

# The Predictive Validity of the Baccalaureate Exam English Test

Bamogo Yacouba<sup>1</sup> ; Dr. Sanon/Emilie Ouattara<sup>2</sup> ; Dr. Kambou Kwado Moses<sup>3</sup>  
 Universié Joseph Ki-Zerbo/Burkina-Faso

**Abstract:-** The objective of the study is to find out the predictive power of the 2019 and 2020 Baccalaureate exam English test. Using a whole class design the performances of 415 and 62 students in the Baccalaureate have been correlated with their performance in Semester one University studies in the Department of Anglophone Studies. Analysing with the statistical package R, the descriptive statistics of the results shows that for the cohort 2019, 358 students (86.27%) decreased in performance and 43 students (69.35%) for the cohort 2020. The students' mean scores at the Baccalaureate are above the ones in semester one. Comparing them, the difference is statistically significant at  $p < .05$  for the cohort 2019 and  $p < .05$  for the cohort 2020. This means that the Baccalaureate exams have invalid predictive power. These exams could not foresee candidates' future performance. It is because either they are of bad quality or there is no link between what is taught in the second cycle and what is studied in English at university. When questioned to find out the challenges to cope with university courses, students said that the main hurdle is note taking, then comprehending spoken English from their lecturers, reading novel and understanding them, the quality of the loud speaker sound, and holding a conversation with their teachers. These difficulties entail that they have not been prepared in the second cycle in those language skills. So, suggestions were made to teach the listening, reading and note-taking skills in addition to associating technicians of sound in the auditoriums building design.

**Keywords:-** Validity; Predictive Power; Baccalaureate Exam.

## I. INTRODUCTION

In any education system, students are tested for formative and summative purposes. In Burkina Faso, after the Baccalaureate exam, many students succeed in getting good marks in English. However, these marks seem to be in contradiction with their real performance. Those who are selected to study English at university suffer to cope with courses as the success rate rarely reach 30% according to the Department of Anglophone studies statistics. This is telling that their marks and their performances in the Baccalaureate and in Semester one are not linked. Passing successfully a level certifies your qualification to future tasks, unless the quality of the tests you took is questionable. Here, what is pointed out that tests administered deceive the prediction of candidates' future performances. As Brown (2004) posits, the end of a language learning level is in the education

system the beginning of another level. The inability of the Baccalaureate exams English test to predict the future performances of students in S1, in addition to difficulties that students have to cope with university studies led me into conducting this study. The first question it attempts to answer is: to what extent does students' Baccalaureate performance in English correlate with their university Semester one performances? The second question is: what are the challenges that prevent students from coping with university courses?.

## II. LITERATURE REVIEW

To better understand validity and how to evaluate it, we need to know first what testing is. Brown (2004, p.3) states that "a test, in simple terms, is a method of measuring a person's ability, knowledge, or performance in a given domain". The one concerned in this study is the Baccalaureate exam test for arts section in the second cycle. It is a cycle end summative assessment which aims at telling the students' mastery of the programme for learning set in the whole cycle. A suitable theory that stems for this investigation is the Classical Test Theory (CTT). For Magno (2009) the "Classical Test Theory is regarded as the true score theory" (p.1). The classical theory assumes that in any test, each individual has a true score which would be obtained in a test if there were no measurement errors. That performance would reflect the real competence of the test taker, his/her abilities, what he/she really knows. However, because of measuring instruments imperfections, the score of an individual can be different from his/her true score Awopeju & Afolabi (2016). This is telling us that tests are fallible and imprecise tools. The score achieved by an individual is rarely the individual's true performance. The true score of an individual will be constant when the same test is repeated and normally the difference in scores correspond to the test error. CTT uses tests standard deviation to estimate test error. The wider the standard deviates, the more imprecise the test is. From this theory test developers start to consider the different test items they develop to meet the criteria of good tests. The errors in tests show the reliability and at the same time the validity of the test.

Validity is one of the key qualities of a test. This is the view of Brown (2004) who argued that "By-far the most complex criterion of an effective test and arguably the most important principle is validity". Validity is part of assessment theory expressed by Hathcoat *et al.* (2016) who states that it is not possible to design a good test without considering the notion of validity. Once an instrument is

invalid, trust in it automatically disappears. Trust in the results that will come out from its measurement will have no credit. For Weir (2005)

“Validity is perhaps better defined as the extent to which a test can be shown to produce data, i.e., test scores, which are an accurate representation of a candidate’s level of language knowledge or skills. In this revision, validity resides in the scores on a particular administration of a test rather than in the test per se” (p12)

In this instance, he sees validity in tests’ scores interpretation. This is posteriori evidence of the test quality after administration. The test is then seen to be valid if results statistically highly correlate with some indices of behaviour we want it to correlate with. This approach to validity is what Weir (2005, p18) qualifies like the “suck-it-and-see” approach. An approach to testing balanced with another called the a priori validity. There are many types of validities like face validity, content validity, construct validity and criterion-related validity.

