

# Schema-based English Achievement and Teacher Effectiveness

Ebrahim Khodadady  
Ferdowsi University of Mashhad, Iran

Beheshteh Shakhsi Dastgahian  
Ferdowsi University of Mashhad, International Branch, Mashhad, Iran

**Abstract**—This study explored whether the domain of teacher effectiveness as well as its underlying factors or genera relate significantly to English achievement when it is measured by schema-based cloze multiple choice item tests (S-Tests) and their tailored versions. To this end, the English Language Teachers' Attributes Scale (ELTAS) designed by Khodadady, Fakhrabadi and Azar (2012) and validated by Khodadady and Dastgahian (2014) with 1483 grade four senior high school (G4SHS) students in Iran was employed as a measure of the domain consisting of eleven genera, i.e., Qualified, Social, Proficient, Humanistic, Stimulating, Organized, Pragmatic, Systematic, Prompt, Exam-Wise, and Lenient. The S-Test designed by Khodadady and Ghergloo (2013) was also administered to 440 of 1483 G4SHS with whom the ELTAS had been validated. The participants' performance on the S-Test, the tailored S-Test as well as grade three final English examination were correlated with the domain and its genera. The results showed varying types and degrees of significant relationship between English achievement and the domain as well as its underlying genera. The findings are discussed and suggestions are made for future research.

**Index Terms**—language learning, teacher effectiveness, achievement, S-test

## I. INTRODUCTION

In order to determine what attributes of English language teachers render their teaching effective, Khodadady, Fakhrabadi and Azar (2012) [henceforth KF&A] collected those explored by Borg (2006), Brosh (1996), Elizabeth, May, and Chee (2008), Ghasemi and Hashemi (2011), Hildebrand, Wilson and Dienst (1971), Irby (1978), Park and Lee (2006), Pishghadam and Moafiyani (1998), Sherman et al (1987), Shishavan and Sadeghi (2009), and Suwandee (1995) and added them to the attributes brought up in teacher evaluation forms designed by various schools and universities in Canada and Iran. By resorting to their teaching experiences, KF&A also added a few attributes to their collected pool of features and designed their Effective English Language Teachers' Attributes Scales (ELTAS).

Khodadady and Dastgahian (2014) followed Khodadady (2010) and administered the ELTAS to 1483 grade four senior high school (G4SHS) students in Mashhad, Iran, and submitted their data to Principal Axis Factoring and Varimax with Kaiser Normalization. The results showed that the scale consists of eleven factors, i.e., Qualified, Social, Proficient, Humanistic, Stimulating, Organized, Pragmatic, Systematic, Prompt, Exam-Wise, and Lenient. These factors which were established statistically are viewed as genera in the microstructural approach of schema theory (Khodadady, 2013). The adoption of statistical factors as cognitive genera helps explain the construct of teacher effectiveness as a cognitive domain having a hierarchically perceived relationship with its constituting genera, species and types as shown in Figure 1. [The cognitive structure of single/phrasal words or schemata comprising psychological measures such as the ELTAS was first brought up by Khodadady and Bagheri (2014) and Khodadady and Dastgahian (2013)].

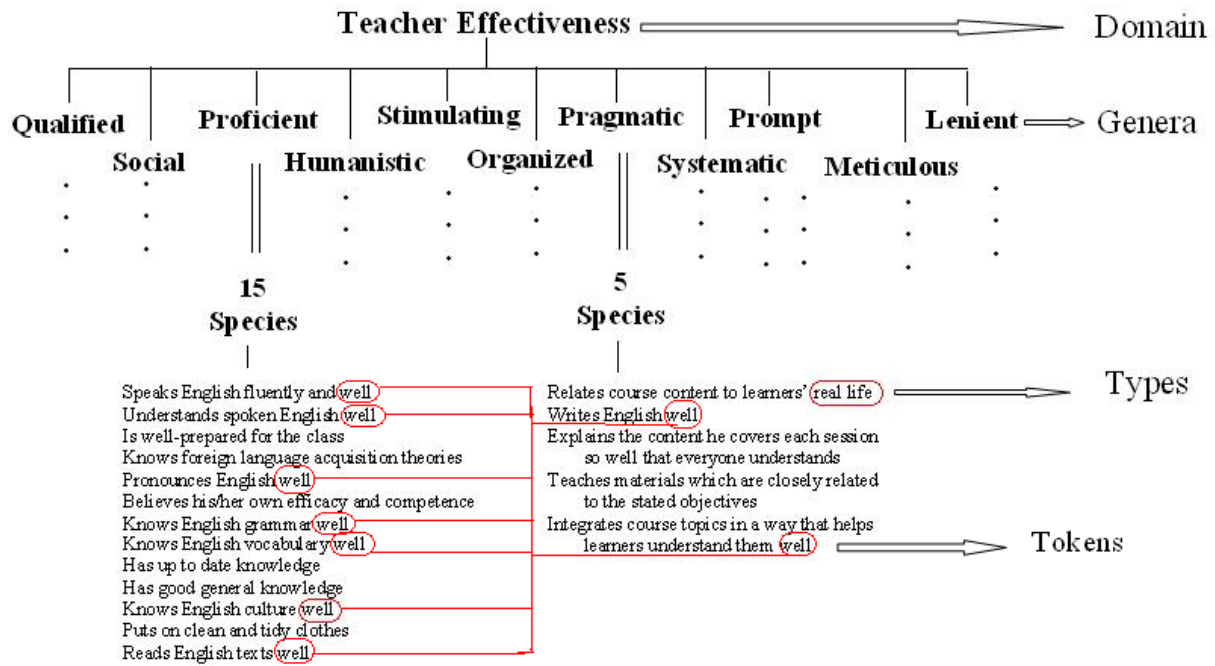


Figure 1. Tokens, types, species and genera constituting the cognitive domain of teacher effectiveness

