

The Relationship between Disciplinary Distance, Language Proficiency and EAP Test Performance

¹Farhad Mazlum*

²Sevda Shamameh

IJEAP- 2004-1532

³Asghar Salimi

Received: 2020-04-18

Accepted: 2020-10-03

Published: 2020-10-09

Abstract

The role of language proficiency in EAP test performance is a hotly contested issue in the field, and, more specifically, disciplinary distance is assumed to mediate such a relationship. The purpose of this study was twofold: a) to investigate the role of English proficiency in teacher-constructed ESAP tests serving achievement purposes; and b) to examine whether with variations in disciplinary distance the role of language proficiency would fluctuate. To this aim, 110 English majors were given TOEFL followed by four ESAP tests: law and psychology as neighbor and chemistry and geology as non-neighbor disciplinary fields. Data analyses revealed that there were statistically significant differences across proficiency levels, and the participants took advantage of their proficiency stock more, and statistically so, when sitting for disciplinarily close rather than distant ESAP tests. Also, the positive effect of disciplinary adjacency was not bound to any linguistic thresholds. Except for geology, moderate to large positive correlations were found between proficiency and ESAP test scores. Findings and implications are discussed vis-à-vis studies addressing the role of language proficiency in EAP assessment in general and ESAP assessment in Iran in particular.

Keywords: EAP/ESP Test, Language Proficiency, Academic Discipline

1. Introduction

Once described as a “fringe branch of English for Specific Purposes (ESP) in the early 1980s” (Hyland, 2006), English for Academic Purposes (EAP) is now a thriving field of study in English language teaching. Driven by pressing demands to fulfill international students’ academic and communicative needs at English-medium universities and colleges, EAP has quickly gained a high profile in both theory and practice (Hyland, 2017; Jordan, 1997). Since its birth and emergence, EAP has expanded markedly from practical considerations (Flowerdew & Peacock, 2001) to genre studies and cultural, social, cognitive and linguistic issues of academic settings (Hyland, 2006, 2015). The further application of this field in socio-political, critical and ideological concerns (Benesch, 2001; Phillipson, 1992) has enriched EAP in terms of professional maturity and sophistication.

Following ESP assessment, EAP assessment is undertaken for proficiency, placement and achievement purposes (Dudley-Evans & St. Johns, 1998; Hutchinson & Waters, 1987). Whereas little is known about EAP tests used for placement and achievement purposes (Dudley-Evans & St. Johns, 1998), EAP assessment literature is relatively rich with studies addressing EAP test development and use for proficiency purposes. In other words, international EAP tests, such as IELTS (International English Language Testing System) and TEEP (Test of English for Educational Purposes) have captured the attention of both researchers and practitioners to a greater extent (Fulcher, 1999, 2000; Clapham, 2000; Alderson, 1993) to name only a few. Such studies, conducted in response to practical problems encountered by English-medium universities in admitting non-native English speakers for

¹Assistant Professor in TEFL, mazlumzf@yahoo.com; English Department, Faculty of Human Sciences, University of Maragheh, Maragheh, Iran.

²MA in TEFL, sevda.shamameh@yahoo.com; English Department, Faculty of Human Sciences, University of Maragheh, Maragheh, Iran.

³Assistant Professor in TEFL, asgharsalimi356@gmail.com; English Department, Faculty of Human Sciences, University of Maragheh, Maragheh, Iran.

undergraduate programs, have led to important changes in IELTS, for example, and its modules (Clapham, 2000).

Then, there were disagreements among researchers as to what extent language proficiency influences students' overall scores in reading texts coming from their specialized fields of study versus texts selected from distant academic disciplines. Such investigations have attempted to explain EAP test takers' performance on reading comprehension tests of IELTS (e.g. Peretz & Shoham, 1990; Alderson & Urquhart, 1985a & 1985b). Evoking these studies were the facts that "language proficiency accounts for most of the variance in EAP test scores", hence, "increased specificity by module proliferation is unnecessary" (Fulcher, 1999, p.232) unless for face and response validity arguments (Alderson, 1993).

Information on ESP/EAP assessment in Iran is scanty. Few local research studies (e.g. Salmani-Nodoushan, 2003; Afghari & Tavakoli, 2004) addressing key issues in the field of ESP/EAP assessment exist. Generally speaking, ESP in Iran is characterized by its unique teaching and testing peculiarities. First, unlike other settings where "only a small proportion of EAP teachers may be personally responsible for creating in-house assessments" (Schmitt & Hamp-Lyons, 2015, p.5), Iranian EAP practitioners are teacher-testers; they are frequently engaged in test construction, too. Second, Iranian ESP and EAP teachers tend to view content technicality and discipline specificity as the prime feature of the courses resulting in substitution of real with carrier content (Mazlum, 2020). Such a deeply entrenched belief affects not only Iranian ESP/EAP practitioners' teaching practices, but also their assessment behaviors. Iranian ESP/EAP practitioners devote a good amount of their time to the explanation (Rajabi, Kiani, & Maftoon, 2012) or translation (Rezaee & Kazempourian, 2017) of carrier content (i.e., highly discipline-specific sentences and technical terms). Subsequently, they, particularly content specialists, design test tasks that primarily require content knowledge leading a test taker to say, "The ESP test I took was not an English test but an anatomy test in English" (Nezaratgoo & Behazdpoor, 2017, p.70). Such tests might be said to take a knowledge test identity (cf. Salmani-Nodoushan, 2003). Whether Iranian EAP teacher-testers develop test items resembling knowledge tests in nature or more general English like item types or manage to strike a healthy balance between focus-on-language and focus-on-content, we know little about key pertinent issues clustering around such tests and Iranian ESAP students' performance on them. Of such key issues are the roles of L2 proficiency and academic distance in ESAP test performance.

