

Investigating the Washback Effect of Comprehensive Weekly Quizzes on Preparatory Year Students' Summative Grades at Umm Al Qura University, Makkah
Iftikhar Alam¹, Abdullah Yassin Abdullah²,¹Deanship of Preparatory Year, Umm Al Qura University, Makkah, Saudi Arabia (iahanan@uqu.edu.sa)²College of Languages, Sudan University of Science and Technology, Sudan

Abstract

The current study is an attempt to investigate the washback effect of comprehensive weekly quizzes on summative exams grades of Preparatory Year scientific stream students at Al Qunfudah Campus, Umm Al Qura University, Saudi Arabia. The quiz is one of the most effective tools of formative assessment for coursework learning. Quiz affects students learning and grades. This effect of a quiz on English language learning, in Applied Linguistics, is called a washback effect. Many researchers have tried different tools of formative assessment to get a positive washback effect. There are many different tools of formative assessment which are applied by teachers during coursework for language learning like class participation, assignments, and quizzes. Quizzes, as a tool of formative assessment, also carry a washback effect on coursework learning progress as well as on final grades. It enhances the classroom learning process, language practice, memorization, feedback, and also influences the students' final summative grades. However, on the other side, some researchers oppose the quizzes' washback effect. They believe that frequent testing might have a negative impact on anxious students. This study used the quantitative- experimental method to know whether students taking weekly comprehensive quizzes performed better in the summative exams. Fifty students of Preparatory Year of First Semester 2016-2017 were selected and divided into two groups. One group of 25 received a routine formative assessment for coursework followed by summative-midterm and final- exams, and the other group received weekly comprehensive quizzes along with the routine formative assessment tools followed by summative -midterm and final- exams. The results analysis showed that the group with weekly comprehensive quizzes performed significantly better than the other group.

Keywords: Comprehensive, Formative, Preparatory, Summative, Washback Effect

المستخلص

تهدف الدراسة الحالية إلى محاولة تقصي تأثير الاختبارات الدورية الشاملة للعملية التعليمية على الدرجات النهائية لطلاب السنة التحضيرية بجامعة أم القرى بمكة لطلاب السنة التحضيرية المسار العلمي بفرع القنفذة جامعة أم القرى بالمملكة العربية السعودية. يعد الاختبار أحد أكثر آليات التقييم الدوري و التكويني فاعلية لمنهج التعلم بما انه يؤثر على تعلم الطلاب ودرجاتهم. يسمى هذا التأثير للاختبار على تعلم اللغة الإنجليزية في علم اللغويات التطبيقية بالتأثير العكسي للعملية التعليمية. لقد حاول العديد من الباحثين تجربة آليات مختلفة للتقييم الدوري للحصول على انعكاس إيجابي. هناك العديد من أدوات التقييم الدوري المختلفة التي يطبقها المدرسون أثناء الفصل الدراسي لتعلم اللغة مثل المشاركة الفصلية والواجبات والاختبارات. بالإضافة إلى ذلك فإن الاختبارات الدورية، كأداة للتقييم التكويني، لها أيضًا تأثير على التقدم في المنهج و العملية التعليمية وكذلك على الدرجات النهائية. فبحيث ان هذا التقييم يعزز عملية التعلم في الفصل، و ممارسة اللغة، والحفظ، والملاحظات، ويؤثر أيضًا على الدرجات النهائية للطلاب. ومع ذلك، و في جهة أخرى، يعارض بعض



الباحثين التقييم الدوري لاعتقادهم أن الاختبارات الدورية قد يكون لها تأثير سلبي على الطلاب الذين يواجهون توتر و قلق. و قد استخدمت هذه الدراسة المنهج الكمي التجريبي لمعرفة ما إذا كان أداء الطلاب الذين يجرون اختبارات دورية شاملة أفضل في الامتحانات النهائية. تم اختيار خمسين طالباً وطالبة من السنة التحضيرية للفصل الدراسي الأول 2016-2017 وتم تقسيمهم إلى مجموعتين. تلقت مجموعة واحدة مؤلفة من 25 شخصاً تقييماً تكوينياً روتينياً للمنهج الدراسي متبوعاً باختبارات نصفية ونهائية تلخيصية ، وتلقت المجموعة الأخرى اختبارات دورية شاملة إلى جانب آليات التقييم التكويني الروتينية متبوعة باختبارات نصفية ونهائية تلخيصية. و قد أظهرت تحليلات النتائج أن المجموعة التي أجرت اختبارات دورية شاملة كان أداؤها أفضل بكثير من المجموعة الأخرى.

