

An Analysis of Proficiency Scores Addressing L1/L2 Use in a United Arab Emirates Reading Class

Amira H. Traish

University of Sharjah
United Arab Emirates

Abstract

This study, an investigation into the effectiveness of using a learner's first language (L1) to teach a second language (L2) in reading comprehension classes, focuses on tertiary students in the United Arab Emirates (UAE). Many theorists argue that a student's L1 contributes to a more successful learning environment and facilitates second language acquisition. Because of the importance placed on proficiency scores, this study uses empirical data, comparing test scores of two classes and exploring and examining their differences. One instructor uses a student's L1 (Arabic) to teach reading in L2, and the other does not. Research questions for this study have employed a quantitative approach which has included assessment data from midterm and proficiency exams. The findings of the study, which indicate negligible differences in overall scores, demonstrate that the pedagogical practice has no significant impact on the scores obtained. However, further study is warranted to accommodate more of the variables that may have affected data outcome.

Key Words: target first second language proficiency exams assessment IELTS TOEFL.

Introduction

1.1 Background Rationale, and Contribution

As the world becomes more globalised, the need to learn a second language, which is of primary concern in the Middle East, is more of a necessity than an option for many (Firth & Wagner 2007; Toohey 2000). The acquisition of English as a Second Language (ESL) is developing into a requirement for upwardly mobile residents of most non-English-speaking countries (Gallagher 2011). Although attainment of the world's most widely spoken academic language (Ammon & McConnell 2002; Hyland et al. 2009) is advantageous on a number of fronts, the issue does present some problems and complexities for those working in educational institutes operating abroad (Cenoz Jessner 2000; Kachru 1992). For an instructor of ESL who has been teaching in the United Arab Emirates (UAE) for the past seven years, the task has been made more complex, perhaps, because not only must an educator be tolerant of students who hail from different cultures in order to understand their educational framework, but in the case of students whose issue is a language barrier, the instructor must determine how best to educate them in linguistic and communicative matters. This is especially pertinent in the context of this study because UAE students must achieve a requisite language proficiency score on the International English Language Testing System (IELTS) or Test of English as Foreign Language (TOEFL) exams to study in their desired accredited English-medium majors. This is a mandate common to most universities.

With such a rich English environment and a vast number of available opportunities, it might reasonably be expected that learners of English in the UAE would have ample opportunity to develop their English language skills. Unfortunately, this is not the reality of most UAE English classrooms. Most students experience the pan-Arab dilemma that is pointed out by Kharma and Hajjaj (1989), wherein students see English as 'a school subject rather than a means of communication'; students in the UAE do receive several years of English in the classroom, but 'the pass mark is often very low.' Consequently, 'learners can proceed to further learning of the language without having first mastered fully what they ought to have mastered' (1989, p. 2). Year after year, they find themselves building on a language they do not really understand. This may be the situation in which students find themselves when it comes to the reading skill. One factor that may influence reading acquisition is the use of L1. Although this is by no means the only influencing factor, it may turn out to be a critical one, especially since the policy decrying its use has been promoted by the Ministry of Higher Education (Ministry).

Therefore, the degree and relevance of L1 use, limited or otherwise, will be investigated in this study. Tertiary education in the UAE has become a concern for the country, with the English programs consuming one-third of the tertiary budget (Bardsley 2009; Fox 2008; Gallagher 2011). With UAE citizens' higher education being funded by the government, the application rates for final-year UAE students at 95% and 80% for female and male students, respectively, are some of the highest in the world (Abu Dhabi Week, 2011). The program's effectiveness has also come into question, and in 2008, the diploma-level students' pass rate for English international benchmark exams was less than 50% at one of the major tertiary institutions in the country (Taylor 2008). Several researchers have investigated motivational as well as emotional and cultural factors that may have influenced L2 acquisition, but with proficiency exams playing such a critical role in a student's academic future, the use of these exams has been selected as a primary tool for this study. While researching this topic, I encountered a perplexing statistic that warranted further study. Students entering the University had averages that were at or above the 90th percentile but could still not pass a proficiency exam. This fact prompted my interest in this topic: These students had come from high school with 'A' averages, yet they were not able to pass the exam, a fact that suggested that they had not been learning English effectively and may have simply been rote memorizing material in preparation for the exam, or that they had just been pushed through the system. After having entered the University, they found themselves unqualified to enter into their English-medium majors and, as a result, their academic progress had been delayed.