For G4SHS students, the first factor represents the genus of a *Qualified* teacher who checks and marks assignments regularly, identifies and solves learning problems, evaluates learners regularly and monitors their progress during the term, assigns tasks requiring group work, knows learners' abilities, talents and weaknesses, identifies and pays attention to individual needs and differences, involves all students in learning and teaching processes, motivates students to learn English and do research, evaluates both qualitatively and quantitatively, tailors teaching to student needs, provides equal opportunities for participation, discussion and asking questions, takes learners attitudes towards learning into account even if they are negative, helps learners in and out of the class, teaches English tailored to students' ability levels, handles discipline through prevention, gives sufficient number of assignments, is demographic in his/her approach, is interested in students, e.g., calls them by their names, and their learning, is willing to negotiate changes to course content, is available to answer questions, specifies methods of evaluation clearly, encourages achievements and discourages unacceptable behaviours, exercises authority to control the class whenever necessary, and encourages and improves creativity in learners.

As the second factor underlying the ELTAS, the *Social* genus specifies a teacher who is cheerful and benevolent, has a good sense of humor, is friendly, is a dynamic and energetic person, establishes strong rapport with students, is good-tempered, caring, and patient, creates a relaxed and pleasant atmosphere in the class, is comfortable interacting with others, teaches English enthusiastically, maintains a welcoming environment for all students, creates self-confidence in learners, and follows social codes and values and treats learners well.

Contributing to his effectiveness, a *Proficient* teacher meets the requirements of the third genus established by G4SHS students when he speaks English fluently, pronounces English well, knows English vocabulary well, understands spoken English well, knows English grammar well, has up to date knowledge of course content, reads English texts well, has good general knowledge to answer the questions not directly related to the course content, teaches English in English, knows English culture well and knows foreign language acquisition theories. He also puts on clean and tidy clothes and is well-prepared for the class.

As the fourth factor underlying the ELTAS, the *Humanistic* genus represents an EFL teacher who respects all ideas, listens to student's opinions, responds logically to suggestions and criticisms, accepts constructive criticisms, is flexible and understands learners well, respects learners as real individuals, pays attention to students of all abilities, helps learners spot and overcome their weaknesses, and avoids discrimination and treats all fairly. The last attribute of *Humanistic* genus, i.e. avoiding discrimination, loaded acceptably neither on this factor nor on any other factors for grade three senior high school (G3SHS) students in KF&A's study, showing that students in senior high schools become more conscious of discrimination when they enter grade four.

As the fifth genus constituting teacher effectiveness, the *Stimulating* factor represents an EFL teacher who employs multimedia materials such as CDs and tapes, arouses interest in learning English through interesting activities, employs interesting learning activities and assignments, teaches how to learn English outside the classroom, e.g., watching certain programs, provides opportunities to use English through meaningful activities, chooses interesting materials to teach, and has creativity in teaching.

For G4SHS students an *Organized* teacher embodies the sixth genus of teacher effectiveness domain when he presents information at the right pace based on students' level of learning, reduces English language learning anxiety, enjoys teaching English, employs methods of evaluation consistent with course outline as initially presented, employs appropriate evaluation techniques, and states course objectives clearly.

As the seventh genus of teacher effectiveness domain, the *Pragmatic* factor specifies an EFL teacher who teaches materials which are closely related to the stated objectives, explains the content he covers each session so well that everyone understands, writes English well, relates course content to learners' real life, and integrates course topics in a way that helps learners understand them well. The schema *Pragmatic* has been borrowed from Applied Pragmatics. According to Crystal (1991) it "focuses on problems of interaction that arise in contexts where successful communication is critical, such as medical interviews, judicial settings and counseling" (p. 271).

Enacting the eighth genus of teacher effectiveness domain, a *Systematic* EFL instructor teaches systematically, organizes course content well in terms of hours and sessions, leaves and enters the class on time and divides class time appropriately for the different language skills based on lesson objectives. The teacher also possesses *Prompt* genus as the ninth factor underlying the ELTAS of G4SHS students when he is prompt in returning test results and returns tests/ assignments in time for subsequent work.

As the tenth genus involved in the domain of effective teaching, an *Exam-Wise* teacher emphasizes important points and materials and answers questions carefully and convincingly to G4SHS students. As the last factor underlying the ELTAS, the *Lenient* genus, however, represents an EFL teacher who ignores cheating, gives good grades, i.e., does not take it hard, and designs simple and easy tests. It is the only genus whose constituting species or attributes remain the same to both G3SHS and G4SHS students.

Upon validating the ELTAS with 1483 G4SHS students, Khodadady and Dastgahian (2014) [henceforth K&D] correlated the scale and its underlying genera with the students' self-reported scores on grade three final English examination (G3FEE) held nationally at the end of grade three high school year. Their results showed that not only the teacher effectiveness domain but also its eleven genera correlate significantly with the G3FEE. The present study is designed to find out whether the domain and its genera relate to English achievement when it is measured by a schema-based cloze multiple choice item test (S-Test) developed on the textbook G4SHS students study for their English course.

