



FORMATIVE TESTING PROGRAM FOR UNDERGRADUATE EFL MOROCCAN STUDENTS

Nadia Nouriⁱ

Mohammed V University,
Rabat, Morocco

Abstract:

While most previous studies attribute academic achievement to external and internal motivation, the present study associates academic achievement to language proficiency of EFL learners. Students' underachievement of summative tests might be explained with their lack of English language proficiency. Students may have some gaps in the language that have been accumulated since elementary or high school. These gaps become fossilized in students' minds if not addressed properly. In this regard, this study presents a testing method that leads to a remedial teaching program at the undergraduate level of the English studies at the Moroccan EFL University.

Keywords: academic achievement, proficiency level, remedial program, summative test, underachievement

1. Introduction

Testing is one of the focal points of university education. It is considered an end and a means of measuring students' learning behaviour at the same time (Sax, 1980). There are two major in-class tests among others in education: formative and summative tests. The former refers to the continuous assessment teachers administer in class to examine students' progress against a learning objective. This type of test develops students' responsibility towards their own learning (Black & Wiliam, 1998; Stiggins, 2002). It also shows them the gaps that they still need to address to reach the learning objectives. Summative tests, on the other hand, evaluate the academic achievement that students have attained at the end of a semester or an academic year (Kibble, 2017).

It should be noted in this context that not all students are capable of attaining the ultimate point of acquiring a skill at the end of a period of time (say the end of a semester). Some students need further time to make additional efforts and to assimilate the bulk of information to which they are exposed. The inability of these students to meet the expected requirements at a specific time of acquisition may not be due to their fixed

ⁱ Correspondence: email nadia.elnouri@yahoo.fr

abilities like intelligence or attention, but it may be related to the gaps or deficits that they may have accumulated up to that moment of assessment. First, the problem lies, therefore, in the time dedicated to acquiring the skill (say writing a monograph). Second, students who have a clear gap in their language acquisition should be highly encouraged by teachers to work on those language problems and be commended on the progress they make in their learning process.

With this in mind, a good question comes to mind: Do all students need the same amount of time to acquire the same language skill? University teachers, obviously, cannot give students different amounts of time to acquire their required skills as final exams are delivered at the same time (usually at the end of the semester, course or the academic year). In this context, the present study attempts to show the importance of an alternative testing approach that informs a remedial teaching program wherein the same content area is taught with different instructions and scaffolding compared to the instructions taught in a non-remedial program.

This remedial approach denotes that EFL students at the same academic level may not have the same level of language proficiency, the thing that could influence their academic achievement. Following the suggested approach, therefore, would satisfy the needs of low achievers and enable all students to reach the ultimate objectives of learning. It would also encourage these students to develop their language skills instead of only being concerned with passing an end of term test regardless of their language deficits that could be fossilized if not treated properly as has been mentioned before. In this regard, the present research paper focuses on the following objectives:

- 1) To investigate whether there is a statistically significant difference between the language proficiency of EFL undergraduate students belonging to two academic levels.
- 2) To examine a testing method that may upgrade the academic achievement of EFL undergraduate students according to their academic level.

2. Literature Review

2.1 Achievement Test

Given the above, the focal point of this research paper is to develop the quality of learning at the university, especially at the undergraduate level of the English studies in the Moroccan university. Educators, teachers, and students are well aware that not all students who pass a test at the end of the academic year or semester acquire the necessary skills taught at that level. Therefore, a critical reflection on this issue is highly required here. It is no doubt that university professors have a number of lessons to cover during a semester or an academic year, the thing that makes it difficult or almost impossible for them to trace individual language deficiencies of their learners. In this regard, one should question what could be done to bridge the gap between high and low achievers without interrupting the flow of the content-area course.