Criterion-related validity is according to Brown (2004) “the extent to which the "criterion" of the test has actually been reached” (p.24). That is to say to find out if the predefined level of performance set for the objective a course is reached. This level of performance depends on schools but should attain at least 80% to attest the real mastery of a teaching. According to Weir (2005), it is quantitative in nature and a posteriori concept, concerned with the extent to which test scores correlate with a suitable external criterion of performance. This type of validity has weaknesses and will give low level of validity if both tests are not testing the same ability. We have two kinds of criterion related validity: the concurrent validity and predictive validity. Concurrent validity refers to how a test confirms the results of another test of the same type administered before. For Taherdoost (2016), It is the extent to which a particular test’s results, corresponds to those of a previously established test for the same construct. According to Oller (1979), a test has concurrent validity, if there is a statistical correlation of the skill(s), or component(s) the test purports to measure with the same skills in other tests. This will not be the concern in this study.

The second type of criterion validity is predictive validity which is related to the predictive power of the test. It tells how well someone is suitable for a future task or his/her likelihood of future tasks. This kind of tests are used for decision making. For students’ admission to an academic course in English for instance. It tells whether the student has the required level to take a course. For Al-Adawi & Al-Balushi (2015) “having invalid, unreliable test that misplaces students in wrong levels may create negative attitudes towards university studies or even destroy the students’ proficiency altogether” (p107). A test which does not measure a job-related content or competencies can be considered a poor test. These low-quality tests result cannot be used for decision making. They are unreliable because they are creating more problem instead of helping to reach

the desired objective. As stated, the criterion validity has weaknesses but it will not hinder my work as it is just for a predictive judgement of the students’ performance. Wier (2005) recognized it saying that “if all one wants is to make certain predictions about future performance on the basis of results, this might entail a radically different test from that where the interest is in providing information to allow effective remedial action to be taken” (p p.36).

Many studies were carried out on the predictive validity. Stuart & Clare (2011) worked in a case study approach on the predictive power of the Cambridge Assessment in relation to students’ Academic success in at Florida State University. Data were collected from a population of 521 students of three cohorts. Examining the correlation of performances, the results show that students who followed acceleration programmes in the U.S. have better university success than those who did not follow any extra-credit programme during their college studies. This study is relevant as it illustrates the course programmes importance in facilitating future learning in education. The methodology and statistics used are strong enough to validate the results obtained from this study.

Taher (2012) in the same vein as the previous work worked on establishing test quality in Iran. He focused on the predictive power of the final exam scores as a measure of success in the students later performance at university. 42 students randomly selected grades were collected from their schools. Correlating these scores with the students’ university entrance test scores using SPSS, the highest validity coefficient is of .449. this indicates a lack of predictive validity. This must raise awareness in high school and pre-university teachers that the test they construct do not have high relation with the students future University performance.

Wang *et al.* (2007) worked to established the predictive validity of an English Language Art performance assessment implemented in an urban school district in California. The objective is to check the predictive validity as well as the influence of factors like school system ethnicity, family status and students’ language Background on the performance in the California high school exit exam. Using the Hierarchical Linear Model, the statistical analysis informs that only language Background has a significant effect on the students’ performance. As the prediction of students’ future performance is concerned, the results show that the test was able to predict the performance in the exit exam.

These studies use different ways of investigating tests predictive power. Those techniques and tools are sources of inspiration for the methodology in this research.

### III. METHODOLOGY

#### ➤ *Nature of the Study*

In the domain of research, there are many investigation techniques used. This study is a quantitative one. Kumar (2011) and Creswell (2009) suggested that quantitative

research is used when we want to quantify the data and if information is gathered using predominantly quantitative variables. It involves the use and analysis of numerical data using statistical techniques. It is objective and suitable for large quantities of data collection. This technique provides a clear, and constant quantitative measure. First and second year students of the Department of Anglophone Studies at university in addition to their Baccaalaureate performances are the target population. Students were surveyed using questionnaires. The total student population considered in this study is 1694 for the cohort 2019 and 830 for the one of 2020. Through a whole design sampling technique all students who volunteered to participate in the survey were considered. A total of 166 students participated in the cohort 2019 and 143 students' answers were considered. For the cohort 2020, 111 students participated in the survey and 86 students' answers were considered. The others were excluded on the basis of their Baccaalaureate series (D; A1; A2), and the misunderstanding of the instruction to carry out the task correctly.

For their performance in the Baccaalaureate and university first year studies comparison, all the Baccaalaureate 2019 students' marks were considered whereas 62 of Baccaalaureate 2020 students were considered. This is because after seven months of pursuit at the *Office du Baccaalaureate* the researcher could not get the performance of the 2020 Baccaalaureate students' performance. So, students were approached individually to get their transcripts so as to collect their performance.