## II. METHODOLOGY

### A. Participants

Out of 1483 G4SHS students who took the ELTAS in K&D's study, four hundred forty, 297 female (67.5%) and 143 male (32.5%), took the S-Test designed by Khodadady and Ghergloo (2013). Their age ranged between 16 and 21 (Mean = 17.86, SD = .57). They were studying in ten schools among which Rangraz, Hekmat, Malek Ashtar, Imam Ali and Azadegan were public whereas Nokhbegan Toos and Imam Reza were private high schools. Alameh Tabatabaee was the only semi-public school whose G4SHS students took the S-Test. The participants were speaking Armenian (n= 1, .2%), English (n= 10, 2.3%), Persian (n= 416, 4.5%), Kurdish (n= 6, 1.4%), and Turkish (n= 7, 1.6%) as their mother language. Most of them were studying sciences (n=208, 47.3%) followed by humanities (n=184, 41.8%). Mathematics was, however, studied by just 48 (10.9%) participants.

### B. Instrumentation

A Demographic Scale and English Language Teachers' Attributes Scale were used in the study. Two measures of content-based English achievement were also employed: A schema-based cloze multiple choice item test (S-Test) and grade three final English examination (G3FEE).

#### 1 Demographic Scale

The Demographic Scale (DS) utilized in this study consisted of four questions dealing with the age, gender, and mother language of participants.

#### 2 The English Language Teachers' Attribute Scale

The English Language Teachers' Attribute Scale (ELTAS) developed and validated by K&D was used in this study. It consists of the noun phrase "My English teacher ..." providing the context for the 92 linguistic statements representing certain cognitive species such as "accepts constructive criticisms." Each species was presented along with five choices with which the participants were required to completely agree, agree, to some extent agree, disagree or completely disagree. K&D administered the ELTAS to 1483 G4SHS students in Mashhad, Iran, and extracted eleven genera described in the introduction section of this paper. The ELTAS is a highly reliable measure of English teacher effectiveness domain at G4SHSs ( $\alpha=.98$ ). The alpha reliability coefficient reported for its Qualified, Social, Proficient, Humanistic, Stimulating, Organized, Pragmatic, Systematic, Prompt, Exam-Wise, and Lenient genera are 0.93, 0.93, 0.90, 0.88, 0.84, 0.78, 0.80, 0.80, 0.77, 0.73 and 0.39, respectively.

#### 3 Schema-Based Close Multiple-Choice Item Test

The Schema-Based Close Multiple-Choice Item Test (S-Test) developed by Khodadady and Ghergloo (2013) was employed to measure the participants' EFL achievement in this study. By resorting to the microstructural approach of schema theory (Khodadady, 2008), they parsed, codified and statistically analyzed the schemata comprising the

textbook *Learning to Read English for Pre-University Students* (Birjandi, Sarab & Samimi, 2012) taught to G4SHS students as shown in Table 1. As can be seen, the eight reading passages and vocabulary sections of the textbook consist of 5790 schema tokens and 1578 schema types in general. More specifically, they comprise 2915 (50.3%) semantic, i.e., adjectives, adverbs, nouns and verbs, 2256 (39.5%) syntactic, i.e., conjunctions, determiners, prepositions and pronouns, and syntactic verbs, and 586 (10.2%) parasyntactic tokens, i.e., abbreviations, interjections, names, numerals, para-adverbs and particles. In terms of types, however, the textbook comprises 1174 semantic (74.4%), 209 syntactic (13.2%) and 195 parasyntactic (12.4%) schemata.

TABLE 1  
THE LINGUISTIC SCHEMA DOMAIN TOKENS AND TYPES COMPRISING THE PRE-UNIVERSITY ENGLISH TEXTBOOK

Linguistic Domains	Tokens		Types	
	Frequency	Percent	Frequency	Percent
Semantic	2915	50.3	1174	74.4
Syntactic	2286	39.5	209	13.2
Parasyntactic	589	10.2	195	12.4
Total	5790	100.0	1578	100.0

Khodadady and Ghergloo (2013) chose at least one paragraph from each of the eight lessons comprising the textbook *Learning to Read English for Pre-University Students* (Birjandi, Sarab, & Samimi, 2012) to develop the S-Test. From the nine paragraphs forming the test, they selected ninety single/phrasal schemata, deleted and offered them as the keyed responses. In order to find out whether the test takers could activate their knowledge of the deleted schemata given as the keyed responses and relate them to the schemata comprising the paragraphs, three competitives, traditionally known as distracters (Khodadady, 1999), were chosen from among the 1578 schema types comprising the content of the whole textbook. As the first and second S-Test items given below show, the competitives have syntactical, semantic and discoursal relationships with the keyed response.

#### Two Example S-Test Items

Aerobics is a word for ... (1) oxygen, and aerobic exercise is any kind of activity that ... (2) your muscles use oxygen.

1	A	lacking	B	needing*	C	avoiding	D	missing
2	A	makes*	B	builds	C	produces	D	creates

The three competitives, *lacking*, *avoiding* and *missing*, of the first item, for example, are all syntactically related to the keyed response *needing* in being verbs. Semantically, *needing* shares the semantic feature of "being essential" with the competitives *lacking* and *missing*. However, the contextual schema *aerobics* does not qualify as an animate agent to *miss* or *lack* a substance such as oxygen and thus *lacking* and *missing* become discursively inappropriate to be chosen as the keyed response. Similarly, *avoiding* indicates the necessity of keeping away from something during aerobics, however, it does not relate to the keyed response in that its semantic feature of "keeping away" renders it unacceptable.

