

Relationship between Iranian EFL High School Students' Knowledge of Universal Grammar and their Performance on Standardized General English Proficiency Tests

Mohammad Mahdi Sharifi, Ph.D. Candidate, Department of English, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran

sharifi_m_m@yahoo.com

Ahmadreza Lotfi*, Associate Professor, Department of English, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran

lotfi.ahmadreza@gmail.com

Abstract

This study investigated the relationship between Iranian high school students' Universal Grammar knowledge and their performance on such standardized general English proficiency tests as PET and FCE internationally administered by Cambridge University. To this end, 108 students were randomly chosen from some high schools located in Malayer from Hamedan. Since this study was correlational in nature, and descriptive and hypothesis-testing by definition, the research participants were given no treatment. Three tests were administered to them instead. To measure UG knowledge, a researcher-made UG test was given to all participants. This test which was made both reliable and valid included pied-piping and preposition stranding principle, binding principle, pro-drop parameter, that-trace effect, projection principle, resumptive pronoun and subjacency principle. To assess the participants' general English proficiency, PET and FCE tests were run. All participants took the three tests consecutively at two-week intervals after they were given clear and detailed instructions. The findings were manifold. Firstly, there was a significant relationship between UG knowledge and performance on the proficiency tests. Secondly, there existed no significant difference between the proficiency tests as far as the UG test was concerned. Finally, the UG test scores were, through the Cubic regression model, proved to predict the scores gained on both proficiency tests. Most importantly, this study led to some suggestions regarding the learner variables and the under-explored issue of integration of generative SLA and language testing, more specifically standardized general English proficiency tests.

Keywords: Generative SLA, UG principles and parameters, language testing, language proficiency, learner variable

Introduction

Recent years have witnessed a growing number of standardized English proficiency tests administered by Cambridge University across the world, more specifically in the educational system in Iran as the context of this study. Due to a high amount of importance attached to such Cambridge tests in the current study, an extract from Cambridge English Language Assessment, part of the University of Cambridge (2016) is provided below in Figure 1.1.

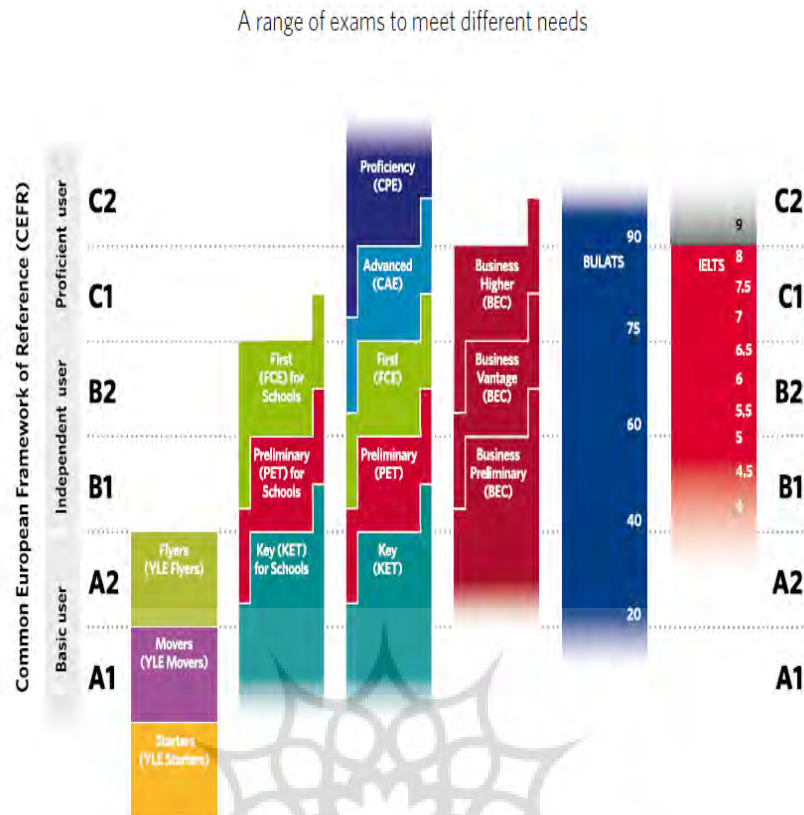


Figure 1. Classification of Standardized Cambridge Tests of English Proficiency (Cambridge English Language Assessment, 2016)

As Cambridge English Language Assessment (2016) points out, these tests categorize the test takers into three groups: basic, independent and proficient users. Basic User tests include Young Learners of English (YLE), Key English Test (KET) at two levels A1 and A2. Independent use tests include Preliminary English Test (PET), First Certificate in English (FCE) at two levels B1 and B2. Proficient user tests include Certificate in Advanced English (PET), Certificate of Proficiency in English (FCE), Business English Certificate (BEC), Business Language Testing Services (BULATS), and International English Language Testing System (IELTS) at two level C1 and C2.

Among many Iranian contexts of education, high schools seem to be of utmost importance due to two reasons. Firstly, there is a very large number of students in such contexts. Secondly, the Iranian high school system has recently shifted its attention from the premises underlying the Grammar Translation Method to those underpinning Communicative Language Teaching. This great shift has led to much emphasis being attached to four major linguistic skills, including listening, reading, speaking and writing in high school textbooks. Owing to the fact that such Cambridge tests can exert influence upon the Iranian system of education at large, tests that are equally reliable and valid in applied linguistics like PET and FCE as two focal points of this study seem to be in high demand, especially in comparison with Universal Grammar (UG) tests that measure UG knowledge like a researcher-made UG test (RMUGT) that was made both reliable and valid by the researchers prior to this study in order to achieve the main goal of this study. The RMUGT comprised 40 Grammaticality Judgement (GJ) tasks including such

Universal Grammar Principles and Parameters (UGPPs) as Piepd-Piping and Preposition-Stranding Parameter (PPAPSP), Binding Principle (BP), Pro-Drop Parameter (PDP), That-Trace Effect (TTE), Projection Principle (PP), Resumptive Pronoun (RP) and Subjacency Principle (SP). As stated by Whong, Gil and Mardsen (2013), “Principles and Parameters is perhaps the best known and most accessible, UG model” (p.4). They add that this model consists of some principles that are crosslinguistically invariant and some parameters whose values are to be set on the basis of the evidence provided by the target language. Given the context of the study, PET and FCE tests shown in the figure above seem to be appropriate for the research participants as high school students. A vigorous discussion of the details regarding these three tests (i.e. PET, FCE, and RMUGT) will be presented subsequently.

