

The Relevance of Students Performance to Intensive English Program Evaluation in Saudi Context

Mubarak S. Al-Braik

**Department of Foreign Languages, College of Education,
King Faisal University, Saudi Arabia**

Abstract :

The recent expansion in scientific specializations in Saudi Universities has led to an increase on the emphasis on teaching English to prepare students for studying in English. The program offered by King Faisal University (KFU) in Saudi Arabia is one example of this new surge in teaching English. The program which has been running in the Colleges of Medicine, Architecture in the Dammam Campus and Veterinary Medicine in Al-Ahsa campus has not been subjected to professional evaluation. The aim of this paper is to evaluate students' performance in this program which is currently offered to four colleges in the main Campus of KFU in Al-Ahsa: College of Medicine, College of Veterinary Medicine, College of Clinical Pharmacy, and College of Computer Science and Information Technology. Only the College of Medicine currently admits girls.

The Michigan Proficiency Test (MPT) was given to all five groups in these colleges in the first and last week of instruction, and the results were compared with the performance of students as evaluated at the end of the program by course instructors. The results show that the achievement in the MPT is rather low and that the instructors' evaluation tended to be higher. The instructors' evaluation was seen to be higher in the term work compared with the final examination in two groups: the Veterinary Medicine and the Computer Science. These are the lower achievers in the five groups, and the lowest in terms of the intake.

The paper concludes that the overall low achievement in the MPT shows that the external general test is not sensitive to the local situation in which the program is run. The low achievement and negative correlation in the Veterinary group reflect week intake, absenteeism and a local classroom culture. These features point to the need for classroom research.

Keywords :

1. Students' performance
2. Intensive English
3. Intensive Program evaluation

Introduction :

Under pressure of accreditation and quality management of academic programs, evaluation has more often than not, been presented as a bureaucratic checking on the work of academic departments and academic staff (CVCP Academic Audit Unit, 1992). In some cases it has evoked a coordinated effort to secure the rights of academics in the dynamics of change from traditional methods of learning, teaching and evaluating, to the technology of the Internet, distance learning, and teachers' evaluation (American Association of University Professors (AAUP) 2007 and Watt 2000). Mackay has put things right when he says that "an additional advantage of the information generated by program-based review is that some of it can be presented in an appropriate and summarized form to meet the interests and concerns of the bureaucracy. A project or program which can show the bureaucracy what its strengths and weaknesses are, and how it plans to reinforce the former and overcome the latter, cannot fail to impress the "powers that be", and runs less risk of having an extrinsically motivated evaluation imposed upon it" (Mackay 1994, p. 149).

Performance is one area in English program evaluation, but it is by no mean the only area (Mackay 1994). In Arabic contexts, there is general awareness of the problems surrounding teaching English (Al-Hajailan 2003, and Al-Saadat, Al-Braik and Al-Shabab 2005), but local, program-specific, data have been scarce in the last twenty years or so. The present paper is part of an effort to evaluate one intensive English program offered to first-year university students at King Faisal University, Saudi Arabia. The paper investigates two aspects of students' performance: 1. learners' performance in two standardized tests used as external reference, and 2. two formal assessments by program instructors used as localized evaluation instruments. Four questions can be posed in connection with students' performance: (1) what is the learners' level when they begin the program? (2) What is the improvement rate at the end of one academic semester? (3) What is the significance of the instructor's grades at the end of the program? (4) What are the factors that contribute to learners' performance in the local situation?

According to Mackay, one of the key areas of program evaluation is Students' performance (Mackay 1994). Program evaluation as perceived and practiced in Mackay (1994) clearly indicates two perspectives: external or internal evaluation.

"It is, I believe, more realistic to distinguish between extrinsically motivated evaluations which are conceived, motivated, and designed at the bureaucratic level, and intrinsically motivated evaluations which are

conceived, motivated, and designed at the programme level, with or without the direct involvement of the bureaucracy."