Introduction

The Preparatory Year assessment scheme of Um Al Qura University (UQU) has 20% marks for course learning (participation marks), 30% for the midterm, and 50% for final exams. Coursework marks are graded through formative assessment, whereas midterm and final exams are based on summative assessment. In Preparatory Year (PY) at UQU, the course marks are collected through different formative assessment tools such as home assignments, class participation, and class attendance. Coursework performance is measured by formative assessment to collect grades out of 20 marks. The university assessment scheme uses both formative and summative assessment jointly for evaluating and grading students' language learning. Summative and formative assessments have different procedures and purposes. Policymakers believe that students' real English language level cannot rely only on a single assessment method. Both assessments are applied together for bearing positive washback on the teaching and learning process (Harlen, W., & James, M. 1997). The university policymakers assign great importance and role to the coursework tasks on students' performance. Sometimes coursework assessment washback effect is

often not as significant as the policymakers hoped it would be.

The washback effect, in applied linguistics, is connected with language learning and assessment. Assessment plays a vital role in language learning. A useful tool for assessment creates positive washback on students' language learning and performance. Therefore, all suitable assignments, tasks, and activities can be applied for success on such tests and exams (Alderson & Wall, 1993; Messick, 1996). The positive washback effect of assessment improves students' learning. There are many kinds of research about the importance of homework, quizzes, assignments, class participation, test, and many other tools to develop students' learning (Smith, Zsidisin and Adams, 2005).

Many researchers pointed out that frequent quizzes were a reliable tool for improving students' performance.

The current study is aimed to investigate the washback effect of the tools of formative assessment that are adopted by teachers and lecturers during coursework to ensure and maximize students learning and grades. In the current study, coursework/ syllabus learning is assessed by certain activities and assignments (in-class and out-class) at UQU, which occur during the whole semester.



Weaver and Qi (2005) believe that active participation of students in the participation works like assignments and activities, learn more than those who do not do that practice. EJ Dallimore et al. (2010) point out that students' participation in all activities is positively associated with the learning process.

This research examines the washback effect on the coursework assessment tools on the performance of the students who are taking English Language courses at the university. Comprehensive weekly quizzes (CWQs) as a formative assessment may affect the students' course and final summative grades. The study took place in the 2016-2017 academic year: the first semester of PY at Al Qunfudah Campus, UQU. The study investigated the Washback effect of CWQs on the final summative grades.

1.1 The Study Problem

It has been frequently observed that the official results of PY students at Al Qunfudah Campus, UQU, have a high inconsistency between the formative marks and summative marks. The previous results of PY show that the majority of students take a high percentage in coursework as a formative assessment. Still, at the same time, they take a low percentage in the summative assessment of the semester. In other words, coursework participation activities and assignments have no positive or desired washback effect on students' final score — assessment rates the outcome of course activities. Therefore, assessment creates a logical and robust association between teaching and learning tasks. The association between assessment and its impact on language learning is termed it washback in Applied Linguistics. The current study applies an additional tool of

CWQs along with the existing tools as a formative assessment for coursework learning. An experimental study is carried out to determine CWQs impact on summative exam scores as compared to the routine coursework assessment on PY students' summative grades.

1.2 Objectives of the Study

The study will determine whether the additional application of graded CWQs in place of routine coursework assignments and activities has any significant influence on the course as well as on the summative score of the students, particularly in the EFL context at Preparatory Year students at Al Qunfudah campus, UQU, Makkah. This study is a significant input to washback study in an EFL Saudi context.

1.3 Research Question/ Hypothesis

The following research question and hypothesis were formulated:

Question: To what extent the comprehensive weekly quizzes (CWQs) impact summative exams score of the Preparatory Year (PY) students.

For investigating the research question, the following three sub-questions may be evaluated:

- To evaluate CWQs washback effect on-course learning,
 - To evaluate the washback effect of quizzes on midterm summative exam score.
 - To evaluate the washback effect of quizzes on the final summative exam score.
- Hypothesis: There is no washback impact of CWQs as a tool of formative assessment on PY students' summative grades.

2 Review of the Related Literature

Washback is comparatively a neutral term, which talks about a useful or positive effect or negative effect (Bailey, 1996).