Some researchers and stakeholders have formed the Committee on Educational Excellence and Test Equity (CEETE) (National Research Council, 2000) to monitor the

impact of education reform on students at risk of academic failure. Although several reasons were attributed to the latter, this research paper focuses exclusively on the lack of English proficiency. Thus, factors such as poverty or the like are not the focus of this research paper. On the other hand, other researchers have associated the outcome of achievement tests, especially underachievement, to the high rates of academic dropout (Smith, 2010). In this regard, the problem of EFL students failing their tests of an academic year seems to be treated as an outcome of a curriculum not a process.

Regardless of one's perspective of this underachievement discussion, two fundamental questions can increase awareness and contribute to some extent to the resolution of this problematic issue: a) is there a statistically significant difference in language proficiency between EFL undergraduate students belonging to two academic levels? b) How can a testing method upgrade the academic achievement of EFL undergraduate students according to their academic level? Investigating the first research question might reveal the extent to which the academic level of students represents their proficiency level. Whereas the second question reveals the way, in which achievement tests can be equitable to all EFL students that have different proficiency levels of English and study at the same academic level of university.

Given these issues, it is necessary to have a critical analysis of summative achievement tests from the perspective of lack of English proficiency rather than addressing the problem of underachievement at the macro level. In other words, addressing all the factors that hinder inadequate testing, especially dropping out of university would generally provide no practical remedy.

The concept of '*underachievement*' has always been defined as the "*discrepancy of actual achievement from the predicted value, predicted on the basis of the regression equation between aptitude and achievement*" (McLeod, p. 46, 1979). Thus, the score of the achievement test is related to the average score of the learner's IQ. Thorndike (1963) pointed out that this equation can be meaningless. Reconsidering this concept, therefore, is necessary since underachieving an English test should not be interpreted in the same way as a math or physics test neither should it be associated with intelligence. Each subject matter has its own specifications since different types of intelligences and abilities are required in several subjects.

According to some authors, the concept of achievement denotes either failing or succeeding a summative test (Cohen, 1998; Delamont, 1999; Gorard, 2001). However, Smith (2010) pointed out that this division does not indicate anything about the reason why some students fail while others succeed. To analyze the latter idea further, one may think of academic underachievement as '*direct*' and '*indirect*'. According to the researcher of the present study, a '*direct underachievement*' refers to the students' incapability to pass a final test of a semester or an academic year. Whereas, an '*indirect underachievement*' means that students succeed the test with language gaps or deficits that can be fossilized one day if not treated properly in the process of learning. These two concepts do not present equal threat to the education system at the university. The reason for this idea is that a considerable number of students, who succeed with language gaps, do not dwell

on these deficits most of the time thinking that they have accomplished the required goals of the educational level they succeeded.

In this regard, a different approach of cultivating 'growth achievement' or 'growth mindset' has to be followed. The latter encourages students to address their language problems more objectively to attain the language proficiency of each academic level they study. This growth mindset has already been mentioned by Dweck in her book 'Mindset: The New Psychology of Success' (2007) where she explains that students should refocus their learning process on some intrinsic motives rather than some external merits like grades. Students, therefore, should understand that struggling to pass a summative test or an academic level with several language problems is similar to failing the test in the first place.

In this context, a good performance test should distinguish between those who have performed well and those who have passed poorly. A focus should be done on the students who did poorly on the test. Perhaps, a standardized estimate should be administered along with an achievement test to measure the extent to which students have achieved the required learning objectives and the language proficiency of the academic level. The following section shows how to discriminate effectively between high and low achievers.

2.2 Standardized Vocabulary Test as a Proxy Measure of Language Proficiency

Several previous studies have proven that a standardized test of vocabulary size is a good indicator of the language proficiency of learners (Noro, 2002; Akbarian, 2010; Nouri et al., 2018). For instance, Noro (2002) and Akbarian (2010) have shown in their studies that the language proficiency of students can be estimated by lexical measures such as Vocabulary Levels Test (VLT, Nation, 1990) in determining sub-groups of lower and higher levels. Learners in both sub-groups have shown high and low performance in reading comprehension depending on the high and low levels to which they belong.