(Mackay 1994, P. 143)

The current pressure on program directors in the Saudi context is to submit to "extrinsically" administered evaluations. Therefore, it may be advisable on the part of independent academic units at the level of centers and departments to undertake a self-assessment exercise which could serve a variety of purposes. The focus of the present work is, however, on students' level and achievement, since, as Mackay himself points out "evaluations which address discrete issues over which program personnel have some or total control are the ones that generate information useful for program improvement" (Mackay 1994, p. 143). This should, in my view, be the prime focus of the personnel of English language centers in Saudi Arabia, i.e. the discrete issues which can be handled at the level of the center, and which are hoped to make a difference in the operation and task-delivery, taking the situation of each center into consideration.

Student performance can be seen as a definitive judgment about the entire operation of the program. For the purposes of identifying weaknesses and progress, and for the sake of presenting "real" rather than "ideal" summary of the effectiveness of a program, students' performance can be adopted as useful indicator of the state of a given program. The evidence obtained from an internal evaluation is valuable for program personnel since it is generally carried out in a congenial atmosphere and has the support of the staff involved, and at the same time it "can be presented in an appropriate and summarized form to meet the interests and concerns of the bureaucracy" (Mackay 1994, p. 149).

One important variable related to students' performance is the level of the intake and the attitude and study habits of program participants. The level of program intake influences students' evaluation of the academic operation as a whole: "Interestingly, in our study, course level is considerably the variable in the analysis that displays the greatest effect on the dimensions of students' evaluation" (Schlenker and McKinnon 1994, P. 19-20). Another variable studied by Schlenker and McKinnon is absenteeism" which also affects students' evaluation of the teaching/learning process (ibid, pp. 15-16). The level of the intake and absenteeism are also relevant practices in the situation under analysis. In fact the high level of attrition may, at least in some colleges of KFU, be another factor that influences students' performance. Therefore, while students' performance can be seen as a definitive statement about the

program, it also reflects the complexity of the situation in which teaching/learning takes place.

Performance as achievement has been investigated by Al-Muhaidib who studies low and high achiever in English as a Foreign language and their female learners' personality traits. Her high achievers are integrative learners rather than instrumentally motivated (Al-Muhaidib 2006, pp. 107-108). In terms of motivational intensity "high achievers spent more effort both inside and outside the classroom to learn English than the low achievers" (ibid.) High achievers also have "positive level of self-esteem" and they are "more anxious" as learners than low achievers (ibid p. 109). . Of course, personality traits and other psychological dimension of second/foreign language learning represent aspects of the complexity of acquiring a second/foreign language. Al-Braik (2001) has studied motivation of university Saudi students majoring in English. He found that the prime motivation for learning English for his subjects was instrumental, concluding that "Motivation would lead to the improvement of the socio-economic status of those subjects, thus providing them with prestige among their acquaintances and in the country at large" (Al-Braik 2001, p. 7). Taking the results from Al-Muhaidib into account, the instrumentalist are not as successful in learning and performance as program providers may wish them to be. The data obtained from students' performance help in "knowing the general trends in learners' performance and offer a concrete measurable manifestation of the learning experience. Such knowledge benefits administrators (Illich et. al. 2004), program designers (Ali 1995), and teachers (Al-Braik et. al. forthcoming), and program evaluators (Shaw 1994).

Examining the situational aspects of learning/teaching general English in the Saudi context leads Al-Saadat, Al-Braik and Al-Shabab to state that

One of the basic aspects of course construction is the establishment of explicit criteria on which learners' levels can be determined. In the local environment, three criteria are particularly needed since much of the dissatisfaction with what is available stems from lack of clarity in establishing the learners' level and lack of will to reinforce the level and evaluation procedures. ... it is not enough to have a target of vocabulary input for a specific level. What is needed is a parallel operational apparatus which activates the vocabulary input and ensures its assimilation and positive use in speech and writing.