The aim of washback to bear a positive impact is rather challenging for test-designers, learners, and institutions. Messick (1996) considers that the effects of a positive washback can be generated if the activities, tasks, and strategies employed in instructions which are expected to be helpful for learning, exam, or test. Similarly, a negative washback will be there if tasks, activities, and methodologies have loopholes or less target oriented. A test in itself cannot be termed good or bad, but its washback effect decides its negativity or positivity. Alderson (1993) accepts that test makers should consider the likely impact of an exam or test.

It is vibrant from the existing literature of washback that assessment methods and tools are likely to influence the way students learn (Alderson and Wall 1993; Green 2007). The washback concept has established the fact that assessment approaches and assessment tools mold learning and grading. Better selection of assessment tools (home assignments and class participation, formal or informal quizzes, mid-exam, and final term exam) affects learners' achievement strategies and techniques. There are many washback studies about testing and evaluation on teachers and learners, but a little investigation is available on the formative assessment method and its tools. Very little literature is available on the washback of formative assessment on students' learning strategies (Newfields, T, 2005).

There are many different definitions of quizzes as a tool for formative assessment. Some termed them as a kind of assessment which is carried out weekly. Some termed them as a kind of examination which is carried out on a daily basis (Dineen, Taylor, & Stephens, 1989), while some researchers linked them as a kind of monthly examination (Kling, Miller, & Reardon,

2005). There are many types of research on quizzes advantages. Quizzes develop students' retention power and make them ready for high stakes assessments (Johnson & Kiviniemi, 2009).

Some other studies claim that quizzes improve the students' performance. Quizzes and tests increase students' classroom attendance (Clump, Bauer, & Alex, 2003). Quizzes also give information about students' current syllabus progress and students' performance in the class.

Bailey (1996) forwarded his findings that washback of assessment was only expected effective power where learners were inspired to do well on the test. Where students were confident about how to be successful and considered that they had enough means to be successful, if the test or evaluation was less valuable for the learners, then there would be a little reason for test preparation.

Roediger and Karpicke (2006) investigated the impact of quizzes on undergraduate university students. The study found out that the students, who were quizzed regularly, remembered the course information better than those who were not given quizzes regularly.

Zarei (2008) conducted a study to investigate the effect of frequent quizzes on Iranian English students' classroom attendance and performance. The results stated that the students who received more quizzes performed better than those who had less exposure to quizzes. The result also showed a positive correlation between frequent quizzes with classroom attendance.

Another study was carried out by Marcell (2008) on frequent online quizzes. The researcher compared two groups of students. One group took an online quiz daily, and the other group was not taking any quizzes either online or traditional.



The study discovered that students with daily quizzes played an active role in the class. The students were more prepared and inquisitive in the class than the group that took no quiz.

Basol and Johanson (2009) conducted a meta-analysis of different quizzes' impact on students' exam performance and learning process. The data showed that frequent tests/quizzes were useful for student academic performance and learning progress.

The literature review gives detail of the washback effect of testing/ quizzes on student learning performance. However, there is very little research on quizzes as a tool for formative assessments on Saudi students. There is an application of quizzes in language learning assessment, but its absence of informal grading in the University Marking Scheme. The current study inducts quizzes as a formal tool for grading and learning performance.

3. Methodology

Experimental – quantitative method was adopted to carry out the washback effect of comprehensive weekly quizzes (CWQs) on course grades and midterm and final grades.

A. Participants

Fifty students of PY at Al Qunfudah Camp took part in the study. The participants of the study were divided into two groups - experimental and control. The experimental group received CWQs and routine coursework activities, but CWQs were only graded for measuring coursework learning. The control group received the routine coursework activities like assignments and class attendance, and the same activities were graded for measuring coursework learning. Both the study groups received equally summative midterm and final term exams. They took the same

materials, contents, and the same teacher for teaching. A random procedure was used for sample selection.

B. Instruments

CWQs and assignments were used as an independent variable for course formative assessment. The CWQs and assignments were composed of integrated skills, including listening and speaking, reading and vocabulary, and grammar and writing. These quizzes were comprehensive due to its range of course and time. It covered the weekly taught syllabus. The unified midterms and final summative exams of the university were used for both groups.

C. Procedure

The experimental group took CWQs along with routine activities and assignments for the whole semester. The control group did not take any CWQs except the routine class activities and assignments. Both groups took midterm and final summative exams during the semester. The study used an experimental design method.

