CSCanada
Canadian Social Science
Vol. 16, No. 11, 2020, pp. 1-6
DOI:10.3968/11905

ISSN 1712-8056[Print]
ISSN 1923-6697[Online]
www.cscanada.net
www.cscanada.org

Backwash in Higher Education: Calibrating Assessment and Swinging the Pendulum From Summative Assessment

Abdallah GHAICHA, Ed.D[a],*; Youssef OUFELA[b]

[a] Associate Professor. Lecturer in Applied Linguistics & TEFL, Educational Assessment, Evaluation, and Policy. Ibn Zohr University, Agadir, Morocco.
[b] MA in Linguistics & Applied Language Studies , IZU, Agadir, Morocco.
*Corresponding author.

Received 6 September 2020; accepted 2 November 2020
Published online 26 November 2020

Abstract

Now more than ever, there exists a plethora of empirical evidence to uphold that examinations used in educational institutions have a backwash effect, a well-recognized phenomenon among applied linguists, educators and teachers, which is the effect of test on teaching and learning (Alderson & Wall, 1993; Bailey, 1999; Messick, 1996; Widen et al., 1997; Hughes, 2003; Yi-Ching, 2009). This article essentially targets this phenomenon in Moroccan higher education. It seeks to provide a concise theoretical framework to render the reader *au fait* with such an unfamiliar term. It aims at examining the extent to which higher education assessments affect EFL students' academic achievements through sketching examples from the summative assessment practices used by faculty instructors at Ibn Zohr University, Agadir, Morocco. It also aims at suggesting some pedagogical implications to harness teaching and learning in Moroccan higher education.

Key words:

Backwash (washback); Summative assessment; Higher education

Ghaicha, A., & Oufela, Y. (2020). Backwash in Higher Education: Calibrating Assessment and Swinging the Pendulum From Summative Assessment. *Canadian Social Science, 16*(11), 1-6. Available from: http://www.cscanada.net/index.php/css/article/view/11905 DOI: http://dx.doi.org/10.3968/11905

INTRODUCTION

Nowadays, there exists an ample amount of evidence to support the fact that public examinations such as tests that are used by educational institutions have a remarkable impact within the educational sphere; students, teachers and society at large (Alderson & Wall, 1993; Bailey, 1999; Messick, 1996; Hughes, 2003; Widen et al., 1997; Yi-Ching, 2009). Test or examination impact is referred to as "backwash" (Hughes, 1989), and it has, recently, sparked the interest of so many applied linguists, educators, policy architects, and test developers. For quite a long time, the impact of examinations on teaching and learning has been well-documented in research. Popham (1987) introduced the term "measurement-driven instruction" to demonstrate the relationship between instruction and assessment, and how the last exerts an influence on the former. Latham (1988) characterized examinations as an "encroaching power" that was influencing education. Shohamy (1993) labeled this effect as the "power of the test".

Taking a bird's eye at the assessment practices undertaken by Ibn Zohr University (IZU) EFL instructors, it is very conspicuous that there is a radical emphasis on summative assessments at the expense of the supportive objective of formative assessment. In effect, end of term summative assessments administered to EFL learners serve the purpose of communicating students learning via grades provision, which renders students' achievements, learning and motivation more problematic.

Though backwash effect is assumed either positive or negative (Hughes, 2003; Yi-Ching, 2009), the current reflection focuses primarily on the negative aspects of higher education examinations. It uses current research data selectively to indicate the flaws and the invalidity of summative assessment practices undertaken by the institution. This is motivated by the desire to bring the issue to light, and to demonstrate its subtly deleterious effect on students' learning, instruction, motivation,

quality education and decision making at large. In this way, educators and policy makers can be informed of the daunting reality of the assessment system that prevails in higher education context.

It is worth mentioning that this article is not based on any empirical study or field research; it is rather a personal reflection that aims at discussing the backwash effect of examination at IZU through betokening that the great emphasis that is placed on summative assessment is lethal in variety of respects. Yet, before conferring the phenomenon, it is ethically mandatory to claim that along the article, samples of ill-practices will be drawn from the institution's assessment practices that are by no means generalizable to all teachers, but unhealthy anomalies that need a total upheaval and moratorium.