Nouri and her co-author (2018), on the other hand, questioned the use of a diagnostic test such as VLT to determine the language proficiency of learners since the latter test uses slices of lexical knowledge. The researchers explained that an achievement lexical measure should be employed instead (e.g., Vocabulary Size Test –VST, Nation & Beglar, 2007) given its wide use of frequency levels and its ability to reveal different lexical profiles of learners. They demonstrated that a standardized measure of vocabulary shows the language proficiency of learners wherein high achievers showed high performance of reading comprehension whereas low achievers did poorly in the same measure of reading comprehension. Hence, it is safe to claim that a standardized vocabulary measurement provides a sound ground for indicating the language proficiency of learners (Nouri et al., 2018).

To go back to the objectives of the present study, it seems quite logical to employ the standardized vocabulary measure as a representative estimate for language proficiency. Thus, a clear idea about the lexical threshold, under which EFL students would be considered underachievers, should be mentioned here.

Several thresholds were given in the literature determining the learners who had problems coping with the difficulties of the language of their academic materials. Schmitt (2014), for instance, showed three frequency bands of lexical thresholds: high frequency band (ranging from 1.000-3.000 word families), mid-frequency band (3.000-7.000 word families), and low frequency band (8.000-9.000 word families). In this regard, frequency then is not only an indicator of language difficulty but also “*a key determinant of acquisition*” (Ellis, 2012, p. 196). In other words, frequency shows the extent to which students have acquired the lexicon of their target language.

Schmitt (2010) explained if learners have the adequate lexical knowledge, they will “*use language without a lack of lexis being a problem*” (p. 70). Considering this claim, one should divide learners depending on their lexical knowledge and thus their language proficiency. This is to provide additional assistance to the students having gaps in their lexical knowledge.

In this context, Laufer and Kalovski (2010) suggests two thresholds of written language 4.000-5.000 word families for minimal comprehension and 8.000-9.000 word families for optimal comprehension. It seems therefore that the 5.000 word family is likely to be the line between low and high performance of written language. In Schmitt’s (2014) classification, 5.000 word families belong to the medium frequency band that Schmitt claims to be the category of academic materials. Given these explanations, the present study will use the 5.000 word families to identify between low and high achievers.

3. Methodology

A number of a hundred and forty students participated in the present investigation. Seventy participants were randomly chosen from each of the second and the third year of BA level of English studies at Mohamed V University in Rabat. In this study, second year students are referred to as group 1 (G1) whereas third year students are group 2 (G2). All participants in both groups meet the ideal sampling criteria. They have similar EFL training, which, as Qian (1998) argues, minimizes the influence of world knowledge and background of academic subjects on their test performance.

In this study, Vocabulary levels Test –VST (Nation & Beglar, 2007) was administered to both groups in order to examine the extent to which students have acquired the 5.000 word families of the VST measure, as has been explained earlier. VST being used as a representative of students’ language proficiency helped to distinguish between low achievers (those who scored below the 5.000 word family) and high achievers (those who scored above the same line). In this context, Nouri et al (2018) noticed in their previous study that EFL students scoring below the 5.000 word family did poorly in the reading comprehension test compared to those who scored above the same lexical average in the VST measure. This might show something about underachievement in this study and the reason why low achievers struggle with the target language of their academic materials.

As all rigorous studies, the present investigation uses statistical analyses to measure the variables in question. Descriptive statistics were calculated using SPSS (22

version) concerning the mean and the standard deviation. A later analysis was further calculated, namely the independent samples t-test.

4. Results & Discussion

Table 1 below shows a numerical difference between VST scores of students in the first group and those of group two. That is, G1 students scored less compared to their counterparts in G2. It is still unclear whether the difference between the scores is statistically significant.