Independent variables were used to determine their effect on dependent variables. The participants of this study were equally distributed into two-group with 25 participants in each group. The samples of both groups had almost the same demographic characteristics. One group was having formal CWQs (supplementary tool of formative assessment), and the other was having class attendance and assignments (routine tools of formative assessment).

The study continued for fifteen-week teaching and assessment. Both experimental and control groups went through the same coursework and all-one in-class activities and teaching material. There was no difference between the teaching materials and timings.



They had unified midterm and final summative exam together. The experimental group had aCWQ at the end of each teaching-week. These quizzes were marked, and the grades of quizzes were accumulated for the whole semester against 20 % participation/course marks. The control group had a daily class attendance and assignments. These two groups used to meet five times a week during the on-campus daytime schedule. The researcher used the same teaching style and lesson plan and the same activities/assignments. All the students who participated in the study were taken their prior consent for agreeing to it. The students were informed that participation in the study was voluntary. They can drop their study at any time of the year and not affect a student's grades. However, it was clear for them that those who took part in the study would be considered for their final course grades.

3.1 Validity and Reliability of the Instruments

The instruments are valid and reliable if they measure what they are supposed to measure. There are always threats or questions to the validity of quizzes and assignments, and even to all tools of formative and summative assessments. These validity threats are construct-irrelevant variance and construct under-representation. The researcher has predefined the components of each weekly quiz and the criteria for class participation assignments to assure the validity and reliability of the variables. For controlling over construct –irrelevance, a standard crafted test was applied. Each bit of a quiz or assignment supported the aims and objectives of the assessment. The researcher crafted quizzes and assignments which were in line with the cognitive domain of the students as required for the current syllabus. Construct-

relevance and representation were achieved by the support of experienced faculty members who were teaching EFL for a long time and were involved in exams making. The researcher used experienced colleagues in the construction and administration of quizzes to make sure the reliability of the quizzes. There was no Hawthorne effect as both groups tried for high grades. There was no researcher bias or influence as the researcher had no control over the final summative exams. The external summative exams were given to both groups.

3.2 Data Collection

There were four instruments employed to collect research data: CWQs, the class activities and home assignments, and midterm and final exams. The experimental group was using the first instrument for the evaluation of 20 percent marks as a formative assessment. The control group was using the second instrument for the evaluation of 20 percent marks as a formative assessment tool. The third and fourth instruments were equally applied to both groups to collect the summative data.

3.3 Data Analysis and Procedure

The experimental design of the study used multiple variables data to examine the washback of CWQs on the Saudi EFL students' summative grades.

1. To make a descriptive analysis of the demographic and academic backgrounds of both samples of students.
2. To examine the washback effect of CWQs on the experimental group as a tool of formative assessment, and the impact of class participation and assignments on a control group as a tool of formative assessment. Range, Mean, Standard Deviation, Percentage, and T-Test analyses were used.

3. To examine the relationship of midterm and final summative grades between experimental and control groups, mean and T-test analyses were carried out to see its washback effect upon learning.

4 Results and Findings

This study investigated the washback impact of CWQs on coursework progress as well as on summative grades of PY students at UQU. The PY students of English Language were divided into control and experimental groups to find the relationship between routine class Demographic Descriptive Information of Experimental (EG) and Control Groups (CG)

assignments and summative final exams scores and the relationship between the comprehensive weekly quizzes and summative final grades, respectively. The data of quizzes, assignments as formative assessment were derived from practical teaching and assessment of the first semester of 2016-2017.

Table: 4.1 describes the demographic information of the samples. It presents the distribution of the students in both samples of the course, English Language for Scientific Stream- PY students.

Category	Respondent description
Gender:	Male only
Male	50 Male students (100% male)
Female	N/A (There is a segregate education system, no female students in the male campus)
Age:	18-19 year
Marital Status:	N/A
Nationality:	Saudi (Arabic native)
Major:	Engineering
Scientific Stream (Engineering)	Preparatory Year (fresh university students)
Coursework/ subject	English Language
Teaching of English	16 hours per week
Previous qualification	Secondary School Certificate

The descriptive data of table-4.1 represent no significant difference in the demographic background of the variables.

They are almost sharing the same age, language, educational level, and ethnicity. There is no significant difference in them.

