

Investigating the Relationship between Proficiency Level and Perceived Test Consequences: A Case Study of Iranian High School Students

Hossein Khodabakhshzadeh¹ and Mahmood Khosravani*²

¹Assistant professor, Department of English, Torbat-e Heydarieh Branch, Islamic Azad University, Torbat-e Heydarieh, Iran

²PhD candidate, Department of English, Torbat-e Heydarieh Branch, Islamic Azad University, Torbat-e Heydarieh, Iran

Corresponding Author: Mahmood Khosravani, E-mail: sadradone@gmail.com

ARTICLE INFORMATION

Received: November 02, 2020

Accepted: December 12, 2020

Volume: 3

Issue: 12

DOI: 10.32996/ijllt.2020.3.12.8

KEYWORDS

Chi Square, Consequential
Validity, Proficiency Level, Test
Impact

ABSTRACT

Aimed at promoting the consequential validity of high-stakes test, unexpected aspects of assessment process which negatively affect the learning process have been investigated by researchers in the last decades. Since discovering those variables that contribute to adverse test consequences is necessary for any successful intervention, the current study was an attempt to shed light on the relationship between proficiency level of examinees and their perception of test impacts. To this purpose, 63 third grade male students studying for a high-stakes exam were selected through convenience sampling from three different high schools in the city of Birjand. A teacher developed proficiency test was then administered and participants were either assigned to high-proficiency group (N=34) or low-proficiency group (N=29). Then, using an interview protocol and conducting a content analysis of the interviews, students were either placed at social or psychological impact groups on the bases of the frequency of their answers to different items of the interview. The obtained data were entered into SPSS (version 21) and analyzed through Chi Square test to determine the goodness of fit. A significant interaction was found ($X^2(1) = 6.647, p = 0.01$). It was concluded that high proficiency students are more likely to perceive test consequences (54.0%) than low-proficiency group (46%). Besides, on the basis of Phi and Crammer's V symmetric measures, the strength of the association between the variable was 0.325 which shows a medium effect of proficiency level on type of perceived test consequence. These findings have implication for language teachers, language learners, researchers and educational decision makers.

1. Introduction

According to Bachman and Palmer (1996), under the umbrella term of consequential validity of large scale and high-stake test, in particular the test scores and the decisions made on the basis of these scores, a number of phenomena affect the interests of stake holders and those who are directly related to these tests which are categorized in two major consequences known as wash back and test impact. These terms (washback, impact, and consequences) are used in different fields of research, and encompass different dimensions of the research in testing discipline. In this conceptualization, washback is more frequently used to refer to the effects of tests on teaching and learning at the classroom level while impact entails the effects that a test may have on individuals, policies, or practices, within the classroom, the school, the educational system, or the society as a whole. A brief review of the literature in the last three decades shows that these terms are seen as being inherently one (mostly in traditional views), as some believe that washback is part of impact or a dimension of it, however in more recent literature the effects of testing on teaching and learning have been traditionally associated with test validity (consequential validity) where washback is considered as only one form of testing consequences that need to be weighted in evaluating validity. Since impact and washback are established disciplines of research in language testing (Cheng, et al, 2015), consequences are used a general term under which impact and washback will be discussed in our research. In other words, impact will refer to the influences of testing on individuals, groups, institutions, and society at large which is immediate and

delayed effects out of the classroom context (Chalhoub-Deville, 2009) and washback will refer to the influence of testing on teaching and learning (Wall, 1997). He shed more light on the distinction by defining impact as any of the effects (positive or negative, intended or unintended) that a test may exert on individuals, policies or practices, within the classroom, the school, the educational system or society as a whole, in contrast to washback, which is more frequently used to refer to the effects of tests on teaching and learning practices inside the classroom (p. 291).

According to Kiani, Alibakhshi and Akbari (2009), many psychological, social and ethical issues and variables might interact with test consequences. One of the variables that is directly linked with educational and evaluation processes is the proficiency level of the test takers. Thus, any educational or policy making aimed at mitigating negative side of test consequences is required to take learning variables, such as proficiency level, into account. Shedding light on any possible relationship between proficiency levels of test takers and different dimensions of test consequences might help the decision makers and language teachers to select the most appropriate intervention. Thus, the current research is an attempt to investigate the goodness of fit between proficiency levels (high and low proficiency) and two major dimensions of test consequences (social and psychological). To this aim the following research question was formulated:

Is there any significant relationship between proficiency level of Iranian high school language learners and their perceived test consequences?