1.2 Criticism of Proficiency Exams in the Literature

Proficiency exams and more specific language assessment exams are useful in that they '... yield information that can help decision makers allocate resources on the basis of merit, rather than lineage or patronage' (Spolsky & Hult 2010, p. 456). They are given by outside organisations that should not have a conflict of interest and should therefore be fair and unbiased in scoring. While some advocate the idea of proficiency exams, others have pointed out these same exams can be used for '... purposes other than those for which they were designed, often with unintended negative consequences to various groups of test takers' (Spolsky & Hult 2010, p. 456). The actual way an exam will be used is dependent on a variety of factors, some of which include the financial or political atmosphere. All exams have their detractors and shortcomings; however, doing away with these critical assessment tools would yield far more negative consequences. The proficiency exams have experienced their fair share of criticism in reference to whether multiple-choice test items are chosen for convenience and question context (Farhady 2005), which underscores the critical question of whether content or language is being tested (Butler & Stevenson 2001; Carrasquillo & Rodriguez 1995; Mohan 1986).

One explanation that attempts to respond to the content requirement is from Short (1993), who contends that 'language and content are intricately intertwined; it is difficult to isolate one feature from the other in the assessment process. Thus, teachers may not be sure whether a student is simply unable to demonstrate knowledge because of a language barrier or whether, indeed, the student does not know the content material being assessed' (p. 629). Criticism of these proficiency exams notwithstanding, they retain their value as a measure of standardization. Given the globalization of the academic marketplace, such tests are needed to ensure general linguistic competency within the student populace. The ease of access to education available to all UAE citizens may be a contributing factor to the overall poor quality of English proficiency (O'Sullivan 2010). UAE students do not always feel the need to achieve high grades in English, and most students do not prioritize the need to achieve English competency as they do with the maths and sciences (Farah & Ridge 2009).

As mentioned earlier, students at the Foundation level achieve an average of 95% or better at the secondary level, and yet their English proficiency level is not high enough to qualify them to enter their major area of study (Farah & Ridge 2009; Gallagher 2011). Students have been promoted to the next level for years without the requisite English skills, and this is now coming to a head at the tertiary level, where vast amounts of research are being dedicated to 're-teaching' students the language skills necessary for academic life. Instructors find themselves struggling to prepare students whose language skills are far below college level to take college-level language proficiency exams. Additional research reveals proficiency exam score results in tertiary levels, but most of these have taken place in other countries, such as Australia, for example (Humphreys & Mousavi 2010; Storch & Hill 2011).

The test/retest design has been predominant, and it has been found that studying in an English-medium atmosphere resulted in improved scores (Storch & Hill 2011), an observation that reflects cited studies on the use of L1 to help teach L2 previously mentioned.

Although these studies have been valuable in helping interested parties to better understand the field, they do not provide a comprehensive view of the topic as pertains to this study since they were conducted in an English environment (one in which English is the country's primary language), a situation that stands out in marked contrast to the UAE, where Arabic is the official language.

1.3 Research Questions

The Research Questions (RQs) addressed in this paper investigate the issue from a quantifiable perspective by exploring the concept of use of a student's first language to learn a target language, and specific attention has been paid to the process of reading. This paper is based on a larger study undertaken as part of a doctoral thesis (Traish 2014). The study examines whether there is any significant difference between the scores of students exposed to L1 in reading comprehension classes and those not exposed by using the TOEFL and IELTS, the only two proficiency exams currently accepted in the University, which are provided by outside organizations, as well as the scores of the MT exam, a standardized test given by the Department during the middle of the 17-week term. To establish whether L1 aids students in L2 reading comprehension, the following questions have been posed:

Main Research Question:

Does the instructor's use of the L1 in class aid students in L2 reading comprehension?