Table: 4.2

Washback Effect of Class Participation and Assignments as Formative Assessment Tools on CG Coursework 20% Marks

Variables / Tools	Students	Marks Range	Mean	St: DV:	Av: Percentage
Class participation / Assignments	25	0-20	19	1.0	95 %
Average/ Total	25	0-20	19	1.0	95%

Table-4.2 data show the result detail of coursework marks through various formative assessment tools. Class participation and home assignments are the formative assessment tools that are used for

gathering the whole semester coursework marks of the students. The data show that CG got 95 percent in 20 % formative assessment.

Table: 3

Washback Effect of CWQs as a Formative Assessment Tool on EG Coursework 20% Marks

Variables	Students	Marks Range	Mean	St: DV:	Av: Percentage
Quizzes	25	0-20	17.25	1.2	86.25%
Average/ Total	25	0-20	17.25	1.2	86.25%

Table-4.3 data show the results of coursework marks through quizzes as a formative assessment tool. CWQs are used for gathering full semester coursework

marks of the students. The data show that EG got 86.25 % in 20 percent formative assessment marks.

Table: 4.4

Formative Assessment Tools Washback on CG and EG Coursework 20% Marks

Samples	Students	Participation / Coursework Marks	Mean	St: DV:	Av: Percentage
CG	25	0-20	19	1.0	95 %
EG	25	0- 20	17.25	1.5	86.25%

Table-4.4 (continue)

Samples	Mean	SD	T. Value	P. Value	St: Alpha
CG	19	1.0	0.939	0.0035	0.05
EG	17.25	1.5			

Table-4.4 indicates that the p-value is less than Alpha 0.05 ($p < 0.05$), so there is a statistically significant difference between CG and EG performance. The data table-

4.4 shows that CG has a higher score based on formative assessment tools. The CG has 95 an average percentage, whereas EG has a significantly low percentage of 86.25.

Table: 4.5

Washback of Class Participation/ Assignments and Quizzes as Formative Assessment on Midterm Summative Results of CG and EG

Samples	Students	Marks Range	Mean	Av: Percentage
C G	25	0-30	21	70%
E G	25	0-30	25	83.3%

Table-4.5 (continue)

Samples	Mean	SD	T. Value	P.Value	St: Alpha
CG	21	4.5	-2.952	0.0001	0.05
EG	25	2.7			

Table-4.5 indicates that the p-value is very less than Alpha 0.05 ($p < 0.05$), so there is a statistically significant difference between CG and EG performance and grades. The EG has a statistically significant grades margin over CG in midterm results. CG has

an average percentage of 70, whereas EG has a percentage of 83. It shows a big gap between the midterm grades. The performance of EG is far better than CG in midterm grades.

Table: 4.6

Washback of Class Participation/ Assignments and Quizzes as Formative Assessment on Final Summative Results of CG and EG

Samples	Students	Marks Range	Mean	Av: Percentage
C G	25	0- 50	34.7	69.4%
E G	25	0- 50	45.4	90.8%

Table-4.6 (continue)

Samples	Mean	SD	T. Value	P.Value	St: Alpha
CG	37.70	9.1	-2.924	0.0002	0.05
EG	45.40	4.6			

Table-4.6 indicates that the p-value is significantly less than Alpha 0.05 ($p < 0.05$), and we see a big difference in final term results. The performance of EG is far better than CG in final term result analyses. CG has an average percentage of 69.4,

whereas EG has a percentage of 90.8. It shows a big gap between the final term grades. The standard deviation is 4.6 in EG, whereas 9.1 in CG. The final marks of EG are more in the cluster than the marks of CG.

Table: 4.7

Summary of Summative Results of Mid and Final Terms Marks of CG and EG – Consolidative Analyses

Samples	Students	Tot: Summative Marks (0-80)	Xm +Xf= Mean	Av: Percentage
CG	25	30+50=80	21+37.70=58.7	73.37 %
EG	25	30+50=80	25+45.4= 70.5	88.12 %

Table-4.7 (continue)

Samples	Mean	SD	T. Value	P. Value	St: Alpha
CG	58.7	13.6	-2.960	0.0001	0.05
EG	70.5	7.3			

Table-4.7 indicates that the p-value is significantly less than Alpha 0.05 ($p < 0.05$), so there is a statistically significant difference between CG and EG performance. Table -4.7 data show a big difference between the midterm and final term results of both groups. CG has an

Table: 4.8

Whole Semester Results Analyses of Control and Experiment Groups - A Summary of Total Grades