#### ➤ Research tools

To find out tests predictive power, students' performance in Baccaalaureate and first year university studies were collected and computed using the statistical package R. Statistics is used to short data in a described way. To explain the abilities or inabilities for students who succeeded in Baccaalaureate to cope with their studies at university a questionnaire was devised. Kumar (2011) posits that some weaknesses of a questionnaire is the lack of ability to give clarifications and the limited population as its completion necessitates reading and writing. This problem is solved by making directions very easy for the students' comprehension.

#### ➤ Data Analysis

For the first research question on the test predictive power, students' performance collected on the Baccaalaureate exam and first year university studies were computed using the statistical package R. Statistics is used to process the data in a described way. The distribution of students' grades on the Baccaalaureate exam and in semester one and the mean scores were given. To see if their former performances predict their ability to cope with next programme, a correlation of their performance in the Baccaalaureate exam and semester one was established. Then, the proportion of those who improved, decreased or remain stable are provided. For the second research question, students' opinion regarding their difficulties to cope with the courses in first year is coming to boost this analysis and help to make recommendation for the improvement of the syllabus in the second cycle and at the undergraduate level university studies.

## IV. RESULTS

### A. Predictive Validity

After analysis, the results from correlation and questionnaire are presented in this part. Following the academic years, we had first, the 2019 exam results, then the one of the 2020 exam. They were summarised in tables then commented.

#### ➤ Predictive Validity of the 2019 Baccaalaureate Exam English Test

The predictive power of the exam 2019 was presented in two steps. We presented their performances and explained it, then we compare the Baccaalaureate exam performance to the one of the Semester one so that to perceive the statistical difference.

#### • Students' Performance

In 2019, 415 students' performance were able to be extracted from the data basis of the *Office du Baccaalaureate*. Their performances in four subjects in Semester one were considered for the analysis. They are those taught and tested in English. The descriptive statistics of the performance in the Baccaalaureate exam and in the subjects considered for the analysis are in the following table.

Table 1 The Description the Cohort 2019 Students' Performance N=415

Variables	Missing values	Min	Mean	Max	SD
Baccaalaureate Written test	0	8	13.16	19	2
Baccaalaureate Oral test	0	8	14.38	19	2.5
Mean Baccaalaureate	0	8,5	13.46	18,50	1.72
S1 GEA <sup>1</sup>	25	1	7.80	19	3.87
S1 ILA <sup>2</sup>	35	0	9.03	19	5.03
S1 PPA <sup>3</sup>	53	1	11.41	19,5	4.6
S1 EOA <sup>4</sup>	0	0	7.08	19	4.09
Mean S1	0	0,3	8.311	17,14	4.26

Field Work Data

<sup>1</sup>Grammaire et Expression Anglaise

<sup>2</sup> Introduction à la Linguistique Anglaise

<sup>3</sup> Phonétique et Phonologie Anglaise

<sup>4</sup> Expression Orale Anglaise

In the table 1, we can see that all the students have grades in the Baccalaureate exam as there is no missing data. The minimum score in the oral and the written exam was 08. The maximum score was 19 for the oral and the written exam. The mean was 14.38 in the oral exam and 13.16 for the written exam. The general mean was of 13.46.

As for the semester one performance, many students did not sit for the tests in some subjects. Twenty-five

students in Grammaire et Expression Anglaise (GEA), 35 in Introduction a la Linguistique Anglaise (ILA), and 53 in Phonétique et Phonologie Anglaise (PPA). Their minimum score 00 and the maximum is 19.5. The general min is 08.31. From the minimum score to the maximum, it seems that the difference is big. It is important to understand students' distribution in the Baccalaureate exam and in Semester one before analysis. Let us consider the following table.

Table 2 Results on Students' Distribution

Variable	Statistic	P. Value
Mean Baccalaureate	0.06	0.08
Mean S1	0.06	0.04
(Mean S1)-(Mean BACCALAUREATE)	0.05	0.15

Field Work Data

The results of marks distribution in the Baccalaureate shows that students are not spread apart. There is not a significant distribution as  $p > .05$ . However, in Semester one, the distribution of scores is not normal as  $p < .05$ . It is showing that there is a big gap between students in their semester one grades and this difference in grade is significant. Performances are spread wide apart. For the difference of means, the distribution is normal. This is good as it gives the green light for further tests of comparison.

- *Comparison of performances between the exam 2019 and semester one*

Given that the distribution of the difference of means is normal, a paired t-test is conducted to find out if the difference of means between the Baccalaureate and Semester one is significant. Let us read the results in the following table.

Table 3 Paired T-Test of Performance in 2019

Df	Statistics	P. value
414	23.02	.00

Field work data

The results from the table above shows that  $p < .05$ . there then a significant difference between students' performance in the Baccalaureate exam and in Semester one.

Another comparison of students' performance in the Baccalaureate and in Semester one is done to just see the number of students who improve or decreased from the Baccalaureate exam to university. That would let us have an idea on the variation of performances from both levels. The following table gives us the results.