(Al-Saadat, al-Braik and Al-Shabab 2005, p. 3)

The above concerns echo Nunan's position which differentiates between "ideals" and "reality" in Curriculum design and course delivery. He writes: "The "curriculum" of a given institution or language program ... can be seen

as a statement of intent, the "what should be: of a language program as set out in syllabus outlines, sets of objectives, and various other planning documents. Another perspective is that of the curriculum as "reality" that is, in terms of what actually goes on from moment to moment in the language classroom" (Nunan 1989, p. 9). He goes on to conclude that "the localized school-based approach to the language curriculum outlined in this article attempts to model the curriculum on instances of successful practice and is therefore tied closely to the classroom." (Nunan 1989, p. 25). It is clear from the above that identifying the gap between "what should be the case" and "what the case is" is at the core of localizing Intensive English programs. It also underlines the present endeavor.

Aspects of the problem facing the English language program managers and all other parties concerned in teaching English in the Saudi context have been highlighted in a seminar held in Yamama College in (2005). At the same time, different aspects of the language of Saudi learners have been studied including the errors of major English students, and the acquisition patterns of English (Al-Mahboub 2007). The present paper investigates a different aspect of the situation at hand by concentrating on an intensive program offered to four colleges at King Faisal University. Still, it would be informative to see if the features under study relate to students performance are maintained by the present subjects, and to what extent they are maintained.

Methodology: Situation, Subjects, Data, Instruments and Hypotheses :

The current evolving situation in King Faisal University is typical of recent developments in many similar academic institutions in Saudi Arabia. English language centers are required to start or revive intensive English courses to students in different specializations across the university. The intake is determined on the basis of criteria set up by university admissions and academic committees. Typically, fresh school leavers apply to the prestigious colleges, medicine, computing, and pharmacy, or to colleges where jobs can be easily found, typically education. High fliers tend to go to the above mentioned colleges, and the College of Education has been recently accepting "literary" secondary school section, which generally attract weaker students. But the situation is volatile and changes in intake and admissions can take place within short notice.

One thing is quite stable and that is the main focus of the resent paper, i.e. the level of English given to the intake. The significant performance indicator sought here is the amount of progress achieved by program

participants. Students are required to attend classes (average 18 hour of English per week), but they are allowed to be absent for 24% of course time. Beyond this they are asked for official excuse or being stopped from taking exams. Most students in the above three colleges do attend relatively regularly, but there can be a sizable proportion of attrition, students who decide to take up jobs, transfer to different institutions, or even opt for a career. Students do not pay any tuition fees and they receive a monthly allowance of about two hundred and fifty dollars.

Typically, two or three members of the English language Center teach on each program, and the four skills are taken as a basis for the teaching operation, with a steady movement from general English to more scientific texts. Continuous evaluation is implemented and there is a formal midterm examination and a final examination which cover the four skills and which amount to about fifty percent of the grades.

Learners' level at the starting point and their level after instruction measured by external instruments in the form of a standardized test can be compared with the results of the achievement tests the learners take at the end of the course. The achievement scores are assigned collectively by course instructors in each college. The difference between the achievement score and the standardized test constitutes one indicator of students' performance. The post test scores show the level of students who are allowed to continue their studies in their relative colleges. As such, the improvement in the intensive program should be benchmarked against improvement of these final scores in achievement and standardized tests taken at the end of program.

Thus, students' performance will be measured at two points by using two instruments. The first point is a pre-session test administered by program instructors under the supervision of the researcher. The second point is the last week of instruction. The first instrument is the university of Michigan English proficiency test. Only sixty item which cover grammar, vocabulary, and reading comprehension are used. To maximize comparability, the same text is given before and after instruction. The second instrument utilizes students' achievement as obtained from instructor's evaluation of students throughout the running of the program. Program instructors' are required to give a midterm examination in week 8/9 of a 15 week term and 18 hours of English. Only students of medicine have a yearly program of 13 hours per week. Instructors also give quizzes and a final examination which, in general, accounts for 50% of the program total. The instructors carried on with teaching and evaluation without being told that the results would be

used as data to compare students' achievements, which meant that their evaluation is both normal and bias-free. At the end of the academic year (2006-2007), the researcher collected the scores of the two instruments as shown in Table (1). Table (2) shows the number of subjects.

