

Formative Assessment and Consequential Validity: A Practice Yet to be Effectively Implemented in Saudi Higher Education

Muhammad Umer*
Choudhary Zahid Javid**
Muhammad Umar Farooq***

Abstract

Standardized tests have been found in various settings strongly influencing students' learning processes as well as outcomes, but such tests have been reported being unhelpful in improving student learning. In contrast, formative assessment (FA) is employed to help students know the level of their learning during a course and allow them to increase the standard of their learning, through knowing their learning weaknesses, receiving informative feedback from their teachers and applying adequate learning strategies, in order to achieve the formulated goals of the course. A thorough examination of empirical published research in the field revealed that insightful studies in Saudi academic context are still awaited. Therefore, this exploratory study was an attempt to investigate consequential validity of formative assessment at a Saudi university. To examine the nature of the consequential validity, data were collected from 960 English-major students via a survey (of whom 465 responded), 18 classroom observations and 4 focus-group interviews. The data underwent both statistical and content analysis. The results showed that FA, in the context of this study, seriously lacks intended consequential validity—positive impact on students' learning. Besides, the data indicated a strong unintended consequential validity i.e., negative bearing of FA on what the students learnt, how they learnt it and the depth of their learning. Therefore, it is recommended that FA practices should be urgently revisited with expert understanding for the

* Foreign Languages Department, Taif University, Taif, KSA.

** Foreign Languages Department, Taif University, Taif, KSA

*** English Language Centre, Taif University, Taif, Saudi Arabia

sake of students' improved learning. Finally, teachers should be trained in how to employ FA tools more skillfully and effectually.

Keywords: Consequential Validity, Higher Education, Formative Assessment, Student Perceptions, Test Impact

Introduction

The concept of consequential validity is not new in assessment literature; the term may be new for many though. It is an in-built characteristic of every assessment tool. In other words, this form of a test's validity implies either or both positive and negative influence, the test has on teaching and learning. The term 'consequential validity' is generally used to refer to the "study of the consequences of assessment use" both intended and unintended (National Center on Educational Outcomes [NCEO], University of Minnesota, p. 8). Whereas, according to the center, the intended consequences refer to the desirable effects of course content on learning, the unintended effects of an assessment tool may include:

- (a) narrowed curriculum and instruction to only content tested;
- (b) intensive use of test preparation materials that are too closely tied to the test;
- (c) unethical test preparation;
- (d) differential performance gains for some subgroups;
- (e) inappropriate or unfair use of test scores; and
- (f) decreased student confidence and motivation to learn and perform well on the test. (Lane & Stone, 2005, cited in NCEO, p. 9)

To study the consequences of a test is as much significant as the test's internal features such as validity and reliability (Shepard, 1997). Therefore, in the same article, she contends that the measurement of a test's validity should include how the test affects its users (the learners in the context of this study). Camara and Lane (2006) argue that the consequences of a test on teaching and learning are strongly linked with the test design and are connected with the validity of the assessment instrument. To advance the discussion, it is argued that an assessment tool may be technically highly valid but might not trigger the consequences required. Therefore, Cheng and Curtis (2004) hold that at times valid assessment tools may not necessarily guarantee that it serves the purpose it is intended to serve. In the context of this research, a formative assessment tool to be consequentially valid will be seen through its impact on students' learning, i.e. the extent to which it raises

the level of students' learning. Formative assessment is considered more as a teaching and learning tool than an assessment instrument. It is required to increase the level of students' level learning.

Test validity is quite broad a term. As referred to earlier, a test valid in one respect may not ascertain its other forms of validity. Therefore, an assessment instrument is required to possess validity in all forms. However, the discussion intrinsically centered on test validity should not eclipse the reliability dimension of the test; validity and reliability are interconnected. Only those tests valid in all forms can be considered reliable and vice versa as shown by Winke, (2011: 628) in the illustration below. Winke, in the same article, has explained that a test can only measure adequately if it successfully achieves its smart objectives leading to the attainment of the broad goals of the test, i.e., the positive impact on learners.

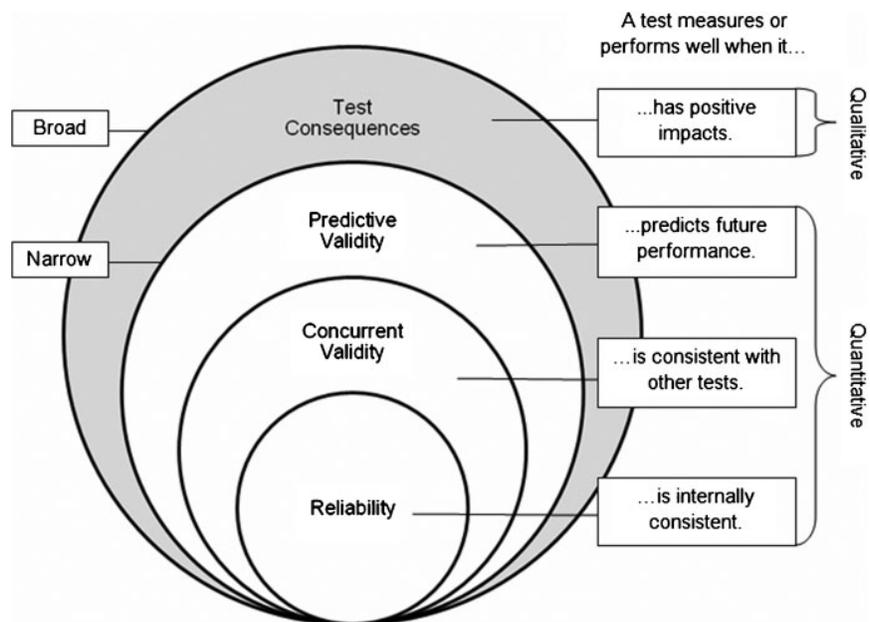


Figure 1. Adapted from Winke(2011: 628)