Khodadady and Ghergloo (2013) administered the 90-item S-Test to two hundred eighty three G4SHS students in Dargaz, Iran, in order to explore its empirical validity with a C-Test developed on the same text. The results showed that the S-Test is a highly challenging measure of content-based English achievement because its mean score was 39.90. The high difficulty level of the test resulted in having almost one third of items, i.e., 29 out of 90, function well, i.e., they had the acceptable item facility indices falling between .25 and .75 and item discrimination indices of .20 and higher. The alpha reliability coefficient of the test is .75. It enjoys high empirical validity because it correlated at the highest possible level with the 100 item C-Test when it was concurrently administered to the same test takers, i.e.,  $r = .99, p < .01$ .

#### 4 Grade Three Final English Examination

The grade three final English examination (G3FEE) designed by the Ministry of Education was also employed to measure the participants' English achievement. It consists of 13 sections containing various types of questions ranging from restoring one missing letter to writing complete response to questions based on pictorial inputs. It is held nationally and the students' responses are marked by two teachers and their scores are reported out of 20. The cut-off scores of 10 and higher are adopted as indicators of passing the English course successfully. According to Dastgahian (2014), the KR-21 reliability coefficient of the G3FEE taken by 1483 students is 0.73.

#### C. Procedure

While administering the ELTAS to 1483 G3SHS students, the researchers brought up the relevance and necessity of administering the S-Test designed by Khodadady and Ghergloo (2013) as a measure of English achievement and asked their teachers whether they could provide them with another session to administer the test. The teachers of 440 students in the ten schools mentioned earlier in section 2.1 agreed to hold the test provided that the researchers supplied them with their students' scores on the S-Test. Upon coordinating with the teachers the test was administered to their students in a single session after two weeks. Both the ELTAS and S-Test were administered under standard conditions lasting for about 30 and 90 minutes, respectively. The participants' scores on the S-Test were reported to the teachers as agreed.

#### D. Data Analysis

For determining the internal validity of S-Test and establishing its tailored version (e.g., Kamimoto, 1993) the functioning of its items were statistically determined by employing two indices. Point biserial correlation coefficients were estimated as the item discrimination (ID) values by correlating each individual item with the total score obtained on the S-Test. Following Thorndike (2005), items whose IDs fell below 0.20 were considered as malfunctioning and removed from validity analyses for the tailored version of the S-Test. The number of correct responses given to each item was also divided by the total number of answers to obtain item facility (IF) indices. Baker (1989) was followed and the IFs falling below 0.25 and above 0.75 were considered malfunctioning. Cronbach's Alpha was estimated to determine the internal consistency of the 90-item S-Test and its tailored version. Finally, Pearson correlations were employed to explore the relationship between the ELTAS as well as its genera with both the S-Test and its tailored version. All statistical analyses were performed by using IBM SPSS statistics 20.0 to explore the hypotheses below.

H1. The S-Test and its tailored version will correlate significantly with the ELTAS.

H2. The S-Test and its tailored version will correlate significantly with the genera underlying the ELTAS.

### III. RESULTS

Table 2 presents the descriptive as well as the reliability estimates of the ELTAS and its eleven underlying factors as established by 1483 G4SHS student in K&D's study and 440 of those student in the present. As can be seen, the mean score of the scale taken by the latter (360.60) is slightly higher than that of the former (349.31). The higher magnitude of the mean scores of the latter group holds equally true for the factors as well, e.g., 89.34 for 440 G4SHS student vs. 87.39 for 1483. As it can also be seen, the participants in this study have reliably evaluated their English teachers' effectiveness as a domain measured by the ELTAS ( $\alpha=.98$ ) and their Qualified ( $\alpha=.94$ ), Social ( $\alpha=.94$ ), Humanistic ( $\alpha=.91$ ), Proficient ( $\alpha=.88$ ), Stimulating ( $\alpha=.85$ ), Systematic ( $\alpha=.81$ ), Pragmatic ( $\alpha=.79$ ), Organized ( $\alpha=.76$ ), Prompt ( $\alpha=.76$ ), Exam-Wise ( $\alpha=.71$ ), and Lenient ( $\alpha=.40$ ) genera.

TABLE 2  
DESCRIPTIVE STATISTICS AND RELIABILITY ESTIMATES OF ELTAS AND ITS UNDERLYING FACTORS

No	Factor	No of items	K&D study (n=1483)			This study (n=440 of 1483)		
			Mean	SD	Alpha	Mean	SD	Alpha
1	Qualified	24	87.39	15.912	.928	89.34	16.357	.935
2	Social	14	55.55	10.480	.930	57.97	9.750	.935
3	Proficient	15	62.76	9.056	.901	65.31	7.632	.877
4	Humanistic	9	34.51	6.832	.883	35.20	7.307	.907
5	Stimulating	7	22.55	5.890	.844	23.83	6.007	.853
6	Organized	6	22.88	4.381	.782	23.69	4.037	.758
7	Pragmatic	5	18.97	3.969	.802	19.63	3.803	.789
8	Systematic	5	20.53	3.716	.801	20.68	3.843	.806
9	Prompt	2	7.57	2.005	.772	8.01	1.909	.758
10	Exam-Wise	2	8.42	1.569	.732	8.66	1.455	.710
11	Lenient	3	8.16	2.514	.391	8.29	2.520	.402
	ELTAS	92	349.31	54.345	.976	360.60	53.242	.977

Table 3 presents the psychometrics of the 90 items comprising the S-Test. As can be seen, out of 90 items 39 (43.3%) have functioned well, i.e., 6, 7, 23, 32, 40, 42, 45, 46, 47, 49, 51, 52, 53, 54, 56, 57, 59, 60, 62, 63, 65, 66, 67, 69, 70, 73, 75, 76, 77, 78, 79, 81, 84, 85, 86, 87, 88, 89, and 90. These results show that the S-Test enjoys a higher level of internal validity than a language proficiency test such as Test of English as a Foreign Language (TOEFL) does. Khodadady (2012) administered a disclosed TOEFL consisting of 115 multiple choice items i.e., 30 structure, 25 written expressions, 30 vocabulary, 30 reading comprehension items, to 430 university students and tailored the items. His findings showed that only 47 items (40.9%) had functioned well.