Table (1)
Tests and Subjects of the IEP.

Instruments	Subjects				
	MedF	MedM	Vetr	Phar	Comp
Test1	1 st week				
Test2	End of term				
Term	Accumulative	Accumulative	Accumulative	Accumulative	Accumulative
Final	End of year	End of year	End of term	End of term	End of term
Total	End of year	End of year	End of term	End of term	End of term

Table (2)
Number of subjects of the IEP.

College	MedF	MedM	Vetr	Phar	Comp	Total
Numbers of subjects: Test1, Test 2, Total	57	69	26	14	36	202
Numbers of subjects: Term, Final & Total	73	71	55	17	64	280

Terms and abbreviations :

KFU: King Faisal University;

IEP: Intensive English Program.

MedF: Female students of the College of Medicine.

MedM: Male students of the College of Medicine.

Vetr: Students of the College of Veterinary Medicine (Male only).

Phar: Students of the College of Clinical Pharmacy (Male only).

Comp: Students of the College of Computer and Information Technology (Male only).

Test1: The University of Michigan English proficiency Test.

Test2: The University of Michigan English proficiency Test.

Term: Term grades (usually 50 points out of a 100) given by program instructs' for Midterm examination, quizzes and participation, and attendance.

Final: The examination given by program instructors at the end of the program to test the skills and material taught throughout the program.

Total: The overall grade obtained by a student calculated by adding the term score to the final score.

It is clear that not all students who attended the program and obtained instructors' evaluation took the external tests. This could be useful in explaining study habits and absenteeism (see section 4.3 below).

The data collected and the objectives of the paper enable us to investigate students' performance as measured by an external (standardized test) and by program instructors' evaluation. Three hypotheses can be suggested in an attempt to answer the four questions posed in section one earlier. The hypotheses are:

- 1- There is a correlation between scores obtained from Test1 and scores obtained from Test2.
- 2- There is a correlation between scores obtained from Test2 and the Total score obtained from program instructors.
- 3- There is a correlation between Test2 and Term scores obtained from program instructors.

The three hypotheses will provide information about the level of the intake to each of the four colleges offering the Intensive English Program (IEP). This is obtained from the first hypothesis. By comparing the pre-test scores and the post-test scores, the first hypothesis measures the performance after instruction, i.e. the improvement brought about by instruction. The significance of Test2 is seen when its results are compared not only with Test1 to externally measure performance after instruction, but also when its results are used to investigate performance in instructors' evaluation. Correlations between instructors' evaluation and external evaluation of performance will verify the level of students' performance and achievement. Lack of correlation will call for explanation and may help in understanding the complexity of learning a foreign language in a specific locality. Factors influencing instructors' evaluation will help program directors and program staff in seeking ways of improvement in teaching and policy making for the future.

Results :

The results are obtained by using the SPSS to analyze the scores obtained from the proficiency test and the instructors' evaluation. The numbers included in the texts are those who took Test1 and Test2 and have a total score assigned by the instructor at the end of the first term. The tests

were administered by the instructors in the first week (Test1) and the last week of instruction (Test2). In the sections below the results of the analysis are reported, including the Means, Pearson correlations and Standard Deviation.

The Means :

Comparing the means obtained from the analysis of the data at hand reveals the general trends in students' performance. The first performance indicator is seen in the difference between the mean scores of Test1 and Test2 (Columns 4 and 5 Table 3). The lowest mean in Test1 is in Veterinary Medicine Group 24.000 and this group is the lowest in achievement, only 26.692 in Test2. The highest means in Test1 is the Medicine Female group (39.684) and their achievement is the second best (Test2 47.263), next to the Medicine Male group which achieved lower in Test1 (34.790) and higher in Test2 (50.666) , a fact which indicates higher achievement. This is an improvement of about 16 out of a hundred, the highest in all groups.