Insights from previous research

Sergers, Dierick and Dochy (2001) studied consequential validity of Over All Test—the name of a specific test in the context of the study. In the study, the researchers looked at the influence of the test

on students' learning. The data gathered, both quantitative and qualitative, indicated a mismatch between the program's goals and the test's characteristics. The students had to focus on a part of what the test was supposed to measure. In addition, students resorted to memorization and reproduction of knowledge, the course goal was application of knowledge though. This shows the test being consequentially invalid which is considered a serious drawback of an assessment tool.

Sambell, McDowell and Brown (1997) carried out a qualitative investigation of students' perceptions of formative assessment, particularly with regard to its consequential validity to find out how the assessment practices affected the students' learning. They collected data from students and teachers through observation, document analysis and semi-structured interviews; however, the focus of the study was on the students' perceptions. The respondents reported that formative assessment had a highly positive impact on the quality of their learning as compared to the traditional or summative assessment methods. In addition, they believed that formative assessment instruments were more authentic than their summative counterparts. In that, they assessed students' skills and competences that were contextually relevant or close to real life situations. Students valued such assessment more than the traditional examination because it allowed them to apply what they learnt during lessons. Furthermore, formative assessment was considered fairer than summative assessment, in that, it measured their performance over a period of time, thus a true reflection of their knowledge and skills unlike summative or one-day examination at the end of a course which depended on students' last minute efforts or, sometimes, on their luck.

Another example of how tests tend to bear positive consequences on students' learning strategies is the study of Saif (2006). The paper has reported the findings of a quasi-experimental study conducted in the University of Victoria, Canada. The participants of the study were 47 International Teaching Assistants (ITAs). The ITAs in the university were noticed to have communication problems during their interaction with the undergraduates whom they taught in make-up classes during lab sessions, and teachers' office hours. These ITAs were usually referred to the English Language Centre of the University to undertake English language courses to help them improve their communication skills.

The study was intended to investigate the impact of a purposefully-designed test on the ITAs. Furthermore, it aimed to examine the consequences the test had on teaching and learning activities in classrooms. The ITAs were divided into two groups: a control and an experimental group. After one semester, the experimental group showed

better results than the control group. The control group was taught in the normal way and was given the usual kind of exam where their speaking was not central to the assessment. On the other hand, the experimental group was taught by a teacher from the ELC who had been involved in the course process and knew the new program, the test itself, and the program's importance. The findings of the study indicated positive impact on teaching and learning activities, in that, both teaching and learning activities were in line with objectives of the test. However, the study has not been able to offer sufficient explanation as how and why the learners in the experimental group learnt better.

On the other hand, there are studies which show that a highly standardized and valid test may not trigger desirable consequences. For instance, Tzagari (2009) investigated the influence of the First Certificate in English (FCE) on classroom activities in private language schools in Greece. She interviewed 15 FCE teachers in order to answer some general research questions which indicated a close association between the examination and the teaching and learning that took place in their classes. This preliminary study led to her main study where she used mainly documentary method by analysing FCE textbooks and a couple of general English textbooks.

The findings of the study confirmed the notion that test impact is a very complex phenomenon. Therefore, she has recommended that in order to make an exam having positive effect on teaching and learning, several other variables need to be taken into consideration such as "teachers, students, schools, parents, local educational systems and local society" among others (2009, 8). Her study showed that the test influenced the teachers in selecting and narrowing their syllabus, students' learning styles and preferences, the materials developers, the school itself, the society and parents who made the students work hard. She has remarked that:

Students believed that the immediate goal of learning English was to achieve success in the FCE. What mattered more to them was the acquisition of the qualification itself rather than learning the language....students came to consider studying for the test as language education, and passing the exam as the aim of their schooling and evidence of educational excellence (p. 7).

Another empirical evidence of how a very reliable and valid test might result in undesirable consequences for learners is the study conducted by Gijbles, Segers and Struyf (2008). They investigated the

perceptions of students with regard to assessment demands. The participants were students of Education and Psychology course at the University of Antwerp, Flanders, Belgium. The instrument applied to measure their perceptions was a survey questionnaire. The study found that "short answer examinations involve lower level of intellectual abilities compared to assignment essay which promotes higher levels of intellectual abilities and skills such as analysis and synthesis" (2008, 433).

The authors have divided learning into two kinds: surface level learning and deep learning. The result of their study suggest that "the initial learning approaches students use seem to have a relative strong impact on how they approach their learning in a specific course, no matter the characteristics of the course" (2008, 439). Their findings have also indicated that only "constructivist" approaches to assessment can hardly change students' approach from surface level learning to deep learning. Finally, their findings have also supported the view that factors other than assessment that constitute the teaching and learning environment also play vital role in shaping students approaches to learning. However, this study has only relied on a quantitative data collection tool. Hence, it is not clear how the learners' surface level looked like.

Despite the fact that the formative assessment is a part and parcel of the examination component of Saudi education system at all levels, there has been no published empirical research having investigated consequential validity of formative assessment particularly in higher education context to show the extent to which it plays its role successfully. Therefore, the current study was an attempt to explore both intended and unintended effects of formative assessment in Saudi higher education context. Thus, in relation to the research problem stated, the following three questions were asked and were sought answers for through empirical evidence.