Although the effect of L2 proficiency in disciplinarily close and distant ESP/EAP tests of reading comprehension has been reported in some studies (e.g. Usò-Juan, 2006; Salmani-Nodoushan, 2003), little is known about the role of English proficiency in teacher-made discrete-point tests serving achievement purposes. In fact, most of the existing studies tend to focus on L2 proficiency and performance on integrative tests of ESP/EAP (i.e. reading comprehension). This study, however, aimed at studying the role of L2 proficiency in ESP/EAP tests that are not integrative but discrete-point. At the same time, the present study was an attempt to take Bachman's consideration into account, too. In his framework of test method facets, Bachman (1990) argues that the nature of the input test takers receive can influence the performance of test takers.

The input feature in this study was fundamentally different; it was not global academic reading comprehension texts but discrete-point multiple-choice items. This is important since changes in test method facets might have influences on test takers' performances in general (Bachman, 1990). Therefore, the role of L2 proficiency in ESP/EAP tests coming from disciplinarily near and remote territories needs to be seen in light of important test facets (e.g. the nature of input) proposed by assessment scholars (Bachman & Palmer, 2010). The study, therefore, made an effort to contribute to knowledge about ESAP testing by catering for one of the key facets of language assessment (i.e. the input test takers receive). A brief literature review on ESAP testing, the role of L2 proficiency and academic neighborhood or background familiarity in such tests is provided first. Next, research methodology provides information on participants, data collection instruments, procedures and analyses. Then, findings are given and discussed vis-à-vis the existing literature. Finally, conclusions and implications are presented.

2. Literature review

Factors contributing to students' performance on LSAP (Language for Special Academic Purposes) tests have been investigated in several studies. Such studies, it should be noted, were primarily encouraged by validity issues pertaining to academic reading components of IELTS or "other large-scale standardized examinations" (Schmitt & Hamp-Lyons, 2015, p. 5). It was then assumed that EAP students' reading comprehension ability could be affected by L2 proficiency as well as background knowledge or prior familiarity with test content (e.g. Clapham, 2000). Krekeler (2006) gives a long list of studies in which English proficiency, compared to topic familiarity, accounts for more of students' performance on not only EAP tests of reading comprehension, but also tests of other language skills, such as listening, speaking and writing.

Drawing upon the Threshold Hypothesis, some researchers (e.g. Clapham, 2000; Ridgway, 1997) argue that the effect of background knowledge should be insignificant for two groups of EAP test takers: advanced and low proficiency students. This is so, it is argued, because students with high proficiency levels are "able to make maximum use of their linguistic skills" so that they do not need to "rely so heavily on their background knowledge" (Clapham, 2000, pp. 515-516), whereas lower proficiency students are too engaged with bottom-up skills to take advantage of their background knowledge. In her now classic study, Clapham (2000) and Ridgway found some empirical support to this two Thresholds Hypothesis. Both researchers found that when low proficiency subjects answered reading comprehension questions from inside and outside their fields of study, the mean differences were insignificant. However, the mean differences for advanced level students were statistically significant. Their studies, thus, provided partial evidence only to the Threshold Hypothesis. They concluded that while a lower threshold seemed to be founded, an upper threshold sounded untenable.

In a more comprehensive study, Krekeler (2006) used C-test as a measure of language proficiency and gave his participants reading texts from inside and outside their disciplines. In accord with Clapham's view (2000), he found that, "Students with low levels of language proficiency scored similarly on texts from within and from outside their future disciplines" and that "Students with high scores on C-test did not rely heavily on background knowledge" (Krekeler, 2006, p.121). Despite this, he did not interpret his findings as strong endorsement of Clapham's view because "The vast majority of test-takers were able to make use of their background knowledge regardless of the level of L2 proficiency" (Krekeler, 2006, p.122), particularly when the degree of text specificity was low. His study, however, did not "disprove the two thresholds hypothesis" (Krekeler, 2006, p.123). Unlike Krekeler (2006), Usó-Juan (2006) found that low proficiency students could not take advantage of academic adjacency when they took EAP tests of reading comprehension. She concluded that the positive role of academic neighborhood in EAP test performance was bound to L2 proficiency levels. Additionally, she suggested that L2 proficiency accounted for more of her participants' EAP test performance than disciplinary knowledge.

Iranian researchers have also taken an interest in the topic. Studies by Salmani-Nodoushan (2003) and Taghizadeh Vahed and Alavi (2019) are related to the role language proficiency and disciplinary distance play in Iranian students' performances on ESAP tests. While Salmani-Nodoushan (2003) found significant differences across all proficiency levels "with totally familiar, partially familiar and totally unfamiliar propositional content" (Salmani-Nodoushan, 2003, p.8), Taghizadeh Vahed and Alavi (2019) concluded that students with low L2 proficiency levels could not benefit from disciplinary adjacency and background knowledge. Their findings were in line with that of Ridgway (1997) reported above. In other words, their study lent partial evidence for the Threshold Hypothesis since disciplinary knowledge significantly facilitated intermediate and advanced students' performances only. Besides, both studies gave further support to the more forceful explanatory power of L2 proficiency in ESAP test performance. Salmani-Nodoushan (2003) interpreted his findings as somehow contradictory to Clapham's views (1996) and concluded that, with less content specific reading tests, "language proficiency exerts such a great influence on test performance that the impact of text familiarity is almost negligible" (Salmani-Nodoushan, 2003, p.8). The interesting common point about these studies was that in both the nature of test task output was brought to the focus. In the former, the effect of task type (true/false, skimming, sentence completion,

outlining tasks), and, in the latter, the effects of elicitation method (objectively versus subjectively scored test items) were taken into account.