Table 1: Descriptive Statistics of the VST scores (N = 140)

Group	Test	Mean	SD
1	VST	58.90	14.30
2		62.80	16.30

1: Second year of BA; 2: Third year of BA.

This subsection presents results of whether there is a statistically significant difference between scores of the first group referring to the second year of BA degree and that of the second group indicating the third year. By comparison, the mean of VST in G1 ($N = 70$) is numerically lower $M = 58.90$ ($SD = 14.30$) than that of the second one $M = 62.80$ ($SD = 16.30$), as Table 1 shows above. To answer the question of whether G1 and G2 EFL students were associated with a statistically significant difference of vocabulary size, an independent samples t-test was performed.

Table 2: Independent samples t-test results whole sample (N = 140)

VST	Levene's Test		t-test			Skew	Kurtosis
	F	Sig.	t	df	Sig. (2- tailed)		
	1.515	.220	-1.504	138	.135	-.124	-.584

Significance Level .05.

The normality assumption was examined for the execution of the t-test, showing a sufficiently normal distribution (i.e., skew $< |2.0|$ and kurtosis $< |9.0|$; Schmider, Ziegler, Dancy, Beyer & Bühner, 2010) (see Table 2). For the homogeneity of variances, the Levene's F test is satisfactory, $F(138) = 1.15$, $p = .220$. However, Table 2 shows a statistically insignificant effect, $t(138) = -1.50$, $p = .135$. Thus, no statistically significant difference was found between G1 and G2 learners. That is, the higher academic level of group 2 students did not have any effect on their lexical knowledge compared to G1 students belonging to a lower academic level. The employed measure of lexical knowledge (i.e., VST) is likely to be a representative skill of language proficiency, as has been shown before.

This result, therefore, has demonstrated that G2 students have not shown any improvement in the size of their lexicon given the two semesters of study (i.e., one academic year) despite the numerical difference found between the means of scores (see Table 1). Is, therefore, the academic level of EFL students representative of their proficiency level? The answer might not always be positive given several factors, namely

the intake of students, the reading input, and the efforts they make to learn new information. However, the mean of scores in both groups has reached the difficulty of written materials (see Table 1). To explain, the average number of lexical size in both groups is categorized in the medium frequency band (i.e., 3.000-7.000 word families), according to Schmitt's (2014) classification. The latter researcher explains that academic materials belong to this mid-frequency band.

Representative measures of language proficiency such as VST should be administered nationally to EFL undergraduate students at each academic level. This is to demonstrate whether students accomplishing a degree have the same language attainment. However, judging from the average number of all lexical profiles of learners would not provide much insightful results regarding the number of students who have or have not attained the same level of proficiency. A further analysis is needed to distinguish between individual achievements of students. The second question of this research paper investigates the way in which a testing method can upgrade the academic achievement of EFL undergraduate students according to the academic level they study.

After administering the VST measure to all EFL students investigated in the present study. Sub-groups of low and high achievers were determined in each group. The cutting line between the two proficiency levels is the acquisition of the first 5.000 word families in VST. Table 3 shows the number of students of each sub-group as well as the mean, standard deviation, minimum and maximum scores of their lexical knowledge at both academic levels.

Table 3: Descriptive Statistics of lexical size below the 5.000 word families

Groups	Test	N	Percentage	Minimum	Maximum	M	SD
1 ^a	VST	19	27%	20.00	49.00	41.89	7.69
2 ^b		17	25%	30.00	49.00	41.00	7.24

^a: second year of BA; ^b: third year of BA.

At the first group of academic level, 27% of EFL students were found below the 5.000 word families whereas 25% below the same average were found in the second group. The number of students below this lexical average is quite low indicating that EFL students in both groups have acquired the necessary lexis to deal with academic written materials. Another remark is that both academic groups have almost similar percentages of students below the aforementioned cutting line. As has been mentioned before, one of the previous studies have shown that students' performance in reading comprehension was drastically different from students scoring above the 5.000 word families compared to those scoring below (Nouri et al., 2018). The present study is also taking this average as a cutting line between the two proficiency levels because it belongs to Schmitt's classification (2014). The latter researcher explained that academic materials need a lexical knowledge of 3.000 to 7.000 word families.