1. Does formative assessment play its role of a teaching tool in the context of this study?
2. To what extent does formative assessment influence English-major students' scope of learning materials?
3. What is the impact of formative assessment on the learning of English-major students?

Research context

Formative assessment is an essential component of the assessment regime of undergraduate courses in Saudi higher education (Al-Seghayer, 2011). Different instruments are in use, namely quizzes, presentations, research assignments and midterm examinations. Each semester comprises 16 weeks teaching duration. The midterm examinations are given during between weeks eight and ten. Two quizzes (for which students are informed in advance), one before and one after midterm examination, are served to measure students' achievement of course-learning outcomes. However, there are no fixed dates for students' presentations. Midterm examinations which carry 20 marks, lasts for one hour, but the duration of quizzes and presentations is up to the teacher to decide depending on the nature of the assessment tasks. Concerning the preparation of tests/assessment tasks, teachers are autonomous. Every teacher makes his/her own assessment tasks. No measures are taken for checking the reliability or validity of the assessment tasks.

Methodology

Research Design

To answer the questions of the study, the researcher employed mixed-method approach (Creswell & Clark, 2011), which includes a mixture and integration of both quantitative and qualitative information (Creswell, 2013). Creswell and Clark (2011) have provided a comprehensive overview of mixed-method research definitions, based on the work of twenty one highly published mixed methods researchers (p.3). In the conclusion of their overview of mixed-methods-research literature, they have selected the definition of Johnson, Onwuegbuzie, and Turner (2007, p. 123) who have defined the research method as follows:

Mixed methods research is the type of research in which a researcher or team of researchers combines elements of qualitative and quantitative research approaches (e.g., use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) for the purposes of breadth and depth of understanding and corroboration.

Most consequential validity research has employed mixed methods. That is, quantitative and qualitative data have been used to answer research questions. The most common data collection methods employed are surveys, interviews, focus groups and classroom observations (see for example, Cheng, 1997; Watanabe, 1997,2004; Sergers et al. 2001; Chen, 2002; Hayes and Read, 2004; Chu and Gao, 2006; Green,2007; Muñoz and Álvarez, 2010 among others). The rationale behind converging quantitative and qualitative data was to find out not only how the participants perceived the impact of formative assessment on their academics but also to observe how the impact looked like inside classrooms when they were actually involved in learning. Thus, when put together, quantitative and qualitative data were expected to yield a more wide-ranging picture of the phenomenon under investigation.

Participants

The population of the study included all undergraduate students of the university who pursued four-year courses in different colleges with the same assessment scheme at the time of the study. However, only English-major students were selected as a purposive sample for the study (Denscombe, 2007). Thus, the student-survey questionnaire participants included both male (N=360) and female (N=600). Non-probability purposive sampling technique was used for the survey-questionnaire, whereas, five volunteer teachers teaching the male students allowed the researcher to observe their classes. Furthermore, sixteen male volunteer student-participants from different levels are recruited for the focus-group interviews.

Research Instruments

The student-survey questionnaire, a self-developed instrument, was divided into four parts. Its first part included four demographic questions regarding their age, sex, level/year of study and their current GPA. The second section consisted of 18 items concerning research assignments in order to explore the students' perceptions of the impact of research assignments and their approach to writing research assignments. The penultimate part of the instrument had 24 items to investigate the type of examination they preferred and why they preferred it, and to look at their common learning strategies toward formative assessment. The last part, i.e., from item number 43-52 aimed to explore the level of interactions between teachers and students i.e., the type of feedback students received from their teachers. In addition, the last part included a box with blank

space inviting the participants to write their comments about the questionnaire items. The participants' responses were received on a 5-point Likert type scale. All statements on the scale were coded as 1= Strongly Disagree, 2= Disagree, 3= Undecided, 4= Agree, and 5= Strongly Agree. Second instrument used was classroom observation. The rationale behind using this instrument was to observe the impact of formative assessment activities on classroom events, particularly students' learning. The observation instrument was adapted from Watanabe (1997) and comprised the following components:

- Classroom information description sheet
- Teacher's background description sheet
- Materials description sheet
- Types of exercises
- Notes-taking sheet
- Impression description sheet

Third instrument employed was focus-group interviews. In total, four interviews were conducted. Each group included four participants. The questions included in the instrument by and large flowed from the preceding two protocols. The title objective of using it was to double check the respondents' opinions gathered via the questionnaire and their learning behaviors observed in classrooms.

Data Collection Procedure

In the first part of the research, the cross-sectional student-survey design was used. The second instrument i.e., the classroom observation was used after the quantitative data were analyzed. When the classroom observations were finished, the third protocol, i.e., the focus-group interviews, was administered to find explanation for discrepancies found, if any, between what they reported in their survey responses and what was observed during the lessons.

Data Analysis

The respondents' answers to the student-questionnaire survey were analyzed with the help of SPSS (20). At the first stage of this quantitative analysis, the participants' demographic data i.e., sex, age and year of study, were looked at. In addition to exploring the demographic information, descriptive analysis was carried out to get familiarized with the prominent characteristics of the respondents' answers to all items in terms of frequencies, percentages, means and standard deviation.