TABLE 3  
PSYCHOMETRICS OF 90 ITEMS COMPRISING S-TEST

No	IF	ID	No	IF	ID	No	IF	ID	No	IF	ID	No	IF	ID
1	.96	.366	19	.78	.408	37	.90	.470	55	.82	.622	73	.57	.506
2	.81	.248	20	.28	.146	38	.87	.416	56	.48	.411	74	.77	.565
3	.91	.224	21	.80	.329	39	.76	.493	57	.70	.458	75	.68	.495
4	.93	.289	22	.77	.278	40	.53	.383	58	.87	.556	76	.68	.467
5	.91	.207	23	.75	.300	41	.76	.429	59	.61	.453	77	.71	.427
6	.75	.203	24	.87	.281	42	.54	.416	60	.72	.522	78	.54	.355
7	.70	.278	25	.77	.413	43	.81	.473	61	.78	.524	79	.58	.424
8	.79	.361	26	.81	.383	44	.89	.541	62	.48	.507	80	.85	.537
9	.86	.365	27	.80	.486	45	.63	.426	63	.48	.302	81	.67	.578
10	.78	.410	28	.79	.491	46	.58	.382	64	.79	.481	82	.76	.556
11	.83	.440	29	.82	.426	47	.61	.510	65	.42	.332	83	.76	.551
12	.78	.319	30	.82	.446	48	.79	.562	66	.74	.549	84	.58	.497
13	.87	.360	31	.76	.454	49	.69	.255	67	.65	.428	85	.74	.556
14	.89	.403	32	.69	.395	50	.83	.524	68	.88	.534	86	.70	.601
15	.80	.362	33	.80	.489	51	.57	.362	69	.53	.362	87	.65	.457
16	.87	.454	34	.88	.374	52	.50	.395	70	.75	.602	88	.70	.500
17	.90	.376	35	.76	.307	53	.55	.456	71	.88	.593	89	.62	.509
18	.84	.328	36	.90	.454	54	.75	.476	72	.86	.568	90	.70	.410

Table 4 presents the descriptive statistics as well as the reliability estimates of the original S-Test consisting of 90 items and its 39-item tailored version. As can be seen, in addition to having internal validity, the S-Test enjoys a high level of reliability ( $\alpha = .95$ ). As it can also be seen, its tailored version is highly reliable ( $\alpha = .91$ ), too. These results are compatible with other studies in which S-Tests have been employed for research purposes. Khodadady, Pishghadam, and Alaei (2012), for example, designed a 43-item S-Test on *English Book 3* (Birjandi, Nouroozi, & Mahmoodi, 2010) and administered it to 477 G3SHS students and reported an alpha coefficient of .94.

TABLE 4  
DESCRIPTIVE STATISTICS OF ORIGINAL AND TAILORED S-TESTS (N = 440)

Tests	Mean	SD	Mean IF	Mean ID	Skewness	Kurtosis	Alpha
S-Test 90	66.32	16.090	.74	.42	-1.021	.929	.95
S-Test 39	24.55	8.625	.63	.43	-.245	-.552	.91
G3FEE	16.69	2.69	-	-	-.886	.515	-

As it can also be seen in Table 4 above, neither the alpha reliability coefficient of the G3FEE nor the mean IF and ID could be estimated because the researchers could not have access to the participants' responses as regards the items comprising the examination. KR-21 formula provided by Hatch and Lazaraton (1991) was therefore employed and the coefficient of .66 was obtained. In addition to having a relatively low value, it does not reveal the reliability level of the examination because the assumption upon which the formula is based, i.e., having items of the same difficulty, is violated (see Thorndike, 2005)

Table 5 presents the correlation coefficients obtained between the ELTAS as well as its eleven underlying factors, the G3FEE, S-Test and its tailored version. As can be seen, the ELTAS correlates significantly with the G3FEE ( $r = .27, p < .01$ ), explaining 7.29% of variance in each other. Among the eleven genera forming teacher effectiveness domain, the *Pragmatic* and *Qualified* genera correlate the highest with the G3FEE ( $r = .29, p < .01$ ) followed by *Exam-Wise* ( $r = .28, p < .01$ ). *Social*, *Organized*, *Humanistic*, *Systematic*, *Lenient*, and *Proficient* genera. They reveal successively lower significant relationships with the G3FEE, i.e., .25, .25, .23, .19, .18, and .17,  $p < .01$ , respectively. These results are largely in line with K&D's findings and establish stronger relationships between teacher effectiveness domain and English achievement on the one hand and nine genera of the domain and English achievement as measured by the G3FEE on the other.