Table 4 Variation of Performance

Performances	Number of students	Percentage
Increased	56	13,49
Decreased	358	86,27
Stable	1	0,24
Total	415	100

Field Work Data

The results from this table shows that 57 students' performance in the Baccalaureate exam and Semester one first year university studies improved. This corresponds to 12.49%. 358 students (86.27%) decreased in performance and only one student (0.24%) remained stable.

- *Predictive Validity of the 2020 Baccalaureate Exam English Test*

- *Students' Performance*

A total of 62 students' performance were used in the analysis. This is because as explained in the data collection procedure, due to the inability of the *Office du Baccalaureate* to provide the researcher with their performance in the exam. The descriptive statistics of the performance in the exam and the subjects considered in first year are in the following table.

Table 5 Students Performances N=62 Observations

Variables	Missing values	Min	Moyenne	Max	SD
Written	0	8	15.08	19	2.46
Oral	0	10	15.53	19	2.36
Mean-BACCALAUREATE	0	9,25	15.19	19	1.99
GEA	0	0	13.79	19	4.26
ILA	0	0	12,22	19.5	4.83
PPA	0	0	14.57	20	4.09
EOA	0	4	14.31	19	2.84
Mean-S1	0	4,57	13.64	18	3.15

Field Work Data

The minimum score in the oral exam was 10 and 08 in the written one. The maximum score was 19 for both. The mean score in the oral exam was of 15.53 and 15.08 for the written exam. The overall mean score was 15.19. In semester one, the minimum scores in GEA, ILA, PPA were zero (00), and four (04) in EOA. The maximum scores were 19; 19.5; 20 and 19 respectively.

The gap between the minimum and maximum scores compels us to see if there is a normality in scores distribution. For that, the distribution is calculated using Kolmogorov as the number of observations is more than 50. The table below gives the results of the students' distribution.

Table 6 Results on Students' Distribution in 2020

Variable	Statistics	P. Value
Mean BACCALAUREATE	0.10	0.46
Mean S1	0.14	0.13
(Mean S1)-( Mean BACCALAUREATE)	0.08	0.70

Field Work Data

We can observe from the table that  $p > .05$  for the students' exam grades. The value is statistically not significant to confirm a difference. So, the distribution is normal in the Baccalaureate exam. It is the same for the semester one grades,  $p > .05$  which shows that the distribution is normal. The most important is that we can carry on with comparative test as the difference between the mean in the Baccalaureate and the mean in Semester one has a normal distribution.

- *Comparison of Performances between the 2020 Exam and Semester One*

A paired t-test is conducted to perceive the statistical difference. Let us consider the result in the following table.

Table 7 Paired T-Test of Performance in 2020

Df	Statistics	P. Value
61	3.67	0.00

Field Work Data

The result from the table shows us that the difference between students' performance in the Baccalaureate is significantly different from their performance in Semester one as  $p < .05$ . Students' performance could thoroughly be understood at both levels. From a comparison of students' performances at both levels the number of those who improved or decreased was indicated in the following table.

Table 8 Students Variation of Performances from the Baccalaureate to Semester One

Performances	Number of students	Percentage
Increased	19	30.65
Decreased	43	69.35
Stable	0	0
Total	62	100

Field Work Data

The results in table 8 shows that out of 62 students, 19 improved which corresponds to 30.65%. None had a stable performance and 43 (69%) decreased from the Baccalaureate to Semester one. Even though these statistics do not tell if the individual increased or decreased was significant, it is clear that the number of negative results is high. This situation pushed to the next point to find out some explanations with students.

### B. Difficulties to Cope with University Studies

#### ➤ Cohort 2019

Trying to find out why so many students failed to improve, a list of the possible difficulties that students have to cope with English classes in Semester one was given to them so that they rank the possible impediments to their success. It contains eleven items. Consider the following table for the results.

Table 9 Results on Students' Difficulties to Cope with University Courses

Difficulties	Number	Percentage
Note taking in class	46	33.82
Understanding of teacher when he speaks	26	19.11
Reading novels	16	11.76
Quality of the loud speakers' sound	10	7.35
Speaking to teachers	10	7.35
Vocabulary	10	7.35
Writing in test	07	5.14
Speaking to peers	05	3.67
Taking the floor in the classroom	03	2.20
Grammar	03	2.20
Understand the course	02	1.47
Total	136	100%

Field Work Data

The information contained in the table above does not include the other possible variable that can explain the difficulties to cope with the classes such as students' socio-economic status. Only classroom context language use and teaching/learning practices were considered. The biggest impediment for students was note taking, followed by listening comprehension, then reading (novels). Speaking to lecturers, the quality of the auditorium speakers and vocabulary are at the same level of difficulty with 7.35% each. The other minor difficulties for them were consecutively writing in test; Speaking to classmates; taking

the floor in the classroom; grammar and finally understanding the course.

#### • Cohort 2020

The desire to understand why do students decrease significantly in performance at university Semester one is important. To find out the possible difficulties students have to cope with English classes, the students classified the possible impediments of their success. Eleven elements were labelled. Look at table 10.

