

The Washback of Formative Assessment on Summative Assessment: Students' Perception

KANAR ZIRAK HASEEB CHICHO

English Language Teaching Department, Faculty of Education,
Tishk International University, Erbil, Iraq.

MOWAFAQ MUHAMMED AL-KASSAB

Mathematics Education Department, Faculty of Education,
Tishk International University, Erbil, Iraq.

SOMA HASSAN HUSSEIN

English Language Teaching Department, Faculty of Education,
Tishk International University, Erbil, Iraq.

Abstract

Exams are a crucial component of the educational system. The significance of exams must be considered by stakeholders. As exams are a part of assessment and evaluation, and they impact the study program, students, and teachers. Testing and evaluation's impact on students and teachers is known as washback. In many regions Exams are tools for deciding on students passing from one grade to another. Therefore, students study to pass examinations rather than learn the course goals because of the negative washback from exams. Positive washback, on the other hand, encourages pupils to study harder and learn more. This study explores students' perceptions regarding the washback of the midterm exam (formative) on final exams (summative). A quantitative methodology was chosen to fulfil the aims of this study by distributing an online survey. With the responses from 108 participants the impact of these variables gender, age, and department were the focus of this paper. The findings showed that there are no significant differences between these variables gender, age, and department. In brief, students will always have the same washback positive or negative no matter no matter their ages, gender and department. Thus, this study suggests that the student's perception on exam washback should be considered rather their age, gender or department.

Keywords: Assessment, Test Paper, Formative Assessment, Summative Assessment.

Introduction

Assessment is one of the most crucial parts of the academic life of teachers and students. It deals with the outcome of education. Brown (2004) indicates that teachers continuously assess the students, intentionally or unintentionally, by answering questions, commenting on the student's work, and restructuring teaching styles by trying new solutions for problems. Thus, educators must focus on how, when, and why they assess their students and what they do with the results. In general, assessment has two main types: formative and summative as defined by (Russell & Airasian, 2012). The writers defined both types of assessment as the first type is formative assessment which refers to forming and constructing ongoing classroom knowledge such as portfolios, quizzes class presentations, classroom discussions and seminars.

On the other hand, summative assessment is implemented to evaluate a final product through tests, final exams, projects, or term papers (Russell & Airasian, 2012). Summative assessment provides significant information to the teachers regarding each student's achievement. In this way, teachers have a starting point for the coming year as a placement test or any changes needed in the syllabus (Hidri, 2020).

This study is an extension of the research problem in Chicho & Hassan (2022). They investigated students' feelings about final exams and the type of washback experienced following midterm exams. Bailey (1996) introduced the concept of washback, highlighting its effect on students' psychology, which includes academic performance but also encompasses broader emotional and motivational aspects. The findings indicated a positive feeling towards final exams and a positive washback from formative to summative assessments. They referred their positive feeling to the fact that they are used to the style and atmosphere of the exams. Human's perceptions change according to

their age, gender and the place they study. Due to this reason this study further sought to determine whether the participant's age, gender, or department affected formative assessment washback by using the following research questions:

Is there a significant difference between the participants' genders regarding the midterm examinations' effect on the finals?

Is there a significant difference between the participants' ages regarding the midterm examinations' effect on the finals?

Is there a significant difference between the participants' departments regarding the midterm examinations' effect on the finals?

These variables are crucial to be investigated because these variables can provide data about students' point of view. Age affects physiological changes, gender affects psychological well-being, and departments can affect students' background on specific topics (Watters & Rourke. 200; Baibergenova, Thabane, Danesh, Levine, Gafni, Moineddin, Pulcins. 2005).

Literature Review

Assessment

According to (Kizi, 2022), assessment is a significant part of education that allows teachers to discover the student's skills and progress. Teachers and students can collaborate to create a more effective learning environment by collectively deciding on the educational goals and objectives. During the assessment, the students can be aware of and regulate their weaknesses and strengths. Moreover, assessment assists the teachers in managing the student's development and improvement. In other words, the assessment shows how education functions during teaching and learning (Luitel, 2022). For instance, summative and formative assessments may confirm that teachers and students have the same educational goals. Summative assessment, as a certification tool, plays a crucial role in achieving educational aims such as professional development, certification, promotion, and declaration of higher education levels, and formative assessment helps the students to take the adventure of learning.