DEFINITION OF BACKWASH

The issue of Backwash has been an important interest for many applied linguists and educators. Anderson and Wall's study: "Does Backwash Exist?" is, perhaps, the most prominent among all other studies in that it illustrates the delineated connection between testing, teaching and learning (Alderson & Wall, 1993). According to Alderson & Wall (1993, as cited in Yi-Ching, 2009, p. 258), "Backwash compels teachers and learners to do things they would not necessarily otherwise do because of the test". This implies that one indirect potential effect of testing is that it forces students to *learn to the test* and teachers to *teach to the test*. Although this might sound to contradict the pedagogical values of teaching as it renders instruction more selective in content, abilities and skills, it actually exists (Bachman & Plamer, 1996). This leads to counterproductive impact on quality instruction and assessment. In accountability programs, the overuse of high-stakes testing is well-known, though these tests usually exert a negative influence on students' learning and teachers' teaching practices. In terms that are bald and clear, teachers in such programs are hold accountable for students learning. Tests' results are used to examine how much students are learning and how well teachers are teaching. Under such pressure, teachers, more often than not, prepare students for such exams through exposing them to examples of tests and also to strategies and techniques to answer questions and problems that are likely to be on the test.

The definition provided by Wall and Anderson (1993) is more restricted to high-stakes testing situations. Two definitions are considered for the purpose of this reflection: Bachman and Palmer (1996) and Pearson (1998). Bachman and Palmer (1996) assert that "Backwash is the direct impact of testing on individuals and it is widely assumed to exist" (p. 30). On the other hand, Pearson (1998, as cited in Yi-Ching, 2009, p. 258) claims that "public examinations influence the attitudes, behaviors and motivation of teachers, learners and parents, and because examinations often come at the end of the course, this influence is seen working in a backward direction, hence the term backwash". The proposition of each definition will be elucidated in the next section.

Potential Backwash at Ibn Zohr University

The backwash effect is apparent through the assessment policy that is adopted by the institution, and which places a radical *emphasis* on *summative examinations*. It is strongly observed that such tests lead to *problems of validity and reliability*, which put the assessment system at disadvantage. Worse than that is the thoroughgoing *emphasis on multiple choice examinations* pokes examinees and pedagogues' discomfort. These three undermining issues will be considered in detail, in what remains from the discussion.

Obnoxious Emphasis on Summative Tests

The assessment policy at IZU, like the remaining Moroccan ones, overuses and abuses summative assessment. The latter, and if it is used by many institutions, has been criticized for a multitude of considerations. There is now some evidence to show that summative assessment becomes precarious when it is solely used to gauge students' learning and achievements (Black & Wiliams, 1998; Black, 1999; Black et al. 2004). Summative assessment does not support learning (Black & Williams, 1998) as is the case at IZU. Tests do not provide students with a meaningful supportive feedback. When students fail a summative test at the University, they are expected to take amake-up exam. However, they do not receive feedback on their performance to know their weaknesses and their deficiencies, which is likely to affect negatively their performance on the make-up test.

One summative assessment at the end of the semester results in problems of inaccuracy and invalidity (Black, 2000). One exam cannot always measure what students can do and know. There are cases where some students are diligent, attend all the courses, take part in the class participation and discussion, and keep track with the teacher throughout the whole semester; yet, at the end of the semester they flunk, due to psychological, physiological or social constraints. Therefore, it is harmful, unfair, unethical and unaccountable to students to use one single formal examination throughout the whole semester to measure their achievement especially that the resultsare used to make some important decisions about students.

Summative assessment affects students' motivation for learning (Harlen & Deakin, 2002; Black & Williams, 1998). Motivation for learning is defined as "the will to learn and the desire to maintain this will" (Johnston, 1996 as cited in Harlen & Deakin, 2002, p. 11). The will to learn is related to the degree to which the student is ready

to invest some effort into the process of learning. It is safe to assume that all students come to the university with this will to learn, however, this "will" is sometimes affected by the assessment they undertakeatuniversity. Black & Williams (1998, p. 144)state that

> If students are given only grades of marks, they do not benefit from feedback. The worst scenario is one in which some students who get low marks this time, also got low mark last time and come to expect to get low marks next time. This cycle of repeated failure becomes part of shared belief between such students and their teachers.

