Research Paper

Journal of Language and Translation
Volume 14, Number 4, 2024, (pp.103-119)

Developing a Native Model for Iran's National University Entrance Exam: Washback Effects on English Language Education at Iranian High Schools

Farhad Fathinejad[1], Behdokht Mall-Amiri[2*], Hamid Marashi[3]

[1]Ph.D. Candidate, Department of English, Central Tehran Branch, Islamic Azad University, Tehran, Iran
[2*]Assistant Professor, Department of English, Islamic Azad University, Central Tehran Branch, Iran
[3]Associate Professor, Department of English, Islamic Azad University, Central Tehran Branch, Iran

Received: April 02, 2023 Accepted: May 25, 2024

Abstract

The current study was an attempt to develop a local model and scale of washback effects of the English language module of the National University Entrance Exam on English language education at Iran's High Schools. To this aim, a mixed methods study was adopted. The qualitative phase involved data- driven from interviews with twenty experts in English education concerning NUEE washback effects which led to the formation of a tentative scale. Then, the quantitative phase included piloting and reliability estimation as well as factor analyses and Structural Equation Modeling for validation of the developed questionnaire. To this purpose, 570 participants were selected through convenience sampling from Tehran, Qom, Varamin, and Gharchack amongst teachers and students as respondents to the questionnaire. Data analyses showed that 58 items were loaded under five dimensions: educational process, attitudes and perceptions, educational policies, emotional and consequential, and socio-cultural dimensions. The model was shown to enjoy acceptable fit indices.

Keywords: Assessment; Learning; NUEE; Teaching; Washback effect

INTRODUCTION

Among major critical shifts in education in general and language teaching, in particular, have been those approaches aimed at promoting learning and teaching through assessment. The logic behind this endeavor was mainly associated with approaches such as measurement-driven instruction elaborated on by Resnick and Resnick (1992) who claimed that well-developed high-stakes tests enhance effective teaching and learning because both instructors and learners inclined toward the assessment due to its critical consequences (Chapman & Snyder, 2000). This importance changed tests not only as a tool for measuring achievement but as leverage for inducing intended changes in teaching and learning (Linn, 1992). According to Elford (2002), this movement brought about some successful and unsuccessful educational reforms in some countries such as the United States.

From Bailey (1996) and Messick (1996) to Cheng and Curtis (2004) and Spratt (2005) and more recently Wang and Huang (2020), the term washback is defined as the influence that tests have on teaching and learning. Messick (1996) stated that the washback effect encompasses test rehearsal behavior, where a significant amount of time in classrooms is spent on making the learners ready for the tests in a way that learning and teaching goals and objectives, and even teachers and students' attitudes are affected by the curriculum and assessment criteria that serve for doing well in high stakes tests.

According to Bailey (1996), educational testing has become an important topic that has attracted a huge bulk of research in the past few

*****Corresponding Author's Email:
beh.malamiri@iauctb.ac.ir

decades. More specifically, a non-exhaustive review of literature on language teaching shows that the relationship between teaching and testing has attracted researchers, practitioners, theorists, and educational and policy decision-makers from the very beginning, acknowledging the fact that tests affect classroom teaching, the syllabus, teaching practices and students' practices and activities (Jacoby, Wahlheim, & Coane, 2010). According to Bachman and Palmer (1996), the uses made of test results imply values and goals, and for the same reason, bring about consequences for society and the educational system in general and individuals in particular. One outcome related to testing is called washback and test impact which has been the subject of research and encompasses various dimensions of the research in the testing discipline.

According to Cheng (2005) and Hughes (1989), the influence of washback effect is vast in society. It affects the individuals as test takers and their teachers who help them to become prepared for the tests. Also, its impact is reflected in the ways the test takers study for a test, and the way teachers try to teach is commensurate with the prospective test to improve students' scores.

Washback is a simple, yet intriguing concept which is concerned with the effect of tests on instruction and learning but this simplicity is sometimes misleading because many concepts, factors, and variables inside and outside the field of education are involved (Alderson & Wall, 1993). Not unlike the field of education, in applied linguistics, the relationship between different forms of assessment, instruction, and learning is strong and highly correlated; despite that, test influence research did not attract applied linguistics until the last decade of the 19th century (Andrews, 2004). Elder and Wigglesworth (1996) state that some variables were involved in test effects and consequences, and therefore, identifying these variables and conceptualizing the relationships between them was a formidable task for researchers to enter into this area of research. As it is stated above, the logic behind conducting the present study is the presence of this gap in the literature, especially in Iran.