Furthermore, one advantage of the summative assessment is that the teacher indicates the student's performance, which is vital for educational aims such as professional development, certification, promotion, and declaration of higher levels of education. This type of assessment has advantages for job marketing as well. For example, regarding achieving the learning outcomes summative exam results takes priority for decision making. the majority of marks (in Kurdistan-Iraq %60 of total mark) are dedicated to the final summative assessment, that is the major decision maker for passing a grade. The primary tool of the assessment is tests. Merriam-Webster's Collegiate Dictionary (1999) defines tests as a set of questions, inquiries, and tasks that are conducted to appraise a group of student's aptitude, capacity, intelligence, knowledge, or skills. Educators use these tests for teaching and evaluation tools. Different types of tests have different applications and objectives. Tests facilitate the teacher's duty to determine the learner's ability. A significant relationship exists among the types of tests; by analyzing their relationship, teachers can diagnose the class and students' needs (Hilliard III, 2000). Due to the significant of tests there are some criteria to consider while designing a test, such as validity and reliability (Carmines & Zeller, 1979). The validity of a test means how logical and reasonable the test is. Similarly, reliability belongs to one of the principles of designing a test, and the reliability of a test means constructing trustworthy questions.

Washback

In literature, different terms have been used for washback, such as "backwash" and "test impact" (Cheng, 2005). Bailey (1996) defines washback as the impact of the exams on students' psychology, including the student's performance. According to the test impact, two types of washback exist: one is positive washback, and the other one is negative washback (Cheng & Curtis, 2004). As the matter of fact, the role of washback in teaching and learning is sufficient. It is significant for both teaching and learning. The teachers gain opportunities to make changes according to the washback students encounter, and in this way, the educators can develop their teaching skills.

Educators need to take washback into their concentration; therefore, washback shows the relationship between assessment and teaching and learning. 11 empirical studies on washback have been reviewed by Spratt (2005), and he also analyzed the factors that affect the type and the range of

washback. In regards of the studies the researcher reviewed he came to the following conclusions about washback of the exams, if the students feel well and more motivated after the test or the exam, the washback is positive. If the students feel down and demotivated, this indicates a negative washback. Therefore, the positive washback is the target of every teacher. Positive washback supports a well-constructed teaching and learning process. To conclude washback the power of creating a strong relation between education and assessment. it can be said that they are two sides of one coin.

Additionally, a well - constructed exam paper provides a positive washback (Taylor, 2005). According to Taylor (2005), negative washback can happen if a test is created only to measure a specific set of linguistic abilities or if it limits the learning environment. Furthermore, negative washback occurs if a test is constructed in a narrow format or a narrow content. If only multiple-choice exam questions will measure a writing test, then there is "great pressure to practice such items rather than to practice the skill of writing itself" (Davies et al. ,1999: 225). Another reason for negative washback is implementing traditional ways of examining the students' knowledge of a subject (Bailey, 1996). For example, paper and pencil exams may cause negative washback due to the fact that students have been involved with this type of exam consistently. As a result, educators can employ technology to design gamified tests that make students feel more at ease during assessment. By considering washback of exams the teachers can be more mindful in terms of designing their exam papers. for instance, when they decide on the number of the questions in the exam paper they will try to keep students understanding and time restriction in mind. They also become more cooperative with the students the washback and feedbacks make the teacher students stronger. To conclude, educators need to consider the students perception and psychology when designing exam questions by designing a well-constructed exam question. These questions need to be SMART (Specific, Measurable, Achievable, Relevant, and Time-Bound).