The studies reviewed so far were mainly concerned with the role of English proficiency in EAP test-takers' performances on integrative tests, i.e., reading comprehension tests. This focus on EAP reading comprehension and how much it is affected by L2 proficiency level traces back to the roots and origins of earlier IELTS-aligned studies or any other international EAP test for proficiency purposes resulting in the minimal attention paid to EAP tests for placement and achievement purposes (Dudley-Evans & St. Johns, 1998) or teacher-designed ESAP tests that are not integrative but discrete-point. The latter point warrants attention due to not only the frequent use of multiple-choice questions among Iranian ESAP teachers, but also the documented effects that changes in test method facets (e.g. the nature of input) bring about in language test performance (Bachman, 1990). The question remains as to whether language proficiency still contributes as strongly to ESAP test-takers' performances on teacher-constructed discrete-point tests as it does to global/integrative reading comprehension tests. Motivated by the dearth of research studies on teacher-made ESAP tests in general, inspired by the seemingly overriding impact of language proficiency on EAP test performance, and stimulated by the fundamental differences between integrative and discrete-point testing, the current study intended to answer the following questions:

Research Question One: Are there statistically significant variations in students' performances on different ESAP tests across proficiency levels?

Research Question Two: Do students perform better on ESAP tests coming from inside (i.e., psychology and law) versus outside (i.e., chemistry and law) their academic neighborhood?

Research Question Three: Is there a significant relationship between students' English proficiency and ESAP test scores?

3. Methodology

3.1. Design

The study is descriptive in nature. According to Cohen, Manion and Morrison (2007), descriptive studies aim at describing a population, situation or phenomenon and providing information and description about relationships between elements, factors or variables in social, educational, cultural and political settings. In this study, it was intended to describe the relationships between L2 proficiency, academic and disciplinary distance and ESAP test performance and describe the existing situation under investigation.

3.2. Participants and Setting

One-hundred fifty male and female students majoring in English Language Teaching (ELT) from a state university in Northwest of Iran participated in this study. They were sophomore, junior and senior students whose age ranged from 19 to 23. Forty students were excluded from the study due to incompleteness of their responses and the final sample consisted of 110 students: 73 females (66.6%) and 37 males (33.6%).

3.3. Instrumentations

3.3.1 Test of English as a Foreign Language (TOEFL)

A paper-based TOEFL (2004) was used in order to measure the participants' proficiency level and divide them into low (N=39), intermediate (N=52) and advanced (N=19) groups. Students took its Reading Comprehension section with 50 items and Structure and Written Expression section with 40 items in 85 minutes. Cronbach's Alpha yielded a reliability index of 0.86.

3.3.2 ESAP Tests

Four in-house ESAP sample tests were used. Each sample consisted of 40 multiple-choice test items of psychology, law, chemistry and geology. The tests had been developed and used by content specialist ESAP teachers. The distinctive feature of the tests was that they included discrete-point

multiple-choice test tasks only. The rationale to choose them was based on the research purpose: whether language proficiency was associated with performance on EAP tests that were not integrative but discrete-point. The reliability indexes, using Cronbach's Alpha, were 0.9, 0.85, 0.81 and 0.73 for ESAP tests of psychology, law, chemistry and geology, respectively.

3.4. Data Collection Procedures

All participants took TOEFL and were divided into low (N=39), intermediate (N=52), and advanced (N=19) groups first. Next, they took ESAP tests of psychology, law, chemistry and geology in two consecutive weeks. Following test designers, the participants were given 45 to 50 minutes for each test. Each test item was weighed one score making the total 40 for each ESAP test.

3.5. Data Analysis

Making sure of normality of data distribution, homogeneity of variance and after checking for test of sphericity, one-way and repeated-measures ANOVAs and Pearson correlations were carried out.

4. Results

4.1. Students' Performances on Different ESAP Tests Across Proficiency Levels

The results of ANOVA indicated that there were significant variations in the performances of students on different ESAP tests across language proficiency levels, $F_{psychology}(2,107)=21.66$, $p=0.00$; $F_{law}(2,107)=17.1$, $p=0.00$; $F_{chemistry}(2,107)=24.22$, $p=0.00$; and, $F_{geology}(2,107)=7.29$, $p=0.00$ (Table 1). The same analysis through General Linear Model revealed a very high Adjusted R Squared index (0.859) but no statistically significant interaction effects. The effect size (here eta-squared) was also calculated to see how different the means were (Table 1).

Table 1: ANOVA Results for ESAP Test Performances by Proficiency Level

		Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
psychology * ProfLevel	Between Groups	.052	2	.026	21.666	.000	.288
	Within Groups	.128	107	.001			
	Total	.180	109				
law * ProfLevel	Between Groups	.039	2	.019	17.102	.000	.242
	Within Groups	.121	107	.001			
	Total	.160	109				
chemistry * ProfLevel	Between Groups	.080	2	.040	24.226	.000	.312
	Within Groups	.176	107	.002			
	Total	.256	109				
geology * ProfLevel	Between Groups	.035	2	.017	7.292	.001	.120
	Within Groups	.255	107	.002			
	Total	.290	109				

Note. ProfLevel=Proficiency Level

The post-hoc tests were carried out for all ANOVAs above followed by a general one. It revealed that the mean differences between advanced, intermediate and low proficiency participants were significant indicating that the upper-level students outperformed intermediate learners and they, in turn, excelled lower-level participants in ESAP tests (Table 2).

Table 2: Duncan's Post-hoc Test for Advanced, Intermediate and Low Proficiency Students' on ESAP Tests

Level	N	Duncana		
		Subset		
		1	2	3
Low	168	.7667		
Intermediate	180		.7969	
Advanced	92			.8223
Sig.		1.000	1.000	1.000

a. Alpha = .05.

4.2. Performance on ESAP Tests from Different Academic Neighborhoods

The second question was concerned with the effects of disciplinary distance—another pertinent issue in ESP/EAP test performance— on test takers' performances on different ESAP tests. Calculation of repeated-measures ANOVA suggested that the participants' mean scores on four ESAP tests were statistically significant (Table 3). In other words, they performed differently on disciplinarily close versus distant ESAP tests.