The same assertion is embodied in Noble and Smith (1994) who introduced assessment as top-down positive reform when there is an overlap between the various contents and the formats of the tests, and the corresponding contents and formats of the curriculum. Underscoring the importance of this match in the aforesaid areas, Cheng and Curtis (2004) call it "curriculum alignment". In sharp contrast with this line of thinking, Goertz and Duffy (2003) had previously stated that it is testing that determines teaching and learning and their qualities rather than curriculum because it is the assessment that possesses value and then becomes what is taught in the curriculum.

Among the earliest empirical studies of washback was a two-year-long research performed by (Alderson & Wall, 1993) aimed at investigating the effect of a large-scale assessment modification program on teaching and learning practices in the Asian country of Sri Lanka where they found that the constructive type of washback effects turned out to much more limited than predicted. According to Wall and Horák (2006), the subsequent empirical studies on washback can be categorized into two main categories. The first category investigates the effect of the test that already exists including high-stakes exams such as IELTS (Hayes & Read, 2004), TOEFL (Alderson & Hamp-Lyons, 1996), and university entrance exam (Watanabe, 1996). However, the second category of studies inquired about the intended positive washback by making minor changes to a ready test or the administration of a novel test. Studies conducted by Cheng (1999) and Shohamy, Donitsa-Schmidt, and Ferman (1996) which were aimed at disparaging conventional instructor-centered grammar lessons and promoting L2 instruction and learning to enhance communication skills are instances of this category.

The results of the studies conducted in the domain of applied linguistics showed that not only they are faced with washback as a phenomenon but also it is a complex one (Alderson, 2004). According to Cheng and Curtis (2004), researchers changed their angles of look to shed light on the effect of intricate causes of washback on instruction and learning

instead of seeking to determine whether or not the impact identified is positive or negative. McNamara (2000), for example, stated that washback is a complex phenomenon and therefore, mere assessment fails to create desired changes in instruction and learning intended by educational policymakers.

The complexity of the washback phenomenon was asserted by Fullan (2001) too. Many research findings showed that the positive washback effects could be enhanced or deterred by multiple variables, inside and outside the test itself including test variables, prestige variables, personal factors, micro factors (variables associated with school setting), and, macro factors (variables associated to socio-cultural settings) such as beliefs and attitudes (Watanabe, 1996).

Impact vs. Washback

Cheng (2005) believed that the earliest significant study which delved into the concept of washback was that of Alderson and Wall (1993) which resulted in the earliest hypotheses about washback for those areas of education which was possibly influenced by tests including the materials and content of teaching, the quality of teaching, what learners learn, the quality of their learning, the pace and order of instruction and learning, the extent and breadth of instruction and learning, and the perceptions of the content or teaching methods and learning. These findings were a critical development toward the construct of washback.

Bailey (1996) developed a basic model of washback in which a clear distinction was made among the following: washback to the students; the direct influence of the assessment on the test-takers; and washback to the program or the effects of the test on instructors, administrators, and curriculum developers. This basic model served as a stepping stone for future research, at least, for a couple of decades.

Madaus (1988) carried out research to shed light on the logic behind the teachers' preference to teach for the test rather than for the curriculum. The findings showed that this preference emerges from the attitude and perceptions of teachers toward tests. Also, some of this preference is shaped by the society in which test results are used. Investigating the washback effect of a high-stakes English test in Greek context, Tsagari (2011) concluded that the participants of the study, due to exam pressure, concentrate mainly on making learners ready for meeting the criteria of the exam since the exam is claimed to have a negative effect on the instructors. This is because the assessment results were also used to hold them accountable and judging the efficacy of their teaching. It was also concluded that the washback effect is negatively associated with emotional states of the participants since its pressure resulted in an increased levels of stress and anxiety.

Several similar studies were conducted in the Asian context. For instance, Ahmad and Rao (2012) conducted a study in Pakistan and found that the instructors' main objective for teaching is preparing students for the requirements of the test package rather than real knowledge and practice of language use because students' failure in the exam is interpreted as their teacher's inadequate practice or knowledge.

Models of Washback

Modeling a complex phenomenon, such as washback in which many variables are involved, provides information that not only sheds light on the interplay among and between factors but can inform policy makings at the macro level in any educational system. For the same reason, many attempts are made to develop a model that captures how a test affects instruction and learning, indicating the difficulty of finding patterns of the way tests influence teachers and students (Mizutani, 2009). In addition to that, there are some variables beyond mere teaching and learning which also affect the process and make model-making a formidable task for applied researchers. In other words, the effect of a test varies based on a combination of the assessment quality and how the test results are practically applied and interpreted. It also depends on the amount of test use and the perceptions held by stakeholders, including teachers and students.