Methodology

The washback of formative assessment at Tishk International University- Faculty of Education was investigated through this study. A quantitative research study was implemented to achieve a higher reliability rating by collecting a large amount of data from the population (Dornyei, 2007). As (Rashid & Sipahi, 2021) pointed out in their research paper that the results of qualitative research are more representative and generalizable, and it also aids in identifying connections and trends within the data. Additionally, an adequate number of samples is used, allowing for the establishment of prediction. The researchers created the questionnaire through Google Forms. One hundred and eight had responded. The questionnaire comprised three demographic questions and nine close-ended and rating questions (see Appendix 1). The population of the present study was first-grade students from the Faculty of Education at Tishk International University. The Faculty of Education has five departments. The departments are Biology Education, Physics Education, Mathematics Education, Computer Education, and English Language Teaching. Thus, of the 108 students who participated in this study, with 13% were from Physics Education Department, and 31% from English Language Teaching Department (see table 1). Male participants were 39% of the students from all departments, and 61% were female (see table 2). The participants' age ranges revealed that 48% were aged between 16 and 19 years old, and 7% were aged between 24 and 26 years old (see table 3).

Table 1: Participants departments

Department	Percentage	Number
ELT	31	33
Biology Education	22	24
Computer Education	20	22
Physics Education	13	14

Mathematic Education	14	15
Total	100	108
Gender	Percentage	Number
Male	39	42
female	61	66
Total	100	108

Table 2: participant's gender

Table 3: Participants age

Age	Percentage	Number
16-19	48	52
20-23	45	49
24-26	7	7
total	100	108

The data analysis process contained three types of test analysis: one-sample t-test, two-sample t-test, and one-way ANOVA. The one and two sample T-test are "used to compare differences between two independent groups when the dependent variable is either ordinal or continuous, but normally distributed" (Leard Statistics, n.d., para.1). ANOVA test is "is a rank-based nonparametric test that can be used to determine if there are statistically significant differences between two or more groups of an independent variable on a continuous or ordinal dependent variable" (Leard Statistics, n.d., para.1). The tests were administered to analyze the effect of gender, age, and department on students' test washback. This study is limited to only Tishk International University and it cannot be generalized to the whole community of Kurdistan- Iraq.

Results and Discussion

The questionnaire was designed to collect information on the student's perceptions, and the results were analyzed using Minitab software. The data was categorized into three types based on the questions: the first was characterized by positive questions, the second by negative questions, and the third by a combination of positive and negative questions. This technique was implemented to ensure the reliability of the questionnaire by asking the same questions in negative and positive form. The positive and negative questions were analyzed separately. This separation has been made in grounds of having both types of question to ensure the reliability of participants' answers. If the participants answered both negative and positive questions consistently engaging this means that their answers were reliable and reflects their true feeling.

Afterward, the negative and positive questions were analyzed together with both positive and negative ones. Therefore, the one-sample t-test was used to see whether the participants agreed or disagreed with these questions.

Table 1: One sample t-test

Type of questions	Mean	Standard Deviation	T- value	P- value
Positive	12.583	1.492	4.06*	0.000*
Negative	12.602	2.242	2.79*	0.006*

Table (1) shows that the participants agree with these questions. A Two-sample t-test was carried out to see if there was any significant difference between the participants according to positive and negative questions (Table 2).

Table 2: Two sample t-test

Type of questions	Mean	Standard Deviation	T- value	P- value
Positive	12.58	1.49	0.07	0.943
Negative	12.6	2.24		

Table (2) shows the non-significant difference between the positive and negative questions. The researchers separated the negative and positive questions to see if there was a significant difference between genders in both positive and negative questions. The participants' tendency for each question is illustrated in Table (3).

Table 3: Number of times occurs for each question according to choices.

Questions	Choice	Count	Percent
My participation changed to better after doing my midterm examinations.	1	2	1.85
	2	8	7.41
	3	25	23.15
	4	62	57.41
	5	11	10.19
I have a better feeling about my final after midterm exams.	1	24	22.22
	2	61	56.48
	3	18	16.67
	4	4	3.70
	5	1	0.93
I feel worse about my finals after doing midterms.	1	26	24.07
	2	67	62.04
	3	8	7.41
	4	5	4.63
	5	2	1.85
I feel anxious about my finals because of midterm.	1	7	6.48
	2	28	25.93
	3	36	33.33
I feel less anxious about my finals because of midterm.	4	25	23.15
	5	12	11.11
	1	2	1.85
	2	33	30.56
	3	34	31.48
	4	35	32.41
	5	4	3.70
	1	23	21.30