This study is based upon two theoretical frameworks as two highly important topics dealt with in language studies including generative SLA and Language Testing (LT), in general, PET and FCE as two Standardized English Proficiency Tests (SEPTs), in particular. The plan, in this study, was to make a bridge between these two camps (generative SLA and LT) which has recently gained much recognition in the related literature (e.g. Bachman & Cohen, 1998; Shohamy, 2000; Douglas, 2001; Yoshida, 2006; Gu, 2011, 2014; Cicerchia, 2014, to mention a few). Therefore, the general framework within which the current investigation proceeded has been presented in many places in the related literature showing that there exists a gap between these two paradigms. It is worth emphasizing here that there are many approaches to SLA one of which is called generative SLA or generative grammar SLA. This approach is the focal point of this study, though it is not the most dominant approach to SLA. Most importantly, this approach is supported by Hawkins’ (2001) study on the theoretical significance of UG in SLA published by Sage Publications Ltd. Moreover, White’s (2003) study on second language acquisition and universal grammar lends much support to the approach to SLA adopted by this study. Such studies are believed by the proponents of generative SLA to theoretically support the claim made by this study that UG could be related to foreign language learning contexts, on the one hand, and such Cambridge University tests of proficiency as FCE and PET, on the other hand.

UG is formulated as the content of a universal language core while the periphery of any language comprises language-specific rule, vocabulary as a list of exceptions, and etc. The way core and periphery are related to each other is still shrouded in history. However, it sounds a reasonable (though still too speculative) proposal that the core is causally related to, and in a sense, scaffolds the periphery. Standardized proficiency tests target the periphery of a language. A significant correlation between the UG test as employed in this study and commercially produced English proficiency tests (i.e. PET and FCE) could direct us toward a possible relationship between core and periphery knowledge of language although such a relationship does not need to be a causal one. Whatever the case, the fact that UG tests are more economical both in construction and administration, such a high correlation makes UG tests greatly desirable. To bridge this gap in the literature related to the relationship between the core and periphery, this study investigated the relationship between UG knowledge and performance on PET and FCE tests of English proficiency in order to determine the extent to which UG accessibility exerted influenced upon performance on the proficiency tests.

After much consideration was given to the context and theoretical framework of the study, it is now plausible to state the problem at which this study aimed at solving through designing the RMUGT that is expected to lead to some suggestions regarding the learner variable and a new trend in the studies pertinent to the under-explored issue of marriage between generative grammar SLA and LT. A careful review of the related literature on LT and test construction research paradigm, especially SEPTs, indicates that there is a multitude of

contributors to this area of enquiry probably best characterized by “steps for constructing tests like planning for the test, preparing the test, analyzing and revising the test” (Corpuz, 2013, p. 1), “phases like decision to provide test, planning, design, tryout, (informing stakeholders), and final test specifications” (Council of Europe, 2015, p.20), “modules like test planning stage, test item development, test item analysis, development of marking scheme” (Olubodun, 2007, p.5), “general standards for test items” (Michigan State University, 2011, p.7), “planning the test (Clay, 2001, p.6), “planning the test” (Lucy, 2004, p.1) and “keys to effective testing” (Zimmaro, 2004, p. 3). In this account of test construction, many issues have been taken into consideration: the characteristics of the test takers (e.g. age, gender, social and educational situation, and etc), the purpose of the test and the standard(s) for the proposed test, the relation of the test to an educational context, the use of the test results, the stakeholders, the time and place of test administration, the delivery mode of the test, the implications of the test, test performance monitoring, the design of the test, the content of the test, test duration, test length, test type, scoring scheme and criteria, number and types of items or tasks, recruiting and/or training test raters or scorers, cultural factors, developing assessment specifications, domain of knowledge and skills, relative weights of tasks or skills, assessment and response forms, using appropriate directions and accessible language, using statistics to evaluate the assessment and scoring, the cornerstones of testing (i.e. usefulness, validity, reliability, practicality, washback, authenticity, transparency and security), grading procedures, and so forth. However, UG knowledge seems to be one of the characteristics of the test takers which has gone underexplored or even under-noticed. Once unchecked, this factor was hypothesized prior to this study to influence proficiency test performance to the effect that test takers were hypothesized to draw upon their knowledge of UG principles and parameters while taking SEPTs, irrespective of their general knowledge of English. This apparently neglected issue of UG knowledge and availability and the role it was hypothesized to play in taking such tests as PET and FCE constitutes the main problem of the current study.

This problem, to the best of the researchers’ knowledge, seems to stem from the point that the gap between generative grammar SLA and LT has not been fully bridged yet, in spite of the fact that many research studies place emphasis upon the synergy of these two areas of applied linguistics (e.g. European Second Language Association, 2010; Cenoz & Ulrike, 2001; Khany, Youhanaee & Barati, 2008; Leung, 2003; White, 2003; Gu, 2011; and Jordan, 2004). Bachman and Cohen (1998, p.98) believe that “the state of affairs related to the segregation of SLA and LT, known as two of the fastest growing and most technical areas of applied linguistics, is clearly unproductive, though many journals emphasize the need for their integration.” They also emphasize that dialogues between the practitioners in the two fields have been sporadic at best, and non existent in many ways. The RMUGT was expected to solve this problem. Given the problem and its corresponding solution stated above, the current study aimed at taking into consideration UG knowledge and accessibility as a hypothetically important but underexplored or even unnoticed factor contributing to achievements on SEPTs. As for the categorical nature of the sentences included in the RMUGT, it is worth stating that being categorical entails being either grammatical or ungrammatical. However, grammaticality is replaced by acceptability when it comes to utterances because the sentence is associated with competence, while the latter is pertinent to performance. Therefore, the likert scale which deals of degrees of grammaticality has gained much recognition in the literature related to UG levels of knowledge and availability. Interestingly enough, even the native speakers of English do not necessarily pass the same judgements on the grammaticality/acceptability of the sentences/utterances. For example, some linguists may mark ungrammaticality by * (asterik), but other linguists mark it by ?? which

means that it is more acceptable. Moreover, since the emergence of the Principle and Parameter paradigm, it has been the case that the violation of a principle or parameter results in ungrammaticality the degree of which varies depending on the number of violations. Therefore, the issue of degrees of grammaticality which is the cornerstone of the RMUGT seems legitimate in the current study on the integration of generative SLA and LT.

