

Prevalence and Impact of Backwash on Learning in Undergraduate Medical Students

Afshan Sumera¹, Ankur Barua², Arshad Jawed^{3*}

¹Assistant Professor, Lecturer Faculty of Medicine, International Medical University, Kuala Lumpur, Malaysia

²M.B.B.S, Lecturer Faculty of Medicine, International Medical University, Kuala Lumpur, Malaysia

³Assistant Professor/Researcher, Research & Scientific Research Unit, College of Nursing and Allied Health Sciences, Jazan University, Jazan

Abstract

Background: Backwash or washback is a common phenomenon that exists in almost all the educational institutes in the world. It directly correlates itself to teaching, learning and student assessment. It's a prevailing conception that conducting exams / tests could drive teachers to teach and students to learn (also called as measurement-driven instruction).

Methods: A questionnaire based tool was developed and named the Backwash Assessment Questionnaire (BAQ). It was validated to evaluate backwash among medical students in Malaysia. A preliminary version, consisting of 15 structured items was developed based on an exhaustive and focused group discussion with students, content experts and extensive literature reviews. Multi-stage refinement narrowed it down to 10 items after the process of testing. The final version was analyzed using reliability testing. For testing, refinement and review process, 90 students and 10 lectures were selected to assess the acceptability, time-management and reliability of this questionnaire. A fresh sample of 105 undergraduate medical students from semester three to five from the same university was used for the test of the final version.

Results: The overall internal reliability (Cronbach's alpha) of the final version of the BAQ was found to be 0.874. The test-retest reliability from kappa statistics was found to be high with $k=0.92$ with $P=0.0001^*$.

Conclusion: The study showed that the BAQ was a valid, powerful and reliable tool for measuring backwash effects on education and assessment process. We found a high prevalence of backwash among the medical students in the undergraduate course.

Keywords: BACKWASH, MEDICAL STUDENTS, ASSESSMENT, LEARNING, TEACHING

Journal of Medical Education Winter 2018; 17(1):37-43

Introduction

Backwash or washback is a common phenomenon that prevails in educational institutes all over the world. It refers to the influence of student assessment on learning and teaching. It's a common conception in the academic arena that conducting tests could

drive teachers to teach and students to learn (also called as measurement-driven instruction). A set of questions are laid out to students after a specific time interval, response to which is assessed. This assessment is evaluated, and what is evaluated becomes what is taught and should be learnt (1). Learning approaches have been divided into two broad categories according to the Deep Learning Approach (DLA) and the Surface Learning Approach (SLA) (2). The DLA is mainly focused on engagement with content or learning task and creating an understanding of the content being

*Corresponding author: Arshad Jawed, Assistant Professor, Research and Scientific Studies Unit, College of Nursing and Allied Health Sciences, Jazan University, Jazan, KSA. 45142

Phone\Fax: +966 (17)) 317 4383

Email: arshadjawed29@gmail.com

learnt (3). The learners using DLA are assumed to be self-motivated and self-directed. While the SLA is not focused on understanding but mimics or reproduces learning content in the absence of understanding. SLA shows little to no evidence of integration of new knowledge with old knowledge (4) and is instead directed on assessment, rather than knowledge build-up. Backwash, by its nature, does not necessarily brings a negative impact on learning and teaching. The key is the ways in which an assessment task functions as a constructive component rather than an end on its own in the iterative cycle of learning and teaching. The positive backwash effect can be materialized via pedagogically sound and practically feasible assessment tasks such that the way(s) teachers teach, and students learn are advantageously informed by these tasks. The positive backwash effect can even facilitate students to become lifelong learners in the pursuit of their disciplines' understanding. The factors that influence backwash are multifactorial. To modify an assessment that leads to influential learning (5), the most effective way is to have changes in the system that can transform student's learning (6-8). To measure the learning activities done in medical sciences and to acquire data, proper data collection tools are considered as an essential way. To collect data, a well-structured questionnaire is a lucid and effective approach (9). Most of the literature available on backwash is in language studies (10). To the best of our knowledge backwash has not been assessed by a quantitative questionnaire in medical students.

The aim of this study was to develop and validate a tool which can assess backwash in educational settings to improve assessment techniques based on results.