1.1 Is there any significant difference between the TOEFL reading scores of students who are exposed to L1 in reading comprehension classes and those who are not?

1.2 Is there any significant difference between the IELTS reading scores of students who are exposed to L1 in reading comprehension classes and those who are not?

1.4 Research Paradigm

Quantification and statistics are often considered a standard measure of validity (Henning 1986), and assessment and test results to measure achievement are often crucial (Farhady 2005; Kohn 2000). Consequently, quantitative research for this study has focused on the statistical outcome attained when L1 is used to teach L2 reading. Through use of the quantitative method, this study explores and assesses the effectiveness of L1. The objective of this study is to answer the main RQ: 'Does the instructor's use of L1 in class aid students in L2 reading comprehension?' This question has been addressed using the quantitative approach employing two tools: exam scores at different intervals and non-participant observations of reading lessons. The intent of this study has been to determine if students who are exposed during one term to an instructor who uses L1 during a reading lesson will achieve higher scores than students not exposed. To achieve this, data has been collected and analyzed under controlled circumstances (Duff 2002; Orlikowski & Baroudi 1991). This has been accomplished with the use of naturalistic settings (Johnson 2004) such as a classroom where one group of students has been exposed to an instructor using L1 during reading lessons and one control group has experienced limited exposure, if any. Using quantitative methods, data has been collected and measured to attain a statistically mathematical outcome (Cunning 2012), which would be needed to answer the main RQ.

One important variable that can affect the outcome of the study is a student's reading proficiency in L2. To determine if the variable had influenced test scores, participants have been measured before the variable was introduced, namely before the term began, using a pre-test, and then again after the variable was used. This was the first tool used to collect data, and it has entailed large groups of selected participants for whom a variable was manipulated in the participants' environment (in this case, the L1) in order to determine whether there was any relationship between the independent (manipulated) variable (L1) and the participants' acquisition of L2. The finding has then been generalized beyond the parameters of the individuals who were participating in the study. Two groups have been employed, one with a variable that was changed through the use of L1 in the classroom during a reading lesson, and one without that variable. Then, statistical procedures have been used to determine whether the relationship was significant, and if it was significant, the consideration has been whether the results could be generalized (Creswell 2008) to a larger population beyond the immediate group of participants.

1.5 Research Instruments

1.5.1 First Stage: Pre-test

To ensure that students were not preselected to influence outcome and that all participating students were at the same language proficiency level, the scores of the proficiency exams, which place students in appropriate levels, was used. One of the first tools used in this study, the pre-test analysis, took the form of results of the TOEFL proficiency exam.

This exam includes a Reading section, and results facilitated student placement into the appropriate instructional level. This tool was useful as a basis for the study's data collection. The TOEFL tests three skills: listening, reading, and grammar (structure). The test is given at the beginning of each term over a two-week period to all students who wish to enter an English-medium major at the University. Most students are newly graduated and are taking the TOEFL for the first time; some have taken the TOEFL earlier but did not achieve a 500, so they are attempting it again. The University advertises on its website the dates for the TOEFL, and instructors and administrative staff help to register students for the exam. The paper-based test (PBT) is administered on campus by the American-Mideast Educational and Training Services Inc. (AMideast), an outside independent testing organisation. TOEFL exam results are received by the University three days later in the form of exam report forms, which include information on each section's results (Listening, Structure, and Reading) as well as the overall score.

Like the IELTS, the University is interested only in the overall score and not individual skill scores. Students that receive a 500 overall can register for their majors, while students that do not achieve a 500 can enter either the Foundation programme, an Arabic-medium major, or choose not to register at the University. Participants for this study have received between a 420-460 overall and have been placed in Level 3. The final tool used in collection of student data was the gathering and analysis of the post-test, which took the form of either the TOEFL or IELTS proficiency exam, which provided empirical evidence to help answer all of the questions pertinent to this study. The University accepts only the TOEFL and IELTS as language proficiency exams. Both exams can be taken on campus and are scheduled at least twice during the term. Students are permitted to take the test any time during the term either on or off campus through an authorized exam centre. All results are reported to the University for recording in the student's file. TOEFL results are returned to the University within 4 days, while the IELTS takes 13 days. Once scores are reported, the Department records results for students, keeping a hard copy on file. The vast majority of students graduating from a high school in the UAE are familiar with the TOEFL exam and attempt to achieve a 500 at the beginning of the term.