Table 4: Descriptive Statistics of lexical size above the 5.000 word families

Groups	Test	N	Percentage	Minimum	Maximum	M	SD
1 ^a	VST	51	73%	51.00	91.00	65.23	10.49
2 ^b		53	75%	52.00	93.00	69.79	11.46

^a: second year of BA; ^b: third year of BA.

Table 4 shows high percentage of EFL students scoring above the 5.000 word families (73% in G1 and 75% in G2). These averages are not alarming as more than a half of students scored above the determined average indicating students' ability to deal with academic materials. The VST measure has demonstrated the ability to discriminate between the two sub-groups. The objective is not to single out a group of students from the other but to monitor the students in need of further developing their language skills. To go back to the term direct and indirect underachievers explained above, one should note that it is unlikely for students scoring above the indicated lexical average to underachieve their summative tests unless they have not been prepared for it. However, those who might not succeed in the other sub-group (scoring below 5.000 word family) are called direct underachievers compared to the indirect ones who succeed but still present some noticeable issues in their language compared to high achievers.

5. Implications

The distinction between the two sub-groups should not be attributed to students' intelligence, as what is needed is further language training. Instead, summative tests as well as the representative test of language proficiency might be a formative evaluation program where EFL students are examined for remedial replacement. To explain, after a freshman has been exposed to and familiarized with the academic material of the first year of university, they should be tested and replaced in either a remedial program or a normal one. The former program should be addressed to *direct* or *indirect* low achievers whereas the latter is given to students who have not shown any serious language problems. Day-to-day instructional adjustments (Shepard, 2005) can be helpful, but a remedial program might have a better result since major focus would be on developing the same skills or sub-skills to all students not adjusting instructions based on students' individual needs. The more likely the need is similar, the more effective is the teaching, and the easier is the evaluation of that remedial program.

One wonders how best to test content-area subjects in this case. It should be noted that the knowledge in such courses remains the same, but different instructions and possibly materials should be provided to the remedial program versus those in the non-remedial one. In the meantime, it is up to the teachers to decide whether the tests of the two programs should have the same language difficulty according to the performance of EFL students in continuous assessment. Hence, dividing students into high and low achievers would also inform decision on the type of materials students should study. Following this test/teaching method would consequently cultivate the growth mindset (Dweck, 2007) among students. More importantly, this systematic approach of testing

might enable objective comparison across different programs such as Linguistics and Culture or BA and MA at different Moroccan EFL universities.

6. Conclusion

Remedial teaching programs could function effectively if they meet the needs of EFL learners. Employing this educational approach at the undergraduate level of English studies is not a luxury but a necessity since a considerable number of EFL students might have accumulated language errors since elementary or high school. Thus, these language gaps should be better considered as a process not an outcome of the teaching/learning procedure that has to be treated during a long period. Day-to-day instructional adjustments (Shepard, 2005) can be effective but cannot be traced individually given time constraints facing university teachers. Moreover, underachievement is not a generic term that can be interpreted in the same way across different subjects. For instance, a language course is different from a math subject and so are their requirements. Rather, the term underachievement as was shown in this research paper is not an all or nothing outcome but it could be direct or indirect given the explanations provided in this article.

Conflict of Interest Statement

The current research fills the gap in the literature of what was previously investigated in the formative testing of EFL Moroccan students, and does not have any conflicts of interest with other research studies.

About the Author

Nadia Nouri is a PhD holder in “Applied Linguistics” from the Faculty of Education Sciences, Mohammed V University-Rabat, Morocco. She holds a master degree in Teaching English as a Foreign Language (TEFL) from Ibn Tofail University, Kenitra Morocco, and is currently teaching English at the American Language Center and at the INPT (Posts and Telecommunication National Institute) in Rabat, Morocco.