Methods

Setting and Participants

To test the reliability, acceptability and time-management of the initial version of BAQ, a

pilot study was conducted with 90 first-year medical students at medical universities in Malaysia. We randomly selected 10 lecturers from various disciplines and asked them to review the BAQ. Based on the results of the pilot study and teachers review, BAQ was refined from 15 to 10 questions. The refined/ final version (Table 1) was tested on 105 medical students from the 2nd and 3rd year from the same university. All 195 students (117 girls, 78 boys) completed the study (Questionnaire 1 and 2).

The Development of BAQ

Upon obtaining consent, a mixed method design was employed where both quantitative and qualitative data analysis was carried out. The study was based on a cross-sectional survey. A self-administered survey questionnaire (BAQ) was distributed to identify the negative influence of backwash effect. Five percent of the selected sample was selected for qualitative data collection (focus group interview session), where students were asked several open-ended questions to gauge their feelings towards the backwash effect. To validate the questionnaire, a pilot test was carried out on students from semester 1 and 2, who were not included later in the study. Students of semester 3, 4 and 5 were selected for refinement of BAQ.

We also compared backwash results with the type of learner (strategic and deep learner) and their previous academic performances. Students were divided into three groups.

Group I- consisting of semester 3 students.

Group II- consisting of semester 4 students.

Group III- consisting of semester 5 students.

Each group was further subdivided into three sub groups based on their previous exam scores and results of BA, were compared within the groups; (1) High achievers were defined to have GPA of ≥ 3.5 , (2) medium achievers with GPA between ≥ 3.0 and < 3.5 and (3) low achievers with GPA of < 3.0 . Backwash was studied more in surface learners and medium achievers than strategic and high achievers.

Inclusion criteria: Medical students of

Table 1: BAQ Final Version

Sr. No.	Statements	Cronbach's Alpha after Item deletion
1	I only concentrate on the topics which are important for assessment.	0.761
2	My only aim in learning is to pass the assessment.	0.792
3	I won't be motivated to learn, if there is no assessment.	0.792
4	I usually take cues on what content to learn from the curriculum to pass exams.	0.755
5	I am confident in learning when taking tips from senior students on assessment.	0.773
6	I am confident in learning when I focus on content areas based on previous assessment papers	0.776
7	I feel confident to concentrate on the topics or materials that the lecturers chose to focus on during my learning.	0.756
8	Studying for assessment makes me focus on the subject matter better.	0.758
9	Assessment is the highest motivation for learning to happen.	0.752
10	My learning time and amount increases because of assessment.	0.772

semester 1-5 who willingly participated in this study.

Exclusion criteria: Other disciplines and those who were not willing to participate in the study.

An initial questionnaire with 15 items rated on a six-point Likert scale was administered to assess backwash. Also, lecturers (n=10) from various disciplines reviewed the BAQ. Simultaneously, Bigg's study process questionnaire (SPQ) (2) was also administered to students to assess the type of learner.

Item Generation Phase

The authors generated a list of all items based on a thorough literature search of studies on the same topic. (1, 11, 12)

Types of Questions

Questions in the final version of the BAQ were distributed to the type of learning strategy domain of students as follows.

Domain 1- Compensation strategy for superficial learners

BAQ2, BAQ3, BAQ8, BAQ9, BAQ10

Domain 2- Memory strategy for deep learners who depend on cues and tips

BAQ4, BAQ5, BAQ6

Domain 3- Cognitive strategy for deep learners who depend on assessment and importance

BAQ1, BAQ7

Validity/Testing of the BAQ

Data were then analysed by SPSS software, version 18, for Inter-item correlation matrix (Table 2), Cronbach's Alpha (CA) and Spearman's Rank Correlation (SRC). Based on the results after item deletion, 5-items of this preliminary version were dropped and the final BAQ questionnaire was reduced to 10 items. We calculated complete item statistics (Table 2) for the new version of the BAQ and inter-item correlation was calculated (Table 3). A detailed cross-sectional study was conducted on 105 medical students belonging to the third, fourth and fifth semesters to evaluate the final version of the BAQ (Table 3). Our analysis showed that Cronbach's Alpha based on standardized items was 0.874 (Table 4) which was high and confirms the reliability of the scale. The same can be inferred from the inter-item correlation matrix as well (Table 3). Since none of the Cronbach's alpha values exceeded this in item deletion and the Spearman's Correlation coefficient for any of the two items never exceeded 0.80; all the ten items of the BAQ were retained. The test-retest reliability from kappa statistics was found to be high with $k=0.92$. The p-value was statistically significant ($p=0.0001$).