Samples	Students	Total Marks (0-100)	Coursework Midterm+ Final	Mean+ Final mean	AV: Percentage of total marks
C G	25	0- 100	19+21+37.70=	77.7	77.7 %
E G	25	0- 100	17.25+25+45.4=	87.65	87.65 %

Table-4.8 (continue)

Samples	$X_p + X_m + X_f = \sum \text{Mean}$	$SD_p + SD_t = \sum SD$	T. V	P Value	St: A
CG	$19+21+37.70=77.7$	$0.70+13.6=14.3$	-		
EG	$17.25+25+45.4= 87.65$	$1.6+ 7.3= 8.9$	2.414	0.0003	0.05

Table-4.8 indicates that the p-value is significantly less than standard Alpha 0.05 ($p < 0.05$), and we see that there is a statistically significant difference in the whole result of participation, midterm, and final term. The EG surpasses in score with an evident margin. The performance of EG is far better than CG in the analyses of the total results. Table-4.8 shows the consolidative data of all previous results of the semester (formative and summative). The consolidated data show that CG comes with an average percentage of 77.7, whereas EG has a percentage of 87.65. It shows a big gap between CG and EG "Mean" and "Standard Deviation." The EG comes out with high mean and low standard deviation. The low standard deviation indicates the closeness of marks with each other. It is pertinent to mention here that participation marks of CG are slightly better than EG, but in midterm and final, the marks are low.

average percentage of 73.5, whereas EG has a percentage of 88.12. It shows a big gap between the summative marks of both groups. The performance of EG is far better than CG in midterm and final term exam scores.

4.9 Discussion

The current study shows that the washback of CWQs group is significantly better than that of the control group. It supports the Soehren (1997) study that weekly quizzes improve students' performance. The washback finding of the present research says that weekly quizzes increase students' scores than those who do not have any quiz during the coursework. Comprehensive weekly quizzes cause better learning performance as compare to routine formative assessment tools.

The study did not use weekly quizzes, as pointed by Qi (2005), to allocate more time for preparing students for the high-stakes test and spend more time on teaching of the tested subjects. The research was carried with equal teaching materials and timings and classroom facilities for both groups, which made instructors and students free from exams slaves. No particular artificial environment was created to get unnatural and unscientific results.



The assessment method and its tools can affect the achievement of learners as earlier mentioned by (Alderson and Wall 1993; Cheng and Curtis 2004). The study finding endorses that assessment methods and tools are likely to influence the way students learn (Saif 2006; Green 2007; Zarei 2008; and Basol and Johanson 2009) who investigated the impact of quizzes and found positive for learning and performance.

The results show that the washback impact of class participation and assignments is positive in coursework 20% marks for the control group, as shown in Table-4.2. Still, its effect on the midterm and final summative grades is negative or ineffective, as demonstrated in Table-4.5, 4.6. On the bases of facts and figures, the washback of coursework marks is very conflicting. The data analyses show that coursework assessment has fallen short of its intended role as a positive washback on the students' final grades and performance. It means that the coursework assignments and tasks did not succeed in uplifting the students' performance in summative exams. However, the coursework marks are being used in the university evaluation system long in the assessment system.

The results show that the washback impact of CWQs is positive in coursework 20% marks for the experimental group but little lower in grades than the control group, as shown in Table-4.4, but its impact on the midterm and final summative grades is significantly positive, as shown in Table-4.5, 4.6.

The study declared that routine formative assessment tools for coursework assessment did not meet their objectives for improving the performance of Preparatory Year students at Al Qunfudah Campus- Um Al

Qura University, KSA. It was also pointed out that the routine tools of assessment had a negative impact on summative marks of the students. The study pointed out that the introduction of CWQs had a positive washback on students' midterm and final term achievement scores. Therefore, the hypothesis of CWQs of no washback effect is rejected.

4.10 Limitations and Potential for Further Study

The current study was limited to Preparatory Year Scientific Stream- Students. Only male students of Al Qunfudah Campus at Umm Al Qura University of the first semester (2016-2017) were investigated. Due to the segregate education system for male and female in the Kingdom, the researcher designed method was limited to male students only, and the sample size was relatively small. It was capped just to one-semester academic teaching and assessment. So the findings may be less generalized. However, specific results of coursework assessment were traced very minutely and investigated thoroughly.

Future studies should focus on the other disciplines of the preparatory year of the university. It can be extended to female preparatory year students. Further research is required to investigate the washback of other formative assessment tools in various settings to develop a comprehensive model of language-assessment washback.