Among the earliest attempts for the model-making of test effects is that of Chapman and Snyder (2000) who developed a model which encapsulates four possible uses of

tests (allocating resources to deprived schools or underdeveloped geographical areas, necessitating instructors to use a diverse range of teaching materials and methods so that students can do well on the test, motivating teachers to improve their teaching through scores obtained, motivating governments to spend more on education)as well as the intended outcome (improved educational practice, introducing change in the educational system)and the intermediate events such as community pressure and some other extra factors (community outcry, social pressure, parental concerns, etc.).

Though some later researchers such as (Andrews, Fullilove, & Wong, 2002) submitted some evidence about the practicality of some aspects of the hypothetical model, it did not stay far from criticism since some researchers such as McNamara (2000) argued that that the model is not comprehensive and hypotheses do not mention any factors associated with the way teachers and learners act in the classroom. Then, considering these theoretical flaws and blending the findings of two earlier models, Bailey (1996) developed the most basic model of washback. The innovations of these models are twofold; 1) the inclusion of the role of the researcher in the process of test washback and 2) the inclusion of dotted lines illustrating the direction of the impact. The schematic representation of this model is shown below:

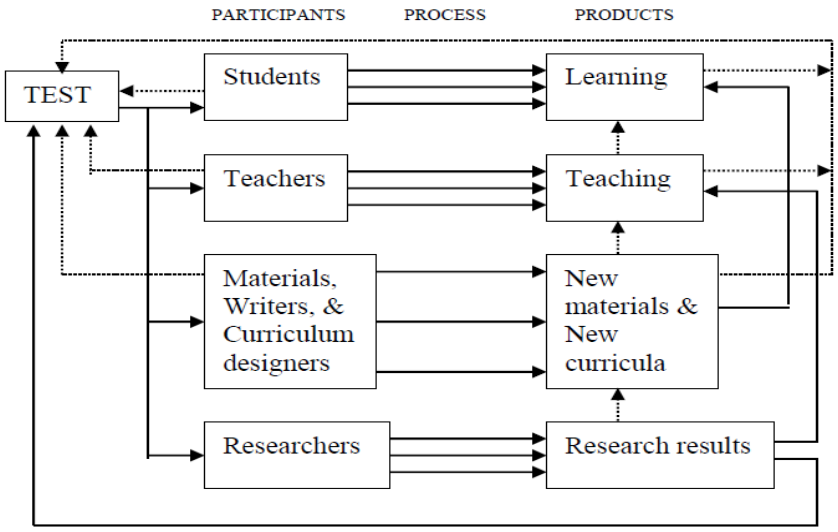

Figure 1

A Basic Model of Washback Effect (Bailey, 1996)

Mizutani (2009) asserted that in this model test exerts a direct influence on test takers while fails to consider the role of beliefs held by them and therefore cannot explain why they perform this action concerning the test. Another drawback found in this model was that it was not consistent with the widespread definition of washback. In other words, it explains test impacts which are beyond the scope of teaching and learning.

LITERATURE REVIEW

Puspitasari (2020) conducted a study in the Indonesian context and investigated the effect washback related to the national examination could have on Indonesian practices in terms of the perceptions and views held by teachers, learners, and parents regarding the effect of the test. The study adopted an interpretive perspective to investigate the teachers, learners, and parents' perceptions and past experiences in the context of the national examination. It especially examined how the national examination influences instruction and learning practices in final-year classrooms. Besides, the investigation examined how the participants' experiences in the national examination year can be developed. The data obtained from the interviews uncovered three main washback themes, including emotion, perception, and practice. Also, findings showed that the exam influences the participants both positively and negatively.

The results revealed the extent to which assessment impacted the role and practices of instructors, learners and parents. In theoretical aspects, too, the findings showed that the graph put forth by Baird, Andrich, Hopfenbeck, and Stobart (2017) which purports to account for how learning is related to assessment theories fails to explain the phenomenon and the envisaged relationship in the Indonesian context. Some key social factors were suggested in this study that implements the application of universal models in the Indonesian context.

Also, aimed at uncovering the relationship between the perceptions held by students toward a test, the nature of learning processes, and learning results, Dong (2020) conducted an investigation to shed light on the possible effect of the washback mechanism on a high-stakes test. The findings showed that the perceptions held by the participants toward test validity, influence, and significance impact the quality of their learning practices. The results also indicated that th English learning practices influenced learning outcomes positively, yet differently. In other words, the results substantiate that learners' attitudes toward test importance significantly impacted instruction-related learning and test preparation.