Table 10 Cohort 2020 Difficulty to Cope with English Courses

Difficulties	Number	Percentage
Note taking in class	23	26.74
Understanding of teacher when he speaks	16	18.60
Reading novels	14	16.27
Speaking to teachers	10	11.62
Quality of the loud speakers' sound	09	10.46
Taking the floor in the classroom	05	5.81
Writing in test	03	3.48
Grammar	03	3.48
Lack of vocabulary	02	2.32
Understand the course	01	1.16
Speaking to classmates	00	00
Total	86	100%

Source : Field Work Data

The information contained in table 10 does not include the other possible variable that can explain the difficulties to cope with the classes such as their socio-economic status. Only classroom context language use and learning practices were considered. The biggest impediment for students was note taking, followed by listening comprehension, then reading and speaking to lecturers. The fifth element mentioned is linked to the quality of the sound from the auditorium speakers. The other minor difficulties for them

were consecutively taking the floor in the classroom; writing in test; grammar; lack of vocabulary; understanding the course and finally Speaking to classmates.

What do these pieces of information for the Baccalaureate exam and the training given to students mean? Next section is devoted to results discussions to show the significance of the data obtained.

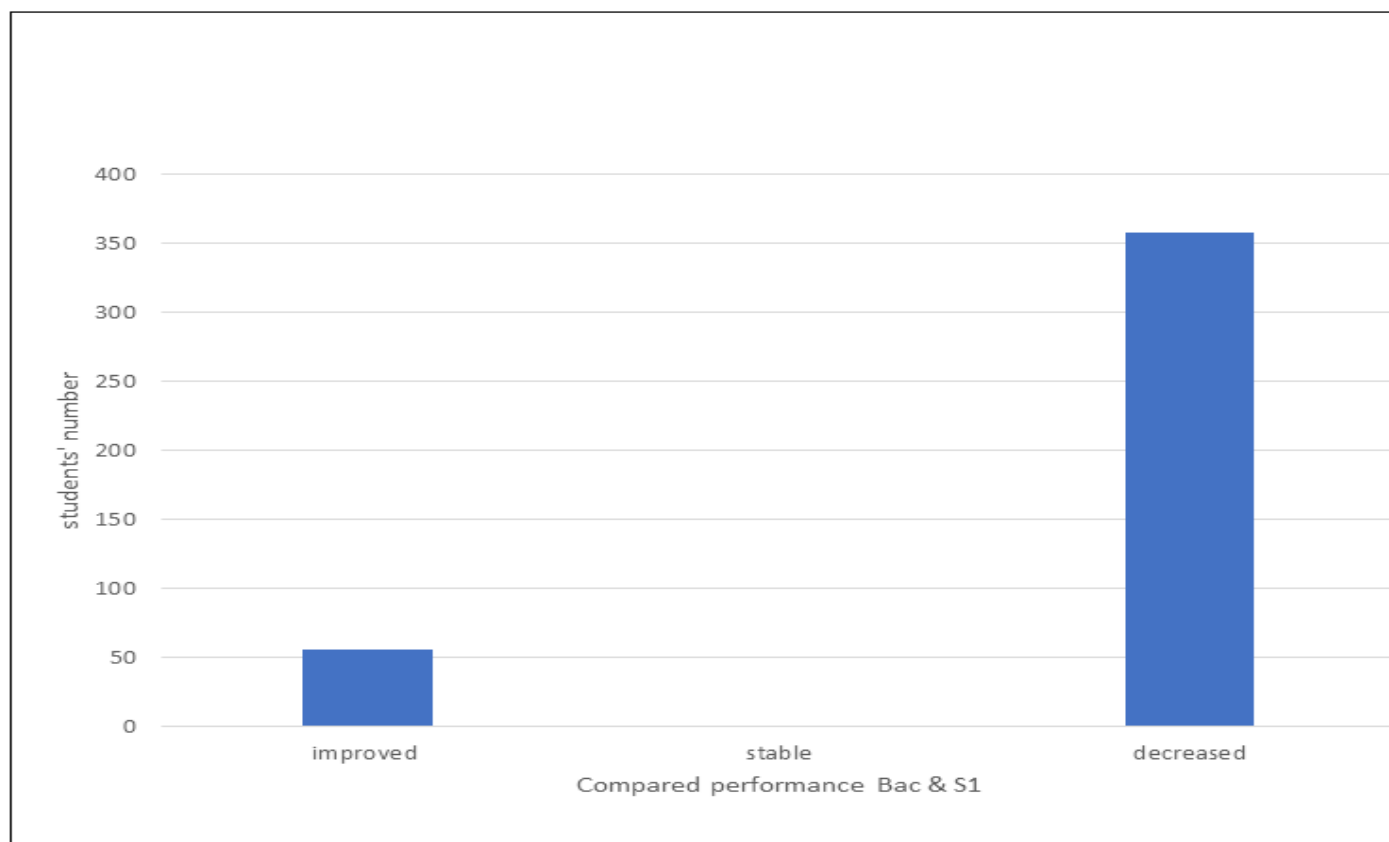
**V. DISCUSSIONS**

➤ *Predictive Validity*

One of the objectives of this study is to determine whether students’ performance in the summative exam at the end of the second cycle qualify them for university studies in English. Normally if the exam tests were valid to tell the actual performances of candidates, they would tell if those who passed could cope with classes in the Department of Anglophone Studies at university level. The results of the correlation of both cohorts first years’ semester one and the Baccalaureate exam performances show a lack of correlation as the majority of students decreased in performance. Let us discuss the results of the two cohorts considering them individually.