The qualitative data collected through classroom observations and focus-group interviews together with the comments in response to the open-ended question of the student-survey questionnaire were analyzed by adapting qualitative data analysis tool, originally developed by Hitchcock and Hughes (1989, c.f. Watanabe 1997). This tool includes the following process:

- Getting familiarized with the transcript of the raw data
- Description and analysis
- Isolating general units of meaning
- Relating general units of meaning to the research focus
- Extracting patterns and themes

Results and Discussion

Questionnaire responses

As one dimension of the study, the respondents were administered a self-developed survey. Out of 600 students, 465 returned it. Mostly, the respondents agreed or strongly agreed with the statements which caused significant contradictions as well. The average mean of their responses was 3 and above, which indicates interesting patterns.

Table 1
Students' perceptions of the impact of research assignments as a formative assessment tool

Statement	N	Minimum	Maximum	Mean	Std. Deviation
Discussing topics of assignment with friends and teachers	464	1.00	5.00	4.2565	.83973
Study several sources or books	465	1.00	5.00	3.2366	1.17442
Study several pages of a single course book	465	1.00	5.00	3.4903	1.02154
Assignment submission dates	465	1.00	5.00	3.8258	1.10727
Submission on due dates	465	1.00	5.00	4.1032	.93200
Number of words specified by teachers	465	1.00	5.00	2.9656	1.30603
Assignments are with clear introduction, body and conclusion	465	1.00	5.00	3.6430	.94543
Writing assignments in my own words	465	1.00	5.00	3.7398	1.06422
Cannot write assignments with over 150 words	465	1.00	5.00	3.0817	1.21671
Grades of assignments are important for me	465	1.00	5.00	4.3118	.91889
I copy the content from other sources	464	1.00	5.00	3.8642	1.01015
I know the meaning of what I write	465	1.00	5.00	4.1871	.88415
I acknowledge the sources of my text if	465	1.00	5.00	3.2602	1.24180

any					
I avoid plagiarism	465	1.00	5.00	3.8688	1.06167
I fail if content is plagiarized	465	1.00	5.00	2.7720	1.17796
I study multiple sources	465	1.00	5.00	3.4753	1.20341
I include bibliography	465	1.00	5.00	2.6925	1.19177
At least one assignment in one semester	465	1.00	5.00	3.3828	1.26104

The statements included in Table 1 were supposed to elicit the participants' opinion with regard to research assignments as a formative assessment tool. As mentioned above, the respondents have mostly agreed with the statements. Therefore, on one hand they claimed that they studied different sources to write their assignments, and that they acknowledged the sources they consulted. On the other hand, they also reported that they studied only a few pages of a single course book to write an assignment. In addition, they made confession with the mean of 3.86 that they copied the content of what they wrote, i.e. plagiarized and at the same time claimed that they avoided plagiarism. But it seemed that they passed their assignments even if their writings were plagiarized. In their response to the item "I fail if the content of my assignment is plagiarized", half of them disagreed indicating that they can get away with it.

Table 2
Students' perceptions of the impact of midterm examinations and pop quizzes on their learning styles

Statement	N	Minimum	Maximum	Mean	Std. Deviation
Using recommended or text books only	465	1.00	5.00	4.2817	.96724
Trying to understand topics deeply	465	1.00	5.00	2.7849	1.36655
Preferring to know the pages or paragraphs to be covered	465	1.00	5.00	4.5290	.77939
Memorizing content even not understanding it	465	1.00	5.00	3.4839	1.29166
Cannot memorize text with more than 75 word	465	1.00	5.00	3.2301	1.31009
I lose grades if write answer in my own words	465	1.00	5.00	2.6581	1.39475
I can pass even if I don't understand anything	465	1.00	5.00	3.2473	1.26675
I expect MCQs	465	1.00	5.00	4.1570	1.07232
MCQs are easy to answer	465	1.00	5.00	4.3032	1.03416
Get good grades through MCQs	465	1.00	5.00	4.2108	1.02905
Getting better grades through memorization	465	1.00	5.00	3.8473	1.08095
Getting a good job is difficult with low	465	1.00	5.00	3.7419	1.28069

grades					
Grades are more important than English	465	1.00	5.00	2.3441	1.36374
Preferring essay questions in exams	465	1.00	5.00	2.5914	1.37587
I fail if I've to write essay questions	465	1.00	5.00	2.9892	1.28900
Essay writing is difficult due to weak spelling and grammar	465	1.00	5.00	3.0559	1.35391
Cannot write essay due to lack of content	465	1.00	5.00	3.0086	1.33743
Preferring questions asking for my personal ideas and knowledge	465	1.00	5.00	3.7699	1.04879
Final exam is more important for me	465	1.00	5.00	3.5849	1.35260
It informs me about my learning and progress	465	1.00	5.00	4.1398	.96926
Applying the same strategies both in mid-term as well as the final exams	465	1.00	5.00	3.7806	1.01245
It helps me improve my reading and writing	465	1.00	5.00	3.9763	1.06950
Not worried about formative assessment	465	1.00	5.00	2.0538	1.15562
More worried about the final assessment	465	1.00	5.00	2.3011	1.31786

Table 2 indicates the students' perceptions of the impact of mid-term examinations and pop quizzes on what they learned and how they learned. A clear majority of the respondents reported that their study was limited to specific pages of course only. In addition, most of them also wanted to know the exact pages and paragraphs to be covered in the assessment tasks. Moreover, they liked the different modes of assessment to include multiple-choice questions (MCQs) only, for the reason that they were easy to answer compared to essay type questions. Essay questions, they believed, caused them loss of grades which triggered difficulty in getting good jobs. Moreover, they stated that essay questions were difficult to answer due to their weak vocabulary, spellings and lack of command on English grammar. In this context, almost half of them were of the view that getting good grades was more important than learning English as they believed that job providers looked at grades mentioned on certificates and not their language proficiency. Concerning the importance of formative assessment compared with the final examination, they indicated equal concern for both of them, i.e. recognizing their importance.