Table 2: Inter-item correlation matrix

	BAQ_1	BAQ_2	BAQ_3	BAQ_4	BAQ_5	BAQ_6	BAQ_7	BAQ_12	BAQ_13	BAQ_15
BAQ_1	1.000	0.184	0.092	0.480	0.461	0.176	0.529	0.318	0.244	0.468
BAQ_2	0.184	1.000	0.273	0.366	0.126	0.155	0.166	-0.046	0.139	0.208
BAQ_3	0.092	0.273	1.000	0.321	0.021	0.181	0.113	0.159	0.316	-0.022
BAQ_4	0.480	0.366	0.321	1.000	0.416	0.245	0.332	0.235	0.370	0.303
BAQ_5	0.461	0.126	0.021	0.416	1.000	0.279	0.319	0.275	0.185	0.157
BAQ_6	0.176	0.155	0.181	0.245	0.279	1.000	0.269	0.346	0.416	0.188
BAQ_7	0.529	0.166	0.113	0.332	0.319	0.269	1.000	0.545	0.377	0.468
BAQ_12	0.318	-0.046	0.159	0.235	0.275	0.346	0.545	1.000	0.653	0.350
BAQ_13	0.244	0.139	0.316	0.370	0.185	0.416	0.377	0.653	1.000	0.288
BAQ_15	0.468	0.208	-0.022	0.303	0.157	0.188	0.468	0.350	0.288	1.000

Table 3: Complete item statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
BAQ_1	34.88	45.385	0.541	0.477	0.761
BAQ_2	36.75	47.613	0.279	0.236	0.792
BAQ_3	36.67	47.750	0.273	0.232	0.792
BAQ_4	35.88	43.810	0.577	0.436	0.755
BAQ_5	35.42	47.472	0.399	0.337	0.776
BAQ_6	35.77	46.232	0.424	0.234	0.773
BAQ_7	35.30	45.061	0.590	0.474	0.756
BAQ_12	35.65	42.829	0.537	0.581	0.758
BAQ_13	35.94	41.059	0.578	0.535	0.752
BAQ_15	35.31	45.291	0.437	0.357	0.772

Table 4: Reliability statistics

Cronbach's Alpha Of initial BAQ	Cronbach's Alpha of final BAQ
0.787	0.874

For the development of the BAQ, three content experts were engaged who agreed and gave equal weightage (W) to individual items on the Likert Scale of the BAQ. The measure of internal reliability (IR), e.g. Cronbach Alpha and well as the Discrimination Index (DI) was directly proportional to the weightage of each response (13). Therefore, the weighted score was calculated for individual scores as shown in equation 1,

$$\text{Weightage of each score}(W) = (\text{Observed response score}) \times (DI) \quad \text{eq.1.}$$

A correction factor (CF) was developed to make amendments to the overall cut-off of the instrument, by obtaining the ratio of the total weighted score and the total raw

score. Subsequently, a correction factor was multiplied by the [mean of the individual raw score – 25th percentile of interquartile range (IQR)] for each item and added together. Cut-off level in an instrument can be elaborated and determined by the following steps.

Content experts assign equal weightage to individual items as per Likert scale questionnaire.

DI calculation for each item = SRC

Weightage of each response in each item of the questionnaire = (OIS) (W) X (DI) X (CA or IR) where OIS is Observed Item Score; CF=(Total Weighted Score) / (Total Raw Score) The instrument cut-off point, sans any gold standard=Σ [(Median of Individual Raw Score –25th Percentile of IQR) X (CI)]

Results

For analysis of the presence or absence of backwash, the students were segregated according to ethnicity, gender, type of achiever, and type of learner. Comparatively, backwash was highest among Indian students (76.2 %) followed closely by their Chinese counterparts (46.6 %). Backwash was almost equal in male and female students. High achievers had the lowest backwash ratio (25%) that increased to 46.3% in medium achievers and 58.3% in low achievers. Deep learners had 75% backwash as opposed to 38% found in surface learners.

