretical implications ▪ appropriate generalizations
Overall Form and Style of Paper	▪ clear and concise ▪ appropriate degree of objectivity ▪ all parts of the paper are properly related to each other ▪ Referencing according to appropriate style

Note. Adapted from *Research in Education* (p. 490), by J. W. Best and J. V. Kahn, 2006, Boston: Pearson. Copyright 2006 by Pearson Education Inc.

Methodology

Participants

In this study, the participants were initially chosen from a sample pool of 52 senior undergraduate students who majored in ELT at Islamic Azad University, South Tehran Branch, on the basis of having experience in ELT. The selected EFL teachers (n = 30; 21 or 70% females; 9 or 30% males) were within the age range of 21 to 33 (M_{age} = 25), and their teaching experience ranged from 6 months to 7 years ($M_{experience}$ = 17 months). The non-participating senior undergraduate students (n = 22) also attended the classes and were exposed to the treatments; however, the data they provided were not used in answering the research question. Therefore, the

participants of the study were selected through implementing the convenience and purposive sampling strategies as only intact groups were initially available (n = 52) from which only EFL teachers were purposefully selected (n = 30). These intact groups were randomly assigned to two groups, i.e. experimental and control. Implementing the principles of ethics in research, the participants exercised the freedom to withdraw from the study at any point in time (Ary et al., 2019 ; due to this, the sample pool shrank to 23 individuals (17 or 74% females; 6 or 26% males) at the post-treatment phase.

Instrumentation

For attaining the objectives of this study, the QRL scale along with the pertinent instructional material were employed whose details and characteristics are explained in what follows.

The quantitative research literacy scale. The QRL scale, developed and validated by (Zaker et al., 2019), is a 50-item Likert-type questionnaire which estimates QRL through addressing four factors and nine sub-factors; these factors, or main areas of QRL, are Developing Research Topic Knowledge (9 items), Research Design Knowledge (19 items), Procedural Knowledge (13 items), and finally Data Analysis Knowledge (9 items).

The participants are expected to respond on the basis of a six-point Likert scale which ranges from *strongly disagree* (1) to *strongly agree* (6), and the allocated time for providing the responses is 25 minutes. The total obtained scores could range from 50 to 300. In the validation study, Zaker et al. (2019) report numerous measures taken for supporting the reliability and validity of this instrument; some of these measures are model development, expert review, initial piloting, revision, administering the instrument, conducting Exploratory Factor Analysis, and finally conducting Confirmatory Factor Analysis, using the MPlus software. In the present study, the data collected on the DAK section of the QRL instrument were analyzed and explored. The calculated reliability/internal consistency index for the DAK section of the QRL scale in this study was estimated to be 0.9 using the Cronbach's alpha coefficient.

The instructional material. The main employed textbook during the instruction in both of the groups was Research Methods in Applied Linguistics 1 & 2 by Farhady (2009), published by Payame Noor University. This book covers almost all of the areas of QRL, including the four factors addressed in the QRL scale. More specifically chapters 1, 2, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, and 15 were used in the instructional process. As a reference, Research in Education by Best and Kahn (2006), published by Pearson was introduced to the participants. Besides, Publication Manual of the American Psychological Association (6th ed.), aka APA, by American Psychological Association (2010) was used as a supplementary material to the main textbook; however, chapter 1 of this book which generally deals with research types, ethical standards, and crediting sources was used as a standard source of the instructed Research courses.

Procedure

In order to conduct this study and fulfill its pertinent objectives, a number of steps were taken which are explained in this section. To begin with, a formal approval from the officials of the Islamic Azad University, South Tehran Branch, was obtained in order to carry out this study. The experiment began with introducing the researchers, research goals, and main characteristics of the study to the participants who had enrolled in the Research course and were chosen based on convenience sampling strategy. The participants were informed that their participation would be purely voluntary, and they were given the right to withdraw from participation at any point in

time, for any reason, and without penalty. Moreover, they were apprised of the fact that the information they supply would be treated as confidential. The two intact groups were randomly assigned to two groups of experimental (n = 15) and control (n =15).