As stated above, the assessment tasks strongly influenced the content the students studied; their learning style also appeared to be under immense effect of the assessment tasks. More than half of the participants employed memorization as a learning strategy with the belief that memorization of data fetched them better grades. Furthermore, they indicated that they used similar learning strategies for both formative and summative testing, i.e. strategies that result in lower order learning,

memorization of materials with little real understanding of application of what they are taught (Gijbles et al. 2008). Therefore, such impact of the formative assessment could be labeled as unintended (National Center on Educational Outcomes [NCEO] and from the perspective of Winke (2011), such assessment tasks, with no or little positive impact on students' learning, are categorized as consequentially invalid despite being reliable. The quantitative findings lend strong support to Black and William's (1998a&b) thesis that *only* improved formative assessment can raise learners' level of learning.

Data from classroom observations

To explore how the participants' responses to the survey were reflected in their classroom behaviors, 18 lessons of 5 different teachers were observed. The number of the lessons of each teacher depended on the time-management factor between the teachers and the researcher. As Table 3 indicates, classes were selected from all levels (year 1-4) of the BA program. It was noted that seats in the classrooms were immovable, apart from one where three lessons of teacher A were observed.

Table 3
Demographic information of the classes observed

	Lessons	Year of Study	Rows	Horseshoe
Teacher A	1 At the Restaurant	1		✓
	2 Describing People	1		✓
	3 At the Bank	1		✓
	4 Past Simple and Past Progressive	1	✓	
	5 Revision	1	✓	
Teacher B	1 Paragraph Writing	1	✓	
Teacher C	1 Free and Literal Translation	2	✓	
	2 Types of Translation	2	✓	
	3 Presentation	2	✓	
	4 Revision	2	✓	
Teacher D	1 Communicative Competence	3	✓	
	2 Teacher Grammar	3	✓	
	3 Monitor theory	3	✓	
	4 Monitor theory	3	✓	

Teacher E	1	Academic Listening	4	✓
	2	Academic Reading	4	✓
	3	Academic Writing	4	✓
	4	Academic Writing	4	✓
Total lessons	18			✓

Five lessons of teacher A were observed; three speaking and two grammar lessons. As can be seen in Table 4 below, a major part of every lesson was spared for students' reading. For instance, out of 140 minutes of the speaking lessons, 69 minutes were used by students' reading; whereas the students speaking time in all three lessons was 54 minutes. Similarly, more than half of the time of the grammar lessons was utilized for the students loud reading of certain items of specific exercise from the book used for the course. Thus, out of 243 minutes only 31% of the total time was availed by some of the students for speaking. On the other hand, more than half of the time of the total time of teacher A's lessons (52%) was consumed by reading of a few students and the teacher's explanation of the content. Nonetheless, the students' speaking time does not mean that every student took part in the reading or speaking activities; rather, it was those students who volunteered to participate in the activities. Similarly, only one or two students were made to read the exchanges of the dialogues of the specific units of the book and the teacher explained its meaning. In addition, the students' reading in the grammar lessons was more of a ticking, underlining, and marking the specific items of particular exercise as important from exam point of view rather than reading for comprehension. On the whole Teacher A's lessons were not reflective of the course goals.

Table 4 shows the time allocated by teacher B for the major activities of his lessons. The course was called paragraph writing. In the beginning of the lesson the teacher announced that the students would learn how to use adjectives. The entire lesson unfolded in the form of teacher's speaking and explanation. Out of the total time which was 43 minutes, the teacher's speaking time recorded was 30 minutes which constituted approximately 70% of the lesson. Only 5 and 8 minutes were used for students' speaking and writing respectively. However, only a small number of students of teacher B class like teacher A's students either wrote something or responded to the teacher's questions and observations. As only one lesson of the teacher was observed, therefore, it might be unreasonable to say that his other lessons will carry the same format.

Table 4 shows the activities recorded in the lessons of teacher C. In total, four lessons of teacher C were observed. The total time of the lessons reordered was 161 minutes. Though two different courses—Types of Translation and Discussion in English—were taught by the teacher, however, his method of teaching was the same. They were purely teacher-dominant lessons. The teacher used about 90% of the total time. The Table shows that very little time was spared for students' activities. In one of the translation lessons, the students wrote, in fact copied, a few sentences written by the teacher on white board. Similarly, in one of the Discussion-in-English lessons, only five students read some sentences from the pieces of paper in their hands loudly turn by turn. This 'group presentation' lasted for about seven minutes. Teacher C mentioned mid-term examination many times during the lessons. From his teaching strategy, it was evident that he focused on what was expected to be given in forthcoming assessment of the course.

Like Teacher B and E, Teacher D was a senior teacher with more than 20 years of teaching experience and had a PhD in Applied Linguistics. Four lessons of Teacher D were observed. The total time recorded of his lessons was 163 minutes. All of his lessons were of the course Applied Linguistics. All of them were purely theoretical lessons as their names suggested. In all of the lessons, the teacher used a single teaching method. He had some students to read a few lines from the book followed by his explanation. Thus in each of the lessons, he made 8-12 students read the sentences the teacher pointed to. Though teacher D's lessons did involve students' reading but in fact it was only to vocalize the words and then explained by the teacher. Like his other colleagues, the teacher's lessons were completely teacher-centered.