Once viewed from the perspective of construct validation research, this study was expected to lead to a reconsideration of the construct validity of the tests of proficiency. This study also was an attempt to provide validity evidence for the internal structure of the standardized proficiency tests and their external relationship with the test takers' UG knowledge and accessibility as a newly emergent type of test takers' characteristics and/or learner variables. Needless to say, this issue constitutes the most significant and innovative aspect of this study. As argued by Gu (2011), FL learners come into a language testing situation as complex human beings, characterized not only by their prior target language achievement but also by their native language background, gender, past and current learning conditions, and many other characteristics. Test-takers' identities and life experiences are also valuable information for us to understand their current learning profiles, and how they have arrived at where they are. The research community has gradually embraced the idea that treating test takers regardless of their identities and life experiences will give us an over-simplified picture of their test performance. Sharing the same thought, this study argues that Iranian EFL test takers bring UG with them to the testing situation besides other items included in their learning profile, so how they perform on tests is not necessarily a function of their language repertoire, rather they can draw upon their UG knowledge to the effect that it can become a new influence on the proficiency test scores

The attempt to find the solution, among many others, to the above-mentioned problem led to the formulation of the following two research questions for the present study:

Q1. Is there any significant relationship between the participants' knowledge of UG and their performance on the PET and FCE tests?

Q2. Can RMUGT scores predict the scores on the proficiency tests?

Literature Review

Once viewed from the perspective of construct validation research, this study was expected to lead to a reconsideration of the construct validity of the tests of proficiency. This study also was an attempt to provide validity evidence for the internal structure of the standardized proficiency tests and their external relationship with the test takers' UG knowledge and accessibility as a newly emergent type of test takers' characteristics and/or learner variables. Needless to say, this issue constitutes the most significant and innovative aspect of this study. As argued by Gu (2011), FL learners come into a language testing situation as complex human beings, characterized not only by their prior target language achievement but also by their native language background, gender, past and current learning conditions, and many other characteristics. Test-takers' identities and life experiences are also valuable information for us to understand their current learning profiles, and how they have arrived at where they are. The research community has gradually embraced the idea that treating test takers regardless of their identities and life experiences will give us an over-simplified picture of their test performance. Sharing the same thought, this study argues that Iranian EFL test takers bring UG with them to the testing situation besides other items included in their learning profile, so how they perform on tests is not necessarily a function of their language repertoire, rather they can draw upon their UG knowledge to the effect that it can become a new influence on the proficiency test scores.

The generality of this FL construct has also been a focus of investigation in the field of language testing. Messick (1989) warned against taking the generalizability of a construct meaning across various contexts for granted. He proposed that context effects, especially different population groups, in score interpretation be systematically appraised. Validity evidence based on a test's generalizability was also proposed by Chapelle (1999) to ensure legitimate test score interpretation and uses across groups of test-takers, time, instruction conditions, and test task characteristics. The idea of a universally applicable construct framework seems especially questionable in language testing, considering the differences in the language to be measured (e.g., English, French, Chinese), and the usually heterogeneous nature of the test-taking population. This line of research helps to answer the question of whether the same construct structure holds across groups of test-takers with UG accessibility which was hypothesized to be an important factor contributing to performance on the above-mentioned FL proficiency tests.

From the foregoing, it can safely be claimed that the apparent separation between LT and GenSLA camps in the existing literature deserves more attention, because UG accessibility and its relationship with proficiency tests, to the best of the researchers' knowledge, has remained unnoticed or under-researched and therefore it is worth a scientific enquiry. Accordingly, this study was an attempt to begin to remedy the situation. This contact across sub-disciplinary borders are thought to be influential in this regard. The construct validity of the proficiency tests was hypothesized to be a problematic issue. It was this concern for solving this problem through a more extensive dialogue between GenSLA and LT researchers and practioners and a deeper understanding and sharing of common research and methodological issues that motivated the researcher to embark on this study as a tool for stimulating discussion among GenSLA and LT researchers by exploring areas of common interest, both substantive and methodological, from a variety of perspectives. The results gained from this study on investigating the internal and external structure of the PET, FCE, and RMUGT could be used to interpret the nature of the FL construct: whether the construct is unidimensional or multidimensional, and what the makeup of a multidimensional construct is.

Methodology

This study is correlational, in nature, and descriptive and hypothesis-testing, by definition, dealing with a significant relationship between the scores gained by the research participants on the RMUGT, on the one hand, and those obtained by the same participants on the proficiency tests. In general, this study is a quantitative method of research in which there are three quantitative variables (i.e. the RMUGT, PET, and FCE scores) from the same group of participants, and the main goal was to find out whether there was a significant correlation or covariance between (a) the RMUGT and PET test scores, and (b) the RMUGT and FCE test scores (as it was addressed by the first research question). In addition to this relational design, a prediction design was also used by this study. The purpose of this design was to figure out whether RMUGT scores could effectively predict the scores on the proficiency tests (as it was dealt with by the second research question).

The participants formed an aggregate of 108 Iranian EFL high school students with an age range of 14 to 17. After they were randomly chosen from some high schools in Malayer from Hamedan, the aim of the study was fully explained to them. Since language proficiency level was not a focal point of this study, the participants were varied in this respect. In other words, since this study aimed at the statistically significant relationship between knowledge of UG knowledge and performance on SEPTs, the research participants' levels of proficiency (i.e. high, mid, low) was not taken into consideration. However, future researchers interested in the line of research

related to generative SLA-LT integration are suggested to investigate any possible effects of proficiency level on the statistically significant relationship between UG knowledge and SEPTs performance.