The Washback of Formative Assessment on Summative Assessment: Students' Perception

How anxious are you about your finals?	2	38	35.19
	3	25	23.15
	4	19	17.59
	5	3	2.78
	1	1	0.93
I feel motivated to do better in my finals because of my midterm exams.	2	7	6.48
	3	14	12.96
	4	68	62.96
	5	18	16.67
	1	4	3.70
I feel demotivated to do better in my finals because of my midterm exams.	2	9	8.33
	3	17	15.74
	4	52	48.15
	5	26	24.07
	1	4	3.70

Table 3 shows the frequency of the choices of each question. On average, the participants chose to "agree" with the positive questions (see Appendix 1, items 1,3,6 and 7). However, for negative questions (see Appendix 1, items 2,4,5, and 8), they chose "disagree." This means they agree with all the positive questions that target the raise for feeling better, being motivated, and less anxious after their midterms. For the negative questions, the participants chose to disagree with the negative questions, which questioned their perception of feeling worse, being demotivated, and being anxious after their midterms. A two-sample t-test is used to see whether the participants' gender affects formative assessment (Table 4).

Table 4: Two sample t-tests for Gender

Type of questions		Mean	Standard Deviation	T- value	P- value
Positive	Male	12.34	1.73	1.32	0.189
	Female	12.73	1.32		
Negative	Male	12.32	1.73	1.03	0.304
	Female	12.78	1.32		
Both (Positive, and Negative)	Male	23.93	3.17	0.07	0.943
	Female	23.97	2.95		

Table (4) shows no significant difference between the gender of the participants and the effect of formative assessment statistically. Fan, Ji, and Song (2014) found that gender does not affect the test washback, whereas Dong, Fan and Xu (2021) found a significant difference between students' gender regarding test washback. The researchers suggest that their findings might be due to the fact that both genders are equally going through the same examination experience, and they share their thoughts. Sharing feelings might create a community for the students, which will make them have less and less

The Washback of Formative Assessment on Summative Assessment: Students' Perception

negative washback. Another variable that was investigated is the age of the participants. The suitable test for this investigation is the ANOVA test to see the effect of the ages of the participants. Table (5) presents the ANOVA test of ages for the three types of variables.

Table 5: ANOVA according to Ages

	Source	DF	SS	MS	F	P
Positive	Factor	2	0.01	0.005	0.002	0.999
	Error	105	238.24	2.7		
	Total	107	238.25			
Negative	Factor	2	13.09		6.55	1.31
	Error	105	524.79	5.00		
	Total	107	537.88			
Both (Positive, and Negative)	Factor	2	15.59		7.79	0.85
	Error	105	961.18	9.15		
	Total	107	976.77			

The P-values of Table (5) show scores higher than 0.05. These scores mean no significant difference between the ages of the participants for the three types of questions in the questionnaire. Şenel and Tütüniş (2011) concluded that there are differences according to the age of the students and test washback. The researchers refer this result to the fact that the participants' ages are very similar. Almost half of the participants are from 16-19 years of age, and the other half are between 20 and 23. Due to these similarities, no significant differences between students' ages and perceptions can be observed. Table (6) gives the findings of the ANOVA test to see if the departments affect students' perceptions.

Table 6: ANOVA according to departments

	Source	DF	SS	MS	F	P
Positive	Factor	4	11.49	2.87	1.31	0.273
	Error	103	226.76	2.20		
	Total	107	238.25			
Negative	Factor	4	13.17		3.29	0.65
	Error	103	524.71	5.09		
	Total	107	537.88			

The Washback of Formative Assessment on Summative Assessment: Students' Perception

Both (Positive, and Negative)	Factor	4	53.62	13.40	1.50	0.209
	Error	103	923.15	8.96		
	Total	107	976.77			

Table 6 gives the findings from the ANOVA test; the findings show that the departments do not have any significant difference regarding their perception of final exams. Positive and negative questions were analyzed. No significant differences were detected regarding the kind of questions in the questionnaire, as shown in Table (6).