Prior to commencing the instruction and delivering the treatments, the QRL scale was administered to the participants in the two groups. The participants were provided with thorough explanation about the instrument and the answering procedure; they were also informed that the answers they provide would not exert any effect on their course scores and how they would be treated. They were given 25 minutes to provide the answers while the researchers were present at the time of administration for resolving the problems. The scores participants obtained at this point were used as the pretest scores and enabled the researchers to inspect the pre-treatment homogeneity of the participants through conducting pertinent analyses (see Results). The treatment phase lasted for 15 sessions of 90 minutes during 15 weeks, and both of the groups received the same method of instruction in Research Methodology; moreover, the same instructional material was employed for both of the groups. At the sixteenth/last session, the QRL scale was re-administered to the participants as posttest. The same procedure as the pretest administration (stated above) was followed for the posttest administration. However, the two classes differed over what happened in the last twenty minutes of the class. This difference was based on the type of treatment in each group which functioned as the independent variable in this study. The instructional procedures in the two groups are explained in the following sections.

The experimental group. The main objective of the researchers in the experimental group was to engage the participants in the CAPR which would enable them to put what they had just learned in the class into test. After taking the pretest, the instructor (one of the researchers) continued the first session with working on the textbook of the course which lasted for 45 minutes (25 + 45 = 70). Thence, during the remaining 20 minutes of the first session, the participants were introduced to CAPR, its definition, and main purpose. Following this, the participants were provided with the guideline by Best and Kahn (2006) for implementing the CAPR (see Table 1). In addition, the instructor supplied each participant with two quantitative research papers, one descriptive/correlational and one experimental research. The participants were asked to bring the guideline and papers each session for the CAPR practice.

The order of instruction topics in the experimental group agreed with the order of topics presented in the CAPR guideline (Best & Kahn, 2006), similar to the control group. From this point onward, the last twenty minutes of each session was dedicated to CAPR after a 5-minute break. Following the break, in each CAPR session the participants were asked to categorize the topic of that specific class session into one of the seven main areas presented in the CAPR guideline. Thence, they were asked to find the location of this topic in both of the articles. Finally, the participants were asked to critically and independently evaluate each paper regarding the specific topic of each class session and highlight the strengths and weaknesses. Considering these strengths and weaknesses, the participants were expected to grade each paper regarding that session's specific point on a scale from 0 to 20. As the term project, the participants were asked to provide two evaluation reports for two new articles, one descriptive and one experimental, and discuss the strengths and weaknesses of each paper in seven categories (stated in the CAPR guideline).

The control group. As stated in the introductory paragraphs of the Procedure section, both of the groups shared the instructional procedure, instructional materials, and instruction length in each session of the class. However, it was explained that the two groups differed in what happened in the last 20 minutes of the class. In the control group, after administering the

pretest (the QRL scale), the first session of the class continued with working on the textbook of the course, lasting for 45 minutes. At this point, the participants took a five-minute break which was followed by reviewing the instructed points in the first session. The instructor asked a number of questions pertinent to the instructed points, and volunteer students were given the chance to answer the questions. As stated earlier, the order of instruction topics in both of the groups agreed with the order of topics presented in the CAPR guideline (Best & Kahn, 2006). However, in the control group, the participants were not provided with the CAPR guideline. This, in fact, had to do with the notion that the absence of the CAPR was the intended treatment in this group.

Similar to what took place at the end of the first class session, in all of the remaining fourteen sessions the last twenty minutes of the class was dedicated to answering the questions asked by the instructor and reviewing the points covered in each session. As the term project, the EFL teachers in the control group were asked to come up with a research topic and prepare a research proposal based on the topic.