TABLE 5  
CORRELATIONS BETWEEN ELTAS, ITS GENERA, G3FEE AND S-TEST

Scale and genera	G3FEE N=440	S-Test 90 N=440	S-Test 39 N=440	Scale and genera	G3FEE N=440	S-Test 90 N=440	S-Test 39 N=440
ELTAS	.272**	-.080	-.113*	Organized	.250**	-.085	-.105*
Qualified	.288**	-.107*	-.153**	Pragmatic	.292**	-.042	-.087
Social	.254**	-.055	-.074	Systematic	.193**	-.127**	-.159**
Proficient	.172**	-.070	-.099*	Prompt	.093	-.100*	-.147**
Humanistic	.226**	-.045	-.080	Exam-Wise	.284**	-.034	-.086
Stimulating	.067	.013	.057	Lenient	.178**	.018	-.011

\* Correlation is significant at the 0.05 level (2-tailed)

\*\* Correlation is significant at the 0.01 level (2-tailed)

However, as it can be seen in Table 5 above, the ELTAS does not correlate significantly with the 90-item S-Test. It does, nonetheless, correlate significantly but *negatively* with its tailored version ( $r = -.11, p < .05$ ) confirming the first hypothesis partially. Five genera forming the domain of teacher effectiveness correlate significantly with the tailored

S-Test as well, i.e., *Systematic* ( $r = -.16, p < .01$ ), *Qualified* ( $r = -.15, p < .01$ ), *Prompt* ( $r = -.15, p < .01$ ), *Organized* ( $r = -.11, p < .05$ ), and *Proficient* ( $r = -.10, p < .05$ ). These results partially provide a positive answer to the second hypothesis that *the S-Test and its tailored version will correlate significantly with the genera underlying the ELTAS*.

#### IV. DISCUSSIONS AND CONCLUSION

Neither learning nor teaching language has been discussed in the literature recently by relating it to testing. Brown (2007), for example, neither treated testing anywhere in his *Principles of Language Learning and Teaching* nor devoted any entry to introduce it in his glossary. Similarly, Richards, Platt and Platt (1992) defined testing as “the use of TESTS, or the study of the theory and practice of their use, development, evaluation, etc” (p. 377) without associating it with *Teaching & Applied Linguistics* as the main theme of their dictionary. The same position is held by Richards and Schmidt (2010, p. 592) in the fourth edition of the dictionary. Khodadady (1999), however, argued that “testing is an indispensable part of language teaching or pedagogy” (p. 62). The findings of the present study support his position.

As a measure of teacher effectiveness domain, the ELTAS explains 7.4% of variance in G4SHS students’ achievement when it is measured by G3FEE whereas similar measures of the domain fail to do so. Feizbakhsh (2010), for example, could not establish any significant relationship between Characteristics of Successful Iranian EFL Teachers Questionnaire (CSIEFLTQ) designed by Moafian and Pishghadam (2008) and English achievement of her 1461 participants, implying that there is no relationship between teacher effectiveness and testing achievement. Feizbakhsh could not establish any significant relationship between the two domains because the CSIEFLTQ does not include the attributes through which EFL teachers test their learners’ English achievement. These attributes are evaluated by their learners through the schema tokens, types and species which constitute the measures such as the CSIEFLTQ and ELTAS.

In addition to establishing a significant relationship between teaching and testing, the ELTAS validated by Khodadady and Dastgahian (2014) highlights the effect of subjective and objective testing on EFL learners’ evaluation of their teachers’ effectiveness as a schema domain. As a subjective test held nationally, the G3FEE explains 7.4% of variance in the domain ( $r = .272, p < .01$ ), showing that the more effective the teachers, the higher the learners achieve English. G3FEE is, however, subjective because no one knows how providing the missing letters of 16 mutilated words given as eight isolated sentences such as “There are two kinds of illnesses. Ph\_sical and m\_ntal” show English achievement and deserves a score of four out of 40. The test takers do, however, receive a score of three out of 40 if they restore six missing words in six isolated sentences such as “A lab is a suitable place to do some ... on acid”.

In contrast to G3FEE, the S-Test designed by Khodadady and Ghergloo (2013) is an objective measure of English achievement whose 90 keyed responses, 55 semantic (61.1%) and 35 syntactic (38.9), are presented along with three competitiveness in nine paragraphs. The G4SHS students had to comprehend the meaning of the keyed responses in the context of paragraphs in order to select them from among 270 competitiveness comprising the materials taught during the school year. While scoring an item such as “A lab is a suitable place to do some ... on acid” is subjective to some extent and depends on the marker, the keyed responses of the S-Tests are determined in advance and can even be machine-marked. The very objectivity of the 90 item S-Test has resulted in its being unrelated to the ELTAS as a measure of teacher effectiveness domain.

It is further argued in this paper that the S-Test does not relate to the ELTAS because it contains 50 items whose answers were known to both low and high achievers. Its 39-item tailored version, however, relates significantly but *negatively* to the ELTAS ( $r = -.113, p < .01$ ) because it could differentiate achievers from each other. The performance of high, middle and low achievers on the test does in fact reveal the nature of negative relationship as shown in Figure 1. As can be seen, the achievers’ mean scores on the domain are different. The difference, according to the One-Way ANOVA analysis, is significant, i.e.,  $F(2, 436) = 8.45, p < .001$ . The Scheffe Post hoc test, nonetheless, showed that the 78 high achievers (mean=341.76, SD=65.68) rated their EFL teachers’ effectiveness significantly lower than their 299 middle achiever counterparts (mean=367.36, SD=47.46)