Table (3)
Means of scores obtained from proficiency tests and instructors' evaluation

College	Students' Numbers		Means					
	Test	Inst.	Test 1	Test 2	Total Inst.	Term Inst.	Final Inst.	Total Inst.
Med. Female	57	71	39.684	47.263	75.578	74.840	75.970	75.408
Med. Male	66	73	34.790	50.666	85.393	82.958	84.219	83.698
Veterinary	26	55	24.000	26.692	59.076	63.454	43.636	53.545
Pharmacy	14	17	33.000	35.714	81.142	80.234	79.764	80.000
Computer	36	64	27.833	34.611	67.472	81.750	46.124	63.937
All Colleges	199	280	33.397	42.603	75.603	76.628	62.234	70.932

The second indicator of achievement is seen in the comparison of the means of Test2 (Column 5) with the means of the instructor's total score for the term (Total Inst., Column 6). The highest mean in the Total Inst. is achieved by the Medicine Male group (85.393) contrasted with (59.076) achieved by Veterinary Medicine group. Medicine Female group achieved (75.578) and the Pharmacy group (81.142). Thus the instructor's total for

pharmacy (81.142) is higher than that of Medicine Female (75.578). A look at the means of Test1 and Test2 shows that Pharmacy group is rather modest (Test1 33.000 and Test2 35.714). This is probably due to the instructor's strict evaluation in the Medicine Female group and to the fact that the Pharmacy group is rather small; only 14 students took Test1 and Test2 and the instructor's total.

Close examination of the means obtained from the instructor' evaluation, reveal remarkable consistency in the term scores, final scores and total scores, except for the Veterinary and Computer groups where the final scores are very low compared with the term scores (Veterinary Term: 63.454; Final 43.636; and Computer: Term, 81.750 and Final 46.124). Thus, the Computer group has in the final achieved only about half the scores of the Term.

Two findings can be reported here: 1) the Means of the Total instructor is higher than those of the external Tests; 2) the term scores are higher than the final in two groups (Veterinary Medicine and Computer), and these two groups are the lowest achievers in Test1, which is the intake score.

Correlations :

The correlations reported in Table (4) are expected to pin down the main performance trends seen in the means. They are examined here in the three main areas specified by the three hypotheses put forward in the previous section. The first correlation corresponds to the first hypothesis. It shows the relationship between the first and the second test performance, and thus it measures students' achievement by using an external instrument (Table 4, Columns 5, 6 and 7). The correlations obtained from MedF and MedM and Phar show some rather high correlations (.722, .705 and .884), which means that the performance is both consistent and the achievement is rather moderate since a remarkably high achievement would show a big positive difference in performance. The correlation of the Comp group shows a lower correlation (.589) which is in keeping with the score indicated by the mean of the group. The Vetr group shows a negative correlation (-.118) which reflects inconsistency in the performance of the group, and again this is in keeping with the low mean of this group. In light of the results above, it can be safely said that the high correlations obtained from using an external measurement show moderate improvement, and the negative correlation shows inconsistency in performance. The first hypothesis is accepted and in external evaluation little improvement is achievement.

The second hypothesis measures the difference between the scores of the external instrument, Test2, and the term total obtained from instructors'

evaluation. If there is a low correlation, i.e. the scores are different, it will show that the instructors' evaluation is not within the same range of that of the external instrument. The results in Table 4, Column 6, show a rather low correlation in the MedF, MedM, and Phar groups (.670, .770, and .691) respectively. The results show a lower correlation for the Comp group (.454) and a very low correlation for the Vetr group (.157). Thus, the low correlations point to the difference between the externally measured scores and the instructors' scores. The second hypothesis is rejected since there is an overall low correlation, i.e. a difference between the two sets of scores.