Table 3: Results of Repeated-Measures ANOVA (Tests of Within-Subjects Effects)

Measure: ESAP Tests						
Source		Type III Sum of Squares	df	Mean Square	F	Sig.
factor1	Sphericity Assumed	3.778	3	1.259	882.272	.000
	Greenhouse-Geisser	3.778	2.837	1.332	882.272	.000
	Huynh-Feldt	3.778	2.977	1.269	882.272	.000
	Lower-bound	3.778	1.000	3.778	882.272	.000
factor1 * level	Sphericity Assumed	.010	6	.002	1.135	.341
	Greenhouse-Geisser	.010	5.674	.002	1.135	.342
	Huynh-Feldt	.010	5.954	.002	1.135	.341
	Lower-bound	.010	2.000	.005	1.135	.325
Error(factor1)	Sphericity Assumed	.458	321	.001		
	Greenhouse-Geisser	.458	303.584	.002		
	Huynh-Feldt	.458	318.542	.001		
	Lower-bound	.458	107.000	.004		

Duncan's post-hoc test was used to discern where the differences were (Table 4).

Table 4: Duncan's Post-hoc Test on Students' on Different ESAP Tests

Duncan ^a					
ESAP tests	N	Subset			
		1	2	3	4
geology	110	.6418			
chemistry	110		.7800		
law	110			.8384	
psychology	110				.9025
Sig.		1.000	1.000	1.000	1.000

a. Alpha = .05.

Post-hoc test showed that all four means were meaningfully different. The first and second best performances belonged to ESAP tests of psychology and law – both neighbors to English majors disciplinarily. It should be noted that English majors' performance on ESAP test of psychology was statistically different from (i.e. better than) their performance on ESAP test of law. From among the ESAP tests of Basic Sciences, the participants' performance on chemistry was better than that of geology. Figure 1 below provides more detailed information since the performances of each proficiency level on four ESAP tests are also represented.

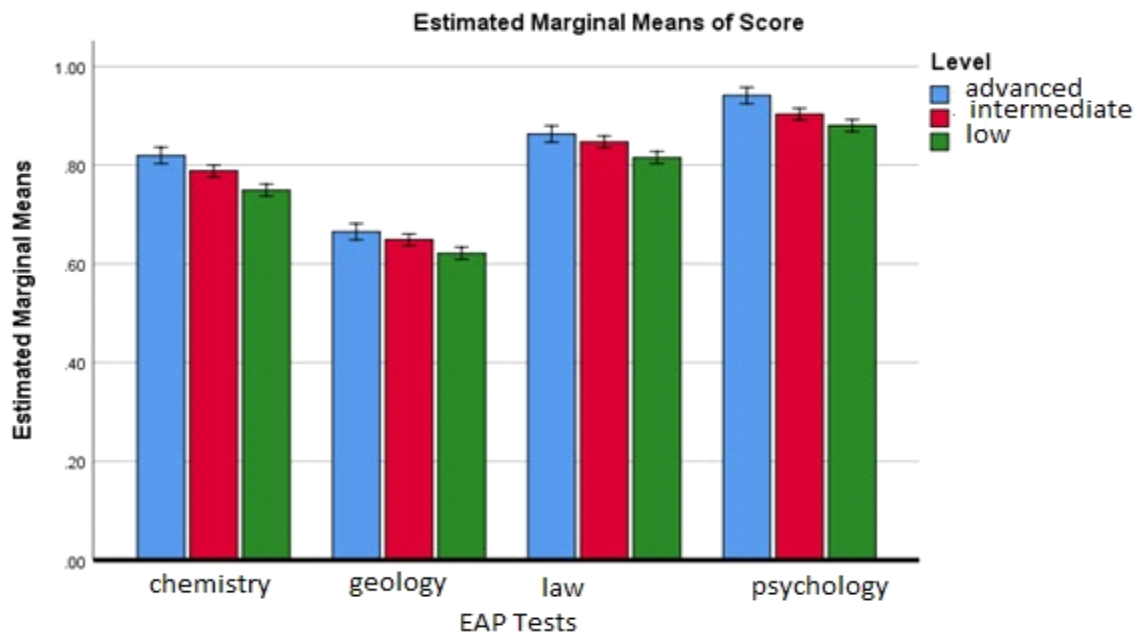


Figure 1: Students' Performances on ESAP Tests of Psychology, Law, Chemistry and Geology

Data analyses revealed that, overall, all the participants scored significantly better when taking ESAP tests from inside rather than outside their disciplinary territory. Further analyses revealed that the differences between advanced, intermediate and low proficiency participants' mean scores on ESAP tests of psychology and law were all different. On ESAP tests of chemistry and geology, however, advanced and intermediate students performed statistically similarly; the differences lied between the advanced and intermediate students on the one hand, and low proficiency students, on the other. Table 5 below summarizes further ANOVA results for English students taking psychology and law (neighbors to English students), on the one hand, and chemistry and geology (non-neighbors to them) ESAP tests on the other. This table indicates that the mean differences between the three proficiency levels were statistically significant for ESAP tests labeled as neighbors. For tests described as non-neighbors, the mean difference between the advanced and intermediate test takers was insignificant, while the association between these levels and low L2 proficiency was found to be significant.

Table 5: Students' Language Proficiency and Performances on Neighboring vs. Non-neighboring Tests of ESAP

English proficiency level	Neighboring fields of study	Non-neighboring fields of study
Advanced	0.90 ^a	0.74 ^a
Intermediate	0.87 ^b	0.71 ^a
Low	0.84 ^c	0.68 ^b

4.3 Relationship Between Students' TOEFL And ESAP Test Scores

Table 6 gives the correlation coefficients of the participants' scores on TOEFL and ESAP tests. Statistically significant correlations were found between TOEFL and all four ESAP tests scores. TOEFL and ESAP test scores of psychology, law and chemistry were significantly and moderately strongly correlated, $r(108) = .46, p < .01$; $r(108) = .40, p < .01$; $r(108) = .49, p < .01$ respectively. The correlation coefficient of .49 can be interpreted as 'largely positively correlated' since it is just 0.01 away from 0.5 (Cohen, 1988). The correlation between TOEFL scores and ESAP test scores of geology was slightly less than moderate, $r(108) = .27, p < .01$.