Results

In this quasi-experimental quantitative study with a pretest-posttest control group design, DAK was considered the dependent variable whereas the independent variable was the treatment type with two levels (the CAPR and the absence of the CAPR). In simple terms, the impacts of these treatments were to be checked on participants' DAK. Further, participants' gender and age were regarded as the intervening variables. For answering the research question, a number of apposite calculations and statistical routines were conducted; in this section these analyses and the obtained results are presented.

Pre-Treatment Homogeneity of the Participants

In order to verify participants' pre-treatment homogeneity in terms of DAK, it was essential to inspect the pretest scores performing pertinent statistical routines. To do so, an independent-samples *t*-test was to be conducted. However, as a major prerequisite for running this parametric test, it was essential to ensure the normality of the data. For this purpose, the Kolmogorov-Smirnov and Shapiro-Wilk tests were run (Table 2).

Table 2. *Test of Normality on the Pre-Treatment Data*

	Group	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
DAK1	Experimental	.319	12	**.001**	.774	12	**.005**
	Control	.255	11	**.043**	.811	11	**.013**

Note. DAK1 = pre-treatment data analysis knowledge.
[a]Lilliefors Significance Correction. *This is a lower bound of the true significance.

As reported in Table 2, the Sig. values of the Shapiro-Wilk test (which is more appropriate in this case regarding the number of participants) for the scores were below the critical value (.05), indicating the absence of normality of distribution (Tabachnick & Fidell, 2013). Accordingly, it was no longer legitimate to run a parametric test for inspecting the pre-treatment data, and the Mann-Whitney U test, which is the non-parametric alternative to the

independent-samples *t*-test, was employed for this purpose. Table 3 through Table 5 report the results.

Table 3. *Table of Ranks for the Pre-Treatment Scores*

	Group	N	Mean Rank	Sum of Ranks
DAK1	Experimental	12	11.88	142.50
	Control	11	12.14	133.50
	Total	23		

Note. DAK1 = pre-treatment data analysis knowledge.

Table 4. *The Pre-Treatment Median Scores for the Groups on Data Analysis Knowledge*

DAK1		
Group	N	Median
Experimental	12	11.00
Control	11	12.00
Total	23	11.00

Note. DAK1 = pre-treatment data analysis knowledge.

Table 5. *Mann-Whitney U Test on Pre-Treatment Scores*

	DAK1
Mann-Whitney U	64.500
Wilcoxon W	142.500
Z	-.095
Asymp. Sig. (2-tailed)	.924
Exact Sig. [2*(1-tailed Sig.)]	.928[a]

Note. DAK1 = pre-treatment data analysis knowledge. Grouping Variable: Group.
[a]Not corrected for ties.

As presented in Table 5, the Mann-Whitney U test revealed an insignificant difference in pre-treatment DAK levels of control (*Md* = 12 , *n* = 11) and experimental (*Md* = 11, *n* = 12) groups, *U* = 64.5, *z* = -.095, *p* = .924, *r* = -.0198 (representing a very small effect size). As a result, it was concluded that the two groups were homogeneous in terms of pre-treatment DAK. This made it possible to attribute any observed difference in the post-treatment behavior of the participants to the treatments they received. The post-treatment scores are explored and compared in the following sections.

Answering the Research Question

As stated earlier, the participants of the study were exposed to two different treatments in the research courses, and the research question concerned the comparison of the participants in the experimental and control groups in terms of DAK. In order to answer this question, the data were to be analyzed through running an independent-samples *t*-test. However, as a major

prerequisite for running this parametric test, it was essential to ensure the normality of the data. For this purpose, the Kolmogorov-Smirnov and Shapiro-Wilk tests were run (Table 6).

Table 6. *Test of Normality on the Post-Treatment Data Analysis Knowledge Scores*

	Group	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	Df	Sig.	Statistic	df	Sig.
DAK2	Experimental	.116	12	**.200***	.949	12	**.620**
	Control	.231	11	**.106**	.919	11	**.313**

Note. Sig. values in bold indicating the existence of normality. DAK2 = post-treatment data analysis knowledge.
[a]Lilliefors Significance Correction. *This is a lower bound of the true significance.