Table (4)

Pearson correlations of the scores obtained from proficiency tests and instructors' evaluation

College	Students' Numbers		Value of Correlation		
	Test	Inst.	Test1-Test2	Test2-TotalInst	Term-Final
MedF	57	71	.722	.670	.899
MedM	66	73	.705	.770	.819
Vetr	26	55	-.258	.074	.819
Phar	14	17	.884	.691	.872
Comp	36	64	.589	.610	.648
All Col.	199	280	.695	.703	.626

The third hypothesis investigates the relationship between the instructors' term scores and the instructors' final examination scores. The results (Table 4, Column 7) show a trend which is similar to the results above, in the sense that the MedF, MedM, and Phar groups show higher correlations than the two other groups. The correlations of the MedF, MedM, and Phar groups are: .899, .919, and .872, respectively. The correlation for the Comp group is only .648, but the correlations for the Vetr group is surprisingly high (.819). This final result shows that the instructor's evaluation in the term work and final examination is consistent. While the Comp group shows a different in the score for the term work and the final examination. The third hypothesis cannot be accepted since the low correlation in the Comp group shows a difference in the scores of the term work and the final examination.

The correlations in Table 4 can be summarized as follows:

- 1) There is a moderate improvement on the external scale, and no improvement in the Vetr group.
- 2) There is a difference between the externally obtained scores and the instructors' total evaluation.
- 3) There is some difference between the instructors' own evaluation in the term work and final examination, which becomes big in the case of Comp group.
- 4) The low correlations for the five groups are reflected in All Colleges scores, the bottom row in Table 4 above.

Table (5)
Standard Deviation of the scores

College	Student's Numbers		Standard Deviation					
	Test	Inst.	Test 1	Test 2	Total Inst.	Term Inst.	Final Inst.	Total Inst.
Med. Female	57	71	14.715	13.232	12.031	5.948	5.941	11/586
Med. Male	66	73	12.980	13.392	9.188	5.331	7.851	12.594
Veterinary	26	55	5.685	8.521	12.129	8.177	10.804	18.118
Pharmacy	14	17	12.197	11.418	8.198	4.688	4.270	8.664
Computer	36	64	8.923	9.877	15.226	8.863	8.238	15.524
All Colleges	199	280	13.180	14.815	14.650	7.825	11.930	17.901

A look at the Standard Deviation of the scores of the five groups, shows that the Vetr group has the largest amount of SD (18.118) followed by the Comp group (15.524). The number of students in the Phar group is only (14) for the external Tests and (17) for the instructors' evaluation. The rather large amount of SD is reflected in the overall numbers in the bottom row in Table 5.

Discussion and Conclusion :

The findings from the present data show that learners' achievement as measured by Test1 and Test2 is rather low: MedF: 16%; MedM: 17%; Phar: 13%; Comp: 7%; Vetr: 3% (Table 3). In addition to the low achievement, a high degree of discrepancy exists in the scores of the Vetr group, a correlation of -258 (Table 4 above) is observed. Secondly, the difference between the scores of Test2 and the instructors' total is quite big: MedF:

28%; MedM: 35%; Phar: 46%; Comp: 32% (Table 3). There is a marked trend for instructors' scores to be higher than those of Test2, despite the lack of correlation in the performance reflecting discrepancy among the groups and the individual learners within the Vetr group (.074). Thirdly, with two exceptions, little discrepancy in the scores of the instructors' term work and final examination: MedF: 1%; MedM: 2%; Vetr: 20%; Phar: 1%; Comp: 35% (Table 3).

The low achievement in Test1 and Test2 can be viewed from different perspectives. First of all, this external instrument is not sensitive to the situation and it may reflect, in some cases at least, priorities and cultural trends relevant to American usage and culture. The Saudi school curriculum and the intensive program under discussion, address vocabulary and topics quite different from the input of the above Test. From the learners' perspectives, the background of the two situations is different. The objectives, learning habits and classroom culture in which the English Intensive Program is run are different from those assumed by the external proficiency test. Out of 280 students, only 199 took Test1, Test2, and the instructors' examinations. This level of absenteeism and attrition is detrimental to high achievement (Schlenker and McKinnon 1994). It shows lack of motivation to learn English and a possibility of not taking tests which do not contribute to the actual evaluation seriously. It should be noted here that the percentage of absenteeism is higher in Vetr and Comp groups, the lower achievers of the five groups. This may also reflect a deeper reason; namely the weak intake performs less well than the better intake (See Free-Weiss 2004 for the influence if the level of intake).