This was evident from the majority of students that were taking the TOEFL and the minority that were taking the IELTS. In recent years (2010), the University has lowered its requirement score for the IELTS from an overall Band 5.5 to an overall Band 5, the minimum required score from the Ministry. Students found that exiting with the IELTS requirement was at times easier than the TOEFL, and as word has spread, more and more students have started to take the IELTS. Evidence of this is found when comparing the percentage of students that have chosen to take the TOEFL versus the IELTS exams. In autumn of 2010, 9% of students studying in the Foundation programme exited with the TOEFL exam, and 22% exited with the IELTS. In spring of 2010/2011, 48% of students studying in the Foundation programme exited with the IELTS versus less than 1% with the TOEFL (UoS.ae).

1.6 Inferential Quantitative Results Analysis

The analyses of the two previous sections have provided empirical evidence for the study and help answer RQ 1, 1.1 and 1.2. Both the midterm exam (MT) grades and the proficiency exam scores have been used to explore the effectiveness of using a student's L1 to teach L2. A main objective of the data was to answer a primary question: 'Was there any statistical difference in the test scores of the classes?' to answer the main RQ: 'Does the instructor's use of L1 in class aid students in L2 reading comprehension?' If there was a difference, the next question would ask whether this difference was statistically relevant. The last question that was investigated sought to determine which class had better exam score results. After analyzing the MT scores for both sections, it was revealed that the results for the students that did not use Arabic (S52) were higher (13.8200) than those of the students who used Arabic (S51) (11.1181) by 2.7019. After approximately seven weeks of class, S52, the class whose instructor did not use L1, outperformed S51, the class whose instructor did use L1. The MT exam was the first standardized exam that students would have taken in the class, and the numbers showed that in the short term (approximately seven weeks), the students who were exposed to L1 in the classroom did not perform as well as those who were not.

Several factors may explain why students performed as they did on the MT exam. One may be that the MT exam was tied to classroom lessons in that some of the themes and vocabulary were used. This is not to say that the same questions were found on the MT as were given in the class, rather that the theme that was used in class was carried forward to the MT. This may have given students an advantage in that they had readings that were thematically similar and studied certain vocabulary words that would have been practiced in class.

Given that students were reading thematically similar text in the class, it would be reasonable to assume that some of the vocabulary words recurred on the MT. This might have given students somewhat of an outline to study for the exam, which may have led to the higher scores. This surmise may help to explain why some got higher scores; however, it also leaves open the question of why the other class did not. Both instructors followed the same syllabus and covered the same material. This means that both classes should have been exposed to the same material. However, the variable that was not the same was the amount of exposure to L2. The class that received more exposure seems to have benefited more, but this did not answer the question of whether the difference was statistically significant. To answer this question, an independent sample t-test was needed, which revealed a significant statistical difference in MT scores between the two classes.

The independent sample test shows that the non-Arabic group's MT scores had a statistically significant higher score than the Arabic group's scores. The extra exposure may have been a contributing factor to the higher student MT scores. This is not to say that there may not have been other factors that contributed to the results, such as motivation; however, neither the classroom observation nor the student questionnaire has shown any significant difference between the two classes in student motivation. Neither tool was designed to explore motivation; rather, the observation confirmed the use or non-use of L1 in the class and tried to identify patterns in the roles L1 played. The role of the questionnaire was to provide a better understanding of student perceptions and attitudes. Both the TOEFL and the IELTS exams have been used to answer RQ 1.1 and 1.2, and they have been analyzed and have yielded some unexpected results. As was the case for the MT, the TOEFL exam reading scores have shown that the students who were not exposed to L1 in class had a higher mean average than the students who were exposed to L1 in class. From the previous correlation test that was conducted, a prediction proved accurate in that in the TOEFL overall scores for S52, the class not exposed to L1 also had a higher mean average than students exposed to L1.