• *Cohort 2019 Students*

In the results section the performance of the cohort 2019 students was presented. Their mean score in the Baccalaureate exam in English was 13.46 and in semester one at university was 8.31. The students’ performance decreased from the Baccalaureate to Semester one. The descriptive statistics showed that comparing their performance individually at both levels, 56 students out of 415 improved in achievement. That corresponds to 13.49% of the students. Only one student’s (0.24%) performance remained stable. 358 students (86.27%) decreased in performance. Look at the following graph to easily see the repartition of students according to their performance in both tests.



Graph1: The Proportion of Students’ Performances Variation of the Cohort 2019

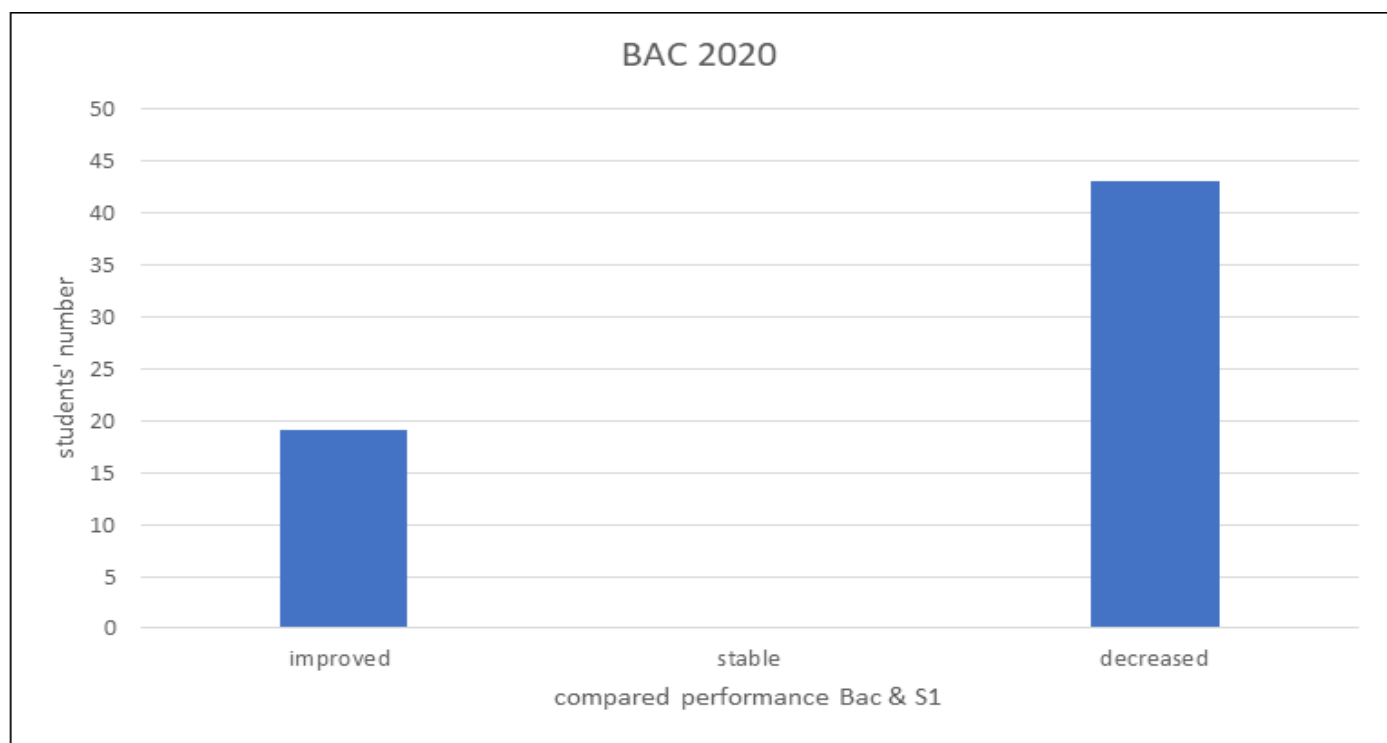
Source: Field Work Data

Compared to those who increased in their performance, the level of students who improved from this visual is small. Those whose performance remained unchanged is insignificant as it is almost invisible in the graph. To strongly ascertain that the difference between the Baccalaureate and Semester one is significant, the mean score of students in the Baccalaureate and S1 is analysed statistically. Their performance indicated that  $p < .05$ . This means that the difference of means is significantly different. So, there is no correlation between students’ performance in the Baccalaureate exam and Semester one first year in the Department of Anglophone Studies. The overall averages comparison state that their performances are significantly different at a level of 5%. We can see that the Baccalaureate performance is far better than the Semester one performance