Conclusion

The study examined the washback of formative assessments on summative assessments based on a participant's age, gender, and department. It was found that there are no statistically significant differences among the mentioned variables and the washback of those types of assessment. This study did not measure the students' performance in their midterm exams and their effect on final exams; it only measured the students' perception. This study contributes to the field of assessment by opening the door for further studies in Iraq. With this study, educators will be more aware of the connection between the types of assessment and washback. The researchers recommend conducting further studies on this topic and the student's performance, not their perception.

References

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

Brown, D. H. (2004). *Language assessment: Principles and classroom practices* (2nd ed). London. Longman.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Sage publications.

Cheng, L. (2005). *Changing language teaching through language testing: A washback study* (Vol. 21). Cambridge university press.

Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning (pp. 25-40). Routledge.

Davies, A., A. Brown, C. Elder, K. Hill, T. Lumley, and T. McNamara. 1999. *Dictionary of Language Testing, Studies in Language Testing, Vol. 7*. Cambridge: UCLES/Cambridge University Press.

Dong, M., Fan, J., & Xu, J. (2021). Differential washback effects of a high-stakes test on students' English learning process: Evidence from a large-scale stratified survey in China. *Asia Pacific Journal of Education*, 1-18.

Dornyei, Z. (2007). *Research Methods in Applied Linguistics*. Oxford: Oxford University Press. Fan, J., Ji, P., & Song, X. (2014). Washback of university-based English language tests on students' learning: A case study. *The Asian Journal of Applied Linguistics*, 1(2), 178-91.

Hidri, S. (Ed.). (2020). *Changing language assessment: new dimensions, new challenges*. London. Springer Nature.

Hilliard III, A. G. (2000). Excellence in education versus high stakes standardized testing. *Journal of Teacher Education*, 51(4), 293-304.

Kizi, G. M. G. (2022). Developing diagnostic assessment, assessment for learning and assessment of learning competence via task-based language teaching. *Academicia Globe: Inderscience Research*, 3(04), 34-38.

The Washback of Formative Assessment on Summative Assessment: Students' Perception

- Luitel, S. (2022). Classroom Assessment Practice of Higher Education: A Case of Tribhuvan University. *Interdisciplinary Research in Education*, 7(1), 49-58.
- Merriam-Webster's Collegiate Dictionary (10th ed.). (1999). Merriam-Webster Incorporated. Russell, M. K.s &, Airasian, P. W. (2012). *Classroom assessment: Concepts and applications*. New York. McGraw-Hill.
- Şenel, E., & Tütüniş, B. (2011). The Washback Effect of Testing on Students' learning In EFL Writing Classes. *Dil Dergisi*, (153), 43-52.
- Taylor, L. (2005). Washback and impact. *ELT Journal*, 59(2), 154-155.
- Zirak Haseeb Chicho, K., & Hassan Hussein, S. (2022). The Washback of Midterm Examination on First-Year Students' Perception Regarding the Final Exam. *International Journal of Social Sciences & Educational Studies*, 9(2).
- Leard Statistics. (10-may-2022). T- test using SPSS statistics. <https://statistics.laerd.com/spss-tutorials/kruskal-wallis-h-test-using-spss-statistics.php>
- Leard Statistics. (10-may-2022). ANOVA test using SPSS statistics. <https://statistics.laerd.com/spss-tutorials/mann-whitney-u-test-using-spss-statistics.php>
- Md., Harun, Rashid., Esra, Sipahi. (2021). The importance of quantitative research in language testing and assessment: in the context of social works. 5:317-330. doi:10.21744/LINGCURE.V5NS1.1413
- Akerke, Baibergenova., Lehana, Thabane., Lehana, Thabane., Noori, Akhtar-Danesh., Mitchell, Levine., Mitchell, Levine., Amiram, Gafni., Rahim, Moineddin., Indra, Pulcins. (2005). Effect of Gender, Age, and Severity of Asthma Attack on Patterns of Emergency Department Visits due to Asthma by Month and Day of the Week. *European Journal of Epidemiology*, 20(11):947-<https://doi.org/10.1007/S10654-005-3635-6>
- James, M., Watters., Keith, O'Rourke. (2001). 16 – Effects of Age and Gender. *Surgical Research*. 167-<https://doi.org/10.1016/B978-012655330-7/50018-6>