2. Literature Review

The history of investigating the effects of tests dates back to 1950s and 1960s when practitioners and scholars in general education started asserting that exams (high-stakes tests) may exert influence of some kinds on teachers, learners, and other stake-holders. In general, studies conducted on test consequences can be divided into two main sections; those carried out on test impacts and those on carried out on washback. Messick's (1989) validity model was the most influential model in language assessment research especially in the field of validation. From the perspective of validity, impact is referred to as a quality which needs to be designed into the test. In addition to that, Bachman and Palmer (1996) place "impact" under the notion of test usefulness to sit beside qualities such as reliability, construct validity, authenticity and interactiveness, arguing that the overall usefulness of a test is a function of these qualities. Thus, a test developer is required to design them into the test and prioritize them based on the practicality of the testing situation. Yet, some pitfalls were reported in these conceptualizations and other innovatory models of test impacts emerged on the scene. Besides, Washback in traditional view was built on a general assumption about the relationship between tests and their impact stating that high stakes test would produce a washback effect, and the direction of washback is determined by the qualities of the test. Alderson and Wall (1993) were among earliest researchers who inquired into the nature of the washback construct. Hughes (1993), Bailey (1996), Green (2007) and more recent practitioners also developed and modified the early model. In line with these theoretical developments, empirical studies were conducted in the field. A brief related literature includes Mahmoudi (2014) that worked on the washback effect of Iranian national university entrance exam on pre-university English teaching and learning. She found many factors which affect the process of English language learning and teaching in pre-university schools. In addition, Sadighi, et al. (2018) examined the possible washback effect of the university entrance exam and concluded that the perceptions of EFL fourth-graders' teachers compel them to use supplementary materials geared toward university entrance exams' demand, contrasting with the main textbook's educational objectives. Furthermore, Kiani, et al (2009) studied the test consequences of an ESP test and found that the test imposes enormously negative psychological and social consequences on the examinees as perceived by them.

3. Method

This study was a qualitative study which utilized interview to investigate the relationship between the proficiency level of students and test consequences as perceived by them. Also, a non-parametric statistical procedure was employed to inquire into the goodness of fit between the variables.

3.1 Participants

To answer the research question, 63 third grade male students were selected through convenience sampling procedures from three different high schools in the city of Birjand. These students were supposed to take part in the university entrance exam in five months.

3.2 Instruments

Two major instruments were employed in this study;

A) A teacher-made English proficiency test developed by the teacher. The content from which the test was selected encompassed vision and prospect English book series developed by Iranian ministry of education for junior high school and high school students. This test was used for determining the proficiency level of the participants. The reliability of the

proficiency test was estimated in pilot study with 80 male and female high school students in the city of Birjand. The reliability estimates are shown in table 1.

Table 1: *Reliability Statistics of Developed Proficiency Test*

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
0.798	0.782	50

B) A structured interview guided by the framework reported by Kiani, et al (2009). This instrument was used to gain the participants' perception of test consequences. The interview had two major axes (psychological and social) covering 9 components. Components of psychological axis included stress and anxiety, depression, disappointment and self-confidence and efficacy while components of social axis included deprivation from academic education, injustice, acceptance of unqualified candidates, financial consequences and family consequences.

3.3 Procedure of the Study

In the outset, 63 male students were selected through convenience sampling procedures. The proficiency test was administered to the participants and the scores were obtained. Setting cut score on the mean, the participants were either assigned to high-proficiency group (N=34) or low-proficiency group (N=29). Then, using the interview protocol, the students were talked to in a face-to-face meeting. The meetings were recorded and transcribed. In the next step, using content analysis, the associations of the answers with axes of test consequences (psychological or social) were investigated. The frequency of each answer in each axis was counted. These data were entered into SPSS (version 21) and analyzed through Chi Square test to determine the goodness of fit.

4. Results and Discussion

The main objective of this study was to investigate the relationship (association) between students' level of proficiency and the test consequence perceived by them. To this aim, a structured interview was conducted and the data were analyzed through Chi-Square. The cross-tabulation of these findings is shown in table 2.

Table 2: *Proficiency/Test Consequence Cross-tabulation*

		Consequence			
		Social Consequence	Psychological Consequence	Total	
Proficiency	Low Proficiency	Count	12	17	29
		% within Proficiency	41.4%	58.6%	100.0%
		% within Consequence	32.4%	65.4%	46.0%
		% of Total	19.0%	27.0%	46.0%
	High Proficiency	Count	25	9	34
		% within Proficiency	73.5%	26.5%	100.0%
		% within Consequence	67.6%	34.6%	54.0%
		% of Total	39.7%	14.3%	54.0%
Total		Count	37	26	63
		% within Proficiency	58.7%	41.3%	100.0%
		% within Consequence	100.0%	100.0%	100.0%
		% of Total	58.7%	41.3%	100.0%

The findings of this table show that there is an association between the variables of the study. High proficiency students' perception of test consequences indicates that they suffer more from social consequences than low-proficiency students who suffer more from psychological consequences. Table 3 also shows whether the relationship is significant or not.

Table 3: Chi Square Test

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	6.674	1	.010		
Continuity Correction	5.414	1	.020		
Likelihood Ratio	6.771	1	.009		
Fisher's Exact Test				.012	.010
Linear-by-Linear Association	6.568	1	.010		
N of Valid Cases	63				

The results of Pearson chi-square show that the p value is less than 0.05 and therefore the relationship between the variables is proved significant. In sum, a chi-square test of independence was conducted comparing the frequency of test consequences in low and high proficiency students. A significant interaction was found ($X^2(1) = 6.647$, $p = 0.01$). It was concluded that high proficiency students are more likely to perceive test consequences (54.0%) than low-proficiency group (46%). Table 4 represents Phi and Cramer's V which are both tests of the strength of association.