A review of related studies also unveils that one cannot predict the influence of high-stakes tests on instructions and learning; moreover, such an effect is not homogenous (Fox, 2005; Tollefson & Tsui, 2003). Considering these two fundamental issues, it seems necessary to bring the phenomenon into light and re-evaluate it from a sociocultural and organizational point of view that resides in Iranian local educational knowledge. Despite many attempts to avoid negative washback, teaching for the test is the major practice of Iranian EFL teachers for preparing students for the university entrance exam. A comprehensive local model of washback which originates from a validated pool of data can address both problems in the Iranian educational context and language assessment setting. Should there be any educational policy reforms, , awareness of the target society's relaities, attitudes and wants would be essential and attainable based on feedbacks driven from administering a newly developed scale which would provide the education policy makers with a realistic and better insight leading to revelation, stabilization and alteration of the covert and overt policies of English language education at Iran's high schools impacted by NUEE. Thus, the current research was an attempt to give birth to a native model for addressing local problems associated with the phenomenon through developing a scale for evaluating the NUEE washback effects. To meet these aims, two research questions were formulated as follows:

RQ1. *What are the main components of the washback effects of the NUEE on English language education at high schools in Iran?*

RQ2. *Does the developed model-based scale of the NUEE washback effect show acceptable fit indices in Iran?*

METHOD

Participants

Based on the nature of the study, several participants with distinct characteristics were selected through a non-random sampling procedure. For the first phase (qualitative), 20 experts in English language education with more than 5 years of experience in teaching the English language at high schools were interviewed. For the second phase (scale development: piloting and validation), 50 participants: teachers (15) and students (25), and other stakeholders, like principals and managers (10), other than the main sample of the study were selected from Tehran through convenience sampling to pilot the initial draft. Then, the initial draft of the scale was administered to 100 English language teachers (32), students (50), and other stakeholders (18) to estimate the reliability thereof. Finally, 150 participants were selected (40 teachers, 98 students, and 12 other stakeholders) for the exploratory factor analysis phase, and 270 participants (35 teachers, 193 students, and 42 other stakeholders) for the confirmatory factor analysis. All were selected on a convenience basis from Tehran, Qom, Varamin, and Gharchak areas.

Instruments

Interviews

This research commenced with gathering qualitative data through interviewing experts in the

field of English education at the high school level to seek their viewpoints about the NUEE washback effects on English education at high schools. To that end, a semi-structured interview was conducted with the twenty informants.

a) In your opinion, does NUEE have any distinct impact on education at high schools?
If yes,

b) In what way does NUEE influence education at high schools?

c) What aspects of education at high schools does NUEE influence?

Procedure

The overall procedure is divided into two main phases. These phases involved 1) interviewing experts to arrive at a tentative scale, 2) piloting and validating the developed scale based on exploratory and confirmatory factor analysis, and Structure Equation Modelling (SEM).

Interviews

For the first operation, initially, in-depth interviews were done with twenty experienced.

English language teachers inquiring about their general ideas regarding the washback effects of NUEE. Based upon transcriptions of the respondents' lengthy answers, the initial themes were formed as aspects of the NUEE washback effect whereby relevant statements were grouped to form major concepts.

Scale Development

The information elicited from the twenty experienced teachers initially interviewed in the first phase of the study was benchmarked with those available in the literature. Based on the collected information, a tentative questionnaire was constructed. Afterward, aimed at developing the items, content sampling and multi-item sampling were used for content selection. The generated items were fit into a questionnaire. The items covering the hypothesized target domains were selected through expert interviews and a review of the literature. In generating the items, the easy-to-grasp themes were generated through the agency of a simple and non-ambiguous language without double-barreling (a single item covering two or more axes). The semantics and language of the items, the length of the items, the possibility of overlapping, similarity, etc. were considered by the researchers and the experts. Also, the Likert scale was selected for the rating scale embracing alternatives ranging from from strongly agree to strongly disagree and receiving values of 1 to 5 respectively for the first and second component and reversely for the third, fourth and fifth components.

Scale Evaluation

After item generation and development of the tentative scale, the next phase consisted of initial piloting, reliability estimation, and factor analysis. No time limitation was imposed as the distribution was done both face-to-face and using the Telegram medium. This resulted in some minor modifications in the questionnaire regarding the wording of the items. In the next step, Cronbach's alpha was used to measure the internal consistency of the hypothesized scale. Afterward, the questionnaire underwent a validation process which was conducted by using a two-phase process unfolded within two distinct administrations (i.e., exploratory and confirmatory factor analyses). Structure Equation Modeling (SEM) was utilized to develop the final model by showing the strength of structural relationships among the variables.