Black & Williams (1998) allude to the fact that testing might remarkably affect students' will to learn if students fail more than one time; a prevailing scenario that occurs at IZU where somestudents change their attitudes towards education;others drop out, when they fail the test more than once. This corroborates the insights driven by the attribution theory which explains that people might attribute personal achievement to different factors. In learning, students might recount their achievement and failure to either ability or effort. There are students who succeed and attribute their success to their ability or effort; others fail and relate their failure to effort. Such students are likely to deal with failure positively as long as they perceive effort as an unstable and controllable factor. However, when students fail and attribute their failure to their ability, which they perceive stable and uncontrollable, summative assessment becomes precarious on students' will and motivation for learning. The students may be no longer motivated, because the test provided some disconfirmation about their perception of their language ability (Harlen & Deakin, 2002).

IZU Exam Construct Validity and Reliability At Stake

According to Bachman and Palmer (1996), validity and reliability are two of the most important qualities of test usefulness. Validity refers to the extent to which the interpretations made upon the test score are appropriate and meaningful (p. 21). However, one cannot make such interpretations as long as the construct to be measured is not defined. Construct validity is defined as the extent to which the test measures what is designed to measure. According to Messick (1989, as cited in Jawhar, 2010, p. 90):

> One of the main elements that put any test's validity at a high risk is construct underrepresentation. According to him, this underrepresentation occurs when the test is constructed in such a way that does not include important dimensions of facets of the constructs.

When one recollects some of the tests used by some teachers at the department of English Studies of IZU, s/he might notice that they are not relatively valid, because they do not accurately measure what they are supposed to measure. For example, in some subjects such as *"Media and Cultural Studies Module"*, the objective is to have students acquire a conceptual as well as topical knowledge of the subject subsuming concepts, theories, and models of Media and Culture in the West including the USA and the UK together with other English speaking countries. Given this, one might easily expect that for the test to be valid, it should measure these constructs. However, some teachers opt for oral examination; a choice which is not justified at all. A worth asking question is what is it that they measure? Is it the topical knowledge of students or oral skills, or both? Speaking of validity, a valid test of '*Media and Cultural Studies Module'* must measure only the topical knowledge of students as well as their potential to act authentically in any of the English speaking countries, or assess the two constructs without taking into account the speaking channel.

Another convincing example has to do with the *"Spoken English Module"*. The basic objective of such a course, as recorded in the Pedagogical Descriptive File that is accredited by the Ministry of Higher Education, Staff Training, and Scientific Research, is to have students develop a practical understanding of phonetics, improve their pronunciation through raising their phonic and phonemic awareness with the assumption that breaching code of appropriateness affect the mutual intelligibility. If teachers teach students how to pronounce sounds in the English language, they are supposed to use the oral test, because the construct to be measured dictates such a choice. However, there are actually cases where some teachers assess this construct through written examination. This foregrounds unethical respect of accountable assessment measures, tools and techniques.

The Overuse and Abuse of Multiple Choice Tests

In the last five years, multiple choice testing has become the most common way of assessing students at IZU, and the number is increasing by the year. The fact that the university is overcrowded as it has been welcoming a substantial number of students each year seems to be one of the strongest arguments that justify its use. Besides this, multiple choice testing is considered an objective tool and can be used to measure a variety of learning points. However, several problems are associated with its use.

Multiple-choice items exams are claimed not provide an accurate measurement of students' knowledge (Roediger & Marsh, 2005). Despite students high rate answers to a certain multiple-choice test,their scores do not reflect a fairly mastery of the knowledge and content they have learnt. When a student answers a particular multiple-choice test and gets 14/20, one cannot confirm that the student's grade is the result of what s/he knows. There exist two hypotheses to explain that. The first hypothesis, the student was able to answer all the 14 questions correctly. The second hypothesis is that the student was able to answer some of the questions and guessed what remained.