The assumption that most would make is that this would be followed through on the IELTS exam. Surprisingly, however, this was not the case in that the IELTS results proved to be more complicated than those of the TOEFL. To establish whether L1 aided in L2 reading comprehension, the IELTS exams results were also analyzed, and this analysis revealed that S52, the class that was not exposed to L1 in class, did achieve higher scores on the IELTS reading section of the exam; however, they did not perform better on the overall. Rather, the class that was exposed to L1 achieved higher scores overall. This pattern was not the same as was found when analysing the TOEFL exam scores in that S52 received higher scores for the TOEFL reading section and for the overall TOEFL exam. One factor that may have contributed to the difference in results is the types of questions that comprised each exam. Both the MT and the IELTS exams tested similar concepts. Students would have had practice/experience in answering these types of concepts in class when they were reviewing for the MT as well as taking the actual exam.

1.7 Quantitative Data Analysis: Inferential Data MT Exam Scores

The next section reports on the results of the Statistical Package for the Social Sciences (SPSS) tests that were conducted on the students' reading MT results, which provided empirical data that was used to answer the main RQ as well as RQ 3. As this research focuses on the impact of L1 in teaching L2 reading, the next section will examine the MT grades of both classes. Rather than just use the data of the proficiency exam, the MT was used to give a more detailed picture of the progression of students throughout the term. This was the first tool that followed the quantitative approach in data collection. The following results assist in answering RQ 1.2: 'Is there any significant difference between the MT scores of students who are exposed to L1 in reading comprehension classes and those who are not?'

Table 1: S51/52 MT Exam Results (2012)

	N	Mean	Std. Deviation
Sec. 51 Controlled (Arabic)	16	11.1181	4.09293
Sec. 52 Experimental (non-Arabic)	16	13.8200	3.21469

Table 1 illustrates the mean scores between S51 and S52. The test shows there exists a difference in mean, a measurement arrived at by taking all the scores of the MT and dividing them by the number of exams, when comparing the mean scores for S51 (Mean = 11.1181, N = 16, SD = 4.09293) and S52 (Mean = 13.82, N = 16, SD = 3.21469) groups. The test shows there is a slight difference in the overall average between S51 (controlled) and S52 (experimental) groups. The results for S52 are higher (13.8200) than S51 (11.1181) by 2.7019.

This test does not answer the question of whether there is statistical difference. In order to determine if there is any statistical difference and if the difference in mean was statistically significant, another test, the independent sample t-test, was needed.

Table 2: S51/52 Statistically Significant MT Scores (2012)

MT Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	.053	.819	-2.077	30	.047	-2.70188	1.30111	-5.35910	-.04465
Equal variances not assumed			-2.077	28.405	.047	-2.70188	1.30111	-5.36537	-.03838

To answer RQ 1.1: 'Is there any significant difference between the MT scores of students who are exposed to L1 in reading comprehension classes and those who are not?' Levene's test for equality of variances was needed. To establish whether this difference is statistically significant, independent sample t-tests have been carried out, and they showed that there exists statistically significant difference in mean when comparing the mean scores of S51 and S52 groups ($t = 2.077$, $df = 30$, $p < 0.05$). The independent sample t-test answered RQ 3 by showing that S52 MT scores were statistically significant higher scores than S51, which indicates that the class whose instructor was not using the students' L1 (S52) was performing better in the reading skill.

1.8 Inferential Proficiency Exam Results

The next section reports on the results of the SPSS tools that were conducted on the students' IELTS and TOEFL proficiency exam results to answer the RQ 1.1 and 1.2 in finding if there is any significant difference between the TOEFL/IELTS reading scores of students who are exposed to L1 in reading comprehension classes and those who are not. Both of these proficiency exam results have been analyzed given that they are the only proficiency exam results that the University accepts for students to enter their English-medium majors. These were independent exams provided by an outside institution, unlike the MTs that were provided by the University.