5. Conclusion

The study applied experimental teaching and class observation, assignments, quizzes, and unified summative exams on randomly selected experimental and control groups for fifteen weeks. It involved 50 students of the preparatory year from the scientific stream.

The control group was using the routine tools of formative assessment for coursework assessment, whereas the experimental group was mandatory to use comprehensive weekly quizzes as a tool of formative assessment for coursework progress. Both groups shared the same coursework or syllabus and materials, attended the same lectures, and received a similar teaching method. The experimental study found very significant pieces of evidence that weekly comprehensive quizzes were highly advantageous for Saudi Preparatory Year students' summative grades and created positive washback on students learning. The current research summarizes comprehensive weekly quizzes as a tool of formative assessment with a positive washback effect on preparatory year students' summative grades. The findings of the present study have implications for teachers, learners, and policymakers. The findings suggest that comprehensive weekly quizzes may be adopted along with other tools of formative assessment for positive washback effect on summative grades as well as on learning performance.

References

Alderson, J. C., & Wall, D. (1993). Does washback exist?. *Applied linguistics*, 14(2), 115-129.

Bailey, K. M. (1996). *Working for washback: A review of the washback concept in language testing*. *Language testing*, 13(3), 257-279.

Başol, G., & Johanson, G. (2009). *Effectiveness of frequent testing over achievement: A meta-analysis study*. *Journal of Human Sciences*, 6(2), 99-121.

Curtis, A., & Cheng, L. (2004). *Washback or backwash: A review of the impact of*

testing on teaching and learning. In *Washback in language testing* (pp. 25-40). Routledge.

Clump, M. A., Bauer, H., & Whiteleather, A. (2003). *To attend or not to attend: is that a good question?*. *Journal of Instructional Psychology*, 30(3), 220.

Dallimore, E. J., Hertenstein, J. H., & Platt, M. B. (2010). *Class participation in accounting courses: Factors that affect student comfort and learning*. *Issues in Accounting Education*, 25(4), 613-629.

Dineen, P., Taylor, J., & Stephens, L. (1989). *The Effect of Testing Frequency upon the Achievement of Students in High School Mathematics Course*. *School Science and Mathematics*, 89(3), 197-200.

Geist, J. R. and Soehren, Stephen, E., (1997). *The effect of frequent quizzes on short -and long-term academic performance*. *Journal of Dental Education*, 61(4), 339-345.

Green, A. (2007). *IELTS washback in context: Preparation for academic writing in higher education* (Vol. 25). Cambridge University Press.

Gholami, V., & Moghaddam, M. M. (2013). *The effect of weekly quizzes on students' final achievement score — International Journal of Modern Education and Computer Science*, 5(1), 36.

Harlen, W., & James, M. (1997). *Assessment and learning: differences and relationships between formative and summative assessment*. *Assessment in Education: Principles, Policy & Practice*, 4(3), 365-379.

Johnson, B. C., & Kiviniemi, M. T. (2009). *The effect of online chapter quizzes on exam performance in an undergraduate social psychology course*. *Teaching of Psychology*, 36(1), 33-37.



Kling, N., McCorkle, D., Miller, C., & Reardon, J. (2005). *The impact of testing frequency on student performance in a marketing course*. Journal of Education for Business, 81(2), 67-72.

Marcell, M. (2008). *Effectiveness of regular online quizzing in increasing class participation and preparation*. International Journal for the Scholarship of Teaching and Learning, 2(1), 7.

Messick, S. (1996). *Validity and washback in language testing*. Language Testing, 13(3), 241-256.

Newfields, T. (2005). *Voices in the Field: An interview with Yoshinori Watanabe*. SHIKEN: JALT Testing & Evaluation SIG Newsletter, 9(1), 5-7.

Roediger III, H. L., & Karpicke, J. D. (2006). *Test-enhanced learning: Taking*

memory tests improves long-term retention. Psychological science, 17(3), 249-255.

Smith, M. E., Zsidisin, G. A., & Adams, L. L. (2005). *An agency theory perspective on student performance evaluation*. Decision Sciences Journal of Innovative Education, 3(1), 29-46.

Weaver, R. R., & Qi, J. (2005). *Classroom organization and participation: College students' perceptions*. The Journal of Higher Education, 76(5), 570-601.

Zarei, G. R. (2008). *The effect of constructivist language teaching/learning on students' conceptions of L2 reading*. Iranian Journal of Language Studies (IJLS), 2(2), 281-298.