The higher scores given by the instructors in the Term work and final examination can be explained in a number of ways. The instructors' work in the real situation, in the sense that they know the exact learning habits, behavioral patterns and gaps in the learners language system. They know the specific needs of certain groups and even certain individuals in the program. Therefore, it is natural to find that the instructors' evaluation is sensitive and the evaluation instruments (teacher-made tests) are localized in such a way as to reflect the rating of individuals in the group and the perception of educational authorities which influence the test constructs and instructors' perception of rating (Hughes 1989/2003, pp 130-134)

Finally, the homogeneity of the scores of term work and final examination in the MedF, MedM and Phar groups, shows consistency in the performance of the high performers. It also reflects emphasis on examining the material covered, rather than language constructs or wider curriculum

issues. Conversely, the low scores in the Vetr and Comp groups may result from a challenging final examination and a generous approach to grading quizzes and the Midterm examination.

To conclude, one can safely say that students' performance in the English Intensive Program like the one offered by KFU, reflects local situation with all its complexities. The findings of the current paper show that teaching English to weak students is an area which needs to be investigated to find methodologies and content that suit this type of learner (see Al-Fraidan 2006 for further discussion). In other words, the English level of the intake is an important factor behind low achievement; the program cannot ignore the level of intake and treat all learners in different colleges in the same way. Finally, as Mackay has observed more classroom research is needed to explain students' performance and help evaluate the program at hand.

References:

1. Al-Braik, Mubarak S. (2001). "Saudi Students' Motives to Learn English as a Foreign Language at King Faisal University. Saudi Arabia". *Journal of Resaerch in Education and Psychology in Minia University*, Faculty of Education, Egypt, pp. (1-11).
2. Al-Fraidan, Abdullah A. (2005). *Direct and Indirect Assessment of Writing with Special Reference to the Development of Foreign Languages at King Faisal University*. M.A. Dissertation, Essex University, UK.
3. Al-Hajailan, Talal A. (2003). *Teaching English in Saudi Arabia*. Riyadh, Adar Alsawlatia.
4. Ali, M. Salaiman (1995). "The Major Quantitative Findings of a Study of the English Language-Based Study Skills problems of Two Groups of Foreign Students at an American University", *ERIC* online Database.
5. Al-Mahboub, Ibrahim (2007). *The Acquisition of the Article System by Saudi Students*. Unpublished PhD thesis, Essex University, UK.
6. Al-Muhaidib, Nouriya S. (2006). *Selected Psychological Traits of High and Low Foreign Language Achievers of Female Majors at King Faisal University: A Contrastive Study*. Ph.D. Thesis, Girls Colleges of Education in Riyadh, Saudi Arabia.
7. Al-Saadat, Abdullah I., Al-Braik, Mubarak S., and Al-Shabab, Omar S., (2005). "Local Needs and the Characteristics of Local General English Language Programs", *Seminar Proceedings: Teaching English in Saudi Arbia: Opportunities and Challenges*. Riyadh, Saudi Arabia, Al-Yamama College, pp.(1-7).
8. American Association of University Professors (2007/1999). "Sample Distance Education Policy & Contract Language". www.aaup.org/AAUP/issued/DE/SampleDE.htm, Retrieved: 21/6/2007.
9. CVCP Academic Audit Unit (1992). "Report of an Academic Audit of the University of Warwick". www.unc.edu/depts/pubpol/Warick.htm, retrieved 25 June 2007.
10. Free-Weiss, Dana (2004). "Community College Freshmen: Last In, First Out?", *Journal of College Students Retention Research Theory and Practice*, Vol.6, No. 2. pp. 137-154.
11. Hughes, Arthur (1089/2003). *Testing for Language Teachers*. Cambridge, Cambridge University Press.
12. Illich, Paul A., Cathy Hgan, and Lislle McCallister (2004). "Performance in College-Level Courses among Students Concurrently Enrolled in Remedial Courses: Policy Implications", *Community College Journal of Research and Practice*, Vol.28, No.5. pp.+ 435-453).
13. Lynch, Brian K. (2003). *Language Assessment Programme Evaluation*. Edinburgh, Edinburgh University press.