at university. This entails that the summative national exam supposed to test students’ achievement throughout the whole second cycle gives results that are misleading for decision taking as to continue the studies in English at university. This lack of correlation explains why the success rate is weak in the Department of Anglophone Studies. According to the department statistics, we have 23.16% in 2017, 22.28% in 2018, 20.70% in 2019, 50.96% in 2020 and 34.84% in 2021. That does not corroborate with what Weir (2005) warns us in test design to build it in identifying what is appropriate for candidates in their future life. Any course or syllabus should be designed considering the learners’ future task, job or studies. The courses and skills tested in the second cycle deviate from the content taught at university.

- *Cohort 2020*

The comparison of the cohort 2020 students' performance in the Baccalaureate and in semester one at university showed that 19 students improved. That corresponds to 30.65 %. None of them got a stable performance. However, the majority (43 students) decreased in performance. The proportion of those who decreased in performance corresponds to 69.35%. The following graph illustrates this comparison.



Graph 2: The Proportion of Students' Performances Variation of the Cohort 2020

Source: Field Work Data

The graph presents the number of students who improved compared to those who did not. It helps to easily perceive the difference. None's performance remained stable, which cannot be seen in the histogram. To scientifically confirm if this difference is significant, a statistical analysis is required. Like for the cohort 2019, the correlation of students' performance in their Baccalaureate exam with the one of Semester one at university was done using the statistical package R. The mean scores were compared to see if the difference is significantly different. As mentioned in table 7 in the results section,  $p < .05$ . This confirmed that the difference between students' achievement in the Baccalaureate is significantly superior to the one in Semester one at university. The Baccalaureate exam then fails to predict the future performance of students. This entails that based on the Baccalaureate performance to come or to send students to anglophone studies is misleading. The nature of the English courses that students receive in the second cycle can account for this factor. This can also explain why we have a lot of failure in first year at the Department of Anglophone Studies.

- *Comparison between Cohort 2019 and 2020 Performances*

The first weakness to make this comparison is that of the difference in the samples. The cohort 2019 students' number is 415 whereas the cohort 2020 has 62 students. Anyway, when we consider the proportion of the increase

and decrease for both groups, we notice that for the cohort 2020, 30.65% of students increased in performance, against 13.49% for the cohort 2019. This is two times higher. Considering those who decreased, 69.35% decreased in 2020 whereas we have 86.27% in 2019. Logically those who decreased in 2019 are more than those in 2020. As stated above the sample is for sure influencing. In addition to the size effect, participants in 2019 marks were obtained automatically from the *Office du Baccalauréat* which include indiscriminately everyone. For students in 2020, as the *Office du Baccalauréat* could not provide the data, students were asked to provide their transcripts so that their performance could be retrieved. Consequently, those who accept to give their transcripts could be supposed to be those who performed well in the Baccalaureate. So, it is perhaps the high-performance students. Their performance was higher in the Baccalaureate exam than their counterparts of 2019. This explains why the high-performance students performed better in semester one. This weakness is at the same time a strength for the study as if the supposed best students have 69.35% of them who decreased in performance. This fact strengthens the results to confirm that the lack of predictive validity is serious as even the best students have about 70% of chances to perform less than their second cycle end achievement in English. For the low performance at university, we can explain it by the fact that the tests lack practicability and many students are not full time. But the gap is too big to state that the Baccalaureate



exam tests design do not have a strong link with the future learning of students.

➤ *Difficulties to Cope with Classes at University*

After cohort 2019 students' classification of key factors to facilitating or impeding their success in first year, the main ones were: note taking, understanding the teacher when he speaks, reading novels to understand them, the quality of the sound of the loud speakers, speaking to teachers and poor vocabulary. They account for 87% of difficulties. The nature of these hurdles means that in the second cycle, students have not been prepared for these future tasks. The basis was not laid to facilitate their future learning. The programme does not include language skills like, reading, listening, speaking in addition to note taking (writing) and vocabulary learning. The quality of the speakers entails that technicians of sound must be associated in auditorium buildings design so that people can have appropriate sound perception in the room.

For the cohort 2020, the main difficulties which came out were note taking, followed by comprehension of what the teacher says in class, reading novels to understand them, speaking to teachers and the quality of the loud speakers. These factors account for 84% of all the hurdles listed. These weaknesses students have automatically affected their performance in their assessments. It means that the second cycles teaching did not clear the forest for them to cope easily with university classroom practices. Again, it entails that the courses designed in the second cycles are not linked to candidates' future use of English at university.