Table 6: Correlation Coefficients between TOEFL and ESAP Test Scores

	TOEFL	psychology	law	chemistry	geology	
Correlation	TOEFL	1.000				
	psychology	.461**	1.000			
	law	.409**	.288**	1.000		
	chemistry	.495**	.264**	.403**	1.000	
	geology	.277**	.193*	.306**	.351**	1.000

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

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

5. Discussion

First, data analyses revealed that test takers with higher proficiency levels excelled their lower level counterparts in general. Otherwise stated, advanced level students did better than intermediate students and they, in turn, outperformed low proficiency students on ESAP tests coming from both neighboring and non-neighboring academic disciplines. L2 proficiency then, is still at work even with discrete-point tests assumedly designed for achievement purposes. Although the beneficial and advantageous effects of higher English proficiency levels have been acknowledged and documented in EAP tests of academic reading in general (Salmani-Nodoushan, 2003) and the findings of this study lend further support to the role of L2 proficiency in ESAP test performance, the tests used in this study were teacher-constructed achievement tests and, by definition, such tests were “internal to the course” (Hutchinson & Waters, 1987, p.147) and “syllabus-dependent” (Moore & Boyle, 1994, p.318). These tests “measure mastery of a syllabus” (Dudley-Evans & St. John, 1998, p.213) through test questions of diverse kinds. The participants in this study answered presumably syllabus-dependent questions designed by content specialist ESAP teachers. They naturally took advantage of the stock available, i.e. their proficiency repertoire, when taking such tests. This gains more significance when all test takers’ performances on all ESAP tests are revisited in Figure 1. All bars in the figure exceed 0.60 indicating that all the participants managed to answer at least sixty percent of the items in all tests. This warrants even more attention when their better performances (e.g. on ESAP tests of psychology, law and even chemistry) are taken into consideration because achievement tests are expected to be inherently syllabus-dependent and measure mastery of syllabi.

One potentially plausible explanation to all participants’ beyond-60% correct answers to such “internal to the course” (Hutchinson & Waters, 1987, p.147) and “syllabus-dependent” (Moore & Boyle, 1994, p.318) test items might stem from a thorough examination and evaluation of test items constructed by Iranian ESAP practitioners. Thus, a follow-up inspection of all test items was undertaken. To a great extent, the participants’ performance becomes justified on the ground that a good, and, sometimes a large number of the items, sound to be typical ‘general English’ questions dyed ESP/EAP and used as such. The larger the number of such items, the better opportunity for the test takers to take advantage of their L2 proficiency. The following examples are given to cast light on this. The first, second, third and fourth items belong to ESAP tests of psychology, law, chemistry and geology, respectively.

The word ‘get used to’ is closest in meaning to

- a. lining b. ridge c. adjust d. facing

Children under 10 years of age not criminally for their actions.

- a. insanity b. mistake c. self-defense d. responsible

We can identify different compounds of some mixtures with a microscope, or even the eye.

- a. unaided b. aided c. unaid d. aid

What is the synonym of the word ‘raise’?

- a. lower b. rigid c. elevate d. alike

Items of this type abound in ESAP tests of psychology, law and chemistry but are fewer in that of geology resulting in less opportunity for L2 proficiency to exercise influence as noticeably as it does in other tests. In other terms, the number of items asking for geology specific content knowledge—the item below, for example—is more, the result of which is minimizing the contribution of English language proficiency.

..... *forms when basaltic magma erupts under water.*

a. Vent

b. Crater

c. Pillow lava

d. Stratovolcano

Such an assessment behavior of the ESAP practitioner of geology (i.e. emphasis on discipline specific technical knowledge) explains the correlation coefficients in Table 6, too. All correlation coefficients between TOEFL and ESAP test scores exceed 0.4 except for that of geology which is 0.27. The gap, then, can be attributed to the content knowledge requirement feature of a good number of items in ESAP test of geology.

The above argument is reminiscent of Salmani-Nodoushan's (2003) conclusion that content specificity and disciplinary knowledge in ESAP testing is a torturing issue. This is because as test designers construct items calling for more of disciplinary knowledge, they get closer and closer to the borders of knowledge tests, and if they move in the opposite direction, L2 proficiency alone accounts for much of ESAP test takers' performance. This seems to be a pertinent point in Iranian ESAP context in general since a similar concern is echoed by Afghari and Tavakoli (2004), too. They hold that when, "a test of highly specific purpose is designed, care should be taken not to split it more from being a language test. In such case, it becomes a knowledge test, which is not in the realm of language testing" (p. 23).

Correlation coefficients are interpretable as measures of effect size for Pearson correlations (Cohen, 1988). According to Cohen (1988), correlation coefficients of 0.10, 0.30 and 0.50 indicate small, medium and large effect sizes, respectively. In this study, there were three coefficients exceeding 0.4 and two of them approximated Cohen's 0.5 (i.e. 0.49 for chemistry and 0.46 for psychology). These coefficients give further credence to the role of L2 proficiency in ESAP test performance. The lowest effect size of L2 proficiency (0.27) belongs to geology which comes close to Cohen's medium effect size (0.30).