Table 4
Time distribution between students and teachers

Teacher	Student speaking time	Student reading	Student writing time	Teacher speaking time	Total time in minutes
A	*78 (31.4%)	*128 (52.6%)	0 (0%)	37 (15.2%)	243 (100%)
B	5 11.4%	0 (0%)	8 18.6%	30 (70%)	43 (100%)
C	7 5%	0 (0%)	5 3%	149 (92%)	161 (100%)
D	0 (0%)	0 (0%)	0 (0%)	163 (100%)	163 (100%)
E	0 (0%)	26 (14%)	30 (16%)	117 (64%)	183 (100%)
	90 11.34%	154 19.41%	43 5.42%	494 8.26 hours	793 13.21 hours (100%)

*Students speaking or reading does not mean the whole class was involved in the activity. Only a few students were invited to take part in the activities. In the reading part, only one student would read throughout the lesson.

Four IELTS (International English language Testing System—a test required by most universities to be taken to demonstrate their English-language proficiency before granting admission in a course) lessons of Teacher E were observed. The total duration of the four lessons of teacher E was 183 minutes. The teacher taught Academic Listening, Reading and Writing modules. The teacher used sample published IELTS tests. The first lesson—Academic Listening—which lasted for about 56 minutes, the students did only 10 minutes listening of the sample recording. The rest of the lesson time was used for teacher speaking. Similarly, less than half of the time was consumed by students in reading in the Academic Reading lesson which lasted for 36 minutes. The total time of the Academic Writing lessons was about 91 minutes. The students took one third of the time to attempt the writing tasks, whereas more than half of the time was spent on teacher's speaking. Thus 5% of the total time was used for listening, 20% for reading and 6% for writing whereas the teacher's speaking time was recorded about 62% of the total time.

In the data from classroom observations, it can be seen that very little real learning took place. The teaching strategies used were indicative of the teachers' expectations from the students in terms of their examination output. In other words, there was an implicit message

for the students to listen passively and reproduce in examination the same content. The students' productive skills, namely speaking and writing are not paid any heed to. Placed in this context, the students' preference reported in the questionnaire data seem to be dependent partly on classroom environment. In a way, the students, by using certain learning strategies, merely respond to their teachers' expectations or it could be vice versa (Tsangaris, 2009).

Student voices

Bearing in mind the students' responses to the questionnaire and their learning behaviors in classrooms, it became indispensable to hear from them the 'why' aspect of the whole episode. For this purpose, 4 focus group interviews with students from different levels were conducted. Their responses were categorized under the following six themes.

The learners' perceptions of quizzes

Most of the informants who had commented on assessment practices expressed their reservations about the pop quizzes (N=12). They viewed quizzes negatively for different reasons. For example, one student said "Minimize the number of quizzes because they create pressure on us" (Male participant, Year 2). Another student commented "There are many short quizzes due to which we can't get good grades" (Male participant, Year 4). Similarly, one of the respondents said "There are many short quizzes and assignments that is why students are unable to perform better in the final exam" (Female participant, Year 4). There was almost no positive comment made about pop quizzes. Rather, they were held responsible for their low grades, disturbing their learning, and a source of anxiety and pressure on students. As one of the participants stated that:

We should not be given a lot of quizzes and assignments.
We have many courses therefore we can't study all of them well. Most of the teachers give us between 60 to 70 out of hundred marks. We suffer (later on in practical life) because of that. We need good GPA because it can help us in future.
(Female respondent, Year 2)

The students' perceptions of research assignments

The University's assessment method of bachelor courses requires the students to be given research assignments. However, in reality students are involved in very little research. Most of the students are not aware of how to prepare a research assignment. Many students (N=20) did not give importance to research assignment; rather, they

considered it a useless activity. For example, one student said “Most of the teachers ask for research papers which wastes our time, particularly at the end of a semester” (Female student, Year 3). How they write the research assignment a student remarked:

Sometimes others write the assignments for me. Most of the time I don't know what my assignment includes. We (students) don't know if our assignments are plagiarised or not.... when I submit my assignment I don't get any feedback from my teachers. I rely on only one book (to write my assignment). Can you tell me what I should do? (Female student, Year 1)

One of the respondents, however, said “Research is good but what is its benefit for students in terms of grades” (Female student, Year 4). Another respondent remarked “We don't know what our teachers want in the assignments...” (Male Student, Year 2). There was another approach some students followed for writing their assignment as one of the student said “Assignments are time and money consuming. We don't do assignments ourselves. We get them done by others (professionals) in the market and pay them money...” (Female Student, Year 3)

The learners' perceptions of examination questions

Another important area the students (N=14) made comment about was the kind of questions they liked in the quizzes, mid-term exams and even the final exams to have. They explicitly showed their disliking for subjective type of questions; rather, they wanted their teachers to include only objective type questions such as MCQs and true-false items. For example, one of the respondents said:

I am serious. The questions should be of true-false type to help us get good grades. Most of the teachers don't do that. They give us essay-type questions in exams. Students forget everything when they see essay questions. (Female student, Year 3)

Similarly, many of the students who asked for objective type questions gave the logic that such kind of questions helped them in getting higher grades. For example, one of the informants remarked:

Multiple choice questions are more helpful for students in getting higher grades compared to essay type questions. We make mistakes in essay type questions and thus we obtain low grades. All teachers should include MCQs in exams because we don't have to memorise for MCQs. (Female student, Year 4)