As presented in Table 6, the Sig. values of the Shapiro-Wilk test for the DAK scores in the two groups were above the critical value (.05), indicating the existence of normality of distribution (Tabachnick & Fidell, 2013). Table 7 reports the descriptive statistics of the scores in the two groups. Moreover, the obtained results of the *t*-test are presented in Table 8.

Table 7. *Descriptive Statistics of Post-Treatment Data Analysis Knowledge Scores*

	Group	N	Mean	Std. Deviation	Std. Error Mean
DAK2	Experimental	12	29.75	7.375	2.129
	Control	11	20.45	6.105	1.841

Note. DAK2 = post-treatment data analysis knowledge.

Table 8. *Independent Samples T-Test for Post-Treatment Data Analysis Scores*

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
DAK2	Equal variances assumed	.406	**.531**	3.275	21	.004	9.295	2.838	3.393	15.198
	Equal			3.30	2	.003	9.295	2.814	3.4	15.1

variances not assumed	3	0.805		39	52

Note. DAK2 = post-treatment data analysis knowledge.

As reported in Table 8, the Sig. value for Levene's test was larger than the critical value (.05) indicating the equality of variances in the two groups. Inspecting the results of the *t*-test indicated that there was a significant difference in post-treatment DAK scores for the participants in the experimental group (M = 29.75, SD = 7.375) and the control group (M = 20.45, SD = 6.105), as reported (t (21) = 3.275, p = .004, two-tailed). The magnitude of the differences in the means (mean difference = 9.295, 95% *CI*: 3.39 to 15.2) was large (*eta squared* = .338, indicating a large effect size). In simple terms, this meant that the CAPR had a significantly better impact on EFL teachers' DAK. Figure 1 presents the comparative charts in order to make an overall visual comparison of the pre-treatment and post-treatment scores in the two groups of the experiment.

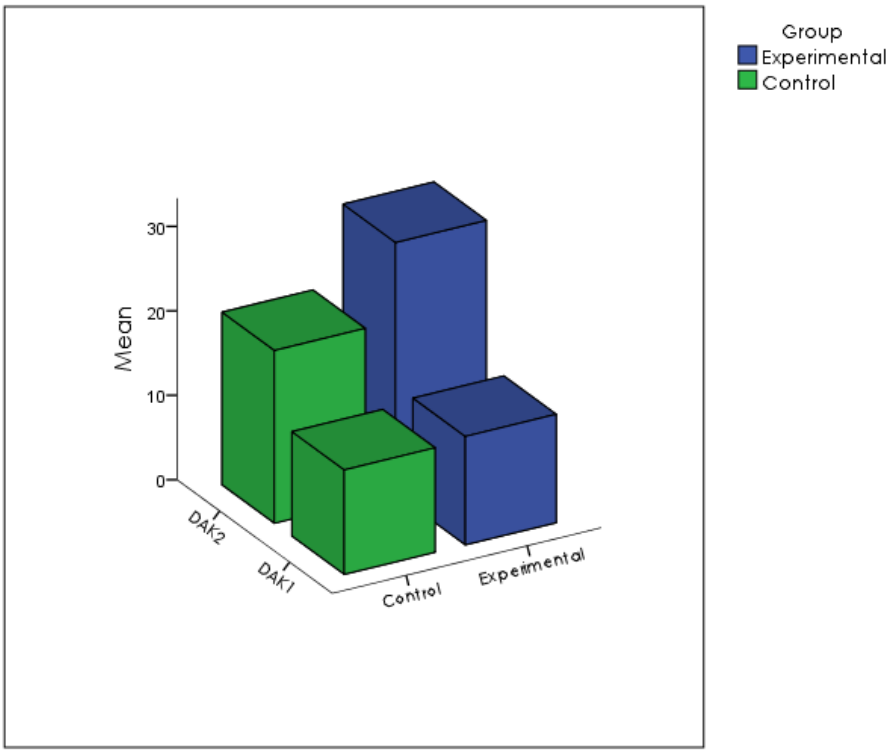

Figure 1. *Comparative three-dimensional chart of pre-treatment and post-treatment data analysis knowledge scores categorized based on the received treatment*

Note. DAK1 = pre-treatment data analysis knowledge; DAK2 = post-treatment data analysis knowledge.