Moreover, eta squared indexes of ANOVA as measures of effect size provide further complementary substantiation to the above argument. According to Cohen (1988), in ANOVA, eta squared indexes of 0.01, 0.06 and 0.14 indicate small, medium and large effect sizes, respectively. Except for geology, all indexes in Table 1 above exceeded 0.14 which is an indication of how large the size of L2 proficiency effect was in ESAP test performance. Following Cohen's criteria, the effect size of L2 proficiency for ESAP test of geology (i.e. 0.12) was slightly smaller than large but much larger than Cohen's 0.06 which was for the medium effect size. These are indicative of the explanatory power of L2 proficiency in teacher-designed ESAP tests for achievement purposes.

Second, disciplinary neighborhood affected the performance of test takers significantly: they did better on disciplinarily close tests. Additionally, intra-proficiency-level differences (e.g. statistically significant differences between low level students' means on four different ESAP tests) suggest that, with discrete point test items, disciplinary distance is equally effective for all the three proficiency levels. Put it differently, the performance of low proficiency students, for instance, is different depending on where the test comes from (distant versus close). This holds true for intermediate and advanced students, too. As a result, no cut-off thresholds of proficiency level sound justifiable even with discrete-point ESAP tests. This is in keeping with some prior studies in which the test takers' performance on EAP reading comprehension tests from inside and outside their subject areas was examined and ample compelling evidence for the threshold hypothesis was not gained (Ridgway, 1997; Krekeler, 2006). As a result, and in line with Krekeler's argument, findings are used to argue that, irrespective of language proficiency level, test takers benefit from disciplinary adjacency in discrete-point tests, too. Our findings are incompatible, though in part, with those of Usó-Juan (2006) and Taghizadeh Vahed and Alavi (2019). In their study, only intermediate and

advanced students benefited from the advantages associated with disciplinary knowledge, and low proficiency students could not. In this study, all participants, regardless of their L2 proficiency level, took advantage of it systematically and meaningfully.

In this study, English majors performed better, and statistically so, when given ESAP tests of psychology and law (disciplinarily close) rather than chemistry and geology (disciplinarily remote). In other words, L2 knowledge contributes more to students' performance when it is coupled with disciplinary neighborhood. This seems to be congruent with earlier studies acknowledging the facilitative role of background knowledge (Chen & Graves, 1995; Ridgway, 1997; Birjandi, Alavi & Salmani-Nodoushan, 2002; & Papajohn, 1999; Afghari & Tavakoli, 2004) in EAP tests of reading and listening comprehension. It is intriguing that English majors have been statistically better in ESAP test of psychology than that of law. In addition to the abundance of the number of test items belonging to general English domain in ESAP test of psychology, hence giving English students the upper hand, it is worth noting that English majors share more disciplinary experience and background with the field of psychology since they also do a good amount of academic reading in psychology; they take courses on 'educational psychology', 'teaching methodology', 'teaching language skills', etc. during which they are exposed to psychology content and discourse.

Third, putting table 5 and figure 1 together indicates one more important fact: L2 proficiency knowledge has more discriminatory power in participants' performance on ESAP tests of disciplinarily adjacent fields and less in ESAP tests of non-adjacent ones. Not only higher English level is more functional in domain familiar ESAP tests, it is also more distinguishing since all proficiency levels are statistically distinct whilst its role as well as discriminatory force fade away when students enter uncharted territories of basic sciences. In domain unfamiliar tests, the distance between intermediate and advanced students collapses statistically; they join together and make up a single group.

Taking the findings to the larger landscape of EAP reveals more. Firstly, Iranian ESAP practitioners' assessment behavior sheds some light on their cognition of what ESAP is and how the courses are run. A large number of items mirror classroom practice of a type Hyland (2006) calls general EAP. As mentioned above, there are many items that are linguistically affordable by students with a relatively good command of English. Proponents of this wide-angle approach to EAP argue that EAP is, "too hard for students with limited English proficiency. Weaker students are not ready for discipline-specific language and learning tasks and need preparatory classes to give them a good understanding of 'general English' first" (Hyland, 2006, p.10). Given the problems Iranian students have with general English proficiency in ESP/EAP classes (Atai & Nazari, 2011; Mazdayasna & Tahririan, 2008), it might be inferred that Iranian EAP practitioners have been inevitably engaged in some general English instruction with subsequent reflections in their assessment practice. That ESP/EAP students' low general English level makes teachers' classes EGP (English for General Purposes) in nature has been reported by Başturkman (2006a), too. This takes us to another point of particular significance. According to Douglas (2005), ESP test performance in general needs to be interpreted in terms of a composite construct of specific purpose language ability that includes both "language knowledge and specific purpose background knowledge" (Douglas, 2005, pp. 858-859). The participants' performance in this study, however, suggests that a good number of the items did not enjoy such a feature; a balanced distribution of language and content knowledge load has not been maintained. Thirdly, and, as for test items asking for discipline specific technical knowledge, it might be argued that content and lexical specificity has been taken for needs and purpose specificity wherein disciplinary, discursal and genre features of students' academic fields are addressed. Such items not only adopt a knowledge test identity (cf. Salmani-Nodoushan, 2003) but also substitute real content with carrier content, thus distorting the essence of ESP/EAP. Finally, English students in this study answered summative, exit-course achievement tests that are assumed to "measure mastery of a syllabus" (Dudley-Evans & St. John, 1998, p.213). Findings suggest that there might be a need to either re-define the tests they take or the syllabi they go through.

6. Conclusion

Alderson (1988a) once remarked that, “it is rather sobering and perhaps depressing to note the minimal attention paid to testing within ESP” (Alderson, 1998a, p.87). Since then, ESP/EAP assessment has been growing and flourishing, thanks to research attempts to substantiate theoretical underpinnings of ESP assessment (Douglas, 2005, 2013) as well as its pragmatic and pedagogical values (Başturkman & Elder, 2006). Much of the effort, nonetheless, has been centered on large scale international (e.g. IELTS & TEEP) and national tests (see Olaofe, 1994 for Nigeria; In’ nami, Koizumi & Nakamura, 2016 for Japan; and, Lumley & Qian, 2003 for Hong Kong) serving academic proficiency purposes elbowing achievement EAP/ESAP tests to the periphery. This study attempted to address the role of L2 proficiency and disciplinary adjacency in teacher-generated ESAP tests in Iran.