There were three instruments used in this study: the PET, FCE and RMUGT. The PET and FCE tests are administered by Cambridge English Language Assessment (For more information see www.cambridgeenglish.org). 100% of UK universities now accept them as proof of English language ability. More than 6,000 educational institutions, businesses and government departments around the world accept them. As for the tests format, they are thorough tests of all areas of language ability including reading and use of English, writing, listening and speaking. As Cambridge English Language Assessment (2016) states, these FCE and PET tests are appropriate for independent users who come in between basic and proficient ones. Therefore, these two tests seem to be appropriate for the research participants who were high school students. The RMUGT contained 40 GJ tasks dealing with a combination of both grammatical and ungrammatical sentences shuffled in the test. Each sentence was characterized with 5 ranks (1=completely ungrammatical; 2=almost but not quite ungrammatical; 3=not sure; 4=almost but not quite grammatical; 5=completely grammatical). The tasks were related to such principles and parameters as PPAPSP, BP, PDP, TTE, PROP, RP, and SP. Besides, some fillers (F) were randomly included in the UG test for distraction purposes. The RMUGT included the UGPPs and Fs as its eight subparts (See Appendix).

As for the reliability of the RMUGT as a purely statistical parameter, the finalized version of the test was administered to 12 randomly-selected students from the high schools twice, and the correlation between the two sets of scores obtained from the two administrations was calculated. During this 2-week time interval, an assumption made by the researchers was that no significant change occurred in the 12 participants' knowledge. As for the time interval, it is worth mentioning that it was set to reduce the effect of the examinees' working memory functioning. This interval was neither too long to bring about drastic changes in the examinees' behaviour nor too short to allow their working memory to function. The Chronbach's α was calculated to estimate the reliability of the RMUGT the results of which indicated that the test was found to be reliable ($r=0.87$).

Dependent on the peculiarities of the test, the validity of the RMUGT, on the other hand, was also taken into consideration. Since the test was supposed to measure the research participants' UG knowledge, the whole test was reviewed by three experts in the field of UG studies. They reviewed both content and appearance of the test, as a whole, as well as the individual items with their Likert scales. In actual fact, the finalized version of the test was given to three experts for their corrective and evaluative feedback that helped the researchers to modify or even exclude some GJ tasks that were found not to measure knowledge of PPs appropriately. Their feedback was about the directions of the test, the difficulty level of the GJ tasks, the structure of the tasks, the arrangement of the items, the content of the tasks, the wording of the tasks, and so forth. The reviewing processes helped maximize the content validity (the correspondence between the content, namely UG knowledge, that was to be tested and the content of the 40-item test) and the face validity of the test.

As for the procedural aspect of the study, all participants of the study were, at first, taught how to approach the research materials. As regards the RMUGT, they received clear and detailed instruction regarding how they were expected to make their own judgements about the sentences included in the test. They were told that they were required to judge each sentence on the basis of what they themselves thought they said under appropriate circumstances. In other words, the frequent dilemma between what they had to say and what they actually said was removed in this

way. They were warned against judging on the basis of what they considered to be “proper English” and/or on the basis of what they had been taught at school because their own patterns of thinking and judgmental preferences were of utmost importance. Also, it was made crystal clear that they were not allowed to reject any sentences because they knew a better sentence to convey the same meaning. Some concrete examples were also used in this introductory phase to make sure that the participants drew on their own grammatical knowledge, rather than the prescriptive rules and regulations existing in English. The participants were also instructed how to use the rating scale of the test. They were made acutely aware that if they considered a sentence as “grammatical,” it meant that they totally accepted it and actually used it under appropriate circumstances. In similar fashion, they received enough instruction on what it really meant when they chose “almost but not quite grammatical,” “ungrammatical,” and “almost but not quite ungrammatical.” It was specifically explained that they were required to select the option “not sure” when they did not have certainty about the grammaticality level of the sentence. As for the proficiency tests, on the other hand, the instructions specified by Cambridge University official website was equally delivered to all participants. Similarly, the three tests were given to the participants at two-week time intervals.

After the administration of the three tests and collection of the desired data, scores obtained from the tests were analyzed through a set of Shapiro-Wilk, Chi-2, Spearman Correlation, ANOVA and regression analysis tests. To determine the type of correlation between the UG and proficiency tests as well as the normality test, the Shapiro-Wilk test was conducted the results of which indicated that the normality assumption was rejected. Therefore, the Spearman correlation test was run.

Results

To answer the first research question, two steps were taken consecutively. Firstly the statistical indexes of the descriptive aspect of the data were calculated. Secondly, the inferential statistics related to the obtained data was calculated. The results yielded from the first phase are tabulated below.

Table 1. *Descriptive Statistics for the RMUG, PET and FCE Tests.*

Test	Minimum	Maximum	Mean	SD
RMUG Test	103.00	163.00	124.4815	11.05361
PET Test	3.15	6.17	4.8214	.76080
FCE Test	3.09	5.74	4.6871	.64381

As it can be inferred from the table above, the mean score of the participants' (n=108) performance on the RMUGT equals 124.48 out of 200, and the standard deviation related to these scores (n=108) is 11.05. This result, in all likelihood, highlights the availability of UG to the research participants. The mean scores obtained by these participants on the proficiency tests (i.e. PET and FCE), on the other hand, reveal that those who took the PET test performed better, though to a small degree, than the FCE counterpart. The minimum and maximum scores obtained on the proficiency tests are out of 9. As for the issue of how the scores (e.g. 124.4815) were calculated, it is worth mentioning that each RMUGT task was weighed from 1 (the totally false answer) to 5 (the totally true answer). Since there were 40 GJ tasks included in the RMUGT, the minimum and maximum scores were 40 and 200, respectively.

To find out whether there was a significant relationship between the UG and proficiency tests (as the answer to the first research question), two chi-square tests were run the results of

which are presented in Table 1 below. Also, to test the correlation between the UG and the proficiency tests, the normality assumption was accordingly checked by Shapiro-Wilk tests the results of which showed that according to the calculated p-values 0.001, 0.02 and 0.000 for the UG, PET and FCE tests, respectively, the normality assumption was rejected, and therefore the Spearman correlation test was conducted the results of which are also presented in the table below.