Discussion

Training the ELT practitioners in employing reflection and critical inspection of the pedagogical events could be justified on the grounds of the constructivist paradigm which highlights the significance of self-discovery and personal reasoning when acquiring knowledge in a specific domain (Creswell, 2014). Rooted in this premise, the present study, carried out in a teacher training context, attempted to inspect the impact of implementing CAPR as a reflective and criterion-guided process on EFL teachers' DAK, a key element of QRL.

In this quasi-experimental study, DAK was considered the dependent variable whereas the independent variable was the treatment type with two levels (the CAPR and the absence of the CAPR). The obtained results indicated that implementing CAPR yielded significantly higher levels of DAK among EFL teachers. This finding brings about a systematic support and confirmation for the notion that ELT practitioners' critical and systematic inspection of pedagogical issues, either when engaged in teaching or when learning how to teach, can considerably develop their capacities (Akbari, 2008; Lightbown & Spada, 2013). Moreover, the obtained results make it more sensible to state that if teacher training programs do not focus on actively involving the prospective teachers in the learning process, the outcome will not be favorable (Farrell, 2012). Basically, this finding is also pertinent to the notion that consciously reflecting on a practice or topic can lead to prevention of burnout, maintaining the internal motiving force, and expecting better outcomes (Dewey, 1933).

The desirability of the results notwithstanding, as it is the case with all studies, this study came across some limitations, imposing unavoidable impediments to the interpretation and generalization of the findings (Mackey & Gass, 2015). The first point to mention in this regard is that the findings could not be directly compared to those of other similar studies as no other studies had previously addressed the objectives of this study. Moreover, there were not equal numbers of male and female participants in the sample, making gender a potential confounder. Also, the sample selection was carried out employing the convenience sampling strategy. The last point to make is that the internal and mental qualities of participants, which are highly sundry and significant, coupled with other context features can exert influence on the findings of studies in the ELT domain (Best & Kahn, 2006; Tabachnick & Fidell, 2013). Therefore, considering the abovementioned points, the obtained results should be interpreted with caution (Creswell, 2014).

Conclusion

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). Consequently, critically and systematically reflecting on the practice of teaching is now a key factor in teacher education (Akbari, 2008). Moreover, there is a unanimous consensus among ELT scholars that one of the major responsibilities of teacher training programs is to prepare the ELT teachers to engage in conducting well-organized and systematic research. Implementing CAPR seems to be the offspring of this emerging trend which deals with one specific area of teacher education, i.e. QRL. However, favoring the implementation of CAPR on its own is the reflection of an ambition to boost EFL practitioners' capability in conducting research in language teaching contexts. According to Farrell (2012) and Jay and Johnson (2002), conducting research can make a significant contribution to ELT practitioners' existing understanding of ELT and the development of pedagogical techniques.

Owing to the emergence of new concerns in the ELT domain (Ellis, 2010; Lightbown & Spada, 2013; Mitchell et al., 2013), there has been a simultaneous emergence of new concerns and techniques in carrying out quantitative research in ELT contexts (Mackey & Gass, 2015). As a result, both professional researchers and TEFL students seem to require advanced research skills in order to address different pedagogical issues (Blessinger, 2015). The focus of the present study was limited only to DAK. In spite of the fact that research in human science is a multidimensional domain (Best & Kahn, 2006), quantitative research has qualified as an absolutely indispensable area of research (Hadi & Closs, 2016). This phenomenon has to do with the fact that ELT research is rooted in positivism through highlighting objectivity, generalizability, and being criterion-oriented (Breen & Darlaston-Jones, 2010; Springer, 2010). Considering QRL, it is believed that DAK 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).