The shift of focus from integrative/global (i.e. EAP reading tests) to teacher-generated discrete-point tests distinguishes this study from prior studies. Findings suggested that students’ performances on such tests vary not only with their language proficiency levels but also with disciplinary distance of ESAP tests they take. Alternatively stated, even with discrete-point tests where advantages brought by more integrative reading comprehension tests (e.g. richer contextual information, top-down processing, guessing from co-textual and contextual clues, expectancy grammar potentials, etc.) are generally gone, disciplinary neighborhood and shared academic territory keeps affecting EAP test performance. It sounds plausible, then, to speculate that the supportive and positive role of English proficiency should be examined vis-à-vis two other relevant and equally important factors: Disciplinary distance and the ‘identity’ of test items in such tests. The farther the academic distance and the more ‘knowledge’ test an item takes in its identity, the lower the profile of L2 proficiency.

Of course, the current EAP assessment behavior in Iran should not be viewed independently and in isolation; rather, it should be seen in connection with pertinent de facto classroom realities (e.g. low general English proficiency level of Iranian students) and Iranian EAP/ESP practitioners, syllabus designers and curricula planners’ cognition and approaches to assessment for which further studies are needed. If the ‘common core’ or ‘study skills’ and the ‘specific’ and ‘academic literacy’ approaches make up two extreme ends of EAP continuum (Cottrell, 2001; Chiu, 2015), it proves challenging to locate where the current EAP programs in Iran stand. This is significant since any orientation adopted will exert influences on assessment. In this study, a good number of items call for general or at best semi-technical vocabulary knowledge reminiscent of a movement in the history of ESP in the 1960s and early 1970s: *register analysis*. Then, it was semi- or sub-technical vocabulary knowledge that was appreciated (Dudley-Evans & St. John, 1998). For instance, in register analysis it was held that sub-technical words such as “‘consists of’, ‘contains’, ‘enables’, ‘acts as’” are “more likely to occur in scientific, technical or academic writing than in general contexts” (Dudley-Evans & St. John, 1998, p.21). Once more, Salmani-Nodoushan’s (2003) concluding remarks sound pertinent. Believing that discipline specific knowledge is a baffling dilemma in LSP (Language for Specific Purposes) testing in general, he concludes that one potential solution might be to “include EGAP [English for General Academic Purposes] tests in the category of LSP tests” (p.8).

Caution should be practiced when generalizing the findings of this study. Although the test task objectivity feature of ESAP tests used in the study catered for the recently raised factor in ESAP test performance (Taghizadeh Vahed & Alavi, 2019), neither students’ L1 proficiency levels nor their homogeneity in L1 backgrounds were attended to. Including students’ L1 proficiency could yield interesting results. It was tenable to expect similar or different results with other test takers who were homogenous in their L1 backgrounds. Additionally, the choice of ESAP tests and academic disciplines was limited since the researchers had to look for test samples with discrete-point multiple-choice items due to the focus of the study. This excluded some academic disciplines in soft and hard sciences from the sample. Similar or different results could have been gained should ESAP tests of geography or biology, for instance, had been used. Only additional inquiries can cast light on these issues.

References

- Afghari, A., & Tavakoli, M. (2004). Learner variables and test performance. *IJAL* (7), 1, 1-24.
- Alderson, J.C. & Urquhart, A. H. (1985a). The effect of students' academic discipline on their performance on ESP reading tests. *Language Testing*, 2, 192-204.
- Alderson, J. C. & Urquhart, A. H. (1985b). This test is unfair: I'm not an economist. In P.C. Hauptman, R. LeBlanc & M. Bingham-Wesche (Eds.) *Second language performance testing* (pp. 25-45) University of Ottawa Press.
- Alderson, J. C. (1988a). Testing and its administration in ESP. In D. Chamberlain & R. J. Baumgardner (Eds.), *ESP in the classroom: Practice and evaluation*. London: Modern English Publications & The British Council.
- Alderson, J.C. (1993). The relationship between grammar and reading in an English for Academic Purposes test battery. In D. Douglas & C. Chapelle (Eds.) *A New Decade of Language Testing Research* (pp. 203-219). TESOL, Alexandria, VA.
- Atai, M. R. & Nazari, O. (2011). Exploring reading comprehension needs of Iranian EAP students of health information management (HIM): A triangulated approach, *System*, 39(1), 30-43.
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachaman, L. & Palmer, A. (2010). *Language testing in practice*: Oxford: Oxford University Press.
- Başturkman, H. (2006). *Ideas and options in English for specific purposes*. NJ: Lawrence Erlbaum Associates.
- Başturkman, H. & Elder, C. (2006). The practice of LSP. In A. Davies & C. Elder (Eds.) *The handbook of applied linguistics* (pp. 672-694). Oxford: Blackwell.
- Benesch, S. (2001). *Critical English for academic purposes*. Mahwah, NJ: Erlbaum.
- Bernhardt, E.B. (1991). *Reading development in a second language*. Albex.
- Birjandi, P., Alavi, S.M., & Salmani-Nodoushan, M.A. (2002). *ESP test performance: A study on Iranian LEP and non-LEP university students*. Unpublished PhD dissertation, University of Tehran, Tehran.
- Chen, H. C. & Graves, M. F. (1995). Effects of previewing and providing background knowledge on Taiwanese college students' comprehension of American short stories, *TESOL Quarterly* 29, 663-686.
- Chiu, T. (2015). Personal statement in PhD applications: Gatekeepers' evaluative perspectives. *Journal of English for Academic Purposes*, 14, 63-73.
- Clapham, C. (1996). *The development of IELTS: A study of the effect of background knowledge on reading comprehension*. Studies in Language Testing, volume 4. Cambridge University Press.
- Clapham, C. (2000). Assessment for academic purposes: Where next? *System*, 28(4), 511-521.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Routledge.
- Cohen, L., Manion, L. & Morrison, K. (2007). *Research methods in education*. London & New York: Routledge.
- Cottrell, S. (2001). *Teaching study skills and supporting learning*. Basingstoke, UK: Palgrave.
- Douglas, D. (2005). Testing languages for specific purposes. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 857-868). Mahwah, NJ: Lawrence Erlbaum Associates.