[ORCID ID: 0000-0001-9775-3102](https://orcid.org/0000-0001-9775-3102)

References

- Akbarian I, 2010. The relationship between vocabulary size and depth for ESP/EAP learners. *System*, 38, 391-401.
- Black P, Wiliam, D, 1998. Assessment and Classroom Learning. *Assessment in Education*, 5:1, 7-71.
- Dweck C, S, 2007. *Mindset: The New Psychology of Success*. Ballantine.
- Ellis N, C, 2012. Frequency-based accounts of second language acquisition. In S. M. Gass, & A. Mackey (Eds.), *The Routledge handbook of second language acquisition*, pp.193-210. London/ New York: Routledge.

- Kibble J, D, 2017. Best practices in summative assessment College of Medicine, University of Central Florida, Orlando, Florida.
- McLeod J, 1979. Educational Underachievement. *Journal of Learning Disabilities*, 12:5, 322–330. doi:10.1177/002221947901200508
- Nation I, S, P, 1990. *Teaching and Learning Vocabulary*. New York: Newbury House.
- National Research Council, 2000. *Testing English-Language Learners in U.S. Schools: Report and Workshop Summary*. Committee on Educational Excellence and Testing Equity. H. Kenji & A. Beatty, eds., Board on Testing and Assessment, Center for Education. Washington, DC: National Academy Press.
- Noro T, 2002. The roles of depth and breadth of vocabulary knowledge in reading comprehension in EFL. *ARELE*, 13, 71-80.
- Nouri N, Zerhouni B, 2018. Lexical Frequency Effect on Reading Comprehension and Recall. *Arab World English Journal (AWEJ)*, 9:2, 234-250. doi: <https://dx.doi.org/10.24093/awej/vol9no2.16>
- Sax G, 1980. *Principles of Educational and Psychological Measurement and Evaluation*. Belmont: Wadsworth Publishing Company, pp 13-25;
- Schmider E, Ziegler M, Danay E, Beyer L, Buehner M, 2010. Is it really robust? *Methodology European Journal of Research Methods for the Behavioral and Social Sciences*, 6:4, 147-151.
- Schmitt N, 2010. *Researching Vocabulary: A vocabulary research manual*, pp. xi-366. New York: Palgrave Macmillan.
- Schmitt N, 2014. Size and depth of vocabulary knowledge: What the research shows. *A Journal of Research in Language Studies*, 64:4, 913-951. <https://doi.org/10.1111/lang.12077>
- Shepard L, A, 2005. *The Future of Assessment: Shaping Teaching and Learning* New York. Proceedings of ETS Invitational Conference.
- Smith E, 2003. Understanding Underachievement: an investigation into the differential attainment of secondary school pupils. *British Journal of Sociology of Education*, 24:5, 575-586. doi: 10.1080/0142569032000127143.

Creative Commons licensing terms

Author(s) will retain the copyright of their published articles agreeing that a Creative Commons Attribution 4.0 International License (CC BY 4.0) terms will be applied to their work. Under the terms of this license, no permission is required from the author(s) or publisher for members of the community to copy, distribute, transmit or adapt the article content, providing a proper, prominent and unambiguous attribution to the authors in a manner that makes clear that the materials are being reused under permission of a Creative Commons License. Views, opinions and conclusions expressed in this research article are views, opinions and conclusions of the author(s). Open Access Publishing Group and European Journal of Education Studies shall not be responsible or answerable for any loss, damage or liability caused in relation to/arising out of conflicts of interest, copyright violations and inappropriate or inaccurate use of any kind content related or integrated into the research work. All the published works are meeting the Open Access Publishing requirements and can be freely accessed, shared, modified, distributed and used in educational, commercial and non-commercial purposes under a [Creative Commons Attribution 4.0 International License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).