Research Design

In this study, a sequential exploratory mixed-methods design that involved qualitative and quantitative stages was employed, as firstly the qualitative data were gathered to answer the research question about the nature of washback effects in the Iranian high school context, and at the second stage, quantitative data collection and analysis took place to examin reliability and validity of the developed scale for measuring the NUEE washback effects**.**

Data Analysis

The current study enjoyed an exploratory sequential mixed methods design in which the qualitative phase preceded the quantitative one. In the qualitative phase, semi-structured interviews were conducted for collecting experts' viewpoints concerning the main factors of washback effects of NUEE. Subsequently,

the derived themes were converted into items in the form of a Likert -scale questionnaire. As for the quantitative data analyses, IBM-SPSS (Version 23) was used for reliability measurement and factor analyses, and IBM-AMOS (Version 14) was employed for path analysis through Structural Equation Modeling (SEM).

RESULTS

Answering the First Question

To answer the first question, hence revealing the main components of NUEE washback effects on English education at high schools in Iran, an interview was conducted with 20 experts in English language education at high schools. The statements together with the number of participants who mentioned them are provided in Table 1.

Table 1
Experts' Statements

Statements	Frequency (From 20)
Educational activities at second-grade high schools aim at both enhancing students' GPA and their success at NUEE**.**	18
Teachers' effectiveness is evaluated dominantly based on their accordance with NUEE rather than students' GPA**.**	19
Only those teachers who focus on and teach the NUEE-preparation –related tests are valued.	12
Apparently, many teachers are happy with the NUEE-oriented English teaching as they wouldn't have to speak in English**.**	10
For teachers, correcting grammatical errors and teaching NUEE-rated areas are more important, and other skills are useless in their classes**.**	20
Only those aspects of the course book which are tested in the final exams are practiced, and therefore, only writing and reading are considered important**.**	16
Students feel more comfortable as they don't learn communicative skills: they simply study English to pass NUEE successfully, hence a lower study pressure.	18
Students pay more attention to instructions when teachers focus on an NUEE-related testing technique or content.	18
By concentrating on the importance of NUEE in their classes, many teachers can draw students to their private Entrance Exam Preparation classes outside the high school, either in private tutoring or at their institutes.	12
Teachers' more success in their students' NUEE performance means other schools' rivalry in recruiting them as teachers, which means a demand for a higher salary.	18
Parents favor and seek teachers with more rate of success at NUEE	17
New decrees on the changes in instructional contents have been enforced to decrease NUEE negative washback effects.	12
Achievement test contents are determined largely by the contents of NUEE.	14
Washback effects on English education may differ in privileged and underprivileged areas due to parental, attitudes, and other factors.	13

Based upon their answers, the themes were formed as a teaching method, teachers' attitudes, test content, learners' preferences, and expectations, learners' attitudes and feelings, parental factors, sociocultural factors, INC (Iran's National Curriculum). In fact, the first research question is answered through revealing the main components of the NUEE washback effects as the main domains impacted by the NUEE: perception and attitude, educational process, policy making, emotional and concequential factors, assessment, and social and cultural factors presented below:

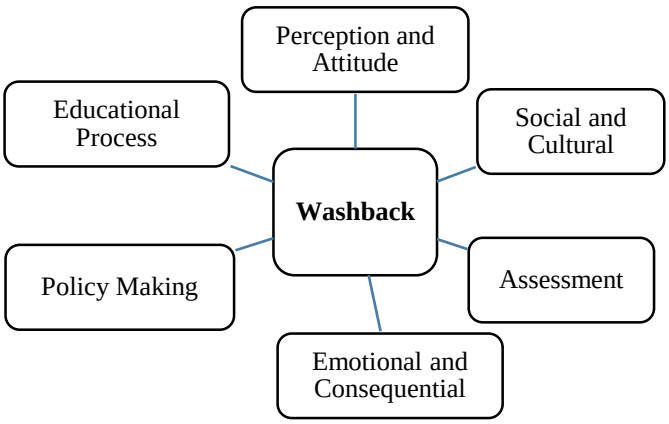

Figure 2
Representation of Washback Scale Components

Answering the Second Question: Questionnaire Development and Validation

Content Selection, Item Generation and Initial Piloting

The data derived from the preliminary interviews with experts underwent thematic analyses and led to the designation of 6 main components 1) Assessment 2) *Education Process: Teaching and Learning 3) Perception and Attitude, 4) Educational Policy-Making, 5) Consequential and Emotional factors* and *6) Social and Cultural factors* as the washback effects of NUEE.

On the basis of the components identified at this stage, a tentative scale was hypothesized including 60 different items in six main components. A five-point Likert type scale was selected for the questionnaire ranging from strongly agree to strongly disagree which were coded from 1 to 5 respectively. The first and second components as well as items 55, 56, and 57 were reversely valued.