Table 2. *Chi-Square and Spearman Correlation Tests for the Significant Relation and Correlation Between the UG and Proficiency Tests*

Tests	Test Statistics	df	P	Spearman Correlation	p
RMUGT & PET	12.21	4	0.001*	0.67	0.000*
RMUGT & FCE	13.22	4	0.005*	0.62	0.000*

As displayed in Table 1, the scores gained on the UG test are compared with those obtained on each proficiency tests (i.e. PET and FCE). Since the p-values (0.001 and 0.005) are smaller than 0.05, the relationship between the UG and proficiency tests are proved to be statistically significant (as it is the answer to the first research question). The p-values obtained from the Spearman test show that there exists a significant dependency between the UG and proficiency tests. To find out whether there was a statistically significant dependency between the two proficiency tests, a Spearman correlation test was conducted the results of which are tabulated below.

Table 3. *Spearman Correlation Test for the Significant Dependency Between the Two Proficiency Tests*

Spearman Correlation	P-value (sig.)
0.66	0.000*

As indicated by this table, the p-value is greater than 0.05 which means that there is a statistically significant dependency between the two tests of proficiency as wholes. To address the second research question, a Cubic Regression model was fit for the FCE test the results of which are provided in the tables below.

Table 4. *Summary of the Cubic Regression Model for Predicting the FCE Scores*

R	R Square	Adjusted R Square	Std. Error of the Estimate
0.959	0.920	0.914	.224

According to Table 3 above, the Cubic model for making a prediction about the FCE scores by the RMUGT scores is truly effective because of the R square and adjusted R square values that show that this model is over 90% able to predict scores gained on the FCE test of proficiency. To find out whether this model was statistically significant, ANOVA was conducted the results of which are presented in Table 4 below.

Table 5. Results of ANOVA for the Significance of the Cubic Regression Model for Predicting the FCE Scores

	Sum of squares	df	mean square	F	Sig.
Regression	13.850	2	6.925	607.456	.000*
Residual	1.199	105	0.0114		
Total	15.049	107			

As vividly indicated by this table, the Cubic model is shown to be significant when it comes to the prediction of the FCE scores by the RMUGT scores. This significance is proved by the significance value 0.000 that is smaller than the set value. On the basis of the obtained coefficients from Table 5, the Cubic Regression model can be extracted.

Table 6. Summary of Significance of the Coefficients Related to the Cubic Regression Model for the FCE Scores

	Unstandardized Coefficients		Standardized Co.	T	Sig.
	B	Std. Error	Beta		
RMUGT	0.476	0.067	6.488	7.151	.000*
RMUGT ²	-.002	.000	-5.609	-6.182	.000*
Constant	-30.714	4.431		-6.931	.000*

According to the results in the table above, the true Cubic Regression model is as follow: $FCE = -30.714 + 0.476 * RMUGT - 0.002 * RMUGT^2$. Since the obtained p-value for the calculated coefficients are smaller than 0.05, they are significant. In a nutshell, the RMUGT scores were revealed to be able to predict the PET scores. The following section is devoted to the same Cubic model for predicting the PET scores.

Table 7. Summary of the Cubic Regression Model for Predicting the PET Scores

R	R Square	Adjusted R Square	Std. Error of the Estimate
0.891	0.880	0.861	.224

The details regarding the Cubic model, as presented in Table 6 above, show that this model is effective as far as the prediction of the PET scores are concerned. This is due to the high values associated with the R square and adjusted R squares (0.88 and 0.86, respectively). To find out whether this model was statistically significant, ANOVA was conducted the results of which are presented in Table 7 below.

Table 8. Results of ANOVA for the Significance of the Cubic Regression Model for Predicting the PET Scores

	Sum of squares	df	mean square	F	Sig.
Regression	12.721	2	6.360	508.8	.001*
Residual	1.315	105	0.0125		
Total	14.036	107			

As shown by this table, the Cubic model is seen to be significant when it comes to the prediction of the PET scores by the RMUGT scores like the FCE ones. This significance is, in actual fact, confirmed by the significance value 0.000 that is smaller than the set value. Also, the

coefficients presented below were revealed to be significant. Therefore, the RMUGT scores were proved to be capable of making predictions about the PET scores.

Table 9. Summary of Significance of the Coefficients Related to the Cubic Regression Model for the PET Scores

	Unstandardized Coefficients		Standardized Co.	T	Sig.
	B	Std. Error	Beta		
RMUGT	0.378	0.087	5.978	6.965	.000*
RMUGT ²	-.005	.006	-5.908	-6.656	.000*
Constant	-27.18	5.012		-6.567	.000*

According to the results in the table above, the true Cubic Regression model is as follow: $PET = -27.18 + 0.378 * RMUGT - 0.005 * RMUGT^2$. The obtained p-values for the calculated coefficients that are smaller than 0.05 show that they are significant. In sum, the answer to the last, but not the least research question is that the RMUGT scores can predict the scores gained on the PET and FCE tests as two SEPTs. This prediction ability is made possible by the Cubic Model for making predictions about the proficiency tests scores.

Discussion

The results gained from the Chi-2 and Spearman tests led to the claim that there exists a statistically significant correlation between the participants' UG knowledge and their performance on the PET and FCE tests. This finding seems to be truly promising in that it can lead to collaboration between GenSLA and LT, in general, and the PET and FCE tests known as SEPTs, in particular. LT, more specifically SEPTs, studies can reap benefits from GenSLA through taking into account the test takers' or learners' UG knowledge and accessibility as a newly emerged set of suggestions regarding the test taker and/or learner variable that can exert influence upon their performance on the tests of proficiency, along with their proficiency in listening, reading and use of English, speaking and writing. This set suggests that GenSLA widen and broaden its scope of enquiry through giving UG knowledge and availability much consideration as far as the SEPTs are concerned. Interestingly enough, this suggestion is supported by Whong, Gil and Mardsen (2013) who argue for recent trends in GenSLA. On the face of it, they believe that there are two considerable trends related to GenSLA: (1) the expansion of its enquiry beyond syntax to other domains, "the interface between these domains" (White, 2011, p.58), and to other cognitive factors such as processing, and (2) the growing emphasis upon the question of which linguistic properties are easily acquirable and which are problematic.