Additionally, multiple choice testing does not measure students' ability to synthesize and evaluate information, apply knowledge and solve complex problems (kuo & Hirshman, 1996). For example, a teacher who teaches linguistics cannot measure students' ability to synthesize the historical development of linguistics or talking over the difference between the descriptive the prescriptive approach using MCT. However, some teachers at IZUabuse MCT with an assessment use argument (AUA) labeled "the classroom is over-crowded"

The continuous use of multiple choice testing endangers students' writing. Even though this is apparent, it is actually being practiced at the university. Some students from the first semester are exposed to MCT and some teachers are trying to figure out why such students have poor writing skills. The predominance of multiple choice testing to the extent to that it is used to measure students' writing skills, has certainly led to a negative backwash. Real life situations or other TLU domains do not involve multiple choice, and the fact that its easiness in administering, scoring and attaining objectivity does not justify its overuse at all. Unfortunately, at IZU, multiple choice testing has been twisted to uses that seem quite inappropriate. When its use invades productive skills such as: writing and speaking, its efficiency in translating students' knowledge and academic progress is minimized to critical levels.

Notwithstanding, MCT leads to false knowledge. (Roediger & Marsh, 2005) conducted a research on the positive and the negative effects of MCT and concluded that it may unintentionally lead to the creation of false knowledge. They explained that exposing students to wrong answers, there is a huge possibility that the statement will be judged true later.

Implications for Language Teaching and Learning

There exists a number of tips to adopt in order to reduce the negative backwash and promote a backwash of a positive nature: teacher training, implementation of formative assessment, and designing useful tests.

TEACHER TRAINING

Integrating a higher education institution should not be granted by holding doctoral degrees as a solitary criterion. Substantial training in the pedagogy of EFL together with assessment literacy and ethics of teaching, assessment and research should be among the compulsory requirements to satisfy by any candidate to be hired for a teaching position in higher education institutions. It is quite rudimentary to train those teachers, to develop their assessment literacy, to provide them some workshops and short-term training that underlies all the important theoretical constructs that define assessment tasks and their design before undertaking any professional work; especially assessing student learning. Yet, for high levels of accountability, assessments, either midterms or end of term examinations, should be delegated to a committee of internal professionals with high credentials and professional expertise in assessment and evaluation.

Implementing Formative Assessment

The effectiveness of formative assessment now is not questionable at all. There is a plethora of evidence to uphold the statement "formative assessment can improve learning and raise standards". Paul and William (1998) reviewed almost 580 study from several countries and concluded that Formative assessment is useful to learning and teaching.

Designing Useful Tests

In this regard, it is important for teachers to draw on Bachman and Palmer's (1996) model of test usefulness that compromises six test qualities: validity, reliability, interactiveness, authenticity, backwash and practicality. The property of authenticity is very important (Bachman an Palmer, 1996). It is very necessary for the test to be as relatively authentic as possible to the test user to make meaningful inferences about students' learning and achievements.

Teaching and learning ⟹ assessment ⟹TLU domain

Chiefly, exams are used to determine the extent to which students have achieved the objective; however the teaching and learning of a particular subject take place because it is assumed that students will need it a particular context outside the test. TLU domain is the situation in which students are expected to perform particular skills outside the exam itself (Bachman an Palmer, 1996). Therefore, it is quite necessary to develop tests that take into account the target language context, especially for language skills. It is the TLU domain that actually determines the nature and the type of the test and the context to which teaching and learning take place and not other variables such as: the number of students.

Conducting Further Research

Science now came to the understanding that the way people behave is directly affected by what they believe (Ghaicha 2008). In this vein, and to this effect, it is highly recommended that more research be conducted to disclose Moroccan teachers' beliefs and conceptions of assessment and how they impact their real practice of assessment. The research should be motivated by questions such as:

What do Moroccan teachers know about assessment?

To what extent their beliefs exert an influence on their pedagogical practices?

CONCLUSION

By large, some of the ultimate objectives of this article are to voice the backwash effect of the examination system being undertaken by IZU, to render teachers *au fait* with their direct and indirect effects on learning, and to provide some implications for amplifying teaching and learning in Moroccan higher education.

It is very apparent that backwash exists and the slightest mistake in the process of designing, administering or taking exams might be a huge consideration and cause radical problems. Exams are ways by which users come to know the extent to which examinees have reached the standards and the expectations. Exams also calculate the value of students' ability to perform in TLU domain. In this respect, it is very rudimentary for IZU teachers to design assessments that relatively correspond to students' TLU domain, and make sure they are administered in the appropriate environment to maximize their reliability. It is worth-mentioning that there is no room for the possibility that the aforementioned statements are generalized to all teachers. Yet, the focus on summative test and MCT is radical which call for urgent interventions and appropriate calibrations.