14. Mackay, Ronald (1994). "Understanding ESL/EFL Programme Review for Accountability and Improvement", *ELT Journal* Vol. 48 No. 2, Oxford University Press, pp. (142-149)
15. Nunan, David (1989). "Towards a Collaborative Approach to Curriculum Development: A Case Study" *TESOL Quarterly*, Vol. 23, No. 1 pp. (9-25).
16. Schlenker, Dale E. and McKinnon, Norma Cole (1994) *Assessing Faculty Performance Using The Students Evaluation*, Research Report, Atlantic Baptist College. ERIC DATABASE. Retrieved 22 June 2007.
17. Shaw, Rose A. (1994). "Performance-Based Evaluation of Student Learning: A Continuum from X-12 Through the University Level", *A Process Evaluation Report of the FIPSE Project*. University of Northern Colorado, Greeley, USA.
18. Watt, David L. E. and M. Lake, (2000). "Canadian Language Benchmarks-TOEFL Research Project: A Comparison Study of the Canadian Language Benchmarks Assessment and the Test of English as a Foreign Language". REIC online Database.

دواعي دراسة تحصيل الطلاب في تقويم برامج اللغة الإنجليزية المكثفة في السياق السعودي

مبارك سالم البريك

قسم اللغات الأجنبية، كلية التربية، جامعة الملك فيصل
الأحساء، العربية السعودية

الملخص:

لقد أدى التوسع مؤخرًا في التخصصات العلمية في الجامعات السعودية إلى التأكيد على تدريس اللغة الإنجليزية لتحضير الطلاب للدراسة باللغة الإنجليزية. ويشكل البرنامج المقدم بجامعة الملك فيصل في المملكة العربية السعودية مثالاً على المد الذي يحصل في مجال تدريس اللغة الإنجليزية. هذا ولم يجر تقويم مهني للبرنامج الذي تدرسه كليتي الطب والعمارة في موقع الجامعة بالدمام وكلية الطب البيطري بالأحساء. ويهدف البحث الحالي إلى تقويم تحصيل الطلاب في أربع كليات في الموقع الرئيسي للجامعة في الأحساء: كلية الطب وكلية الطب البيطري وكلية الصيدلة الإكلينيكية وكلية علوم الحاسب وتقنية المعلومات، وكلية الطب هي الوحيدة التي فيها طلبات حالياً.

أعطي اختبار الكفاءة باللغة الإنجليزية لجامعة ميتشيغان للمجموعات الخمس في الأسبوع الأول والأخير من التدريس في البرنامج وتم مقارنة النتائج مع تحصيل الطلاب حسب تقويم المدرسين في البرنامج. وتبين النتائج أن التحصيل في اختبار الكفاءة باللغة الإنجليزية لجامعة ميتشيغان هو ضعيف وأن نتائج تقويم المدرسين تميل إلى أن تكون أعلى. ويلاحظ أن تقويم المدرسين هو أعلى في أعمال الفصل منه في الامتحان النهائي في كليتين هما الطب البيطري وعلوم الحاسب. وهاتين المجموعتين هما الأقل تحصيلاً بين المجموعات الخمس وهما الأقل درجات من حيث المدخلات في الجامعة.

يستنتج البحث بأن التحصيل العام الضعيف في اختبار الكفاءة باللغة الإنجليزية لجامعة ميتشيغان يبين أن هذا الامتحان الخارجي العام هو غير حساس للظروف الوضعية الاجتماعية التي تقدم فيها البرنامج. بينما في الوقت ذاته، يعكس التحصيل المنخفض والرابط الإحصائي السلبي لدرجات مجموعة الطب البيطري ضعف المدخلات لهذه الكلية والغياب والتسرب منها كما يعكس بيئة الفصل الدراسي في الحالة المحلية. وهذه المظاهر كلها تدعو إلى الحاجة إلى أبحاث ترصد الصف الدراسي.