The learners' perceptions of feedback practices

One of the key themes in the students' comments was the kind of the feedback they received from their teachers. Some of the respondents (N=16) observed that their teachers did not make themselves available to give feedback. For example, one of the students commented "We want help from our teachers. They don't show us our mistakes so we can't improve ourselves" (Female respondent, Year 3). Another student said "Some teachers don't want to hear anything from students about grades or discussions" (Female student, Year 3). One of the male respondents said "our teachers develop negative impression about us if we discuss our grades with them. It is unfair with students". (Male Student, Year 4) A female respondent remarked "I contact my teachers to know my mistakes but they don't give us good responses....for God's sake.....we don't get enough time from our teachers" (Female respondent, Year 2). However, one of the participants wrote that not all teachers behaved in the same way in terms of giving feedback. Some of them provided feedback, but they were not many in number. He said "When I approach any teacher to discuss my weaknesses, grades, mistakes in quizzes or assignments no one pays attention to me but a few of them".

The learners' perceptions of teachers' behaviour

A number of comments were related to teachers' behaviour (N=14). The students were of the view that the relationship between students and teachers were not friendly. They believed the teachers kept distance from the students. For example, one of the participants observed "teachers should minimise distance between themselves and the students" (Male Student, Year 3). Another student from the same group said "We want to have freedom to discuss issues related to any subject with our teachers in classroom". Another respondent argued "Students should feel comfortable with teachers. The teachers should not behave in a strict manner. Students should not be treated as children..." (Female Student, Year 4). A male student commented:

Teachers-student relationship should not be limited to the classroom only. It should be maintained even outside classroom. Students and teachers should have frank relationship so that the problems that we face in classrooms could be discussed with teachers outside the classroom. Some students have the problem of speaking in front of other students, but outside the classroom they can express themselves more easily. (Male Student, Year 1)

One of the male respondents wondered if there was any teacher to guide him to improve himself. He commented:

I am in fifth semester. I fail in some subjects and my colleagues are with me in the same level (semester) have higher GPA than me though my English is better than them. Therefore, I want to know the reason of why this is so.
(Male Student, Year3)

The students' perceptions of their learning strategies

Many respondents (N=15) commented that they used memorisation of materials as a learning strategy due to certain reasons. For instance, one of the respondent said "Some of the students have to memorise (learning materials) for some of the courses where the materials are difficult...it is above the level of students" (Male Student, Year3). Another respondent gave a different reason for why he memorised for exams. He believed that:

Some teachers fail us if we answer from our own. They only want us to reproduce in paper what they teach us in lectures. Some of the teachers like our answers to be exactly what they say in their lectures. (Male Student, Year 4)

Another male respondent referred to the memorisation issue indirectly saying that "Teachers should not expect us to memorise and repeat materials; rather, they should encourage us to be creative" (Male Student, Year 4). A female respondent said "All teachers should focus on revision in the last two weeks of the semester for all courses and don't force us to memorise" (Female Student, Year 3). Another female respondent argued that because of memorisation students passed courses but learnt very little. She said:

We students of English do not need a lot of information. We want our skills like speaking, writing and reading to be improved. Most of the students memorise materials and they pass the courses though they do not understand the materials very well. (Female Student, Year 4)

From the students' comments in the interviews, it emerged that their learning styles and preferences largely depended on the teaching methodologies. Their lack of practice in essay writing seemed to have made them prefer MCQs. The apparent absence of scholarship of teaching seemed to be causing the teacher-centered classroom culture. Thus, in order for the formative assessment to be truly consequentially valid the teaching component of the entire curriculum has to be the center point of focus. The consequences of formative assessment found

indicated that formative assessment on its own may not yield required consequences. It becomes useful if used usefully.

Conclusion and Recommendations

The present research was an attempt to explore whether formative assessment in the given context was consequentially valid. That is, if formative assessment played its role as a teaching tool and influenced what the students learned and how they learned it. Data was collected from English-major students of a single department of a single Saudi public university. Clearly, there are limitations with regard to the extent the findings of this study could be generalized to other similar settings. Further research might help in unfolding the dimensions not investigated in this study, particularly the teacher parameter to seek answer for the 'why' component of the problem. In addition, special attention could be paid to the significant discrepancies found in the participants' responses to the questionnaire. However, evidence gathered from the participants via both quantitative and qualitative tools is indicative of substantive negative impact of formative assessment on learning styles of the learners and on the scope of the learning materials they used. FA is used as a traditional assessment instrument but not as a teaching tool. It does not inculcate any learning autonomy in learners. Advocates of FA hold that it enables learners to grow as self-learners (Black & William, 1998a and b). However, it has to be born in mind that by FA they only mean improved FA that incorporates constructive feedback that bridges the gap between where students are and where they have to be. Concerning the choice and scope of the learning materials the students used, they reported memorization as their main learning strategy and focus on limited amount of content to pass different courses and ascertain obtaining high grades. This finding confirms that if FA is not used in the way its believers believe in, it results in negative impact on learners' learning approaches as well as the scope of what they learn (Gijbles et al. 2008 and Tsagari, 2009). Thus, to answer the question how FA affects students learning in the context of this research, it can be safely concluded that there is a very strong undesirable influence found. Practically, students are involved in surface-level learning. FA only helps them remember limited amount of information, perhaps for a short period of time. No analytical skills are challenged by the assessment tasks. They are not required to produce something on their own.

Though not direct participants and focus of the study, the teachers, whose lessons were observed, happened to share the same teaching approach, i.e., being teacher-centered. None of the lessons included any practice or production activities, as they are referred to Communicative Language Teaching (CLT). The immovable seats might be one of the hindrances in this regard.