➤ *Washback*

Brown (2004) says "Washback also includes the effects of an assessment on teaching and learning prior to the assessment itself" (p.29). At the level of the predictive validity, it rather has a negative impact for education actor. The performance of students is misleading. It gives the impression that they can do, whereas they cannot. Coping with university English classes is a problem. There are a lot of drop-outs as in a cohort half of the students left out studies. Those who held on, in their majority expressed their concern in taking adequately note during lecture, struggling with novels reading, difficulties to handle a conversation and teachers complain about students' inabilities. This ends very often with a failure during tests. In fact, the degree obtained which normally qualifies the students for future tasks is betraying decision making. Those people in general do not have the qualification for their future tasks. This is also impacting badly the society as someone could be placed at the wrong position for a job if he/she succeeds in national competition where he/she will be using English for services.

## VI. CONCLUSION

The aim of this study was to establish the predictive power of the Baccalaureate exam English test with the future performance of students in Semester one, and to find out hurdles that prevent students from coping with university classes. Students from two cohorts were considered in the investigation. After analysis of results, the

majority of students decreased in performance from the Baccalaureate to Semester one (69.35% in 2020 and 86.27% in 2019). The statistical analysis showed  $p < .05$ . We can then conclude that there is a significant difference in the decrease of performance. Among factor that account for students' inability to cope with university classes, the main important ones were, note take, reading, listening, speaking and the sounds quality of the auditorium. From results, we suggest a change in the curriculum to include necessary skills to lay the basis for further studies as the end of the second cycle is the beginning of the higher one. Then the syllabus must be designed in concertation with the university lecturers including the strategies and techniques for teaching all language skills. Also, the exam tests selection teams should include specialists of language testing. Testing is a skill apart and not anybody who teaches can test well. This will bring more scientific design so that the criteria for exam tests selection do not follow personal feelings but rigorous methods. That will improve the validity of the tests selected. Finally, sonority technicians should be associated to auditorium building. This contributes to the sonority quality for students to cope with listening difficulties.

## REFERENCES

- [1]. Al-Adawi1, S. A. S. & Al-Balushi1 A. K. A. (2015). Investigating content and face validity of English language placement test designed by colleges of applied sciences. *English language teaching; Vol. 9, No. 1*
- [2]. Awopeju, O. A. & Afolabi, E. R. I. (2016). Comparative analysis of Classical Test Theory and Item Response Theory Based Item Parameter Estimates of senior school certificate Mathematics examination. *European Scientific Journal*, vol.12, No.28, pp. 263-284. ISSN: 1857 – 7881 (Print) e - ISSN 1857- 7431
- [3]. Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Pearson
- [4]. Magno C. (2009). Demonstrating the difference between classical test theory and item response theory using derived test data. *The International Journal of Educational and Psychological Assessment* April 2009, Vol. 1, Issue 1, pp. 1-11
- [5]. Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE Publications, Inc.
- [6]. Hathcoat, D. J, Sanders B. Courtney & Gregg N. (2016). *Selecting and designing instruments: Item development, development, reliability, and validity*. James Madison University.
- [7]. Hughes, A. (2003). *Testing for Language Teachers*. Cambridge University Press; 2<sup>nd</sup> ed.
- [8]. Kumar R. (2011). *Research methodology: a step-by-step guide for beginners*. 3<sup>rd</sup> edition, Sage
- [9]. Le Pays newspaper n° 7318 du jeudi 29 avril 2021
- [10]. Mary, J. P, John W. Y., Maria M, Teresa C. K, Alyssa B, & Mitchell G. (2009). *Guidelines for the assessment of English language learners*. Educational Testing Service

- [11]. Oller, J. W. (1979). *Language tests at school*. Longman Group Ltd.
- [12]. Siddiek, G. A. (2010). The Impact of test content validity
- [13]. On language teaching and learning; *Canadian Center of Science and Education* pp.133-143
- [14]. Stuart, S. & Clare B. (2011). An American university case study approach to predictive validity: Exploring the issues. *Research Matters: A Cambridge Assessment publication*, issue 12, pp.18-26
- [15]. Taher, A. (2012). The predictive validity of final English exams as a measure of success in Iranian national university entrance English exam. *Journal of Language Teaching and Research*, Vol. 3, No. 1, pp. 224-228, January 2012
- [16]. Taherdoost, H. (2016). Research instrument, questionnaire, survey, survey validity, questionnaire reliability, content validity, face validity, construct validity, and criterion validity. *International Journal of Academic Research in Management (IJARM)* Vol. 5, No. 3, 2016, Page: 28-36, ISSN: 2296-1747 © Helvetic Editions LTD, Switzerland.
- [17]. Wang, J, David N. & Haiwen W. (2007). Predictive validity of an English language arts performance assessment. *National Center for Research on Evaluation, Standards, and Student Testing*, 729 pp.1-23
- [18]. Weir, J. C. (2005). *Language testing and validation an evidence-based approach*. Centre for Research in Testing, Evaluation and Curriculum (CRTEC) Roehampton University
- [19]. Zabramba, J. (2000). *Formative and summative assessment/ evaluation*. The Ohio State University