- Douglas, D. (2013). ESP and assessment. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for Specific Purposes* (pp. 367-384). Chichester, UK: Wiley-Blackwell.
- Dudley-Evans, A., & St. John, M. J. (1998). *Developments in ESP: A multidisciplinary approach*. Cambridge: Cambridge University Press.
- Flowerdew, J. & Peacock, M. (eds.) (2001). *Research perspectives on English for academic purposes*. Cambridge: Cambridge University Press.
- Fulcher, G. (1999). Assessment in English for academic purposes: Putting content validity in its place, *Applied Linguistics*, 20(2), 221-236.
- Fulcher, G. (2000). Computers in language testing. In P. Brett, & G. Motteram (Eds.) *Computers in language teaching* (pp. 97-111). Manchester: IATEFL Publications.
- Hutchinson, T., & Waters, A. (1987). *English for specific purposes: A learner-centered approach*. Cambridge: Cambridge University Press.
- Hyland, K. (2006). *English for academic purposes: An advanced resource book*. London, New York: Routledge.
- Hyland, K. (2015). *Academic publishing: Issues and challenges in the construction of knowledge*. Oxford: Oxford University Press.
- Hyland, K. (2017). English in the discipline: Language provision in Hong Kong's new university curriculum. In E. S. Park (ed.) *English education at the tertiary level in Asia* (pp. 27-45). New York: Routledge.
- In'nami, Y., Koizumi, R., & Nakamura, K. (2016). Factor structure of the Test of English for Academic Purposes (TEAP) test in relation to the TOEFL iBT test. *Language Testing in Asia*, 6(3), 1-23.
- Jordan, R. R. (1997). *English for academic purposes: A guide and resource book for teachers*. Cambridge: Cambridge University Press.
- Krekeler, C. (2006). Language for special academic purposes (LSAP) testing: the effect of background knowledge revisited. *Language Testing*, 23 (9), 99-130.
- Lumley, T. & Qian, D. (2003). Assessing English for employment in Hong Kong. In C. A. Coombe, & N. Hubley (Eds.) *Assessment practices: Case studies in TESOL practices series* (pp. 135-147). Alexandria, VA: TESOL.
- Mazdayasna, G., & Tahririan, M. H. (2008). Developing a profile of the ESP needs of Iranian students: The case of students of nursing and midwifery. *Journal of English for Academic Purposes*, 7, 277-289.
- Mazlum, F. (2020). *ESP assessment in Iran: From syllabi to practice*. Manuscript submitted for publication.
- Moore, M. & Boyle, J. (1994). Testing English for specific purposes in Hong Kong companies. In J. Boyle & P. Falvey (Eds.) *English language testing in Hong Kong* (pp.317-332). Hong Kong: Regal Printing Co., Ltd.
- Nezakatgoo, B. & Behazdpoor, F. (2017). Challenges in teaching ESP at medical universities of Iran from ESP stakeholders' perspectives, *Iranian Journal of Applied Language Studies*, 9(2), 59-82.
- Olaofe, I. (1994). Testing English for academic purposes (EAP) in higher education. *Assessment and Evaluation in Higher Education*, 19 (10), 37-48.
- Papajohn, D. (1999). The effect of topic variation in performance testing: The case of chemistry TEACH test for international teaching assistants. *Language Testing*, 16, 52-81.

- Peretz, A. S. & Shoham, M. (1990). Testing reading comprehension in LSP: Does topic familiarity affect assessed difficulty and actual performance? *Reading in a Foreign Language* 7, 447-455.
- Phillipson, R. (1992). *Linguistic imperialism*. Oxford: Oxford University Press.
- Rajabi, P., Kiani, G., & Maftoon, P. (2012). ESP in-service teacher training programs: Do they change Iranian teachers' beliefs, classroom practices and students' achievements? *Ibérica*, 24, 261-282.
- Rezaee, A. A. & Kazempourian, S. (2017). A triangulated study of workplace English needs of electrical engineering students. *Journal of Modern Research in English Language Studies*, 4 (4), 1-25.
- Ridgway, T. (1997). Thresholds of the background knowledge effect in foreign language reading. *Reading in a Foreign Language*, 11, 151-168.
- Salmani-Nodoushan, M. A. (2003). Text familiarity, reading tasks, and ESP test performance: A study on Iranian LEP and non-LEP university students, *The Reading Matrix*, 3(1), 1-14.
- Schmitt, D. & Hamp-Lyons, L. (2015). The need for EAP teacher knowledge in assessment. *Journal of English for Academic Purposes*, 18, 3-8.
- Taghizadeh Vahed, Sh. & Alavi, S. M. (2019). The role of discipline-related knowledge and test task objectivity in assessing reading for academic purposes. *Language Assessment Quarterly*, 17(1), 1-17.
- Tan, S. H. (1990). The role of prior knowledge and language proficiency as predictors of reading comprehension among undergraduates. In J.H.A.L. de Jong & D. K. Stevenson (Eds.) *Individualizing the assessment of language abilities* (pp. 214-224). NY: Multilingual Matters.
- Usó-Juan, E. (2006). The compensatory nature of discipline-related knowledge and English-language proficiency in reading English for academic purposes. *The Modern Language Journal*, 90(2), 210-227.