The current study was an attempt to detail some cutting-edge issues pertaining to the significant relationship between UG knowledge and performance on SEPTs. To put it another way, the role of UG knowledge in taking SEPTs was viewed from the perspective of generative grammar SLA, in spite of the fact that the participants were thought of as EFL learners who were in the course of Instructed SLA (ISLA) the prototypical contexts of which include the language classroom, the virtual L2 classroom, self-study, study abroad and so forth. In fact, the participants' context of learning was based on ISLA, while this study's context of measurement was focused upon GenSLA. Since language instruction is "a culturally bound endeavor" (Loewen & Sato, 2017, p.4), it is important to note that the current research conducted in Iran as an EFL country is different from North American and Western European contexts in which SLA and ISLA were primarily developed, and this difference can be appreciated by Loewen and Sato's

assertion, “it is necessary to conduct research in different learning contexts that may challenge existing ISLA theories or provide alternative perspectives” (p.5). ISLA research is concerned with L2 learning processes that are hypothesized to be or have been found to be amenable to intervention. Given the issue of non-teachability related to such UG-related processes, it would be legitimate to claim that the designers of proficiency tests, especially SEPTs, are suggested to take into particular account the UG-related issues while constructing their tests to better deal with the considerable importance of UG processes as part of the testees’ mental processes.

The issue of taking into consideration the role that UG knowledge can play in taking SEPTs constitutes the most significant result, among many others, obtained from this study. It can lead to the expansion of the focus in GenSLA as it has, to date, been extended by other results gained by other studies in the related literature. For example, advances stemming from more sophisticated understanding of different domains of language—including the theoretical transition from Principles and Parameters to Minimalism—have also led to expansion of the initial focus within GenSLA. Much research in the late 1990s, for example, explored differences in the development of syntax and morphology, with many concluding that there is a ‘mapping’ problem as learners have to map linguistic features onto particular forms which may or may not resemble those in their native language (e.g. Lardiere, 2000). More recent GenSLA research parallels larger trends in generative linguistics, exploring areas of language beyond core competence. Investigation of the relationships between narrow syntax and the domains of discourse/pragmatics has led to the recent Interface Hypothesis (Belletti, Bennati, & Sorace, 2007; Sorace & Filiaci, 2006; Sorace & Serratrice, 2009) which proposes that constructions that implicate both domains will cause more difficulties for the learner than those that are restricted to one domain of language. Another recent proposal argues that both syntax and meaning are acquirable, in contrast with functional morphology, which is seen as creating a bottleneck for second language acquisition (Slabakova, 2008). GenSLA researchers have also begun to direct attention to issues of processing (e.g. Juffs, 2006) and, more recently, neurolinguistics (e.g. Yusa, Koizumi, Kim, Kimura, Uchida, Yokoyama, Miura, Kawashima & Hagiwara, 2011).

The scores gained on the RMUGT were witnessed to be able to make predictions about the scores obtained on the PET and FCE tests. This helped prove the predictive validity of the RMUGT as a welcome and timely addition to, if not a substitute for, the SEPTs, more specifically the two proficiency tests of this study. In similar vein, this finding lends support to the GenSLA-LT integration and also the new model for considering UG knowledge and accessibility as a newly emerged learner variable and/or test taker characteristic. Most importantly, this set of suggestions has some salient features the most important ones of which are presented below.

Firstly, there seems to be a pressing need for a new agenda and methodology for dealing with the availability of UG in SEPTs in that the current views of GenSLA appear to have taken the effect of UG knowledge on performance on SEPTs for granted.

Secondly, this set of suggestions entails a reconsideration and redefinition of the term “L2 learners’ linguistic knowledge” as far as SEPTs are concerned. Since there are different viewpoints about what constitutes this piece of knowledge, it is all well and good to say that it can include, among many other things, knowledge of UG. Having made a broad claim about the focus of the new set of suggestions, it is important to acknowledge that there are numerous variables, besides the UG-related learner variable as a learner-internal variable, that are both external and internal to the EFL learners, thereby exerting influence upon their performance on SEPTs. Such variables are both interesting and challenging for GenSLA-LT marriage

researchers, test designers, and even teachers. Therefore, this learner-internal variable should receive much more considerable investigation.

Thirdly, it was based upon a kind of interlanguage analysis in the high school classroom context with implicit emphasis being placed upon the internal logic of the research participants' Interlanguage (IL) without being dependent on their errors and the conformity of their IL to the target language. A careful review of the related literature reveals that IL and errors have often been mingled, despite many recommendations about the independence of IL. The participants' performance on the three tests discussed in detail above was not analyzed from the error-counting approach to IL analysis that is commonly implemented. Rather, the participants' IL was analyzed from the perspective of UG-mediated GenSLA. UG-related error types and sources thought of by this study to be timely and welcome additions to the literature related to error analysis, error typology, and error sources.

Finally, it suggests a new protocol for assessing EFL learners' IL with reference to their UG accessibility in major linguistic skills (i.e. listening, reading and use of English, writing and speaking). This newly suggested protocol, instead, diverges from reference to SEPTs as a common tool for measuring their general English proficiency that is prevalent in the relevant literature. This protocol is based on the RMUGT for assessing their IL. The procedure of this protocol begins with the administration of the RMUGT and asking them to judge the grammaticality level of each GJ task on the basis of their own preferences after they receive clear and detailed instruction regarding how to approach the RMUGT. They should, accordingly, mark the grammaticality level of each task on the basis of what they themselves actually say, rather than what they think they should say. They are also warned against rejecting a task because they know a better version of it. Concrete examples of the tasks can be of help and use in this regard. They should also receive clear instruction on how to rate the five Likert scales. The second phase of the procedure deals with the analysis of their judgements in terms of their level of UGPPs accessibility. Finally, predictions can be made about their performance on tests of proficiency, thereby assessing their current level of FL proficiency without recourse to SEPTs. One offshoot of this protocol, among many others, is that the RMUGT seems to be a welcome addition to, if not a substitute for, the proficiency tests. The information that is gained about their level of UGPPs through the RMUGT can be used to construct a Student Portfolio (SP) for the purpose of future analyses and uses. Such a performance and assessment of SP is expected to have its own adequacy for evaluating the skills possessed by an individual in a range of everyday situations in the school contexts, if not overgeneralized to other settings. This linguistic profiling enterprise could be incorporated into language assessment that "stands to benefit from profiling research" (Hulstijn, 2010, p. 234). This constitutes a promising line of research emerging in light of this study in the relevant literature that is quite scant but steadily growing. Alderson (2010) asserts, "linguistic profiles have yet to be achieved through empirical research for any language" (p.244).