Table 3: Independent Sample t-Test for Statistical Significance in IELTS (2012)

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
IELTS Reading	.029	.868	-.578	17	.571	-.1193	.2064	-.5548	.3162
			-.559	13.251	.586	-.1193	.2135	-.5796	.3410
IELTS Avg	.158	.696	.161	18	.874	.0417	.2589	-.5023	.5856
			.173	17.911	.865	.0417	.2409	-.4645	.5479

To answer RQ 1.2, the independent sample t-test was needed, and the results revealed the difference in mean between S51 and S52 IELTS reading. The test shows that the difference in mean between the two groups is not statistically significant ($t = -.578$, $df = 17$, $p = .571$, 2-tailed). The same test also reveals that the difference in mean between S51 and S52 IELTS overall average is not statistically significant ($t = .161$, $df = 18$, $p = .874$, 2-tailed).

The results have no statistical significance and may be due to chance. It can be assumed that using a student’s L1 in the class to teach L2 did not have any statistical significance towards improving a student’s IELTS reading or overall score.

Table 4: Independent Sample t-Test for Statistical Significance in TOEFL (2012)

		Levene's Test for Equality of Variances															
		F		Sig.		t		df		Sig. (2-tailed)		Mean Difference		Std. Error Difference		95% Confidence Interval of the Difference	
														Lower		Upper	
TOEFL Reading	Equal variances assumed	3.234		.085		-1.468		23		.156		-2.833		1.930		-6.825 1.158	
	Equal variances not assumed					1.344		13.879		.201		-2.833		2.109		-7.360 1.693	
TOEFL Avg	Equal variances assumed	.337		.567		-1.368		25		.183		-14.727		10.763		-36.894 7.440	
	Equal variances not assumed					1.346		20.405		.193		-14.727		10.938		-37.515 8.061	

To answer RQ 1.2, an independent sample t-test was used, which helped to establish if there is any statistical difference between S51 - Arabic and S52. In reference to the TOEFL reading and overall scores, an independent sample t-test was used. The independent sample t-test reveals that the difference in mean between S51 and S52 TOEFL reading is not statistically significant ($t = -1.468, df = 23, p = .156, 2\text{-tailed}$). The test also reveals that the difference in mean between S51 and S52 TOEFL average is not statistically significant ($t = -1.368, df = 25, p = .183, 2\text{-tailed}$). The results have no statistical significance and may be due to chance. The tests above have shown that even though there was a slight difference in mean between the TOEFL reading and the TOEFL overall average, it may be due to chance, as well as many variables and factors that may have influenced a class or student’s performance, such as amount of studying, background knowledge, outside help, exposure to L1, or a variety of motivational factors that were not being tested for this study.

1.9 Inferential Quantitative Results Analysis

The analyses of the two previous sections have provided empirical evidence for the study and help answer RQ 1, 1.1 and 1.2. Both the MT grades and the proficiency exam scores have been used to explore the effectiveness of using a student’s L1 to teach L2.

A main objective of the data was to answer a primary question: ‘Was there any statistical difference in the test scores of the classes?’ to answer the main RQ: ‘Does the instructor’s use of L1 in class aid students in L2 reading comprehension?’ If there was a difference, the next question would ask whether this difference was statistically relevant. The last question that was investigated sought to determine which class had better exam score results.

2.0 Discussion and Conclusion

The difference in types of concepts present on the exams may be a contributing factor to the end test results. This may imply that the students that were not exposed to L1 were able to answer the different types of questions more accurately than the students that were exposed to L1. The implication of the results leads to the hypothesis that L1 did not aid in the comprehension of L2 reading, as all three reading exam (MT, TOEFL, IELTS) results reported, and there is a need to establish whether these results are statistically significant or may be due to chance. With the class that was not exposed to L2 achieving higher scores on the MT, IELTS, and TOEFL reading section of the exams, and with the class that was exposed to L1 achieving higher scores in the overall IELTS section of the exams, an independent t-test exam was conducted to determine whether the overall scores were statistically significantly for either one of the classes. The results have proven that there was no statistical significance in overall scores for the classes, and that scores are perhaps due to chance. No statistical significance for the four results was unexpected since initially a pattern seems to have been identified in MT reading, IELTS reading, and TOEFL reading results, showing that S52, the class that was not exposed to L1, was performing more effectively. These results have helped to reaffirm the importance of fully conducting all necessary tests to answer the RQs. Initial response was that there was a difference, and a pattern was found that may prove the difference.