The result of the questionnaire's reliability estimation is shown in Table 2:

Table 2
Reliability Statistics of the Questionnaire

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.903	.911	60

To conduct the exploratory factor analysis, several assumptions (i.e., factorability of data, sample adequacy, and sphericity) were checked as shown in Table 3:

Table 3
Results of KMO and Bartlett's Test

KMO Measure of Sampling Adequacy		***0.798***
Bartlett's Test of Sphericity	*App Chi-Square*	*8211.323*
Df		*612*
Sig		***0.001***

As pointed out by Pallant (2007), the Kaiser-Meyer-Olin index can range from 0 to 1, with values exceeding 0.6 considered acceptable. As Table 3 shows, the value of KMO turned out to be 0.798, which means it is acceptable. Moreover, the sig value of Bartlett's test for sphericity was found to be less than 0.05 ($p = 0.001$), which means that the data qualify for factorability. A scatter plot output was also extracted for controlling linearity assumption as well as factorability which has shown in Figure 3.

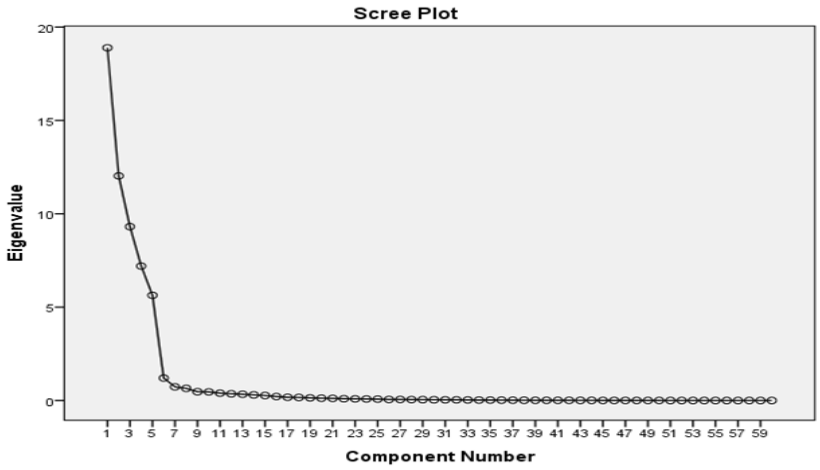

Figure 3
Scatter Plot Output of Principal Component Analysis

As Figure 3 shows, factors 1 to 6 seem to be above the point of change (elbow). The Eigenvalue on the basis of Kaiser's criterion was selected for component extraction. So, with an Eigenvalue larger than 1 or factors that explain a total of 70-80% of the variance are retained (Tabachnick & Fidell, 1996).

Table 4
Variance Explained and Components Extracted (Based on PCA)

Component	*Initial Eigenvalues*			*Extraction Sums of Squared Loadings*			*Rotation Sums of Squared Loadings*		
	Total	*% of Variance*	*Cumulative %*	*Total*	*% of Variance*	*Cumulative %*	*Total*	*% of Variance*	*Cumulative %*
1	**18.892**	**31.486**	31.486	18.892	31.486	31.486	18.458	30.763	30.763
2	**12.029**	**20.049**	51.535	12.029	20.049	51.535	11.438	19.063	49.826
3	**9.312**	**15.521**	67.056	9.312	15.521	67.056	9.775	16.292	66.118
4	**7.200**	**11.999**	79.055	7.200	11.999	79.055	7.413	12.356	78.474
5	**5.635**	**9.392**	88.447	5.635	9.392	88.447	5.984	9.973	**88.447**
6	**1.204**	2.006	90.454						
7	.728	1.214	91.668						
8	.649	1.081	92.749						
9	.467	.779	93.528						
10	.459	.765	94.293						
11	.391	.652	94.945						
12	.358	.597	95.542						
13	.334	.557	96.099						
14	.299	.499	96.598						
15	.267	.445	97.043						
16	.213	.354	97.397						
17	.172	.287	97.684						
18	.162	.270	97.955						
19	.141	.234	98.189						
20	.124	.207	98.396						
21	.114	.191	98.587						
22	.094	.156	98.743						
23	.087	.145	98.888						
24	.082	.137	99.024						
25	.074	.123	99.147						
26	.057	.094	99.242						
27	.055	.092	99.334						
28	.050	.084	99.418						
29	.046	.077	99.495						
30	.040	.066	99.561						