This new set of suggestions has, at least, two implications. Firstly, test designers can, given the results of the current study, be urged to take the effect of UG knowledge on their tests of proficiency into careful consideration, thereby reducing or eliminating the effects. Another option is that they can design their own UG tests or use reliable UG tests available in the related literature, like the RMUGT. Secondly, classroom teachers can apply SLA findings to the four walls of their classrooms due to the "relevance of GenSLA to classroom teaching," as stated by Whong (2011, p.251). This suggestion, however, does not mean that this application has not yet occurred. In fact, the field of applied linguistics, as Whong (2011) claims, grew out of attempts to do exactly this. This study was therefore an attempt to apply research from linguistics, more specifically Chomskyan linguistics dealing with UG, to the confines of the language classroom.

This can be considered as a further step in GenSLA-LT marriage studies and a challenge for GenSLA-LT marriage researchers who have not been able to articulate the usefulness of the corresponding findings for the language classroom.

Conclusion

Conducted within the generative linguistic framework, the current research studying SLA from a linguistic perspective in comparison with SEPTs was carried out to characterize and explain the underlying linguistic competence of Iranian high school learners of English as a Foreign Language (EFL) in terms of the constraints in UG principles and parameters. For the purpose of the study, the participants' UG accessibility was assessed through the RMUGT consisting of 40 grammaticality judgement tasks aimed at tapping their interlanguage competence. The results associated with the GJ tasks revealed that language proficiency bore a statistically significant relationship to the ability to judge different sentence types as grammatically correct/incorrect. This finding confirms that SLA is constrained by UGPPs. It was revealed that the participants' abstract and subtle competence of EFL was constraint by the same universal principles that govern natural language in general. In effect, their non-primary language acquisition is a function of UG availability. This study acknowledges that performance on SEPTs is mediated by UG accessibility.

This study is based on the view that a generative account of SLA can and should engage more with the field of LT, more specifically SEPTs like PET and FCE for mutual effects. This section concludes the whole study by suggesting ways in which this engagement can be achieved. It is argued that this generative account of SLA needs to better articulate the implications of research for LT. Going further, it is suggested that LT should be included as one of the ranges of existing variables in the generative account of SLA. There is also potential for working with other research paradigms that are actively engaged in research in the learner variables, some of which are concerned with questions of learner errors associated with UG availability, and others which are more sociolinguistic in orientation such as the learners' mother tongue, as it included Turkish, Kurdish, Persian and Lori for the participants of the study. However, this study can also be seen as outside the scope of the kinds of questions posed by the generative accounts of SLA allowing for fruitful collaboration at this point. The more closely related line of research is work that investigates UG availability as a newly emerged learner variable exerting influence on performance on SEPTs like PET and FCE. Despite differences in theoretical premise, there, as this study safely claims, is scope for collaboration between the generative account of SLA and SEPTs. This newly emerged line of research is much interested in UG knowledge and accessibility as a learner or test taker variable. Therefore, the issue of the extent to which UG is available in SLA is complicated by giving consideration to the issue of how UG accessibility affects performance on such tests proficiency. It should be emphasized here that this interaction between GenSLA and performance on SEPTs can inform research at the edges of limits of GenSLA related to its overemphasis upon acquisition as the only relevant question for SLA research, thereby making room for UG availability in the major language skills as another line of inquiry within GenSLA. In doing so, this study will suggest a new agenda which engages GenSLA and performance on SEPTs. This agenda addresses the core assumption of applying existing findings related to GenSLA to the context of such SEPTs as PET and FCE with their four sections on the major linguistic skills. Through this application, performance on proficiency tests can be predicted.

References

- Alderson, J. C. (2010). *Language testing-informed SLA? SLA-informed language testing?* Department of linguistics and English language, Lancaster University.
- Bachman, L. F., & Cohen, A. D. (1998). *Language testing—SLA interfaces: An update*. In Bachman, L. F. & Cohen, A. D. (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 1-31). New York: Cambridge University Press.
- Belletti, A., E. Bennati, & A. Sorace. 2007. Theoretical and developmental issues in the syntax of subjects: Evidence from near-native Italian. *Natural Language and Linguistic Theory*, 25: 657–689.
- Cambridge English Language Assessment (2016). *Principles of good practice: Research and innovation in language learning and assessment*. Cambridge University.
- Cenoz, J., & Ulrike, J. (2001). *English in Europe: The acquisition of a third language*. Clevedon: Multilingual matters.
- Cicerchia, M. (2014). *Language Testing and Second Language Acquisition Research: Worlds Apart*. Lingly's blog of SLA inspired tips for language learners.
- Clay, B. (2001). A Short guide to writing effective test questions. Kansas Curriculum Center.
- Corpuz, A. (2013). *Test construction manual*. UTPA Teaching Academy Boot Camp.
- Council of Europe (2015). *Common European Framework of Reference for Languages: Learning, Teaching and Assessment*. Cambridge University Press.
- Douglas, D. (2001). Performance consistency in second language acquisition and language testing research: a conceptual gap. *Second Language Research*, 17(4), 442 - 456.
- European Second Language Association (2010). *Communicative proficiency and linguistic development: Intersections between SLA and language testing research*. ISBN 978-1-4466-6993-8. USA: MIT Press.
- Gu, L. (2011). *At the interface between language testing and second language acquisition: communicative language ability and test-taker characteristics*. Ph.D. thesis, University of Iowa, <http://ir.uiowa.edu/etd/972>.
- Gu, L. (2014). At the interface between language testing and second language acquisition: Language ability and context of learning. *Language Testing*, 31(1), 111-133.
- Hawkins, R. (2001). *The theoretical significance of universal grammar in second language acquisition*. Sage Publications, Ltd.
- Hulstijn, J., H. (2010). *Linking L2 proficiency to L2 acquisition: Opportunities and challenges of profiling research*, in Bartning, I., Martin, M., and Vedder, I. (2010). *Communicative proficiency and linguistic development: Intersections between second language acquisition and language testing research*. European Second Language Association, 233-253.
- Jordan, G. (2004). *Theory construction in second language acquisition*. Amsterdam: John Benjamins.
- Juffs, A. (2006). *Processing reduced relative vs. main verb ambiguity in English as a second language: A replication study with working memory*. In *Inquiries in linguistic development in honor of Lydia White*, ed. R. Slabakova, S. Montrul, and P. Prevost, 213–232. Amsterdam: John Benjamins.
- Khany, R., Youhanaee, M., & Barati, H. (2008). Tense and passive feature/morphology in the transitional English interlanguage of Persian monolingual and Kurdish-Persian bilingual EFL learners. *TELL*, 1(4), 51-74.
- Lardiere, D. (2000). *Mapping features to forms in second language acquisition*. In *Second language acquisition and linguistic theory*, ed. J. Archibald. Oxford: Blackwell.