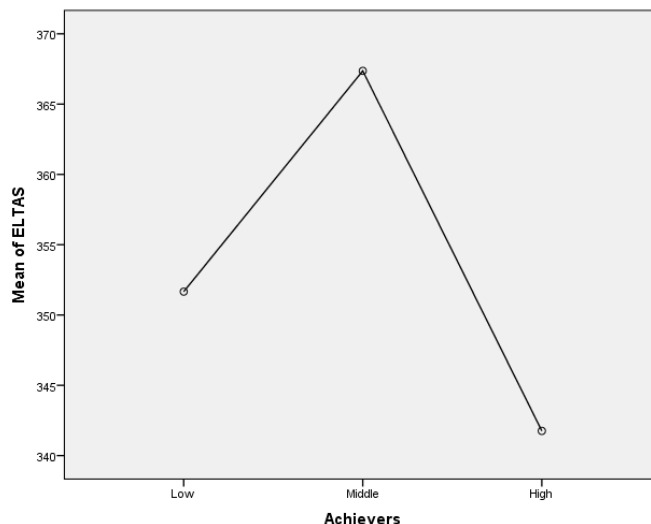


Figure 1. High, Middle And Low Achievers' Evaluation Of Their Teachers' Effectiveness

Similar significant but *negative* relationships were found between the tailored S-Test and Systematic ( $r = -.159$ ,  $p < .01$ ), Qualified ( $r = -.153$ ,  $p < .01$ ), Prompt ( $r = -.147$ ,  $p < .01$ ), Organized ( $r = -.105$ ,  $p < .05$ ), and Proficient ( $r = -.099$ ,  $p < .05$ ) genera of the domain. The relationships are negative because the evaluation of high achievers regarding the five genera of their teachers' effectiveness is lower than their low and/or middle achieving classmates. One Way ANOVA analysis of ratings on Systematic genus, for example, showed the three groups of achievers' evaluation of systematic genus differed significantly from each other, i.e.,  $F(2, 436) = 12.16$ ,  $p < .001$ , as shown in Figure 2. The Scheffe post hoc test showed that high achievers rated the systematicity of their teacher (mean=18.86,  $SD=4.59$ ) significantly lower than both middle (mean=21.20,  $SD=3.46$ ) and low achievers (mean=20.45,  $SD=3.93$ ) did.

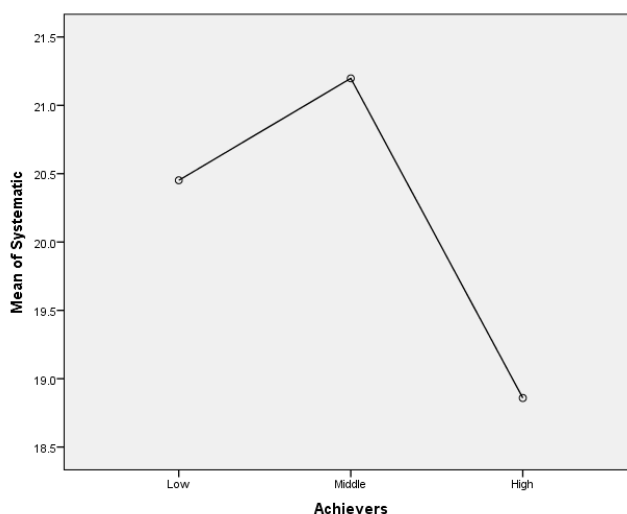


Figure 2. High, middle and low achievers' evaluation of their teachers' Systematic genus

In sharp contrast to G3FEE, the S-Test does not show any significant relationship with Lenient, Social, Humanistic, Exam-Wise, and Pragmatic genera of teacher effectiveness domain, indicating that S-Tests enjoy discriminant validity (e.g. Kline, 2005). In other words, S-Tests measure English achievement as an ability which is different from teachers' being Lenient, Social, Humanistic, Exam-Wise, and Pragmatic in their classes. G3FEE is, however, designed in a way which allows teachers to teach English to their learners on Lenient, Social, Humanistic, Exam-Wise, and Pragmatic grounds. The inclusion of these non-ability genera in their evaluation of learners' English attainment, results in high achievers' negative evaluation of their ability-based Systematic, Qualified, Prompt, Organized, and Proficient genera.

#### REFERENCES

- [1] Baker, D. (1989). *Language testing: A critical survey and practical guide*. London: Edward Arnold.
- [2] Birjandi, P., Nouroozi, M., & Mahmoodi, G. (2010). *English book 3*. Tehran: Ketabhaye Darsie Iran Publication.
- [3] Birjandi, P., Sarab, M. R. A., P., & Samimi, D. (2012). *Learning to read English for pre-university students*. Tehran: Ketabhaye Darsie Iran Publication.
- [4] Borg, S. (2006). The distinctive characteristics of foreign language teachers. *Language Teaching Research*, 10 (1), 3–31.