31	.036	.060	99.621
32	.034	.057	99.678
33	.031	.052	99.730
34	.024	.040	99.770
35	.021	.035	99.805
36	.020	.034	99.839
37	.015	.025	99.863
38	.013	.022	99.885
39	.011	.018	99.903
40	.010	.017	99.920
41	.008	.014	99.934
42	.008	.013	99.948
43	.006	.009	99.957
44	.006	.009	99.966
45	.005	.008	99.975
46	.004	.007	99.981
47	.004	.006	99.988
48	.003	.005	99.993
49	.003	.004	99.998
50	.001	.002	100.000
51	9.790E-5	.000	100.000
52	8.093E-5	.000	100.000
53	2.416E-15	4.026E-15	100.000
54	1.306E-15	2.176E-15	100.000
55	1.050E-15	1.751E-15	100.000
56	7.147E-16	1.191E-15	100.000
57	1.097E-16	1.829E-16	100.000
58	-5.128E-16	-8.547E-16	100.000
59	-1.295E-15	-2.158E-15	100.000
60	-2.633E-15	-4.389E-15	100.000

Component matrix shows the un-rotated loading of each of the items on factors. It also shows the item loading on each factor. As revealed in Table 4, the number of extracted components that enjoy an Eigenvalue of 1 and beyond are six components resembling the tentative and hypothesized model. However, since the data extracted in this table are based on an un-rotated extraction, the sixth factor is deleted in Varimax rotation (Table 5). Yet, the retained five factors explain 88.447 percent of the total variance. The share of each component in the explained variance is shown below:

For easier interpretation of the factors, the rotation of factors is conducted which modifies the effects of the un-rotated factors indicated above in the component matrix, which enhances the awareness of each factor, by presenting the pattern of loadings in a manner that is easier to interpret. Table 6 presents the results of this rotation using Varimax method.

Table 5
Percent of Variance Explained by Each Component

Component	Percent of Variance Explained
1	31.486
2	20.049
3	15.521
4	11.999
5	9.392
Total	88.447

Table 6

Rotated Component Matrix Based on Principle Component Analysis and Varimax Rotation

	Component				
	1	2	3	4	5
ITEM1	.732				
ITEM2	.626				
ITEM3	.832				
ITEM4	.988				
ITEM5	.990				
ITEM6	.673				
ITEM7	.885				
ITEM8	.712				
ITEM9	.641				
ITEM10	.763				
ITEM11	.966				
ITEM12	.855				
ITEM13	.832				
ITEM14	.991				
ITEM15	.969				
ITEM16	.994				
ITEM17	.964				
ITEM18	.978				
ITEM19	.471				
ITEM20		.991			
ITEM21		.977			
ITEM22		.962			
ITEM23		.975			
ITEM24		.981			
ITEM25		.980			
ITEM26		.967			
ITEM27		.964			
ITEM28		.942			
ITEM29		.965			
ITEM30		.973			
ITEM31		.988			
ITEM32			.984		
ITEM33			.953		
ITEM34			.954		
ITEM35			.875		
ITEM36			.894		
ITEM37			.917		
ITEM38			.937		
ITEM39			.912		
ITEM40			.980		
ITEM41			.923		
ITEM42			.957		
ITEM43				.993	
ITEM44				.982	
ITEM45				. 459	
ITEM46				.987	
ITEM47				.983	

ITEM48	.942	
ITEM49	.938	
ITEM50	.927	
ITEM51		.926
ITEM52		.736
ITEM53		.796
ITEM54		.829
ITEM55		.721
ITEM56		.733
ITEM57		.661
ITEM58		.689
ITEM59		.767
ITEM60		.796

As presented in Table 6, Varimax rotation reduced the number of factors from six to five components and two items were removed due to insufficient factor loading values. The main components of the model, their items as well as their factor loading, are presented in Appendix.

The confirmatory factor analysis phase was conducted through SPSS and SEM where the degrees of the model fit were estimated. Both absolute and relative fit indices were calculated.

Table 7
Fit Measures for Extracted Washback Model

Index	**Current Level**	**Accepted Level**	**Conclusion**
Chi-Square (X^2)/df	1.02	<3	Good
NFI	0.90	≥0.90	Fair
CFI	0.91	≥0.90	Good
RMSEA	0.03	<0.05	Good

As indicated in Table 7, all the data obtained were at an acceptable level which is very good for a newly born model. *Hence, the answer to the second question is that the model shows good and acceptable fitness indices.* Model components, variables, and coefficient pathways from each latent variable to other latent or observable variables are presented in figure 4 which submits proof of the strength of correlations between the variables.