Conducted in an undergraduate teacher training program, the obtained results of this quasi-experimental study indicated that the implementation of CAPR yielded significantly higher levels of DAK among EFL teachers. In the wake of this finding, it seems accurate and statistically-supported to argue that teacher training programs should endeavor to involve the students in a mentally engaging process where the content of the course is put into practice by the students. The obtained results provided further support for the premise that actively involving the prospective teachers in the learning process will result in substantial growth in developing their skills (Farrell, 2012).

The implementation of CAPR seems to be a step toward intensifying TEFL practitioners' autonomous practice which is defined by Little (1991) as "[a] capacity – for detachment, critical reflection, decision-making, and independent action" (p. 4). An autonomous ELT practitioner is not dependant on others for the direction and control of their pedagogical decision making (Nosratinia & Zaker, 2017), and CAPR seems to be an invaluable tool for materializing this autonomy. Based on the findings of the present study, EFL teacher trainers in both academic and non-academic contexts are recommended to implement CAPR as a standard part in the training process. Needless to say, this implementation cannot and should not be limited to Research courses. Moreover, it is suggested to consider CAPR tasks and activities one of the key factors in the assessment process.

Due to the fact that DAK is the weakest area of QRL among EFL teachers across different educational levels (Zaker, 2019), EFL students at different academic levels are suggested to engage in the development of their data analysis knowledge through taking statistics courses and self-study. For the latter case, *SPSS Survival Manual: A Step By Step Guide to Data Analysis Using SPSS Program* (6th ed.) by Pallant (2016) is recommended as a reliable source of study. For a higher level, *Using Multivariate Statistics* (6th ed.) by Tabachnick and Fidell (2014) is recommended. Considering the focus, design, and limitations of this study, other researchers are recommended to:

- Replicate this study in graduate levels to inspect if the same results would be obtained;
- Replicate this study in a sample with equal numbers of males and females as the participants, so that the generalizability of the findings is not limited by participants' gender;
- Implement the pure random sampling strategy in order to develop the validity of the findings;

- Replicate this study employing some qualitative instruments and employing triangulation in order to enrich the reliability and validity of the findings and inferences;
- Apply CAPR to other courses in teacher training programs and explore the outcomes; and
- Estimate and inspect the cognitive and mental capacities of the participants simultaneous with estimating their DAK levels.

References

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

American Psychological Association. (2010). *Publication manual of the American Psychological Association*. Washington, DC: American Psychological Association.

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.

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

Borg, S. (2008). *Teacher cognition and language education: Research and practice*. London: Continuum.

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

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. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Boston: Pearson.

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

Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the education process*. Michigan: The University of Michigan.

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.

Farhady, H. (2009). *Research methods in applied linguistics 1 & 2*. Tehran: PNU University Press.

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.

Freeman, D. (2002). The hidden side of the work: Teacher knowledge and learning to teach. A perspective from north American educational research on teacher education in English language teaching. *Language Teaching, 35*(01). doi:10.1017/s0261444801001720

Hadi, M. A., & Closs, S. J. (2016). Applications of mixed-methods methodology in clinical pharmacy research. *International Journal of Clinical Pharmacy, 38*(3), 635-640. doi:10.1007/s11096-015-0231-z

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

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.

Little, D. (1991). *Learner autonomy 1: Definitions, issues and problems*. Dublin: Authentik.

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. (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.

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

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston: Pearson Education.

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

Young, J. M., & Solomon, M. J. (2009). How to critically appraise an article. *Nature Clinical Practice Gastroenterology & Hepatology, 6*(2), 82-91.

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., et al (2019). *Developing and validating a research literacy assessment scale: Investigating the effect of the critical appraisal of published research on EFL teachers' research literacy and reflective teaching* (Doctoral Dissertation), Science and Research Branch, Islamic Azad University, Tehran, Iran. Zaker.