It cannot be conclusively determined whether this was a chance pattern until all relevant analysis has been performed; completion of this analysis is necessary to establish reliable results. In conclusion, the study's results have not validated the UAE's assumption that not allowing the use of L1 in the classroom will aid in the acquisition of L2. However, test scores on MT exams did reveal a statistically significant difference between the two classes; S52, which is the class that did not use L1, scored higher. With the initial findings implying that the use of L1 did not aid in the comprehension of L2 reading, tests given to reveal if the results were statistically significant or if they may be the result of chance revealed no statistical difference between S51, the class that used L1, and S52, the class that did not use L1. The study has not been able to justify the expenditures needed to maintain the program. Instead, the study has found that for both proficiency exams, there is no statistical significance between using L1 in the classroom and not using it.

References

- Abu-Dhabi Urban Planning Council (ADUPC). (2011). Abu Dhabi Vision 2030. [Accessed 7 August 2011] Available at: <http://www.upc.gov.ae/abu-dhabi-2030/capital-2030.aspx?lang=en-US>.
- Ammon, U., & McConnell, G. D. (2002). *English as an Academic Language in Europe*. Frankfurt: Peterlang.
- Bardsley, D. (2009). State pupils' English skills delay degree programs. *The National*. [online] 3 February 2009. [Accessed 20 September 2012]. Available at: www.thenational.ae/news/uae-news/education.
- Butler, F. A., & Stevens, R. (2001). Standardized assessment of the content knowledge of English language learners K-12: Current trends and old dilemmas. *Language Testing*, 18(4), pp. 409-427.
- Carrasquillo, A., & Rodríguez, V. (2002). *Language Minority Students in the Mainstream Classroom*. New York: Multilingual Matters Limited.
- Cenoz, J., & Gorter, D. (2011). Focus on multilingualism: A study of trilingual writing. *The Modern Language Journal*, 95(3), pp. 356-369.
- Cummings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, 28(3), pp. 369-382.
- Duff, P. (2002). [Research approaches in applied linguistics](#). In R. Kaplan (Ed.), *Handbook of Applied Linguistics*. Oxford: Oxford University Press. pp. 13-23.
- Farah, S., & Ridge, N. (2009). Challenges to Curriculum Development in the UAE. *Dubai School of Government Policy Brief* [online]. No16, [Accessed 10 March 2010]. Available at http://www.dsg.ae/En/Publication/Pdf_En/DSG%20Policy%20Brief%2016%20English.pdf.
- Farhady, H. (2005). The effect of coaching on TOEFL type and task-based tests. *Pazhuhesh-e Zabanha-ye Khareji* [online]. no. 9, [Accessed 3 December 2010]. Available at <http://www.aua.am/academics/dep/hf/publications/7%20Coaching.pdf>.
- Firth, A., & Wagner, J. (2007). Second/foreign language learning as a social accomplishment: Elaborations on a reconceptualized SLA. *The Modern Language Journal*, 91(s1), pp. 800-819.
- Fox, W. (2008). 'The United Arab Emirates and policy priorities for higher education,' In Davidson, C. and Smith, P. (eds). *Higher Education in the Gulf States: Shaping Economies, Politics and Culture*, London: SOAS. pp. 110-125.
- Gallagher, K. (2011). Bilingual education in the UAE: factors, variables and critical questions. *Education, Business and Society: Contemporary Middle Eastern Issues*, 4(1), pp. 62-79.
- Henning, G. (1986). Quantitative methods in language acquisition research. *TESOL Quarterly*, 20(4), pp. 701-708.
- Humphreys, P., & Mousavi, A. (2010). Exit-testing—a whole of university approach. *Language Education in Asia*, 1(1), pp. 8-22.
- Hyland, F., Hyland, P., & Kelly, K. (2009). Toward a Culture of Reading. *Journal of Research*, 19. pp. 97-113.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational researcher*, 33(7), pp. 14-26.
- Kachru, B. B. (Ed.). (1992). *The Other Tongue: English Across Cultures*. Illinois: University of Illinois Press.
- Kharma, N. N., & Hajjaj, A. H. (1989). Use of the Mother Tongue in the ESL Classroom. *IRAL*, 27(3), pp. 223-35.
- Kohn, A. (2000). *The case against standardized testing: Raising the scores, ruining the schools*. Portsmouth, NH: Heinemann.
- Mohan, B. A. (1986). *Language and content: Reading*, Massachusetts: Addison-Wesley.