Table 4: Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	-.325	.010
	Cramer's V	.325	.010
N of Valid Cases		63	

On the basis of the standards set for interpreting the effect size, when degree of freedom is 1 ($DF=1$), the effect size between 0.30 and 0.49 indicates a medium effect. So, the strength of the association between the variable of the current research is reported 0.325 which is a medium but statistically significant effect.

These findings support Harris (2014) which reports that any successful intervention aimed at promoting the consequential validity of the tests is required to consider the underlying variables such as test-taking abilities of individual students. Also, these findings are consistent with Kumaravadivelue (2005) who stated that any educational attempt in post-method must be tuned with individual differences. Accordingly, identifying test consequences associated with different proficiency levels paves the way for future researchers and educational policy makers to design an intervention or policy which is attentive to individual differences. The findings of the study also submit evidences on Cheng (2008) which asserts that the form and intensity of test consequences differs greatly across contexts and individuals because the phenomenon has a dynamic and complex nature. These finding are tuned Cheng, Sun and Ma 2015 which highlighted the importance of bearing in mind the factors that may contribute to different consequences in different context. And finally, these findings affirm Di-Gennaro (2017) which concluded that test impacts differ across different proficiency levels.

5. Conclusion

In a long run attempt aimed at promoting the consequential validity of high-stakes test, unexpected aspects of assessment process which negatively affect the learning process has been investigated by researchers in the last decades. However, without the variables involved in the process, an appropriate intervention which reduces negative test consequences is hard to achieve. Thus, the purpose of this research was an attempt to investigate the interaction of proficiency level of language learners with test consequences through adopting a theoretical framework. It was concluded that high proficiency students are more likely to perceive test consequences (54.0%) than low-proficiency group (46%). The findings highlight the effect of localizing and tuning any educational intervention including those conducted in the field of language assessment. A good bulk of literature has underscored the weight and importance of test consequential validity. Delving into the variables affecting this important issue is a must in any educational context. Therefore, the future research is suggested to encompass variables other than proficiency level such as ethnic differences, cultural attitudes, economic level of the test takers and etc. These findings have implication for language teachers, language learners, researchers and educational decision makers.

References

- [1] Alderson, J.C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14 (2), 115–129.
- [2] Bachman, L, & Palmer, A. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- [3] Bailey, K. M. (1996). Working for washback: a review of the washback concept in language testing. *Language Testing*, 13(3), 257-279.
- [4] Chalhoub-Deville, M. (2009). The intersection of test impact, validation, and educational reform policy. *Annual Review of Applied Linguistics*, 29, 118–131.
- [5] Cheng, L. (2008). Washback, impact and consequences. In *Encyclopedia of language and education* (pp. 2479-2494). Springer US.
- [6] Cheng, L., Sun, Y. & Ma, J. (2015). Review of washback research literature within Kane's argument-based validation framework. *Language Teaching*, 48(4), 436–470.
- [7] Di-Gennaro, J.A. (2017). *The washback effects of an English exit exam on teachers and learners in a Korean university English program*. [Unpublished PhD dissertation] University of Exeter, U.K.
- [8] Green, A. (2007). *IELTS washback in context: preparation for academic writing in higher education*, *Studies in language testing* 25. Cambridge University Press and Cambridge ESOL.
- [9] Harris, M.T. (2014). *The effects of a test-taking skills intervention on test anxiety and test performance in 4th graders*. Unpublished Master dissertation, Louisiana states university. U.S.A.
- [10] Hughes, A. (1993). Backwash and TOEFL 2000. [Unpublished manuscript] University of Reading.
- [11] Kiani, G.R; Alibakhshi, G & Akbari, R (2009). On the consequential validity of ESP tests: a qualitative study in Iran. *The journal of applied linguistics*, 2(1), 103-126.
- [12] Kumaravadivelu, B. (2005). *Understanding language teaching: From method to post method*. Mahvah, NJ: Lawrence Earlbaum.
- [13] Mahmoudi, L. (2014). The washback effect of Iranian national university entrance exam (INUEE) on pre-university English teaching and learning (Unpublished doctoral dissertation). University of Malaya, Kuala Lumpur, Malaysia.
- [14] Messick, S. (1989). Validity. In: Linn, R.L. (Ed.), *Educational Measurement* (3rd Edition), (pp.13-103). American Council on Education/Macmillan, New York.
- [15] Sadighi, S. et al. (2018), Wash-back effect of Iranian students' pre-university English textbook and university entrance examinations: Teachers-based perspectives. *African Educational Research Journal*, 6(4), 303-316.
- [16] Wall, D. (1997). Impact and washback in language testing. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education* (Vol. 7, pp. 291–302). The Netherlands: Kluwer: Dordrecht.