Of particular interest is the fact that consequentially validity research like wash back studies is a complex phenomenon. On its own it does not trigger the consequences it is supposed to trigger. Nonetheless, there is still a lot left that teachers can do to make FA in the current setting more effective, but that may not happen without adequate training given to them to implement the assessment tool with the right spirit. Their training might start by taking measures to increase their awareness of the true philosophy of formative assessment and strengthening their belief in it as a teaching tool. In a nutshell, formative-assessment literacy training programs should be launched as soon as possible.

References

- Al-Seghayer, K. (2011). English teaching in Saudi Arabia: Status, issues, and challenges. Riyadh:Hala Printed Co.
- Black, P. and William, D. (1998a). Inside the black box. *Phi Delta Kappan*, 80(2), 139-148.
- Black, P. and William, D. (1998b). Assessment and classroom learning. *Assessment in education: principles, policy and practice* 5(1), 7-75.
- Camara, W., & Lane, S. (2006). A historical perspective and current views on the Standards. *Educational Measurement: Issues and Practice*, 25(3), 35-41.
- Chen, L.M.D. (2002). Taiwanese junior high school English teachers' perceptions of the washback effect of the Basic Competence Test in English. Unpublished PhD Dissertation, Ohio State University.
- Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, 11, 38-54.
- Cheng, L. and Curtis, A. (2004). Washback or Backwash: A Review of the Impact of Testing on Teaching and Learning. In L. Cheng, Y. Watanabe, and A. Curtis, (Eds.), *Washback in language testing; Research context and methods*(pp. 3-17). London: Lawrence Erlbaum Associates.
- Chu, L. and Gao, P. (2006). An empirical study of the washback of CET-4 writing. *Sino-US English Teaching*, 3(5), 36-38.
- Creswell, J. W. (2013). *Research design: qualitative, quantitative, and mixed method approaches* (4th Ed.) London: Sage.
- Creswell, J.W. & Clark, V.L.P. (2011). *Designing and conducting mixed method research* (2nd Ed). London: Sage
- Denscombe, M. (2007). *The good research guide: For small-scale social research project* (3rd Ed.). Buckingham: Open University Press.

- Gijbles, D., Segers, M. and Struyf, E. (2008). Constructivist learning environments and the possibility to change students' perceptions of assessment demands and approaches to learning. *Instr. Sci.*, 36, 431-443.
- Gielen, S. Dochy, F. & Dierick, S. (2003). Evaluating the consequential validity of new modes of assessment: The influence of assessment on learning, including pre-, post and true assessment effects. In M. Segers, F. Dochy & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (pp. 37-44). New York: Springer Netherlands.
- Green, A. (2007). *IELTS washback in context: Preparation for academic writing in higher education*. Cambridge: Cambridge University Press.
- Hayes, B. & Read, J. (2004). IELTS test preparation in New Zealand: Preparing students for the IELTS academic module. In L. Cheng, Y. Watanabe, & A. Curtis, (Eds.), *Washback in language Testing: Research context and methods* (pp. 97-111). London: Lawrence Erlbaum Associates.
- Hitchcock, G., & Hughes, D. 1989. *Research and the teacher: a qualitative introduction to school-based research*. London: Routledge
- Johnson, R.B., Onwuedbuzie, A. J. & Turner, L.A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112-133.
- Muñoz, A.P. and Álvarez, M.E. (2010). Washback of an oral assessment system in the EFL classroom. *Language Testing*, 27(1), 33-49.
- Saif, S. (2006). Aiming for positive washback: A case study of international teaching assistants. *Language Testing*, 23(1), 1-34.
- Sergers, M., Dierick, S. & Dochy, F. (2001). Quality standards for new modes of assessment: An exploratory study of the consequential validity of the OverAll Test. *European Journal of Psychology of Education*, XVI (4), 569-588.

- Sambell, K., McDowell, L., & Brown, S. (1997). 'But is it fair?': An exploratory study of student perceptions of the consequential validity of assessment. *Studies in Educational Evaluation*, 23(4), 349-371.
- Shepard, L. A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice*, 16(2), 5-8.
- University of Minnesota, the National Center on Educational Outcomes. Arizona Consequential validity study for the AZELLA Kindergarten Placement Test: Final report. Retrieved on 4/4/2015 from <http://www.azed.gov/assessment/files/2014/08/kpt-final-report-8-15-2014.pdf>
- Tsagari, D. (2009). Revisiting the concept of test washback: Investigating FCE in Greek language schools. *Cambridge ESOL, Research Notes*, 35, 5-9.
- Watanabe, Y. (2004). Teacher factors mediating washback. In L. Cheng, Y. Watanabe, and A. Curtis, (Eds.), *Washback in language testing; Research context and methods* (pp. 129- 146). London: Lawrence Erlbaum Associates.
- Watanabe, Y. (1997). The washback effects of the Japanese university entrance examination of English-classroom-based research, Unpublished PhD Thesis. Lancaster: Lancaster University.
- Winke, P. (2011). Evaluating the Validity of a High-Stakes ESL Test: Why Teachers' Perceptions Matter. *TESOL Quarterly*, 45(4), 628-660.
- Zhao, Z. (2013). An investigation into the consequential validity of a diagnostic college English test. *English Language Teaching*, 6 (2), 171-178.

Received on: 05.02.2016

Revised on: 07.06.2016

Accepted on: 12.06.2016