- Orlikowski, W. J., & Baroudi, J. J. (1991). Studying information technology in organizations: Research approaches and assumptions. *Information Systems Research*, 2(1), pp. 1-28.
- O'Sullivan, A. (2009). Reading in the Gulf: Is it time for a re-focus?
<http://marifa.hct.ac.ae/files/2009/04/Reading-in-the-Gulf-Is-it-Time-for-a-Re-Focus.pdf>.
- Short, D. J. (1993). Assessing integrated language and content instruction. *TESOL Quarterly*, 27(4), 627-656.
- Spolsky, B., & Hult, F. M. (Eds.). (2010). *The Handbook of Educational Linguistics*. Oxford: Wiley-Blackwell.
- Storch, N., & Hill, K. (2011). What Happens to International Students' English after one Semester at University? *Australian Review of Applied Linguistics*, 31(1). pp. 3.1-3.17.
- Taylor, M. (2008). *Orthographic and Phonological Awareness Among L1 Arabic ESL Learners: A Quasi-Experimental Study* [online]. PhD. Thesis. University of Phoenix. [Accessed 12 June 2009]. Available at: ProQuest.
- Toohy, K. (2000). *Learning English at School: Identity, Social Relations, and Classroom Practice* (Vol. 20). New York: Multilingual Matters.
- Traish, Amira. (2013) *A Comparative Analysis of Proficiency Scores of University Students in the United Arab Emirates: The Effectiveness of Using a First Language When Teaching a Second Language in a Reading Class (unpublished doctoral thesis)*. British University in Dubai, Dubai, UAE.
- University of Sharjah (2010). English Language Centre [online]. [Accessed 1 December 2010] Available at: <http://www.sharjah.ac.ae>.

Appendix

Table 1: S51/52 MT Exam Results (2012)

	N	M	Std. Deviation
Sec. 51 Controlled (Arabic)	16	11	4.09293
Sec. 52 Experimental (non- Arabic)	16	13	3.21469

**Table 2: S51/52 Statistically Significant MT Scores (2012)
MT Independent Samples Test**

	Levene's t-test for Equality of Means								
	Test for Equality of Variances								
	F	Sig.	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	.053	.819	-2.077	30	.047	-2.70188	1.30111	-5.35910	-.04465
Equal variances not assumed			-2.077	28.405	.047	-2.70188	1.30111	-5.36537	-.03838

Table 3: Independent Sample t-Test for Statistical Significance in IELTS (2012)

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
IELTS Reading	Equal variances assumed	.029	.868	-.578	17	.571	-.1193	.2064	-.5548	.3162
	Equal variances not assumed			-.559	13.251	.586	-.1193	.2135	-.5796	.3410
IELTS Avg	Equal variances assumed	.158	.696	.161	18	.874	.0417	.2589	-.5023	.5856
	Equal variances not assumed			.173	17.911	.865	.0417	.2409	-.4645	.5479

Table 4: Independent Sample t-Test for Statistical Significance in TOEFL (2012)

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
TOEFL Reading	Equal variances assumed	3.234	.085	-1.468	23	.156	-2.833	1.930	-6.825	1.158
	Equal variances not assumed			-1.344	13.879	.201	-2.833	2.109	-7.360	1.693
TOEFL Avg	Equal variances assumed	.337	.567	-1.368	25	.183	-14.727	10.763	-36.894	7.440
	Equal variances not assumed			-1.346	20.405	.193	-14.727	10.938	-37.515	8.061