REFERENCES

Amrein, A. L., & Berliner, D. C. (2000). The effect of high-stakes testing on students' learning and motivation. *Educational leadership*. 31-38.

Bailey, K. (1999). *Washback in language testing. TOEFL monograph series*. Princeton, NJ: Educational Testing Service.

Bachman, L. F. (1991). What does language testing have to offer? *TESOL Quarterly*, *25*(4), 671-704.

Bachman, L. & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.

Beardsley, A., & Berliner, D. (2003). The effects of high-stakes testing on student motivation and learning. *Educational Leadership*, *60*(5), 32-38.

Becker, B. (1990). Coaching for the SAT: Further synthesis and appraisal. *Review of Educational Research*, *60*, 373-417.

Black, P. (2000). Research and the development of educational research. *Oxford Review of Education*, *26*(3), 407-418.

Black, P. (1998). Formative assessment: Raising standards. *School Science Review*, *80*(291), 39-46.

Black, P., & Williams, D. (1998). Assessment and classroom learning. *Assessment in Education*, *5*(1), 7-74.

Buck, G. A., & Amy E. Trauth-Nare. (2009). Preparing teachers to make the formative assessment process integral to science teaching and learning. *Journal of Science Teacher Education*, *20*(5), 475-494.

Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language and Education*, *11*(1), 38-54.

Cheng, L. (1999). Changing assessment: Washback on teacher perspectives and actions. *Teaching and Teacher Education*, *15*(3), 253-271. doi:10.1016/ S0742-051X(98)00046-8

Cheng, L. (2003). Looking at the impact of a public examination change on secondary classroom teaching: A Hong Kong Case Study. *The Journal of Classroom Interaction*, *38*(1), 1-10.

Cheng, L. (2004). The washback effect of a public examination changes of teachers' perceptions toward their classroom teaching. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research Contexts and methods* (pp. 147-170). Manwah, New Jersey: Lawrence Erlbaum Associates.

Cheng, L., & Couture, J. C. (2000). Teachers' work in the global culture of performance. *Alberta Journal of Educational Research*, *46*(1), 65–74.

Harlen W., Deakin, C. R. (2002). A systematic review of the impact of summative assessment and tests on students' motivation for learning. *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.

Hughes, A. (1989). *Testing for language teachers*. New York: Cambridge University Press.

Jawhar, S. (2009). Summative & formative assessment: reflection on the Saudi higher education assessment system. *Enletawa Journal*, *2*, 85-98.

Kuo, T. M., & Hirshman, E. (1996). Investigations of the testing effect. *American Journal of Psychology*, *109*, 45-464.

Kuze, M. W., & Almun Shumba. (2011). An investigation into formative assessment practices of teachers in selected schools in fort beaufort in South Africa. *Journal of Social Science*, *29*(2), 159-170.

Latham, H. (1877). *On the action of examinations considered as a means of selection.* Cambridge, Deighton, Bell and Company.

Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13*(3), 241–256. doi:10.1177/026553229601300302

Popham, W. J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan, 68*, 679-682.

Rees, P. J. (1986). Do medical students learn from multiple-choice examinations? *Medical Education*, *20,* 123-125.

Roediger, H. L., & Marsh, E. J. (2005). The positive and negative consequences of multiple choice testing. *Journal of Experimental Psychology, Learning, Memory and Cognition*. *31*(5), 1155-1159. Retrieved from https://doi.org/10.1037/0278-7393.31.5.1155

Shohamy, E. (2001). *The power of tests: a critical perspective on the uses of language tests*. London: Pearson.

Weathly, L., McInch, A., Flemming, S., & Lord, R. (2003). Feeding back to feed forward: formative assessment as platform for effective learning. *Kentucky Journal of Higher Education, Policy and Practice*, *3*(2), 1-29.

Widen, M. F., O'Shea, T., & Pye, I. (1997). High-stakes Testing and the Teaching of Science. *Canadian Journal of Science*, *22*, 428-444.

Yorke, M. (2003). Formative assessment in higher education: Towards theory and the enhancement of pedagogic Practice. *Higher Education*, *45*(4), 477-501.

Yi-Ching, P. (2009). A review of washback and its pedagogical implications. *VNU Journal of Science*, *25*, 257-263.