Discussion

These ratios suggest the applicability and success of SLA and backwash in achieving scores that could have not been achieved otherwise. Students who are inherently deep learners also adopt to SLA when confronted with the pressures to perform better in the exams.

The process of examination for education or employment finds its roots almost 300 years back, based on the currently available literature (1). Examinations as we know have been used universally in almost all the disciplines, e.g., Arts, Science, Mathematics, Engineering, Medical studies, etc., all over the world. Exams were initially started to encourage and nurture an individual's or groups' talent, evaluate and upgrade the status of educational institutions, prevent favoritism, counter regional and religion based bias, curb corruption in education and provide equal opportunities to the deserving candidates (A/B). Apart from the positive implications of examination, it suffers few major drawbacks (14-17). Exams can be used as powerful tools to manipulate educational system, alter subjects and impose new text books and distort teaching methodologies. Examinations thus have become a favorite of the concerned authorities in education systems all over

the world. Examinations too manipulate the behavior of those students or teachers who are directly affected by it. Getting higher scores in exams has become a measure of a students' knowledge, sincerity, overall capability and a criterion for a bright future. Therefore, every student aspires to achieve maximum marks and do the best in exams.

When the future of an examinee depends on the outcome of an exam, student(s) use various means to clear and attain highest marks possible. Backwash or washback refers to the influence of exam(s) on learning and teaching in the currently widespread educational system (1). It is a common observation that a course curriculum, which is followed by a test, has a profound effect on the former; thereby it significantly affects the quality of teaching and learning outcomes. Medical education has not been an exception to the effects of backwash phenomenon. Almost all hospitals and associated universities/colleges have specific course curriculum that's taught all year round and at the end of which an assessment is scheduled. Students are required to secure a specific grade that's higher than the cut-offs set by the educational institute. To secure higher and higher grades, students normally lean towards one of the two learning approaches, namely DLA which focuses on engagement with the subject, content and learning, while SLA takes a shortcut to learning and attempts to mimic curriculum contents as required in the examination. While the earlier approach needs constant attention and labor, the latter proves helpful enough to successfully clear the exams. In our study, we found backwash to be a significant phenomenon amongst strategic learners (Table 5), which is quite unexpected, as strategic learners are keen on learning. The observation supports the notion that exams have profound effect on the students who are genuine but the pressure to perform in exam pushes them towards SLA. Following these students in real time may help predict the long-term effect of exams and their scores on their approach towards learning. A section of

Table 5: Factors associated with backwash

		Backwash present	Backwash absent	P value
Ethnicity	Malay	8	5	0.115
	Chinese	21	45	
	Indian	16	5	
	Others	3	3	
Gender	Male	22	31	0.440
	Female	26	27	
Type of achiever	High	3	9	0.130
	Medium	38	44	
	Low	7	5	
Type of learner	Surface learner	33	53	0.003
	Deep/strategic learner	15	5	

strategic learners may slowly switch to SLA, when continuously confronted with exams, other activities that harp on the time available to them for studies.

Backwash is significantly contributed by examiner or examination managed by authorities not present in the same institute. As this has become a gold standard and a matter of repute for the institution concerned, examinees who focus on DLA also shift their mode of study to SLA as the exam dates approach closer. Slowly but steadily, students shift more and more towards SLA and motivate peers to approach the same. The differences in grades achieved via SLA vs. DLA is not substantial or proportional to the amount of efforts and labor poured in. Teachers tend to ignore the activities or the topics that may look redundant or do not contribute directly towards passing the exam, students too skip these topics to focus more and more on the portions that are considered important and can appear in tests. The learnings consequently become narrow, disjoint and lose the charm; learning becomes focused and turns more into a memorization activity rather than understanding of the content.