Leung, Y. K., I. (2003). *Functional categories in second and third language acquisition: A cross linguistic study of the acquisition of English and French by Chinese and Vietnamese speakers*. Unpublished doctoral dissertation, McGill University, Montreal, Quebec.

Loewen, S., & Sato, M. (2017). *The Routledge of Instructed SLA*. New York: Routledge.

Lucy, C. J. (2004). *How to write better tests*. Indiana University. http://www.indiana.edu/~best/write_better_tests.shtml

Michigan State University (2011). *Handbook of learner evaluation and test item construction*. Office of medical education research and development college of human medicine. Michigan State University.

Olubodun, O. J. (2007). *Test construction techniques and principles*. Retrieved from http://www.ncsc.org/Web%20Document%20Library/EC_StateInterpCert.aspx.

Shohamy, E. (2000). The relationship between language testing and second language acquisition: Revisited. *System*, 28(4):541-553.

Slabakova, R. (2008). *Meaning in the second language*. Amsterdam: John Benjamins.

Sorace, A., & F. Filiaci. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, 22: 339–368.

Sorace, A., & L. Serratrice. (2009). Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism*, 13: 1–16.

White, L. (2003). *Second Language Acquisition and Universal Grammar*. Montreal: Cambridge University Press.

White, L. (2011). *Linguistic theory, universal grammar, and second language acquisition*. In B. Van-Patten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 34–53). New York: Routledge.

Whong, M. (2011). *Seeking consensus: generative linguistics and language teaching*. School of Modern Languages and Cultures Michael Sadler Building University of Leeds.

Whong, M., Gil, K. H., and Marsden, H. (2013). *Introduction: Generative second language acquisition and language pedagogy*, in M. Whong, K.H. Gil and H. Marsden (eds), *Universal grammar and the second language classroom*, 16, 1-13.

Yoshida, K. (2006). Theoretical frameworks of testing in SLA: Processing perspectives and strategies in testing situations. *JALT Testing & Evaluation SIG Newsletter*, 1(10), 2-7.

Yusa, N., M. Koizumi, J. Kim, N. Kimura, S. Uchida, S. Yokoyama, N. Miura, R. Kawashima, & H. Hagiwara. (2011). Second-language instinct and instruction effects: Nature and nurture in second-language acquisition. *Journal of Cognitive Neuroscience*, 23: 2716–2730.

Zimmaro, D. M. (2004). *Writing good multiple-choice exams*. University of Texas.

Appendix

Researcher-Made Universal Grammar Test

Age:	Gender:	LI:	Educational background:
------	---------	-----	-------------------------

Directions: Determine the grammaticality of the following sentences by circling the suitable scale.

1= ungrammatical

2= almost but not quite ungrammatical

3= not sure

4= almost but not quite grammatical

5= *grammatical*

1. Whom did you write to? 1 2 3 4 5
2. The landlord almost always sees himself in the mirror. 1 2 3 4 5
3. It seems that his customers are going to leave the shop. 1 2 3 4 5
4. Which villa has the boss sold? 1 2 3 4 5
5. Who did you say came? 1 2 3 4 5
6. The man was to faint. 1 2 3 4 5
7. He is the person that you may love. 1 2 3 4 5
8. Which should the passengers rent car? 1 2 3 4 5
9. How old is your grandpa? 1 2 3 4 5
10. David believed the story that Mary saw a ghost. 1 2 3 4 5
11. To whom has she written a letter recently? 1 2 3 4 5
12. While shaving, the old man cut herself seriously. 1 2 3 4 5
13. David showed Rita a picture of himself. 1 2 3 4 5
14. Inflation seems that is rising nowadays. 1 2 3 4 5
15. Due to the accident, the driver fainted the children. 1 2 3 4 5
16. The sales may be rising. 1 2 3 4 5
17. This is a man that I just met him in the library the other day. 1 2 3 4 5
18. How are their kids old? 1 2 3 4 5
19. Apparently, Tom and Jerry hate himself in the cartoon. 1 2 3 4 5
20. Those are the books that my classmate will send to me. 1 2 3 4 5
21. Whom do you always talk with on the phone? 1 2 3 4 5
22. Do you ever know that she loves herself? 1 2 3 4 5
23. The teacher has just discovered that they love her very much. 1 2 3 4 5
24. What did David believe the story that Mary saw? 1 2 3 4 5
25. Who did David believe the story that saw a ghost? 1 2 3 4 5
26. What does she say that dies soon? 1 2 3 4 5
27. Sarah said she loved the well-known actor. 1 2 3 4 5
28. How do the tigers run fast? 1 2 3 4 5
29. My mom kept showing Julia some pictures of herself. 1 2 3 4 5
30. Sorry for him to whom I wonder which stories they have told. 1 2 3 4 5
31. The judge just said was not in agreement with the lawyer. 1 2 3 4 5
32. Jack and Rose had loved each other. 1 2 3 4 5
33. Who had your teacher said that failed the exam? 1 2 3 4 5
34. With whom is he going to have a discussion later? 1 2 3 4 5
35. How fast does the plane go? 1 2 3 4 5
36. I wonder if his sister loves himself. 1 2 3 4 5
37. What have they said will appear next week? 1 2 3 4 5
38. Look at those shoes that my mother has bought them for me. 1 2 3 4 5
39. How easy were the teacher's assignments? 1 2 3 4 5
40. The businessmen are devising a plan to rise their sales. 1 2 3 4 5