- [5] Brosh, H. (1996). Perceived characteristics of the effective language teacher. *Foreign Language Annals*, 29, 125–38.
- [6] Brown, H. D. (2007). Principles of language learning and teaching (5<sup>th</sup> ed.). NY: Pearson Education.
- [7] Crystal, D. (1991). A dictionary of linguistics and phonetics (3rd ed.). Oxford, UK: Blackwell.
- [8] Elizabeth, C. L., May, C. M., Chee, P. K. (2008). Building a model to define the concept of teacher success in Hong Kong. *Teaching and Teacher Education*, 24, 623–634.
- [9] Feizbakhsh, M. (2010). The role of emotional intelligence in EFL teachers' success at schools and institutes. Unpublished MA thesis, Ferdowsi University of Mashhad, Iran.
- [10] Ghasemi, B., Hashemi, M. (2011). The study of the characteristics of successful English language teachers from the view point of the English language students of Islamic Azad University, Hamedan Branch. *Procedia- Social and Behavioral Sciences*, 28, 411 – 415.
- [11] Hatch, E., & Lazaraton, A. (1991). The research manual: Design and statistics for applied linguistics. Boston, Mas.: Heinle & Heinle.
- [12] Hildebrand, M., Wilson, R. C., & Dienst, E. R. (1971). Evaluating university teaching. Berkeley, California: Center for Research and Development in Higher Education, University of California.
- [13] Irby, D. M. (1978). Clinical teacher effectiveness in medicine. *Journal of Medical Education*, 53, 808-815.
- [14] Kamimoto, T. (1993). Tailoring the test to fit the students: Improvement of the C-test through classical item analysis. *Fukuoka Women's Junior College Studies*, 30, 47-61.
- [15] Khodadady, E. (1999). Multiple-choice items in testing: Practice and theory. Tehran: Rahnama.
- [16] Khodadady, E. (2008). Schema-based textual analysis of domain-controlled authentic texts. *Iranian Journal of Language Studies (IJLS)*, 2(4), 431-448.
- [17] Khodadady, E. (2010). Factors underlying characteristics of effective English language teachers: Validity and sample effect. *Iranian Journal of Applied Linguistics*, 13 (2), 47-73.
- [18] Khodadady, E. (2013). Research principles, methods and statistics in applied linguistics. Mashhad: Hamsayeh Aftab.
- [19] Khodadady, E., & Bagheri, N. (2014). Development and validation of an Islamic religious orientation scale with pilgrims of Imam Reza shrine. *Pensee Journal*, 76(2), 156-170.
- [20] Khodadady, E., & Dastgahian, B. D. (2013). Logical-mathematical intelligence and its relationship with English language proficiency. *American Journal of Scientific Research*, 90, 57-68.
- [21] Khodadady, E., & Dastgahian, B. S. (2015). Teacher Effectiveness, Educational Grade and English Achievement. Theory and Practice in Language Studies, 5(8), xxx-xxx.
- [22] Khodadady, E., Fakhrabadi, G. Z., & Azar, K. H. (2012). Designing and validating a comprehensive scale of English Language Teachers' Attributes and establishing its relationship with achievement. *American Journal of Scientific Research*, 82, 113-125.
- [23] Khodadady, E., & Ghergloo, E. (2013). S-Tests and C-Tests: Measures of content-based achievement at grade four of high schools. *American Review of Mathematics and Statistics*, 1(1), 1-16.
- [24] Kline, T. J. B. (2005). Psychological testing: A practical approach to design and evaluation. Thousand Oaks: Sage.
- [25] Moafian, F., & Pishghadam, R. (2008). Construct validation of a questionnaire on characteristics of successful English language teachers. *Pazhuhesh-e Zabanheye Khareji Journal* (University of Tehran), 54, 127-142.
- [26] Park, G. P., & Lee, H. W. (2006). The characteristics of effective English teachers as perceived by high school teachers and students in Korea. *Asia Pacific Education Review*, 7(2), 236-248.
- [27] Richards, J. C., Platt, J., & Platt, H. (1992). Longman Dictionary of language teaching & applied linguistics (2<sup>nd</sup> ed.). Essex: Longman.
- [28] Richards, J. C., & Schmidt, R. (2010). Longman Dictionary of language teaching & applied linguistics (2<sup>nd</sup> ed.). Harlow, England: Longman.
- [29] Sherman, T. M., Armistead, L P., Fowler, F., Barksdale, M. A., & Reif, G. (1987). The quest for excellence in university teaching. *Journal of Higher Education*, 58, 676-84.
- [30] Shishavan, H. B., & Sadeghi, K. (2009). Characteristics of an Effective English Language Teacher as Perceived by Iranian Teachers and Learners of English. *English Language Teaching*, 2(4), 131-143.
- [31] Suwandee, A. (1995). 'Students' perceptions of university instructors' effective teaching characteristics', *SLLT Journal*, 5, 6–22.
- [32] Thorndike, R. M. (2005). Measurement and evaluation in psychology and education (7<sup>th</sup> ed.). Upper Saddle River, NJ: Pearson Education.



**Ebrahim Khodadady** was born in Iran in 1958. He obtained his PhD in Applied Linguistics from the University of Western Australia in 1998. He holds TESL Ontario and Canadian Language Benchmarks Placement Test (CLPBPT) certificates and has taught English as a first, second and foreign language to high school and university students in Australia, Canada and Iran.

He is currently an academic member of Ferdowsi University of Mashhad, International Branch, Mashhad, Iran. He was invited as a VIP by Brock University in Canada in 2004 and served as the Associate Director of Assessment Center at George Brown College in Toronto for almost a year. His published books are *Multiple-Choice Items in Testing: Practice and Theory* (Tehran, Rahnama, 1999), *Reading Media Texts: Iran-America Relations* (Sanandaj, Kurdistan University, 1999) and *English Language Proficiency Course: First Steps* (Sanandaj, Kurdistan University, 2001). His main research interests are Testing, Language Learning and Teaching.

Dr. Khodadady is currently a member of Teaching English Language and Literature Society of Iran (TELLSI), TESL Ontario and European Society for Translation Studies. He is on the editorial board of *Ferdowsi Review: An Iranian Journal of TESL, Literature and Translation Studies* and has reviewed some research papers for *Iranian Journal of Applied Linguistics* and *TESL Canada Journal* as a guest reviewer.



**Beheshteh Shakhsi Dastgahian** was born in Iran in 1973. She is the MA student in Applied Linguistics in Ferdowsi University of Mashhad, International Branch, Mashhad, Iran. She is currently teaching English as foreign language to high school and pre-university students in the educational district three in Mashhad, Iran. She has participated in all educational programs related to the State Department of Education up to now, and hold a Diploma in Computer Skills. Her main research interests include Language Proficiency and Multiple Intelligences.