In our opinion, examinations should drive the course curriculum with constant refinements and additions/deletions in the course content. The intended direction and function of examinations should be improvement in learning and course curriculum, rather than

just being creating a benchmark to be crossed by the examinees. In the current times, where a huge repository of knowledge is available to everyone at their fingertips is available with almost everyone in the form of mobile internet, awareness is a widespread phenomenon. We tend to know everything superficially, without building the concept of the same. The focus of current curriculum should build a solid foundation to the current generation of learners, enabling then for speculate and solve problems and challenges that do not exist in the textbooks.

Conclusion

Backwash severely affects the curriculum, teaching methods, and approaches towards learning. To avoid this influence, the course should start with an initial assessment and evaluation before the teaching or learning activities start. The learning activities should be designed to improve the learning gaps identified at the start of course by tests. With the help of backwash, didactic learning could be improved, which is not a new idea in education. The need of the more robust type of tests is there; a test that drives schools to achieve more striving goals and a pedagogical curriculum to promote critical thinking and less memorization of facts. Assessment should be based on performance which can be achieved by changing the prevailing attitude

that assessment and learning are only used to pass the examination and secure good grades.

Ethical Considerations

Approval for the present study was obtained from the Research and Ethics committees of International Medical University, in Malaysia. The information obtained during the data collection was strictly kept confidential. A random code was generated for each participant of this study. Before participation in this study, an informed written consent was obtained from all participants.

Conflict of Interest

The author declares no conflict of interest.

References

1. Cheng L, Curtis A. Washback or backwash: A review of the impact of testing on teaching and learning. Washback in language testing: Research contexts and methods. 2004; 27:3-17.
2. Biggs J, Kember D, Leung DYP. The revised two-factor study process questionnaire: R-SPQ-2F. Br J Educ Psychol. 2001; 71(1): 133-49. doi:10.1348/000709901158433.
3. Laird TFN, Shoup R., Kuh GD, Schwarz MJ. The effects of discipline on deep approaches to student learning and college outcomes. Res High Educ. 2008; 49(6): 469-94.
4. Trigwell K, Prosser M, Waterhouse F. Relations between teachers' approaches to teaching and students' approaches to learning. High Educ (Dordr). 1999; 37(1): 57-70.
5. Biggs J. Assessing learning quality: Reconciling institutional, staff and educational demands. Assess Eval High Educ. 1996; 21(1): 5-16. doi:10.1080/0260293960210101.
6. Elton LRB, Laurillard DM. Trends in research on student learning. Studies in Higher Education. 1979; 4(1):87-102. doi: 10.1080/03075077912331377131.
7. Crooks TJ. The impact of classroom evaluation practices on students. Review of Educational Research. 1988; 58(4): 438-81.
8. Frederiksen JR, Collins A. A systems approach to educational testing. Educational Research. 1989; 18(9):27-32.
9. Kazi AM, Khalid W. Questionnaire designing and validation. Journal of the Pakistan Medical Association. 2012; 62(5): 514.
10. Sumera A, Barua A, Navamoney A. Exploring the effect of backwash in first year medical students and comparison with their academic performances. Procedia-Social BehavSci. 2015; 174: 491-5.
11. Biggs J. Assessment and Classroom Learning: a role for summative assessment? Assessment in Education: Principles, Policy & Practice. 1998; 5(1): 103-10. doi:10.1080/0969595980050106.
12. Tiwari A, Lam D, Yuen KH, Chan R, Fung T, Chan S. Student learning in clinical nursing education: Perceptions of the relationship between assessment and learning. Nurse Educ Today. 2005; 25(4): 299-308. doi:10.1016/j.nedt.2005.01.013.
13. Barua A. Methods for decision-making in survey questionnaires based on Likert scale. Journal of Asian Scientific Research. 2013; 3(1): 35-8.
14. Cheng L. Changing language teaching through language testing: A washback study. Cambridge University Press. 2005; 21.
15. Shohamy E, Donitsa-Schmidt S, Ferman L. Test impact revisited: Washback effect over time. Language Testing. 1996; 13(3):298-317.
16. Vernon PE. The measurement of abilities. University of London Press, London. 1961.
17. Fish J. Responses to mandated standardised testing. Unpublished doctoral dissertation, University of California, Los Angeles. 1988.