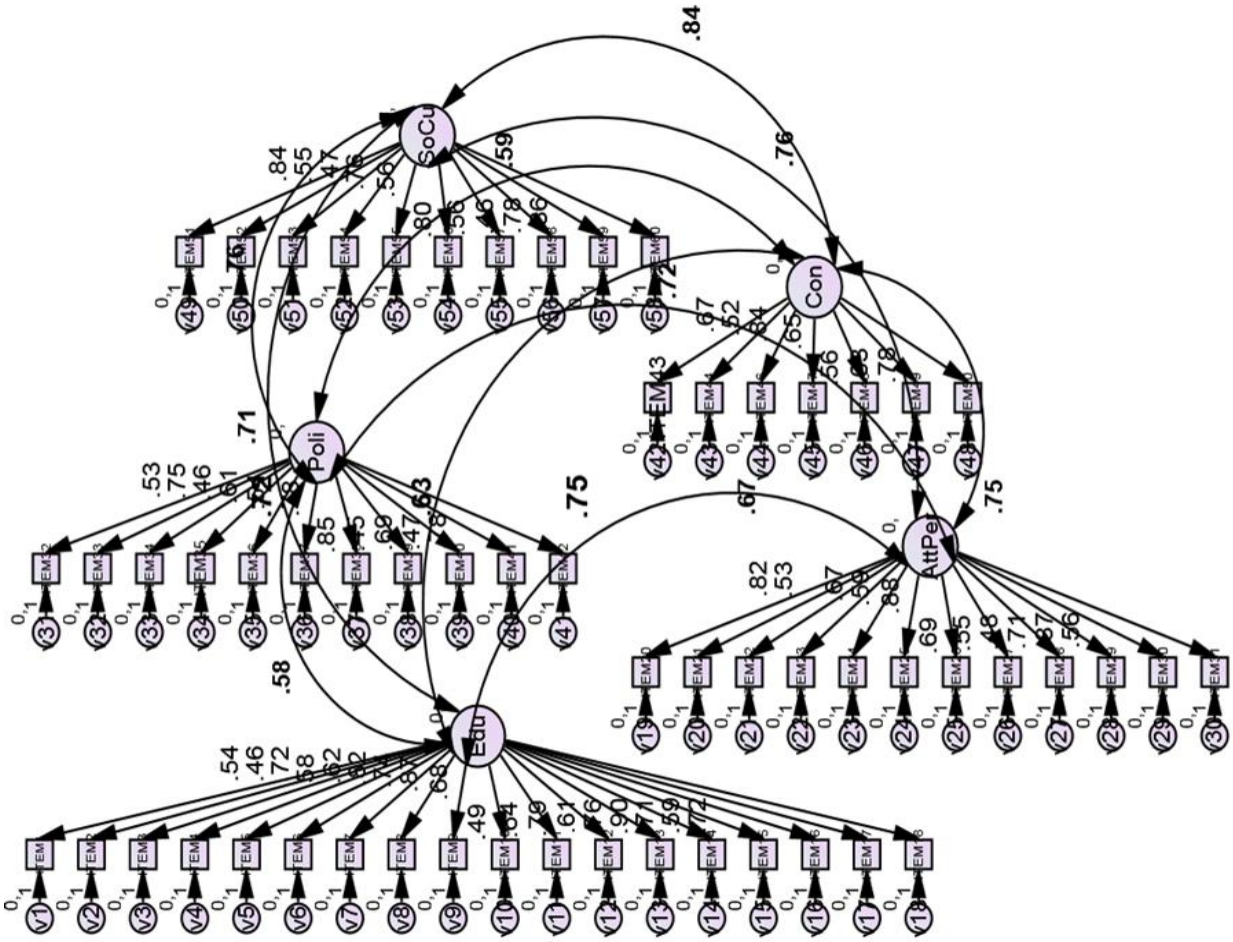

Figure 4
Model of Washback Effects of NUEE

DISCUSSION

The current research was an attempt to delve into the components of the washback effects of the English module of the National University Entrance Exam and explore how this high-stake test affects English language education at Iranian high schools. After a round of expert interviews and statistical analyses, it was found that the English module of NUEE affects English education at high schools in five areas, namely: 1. Educational procedure: teaching, learning, testing, 2. Attitudes and perceptions, 3. Educational policy making, 4. Emotional and consequential, 5. Social and cultural.

In one area of the effects, that is, educational procedure, the results of this investigation showed a rather strong influence of the entrance exam for admission at higher education institutions on the quality of the education process in the Iranian context. These results are mirrored by Ramezaney (2014) who examined the effect of washback related to the English module of the university entrance exam on professors' curricular planning and instruction. More specifically, they concluded that pre-university teachers often use university entrance exam tests and items in their teaching practices and assessments carried out for achievement purposes. Besides, in the model extracted above, it is clearly stated that some areas of language are preferred over some other areas in favor of the exam, where grammar and vocabulary receive more attention in language classes and listening, speaking, and pronunciation, receive meager attention, if at all.

The findings also corroborate those of an investigation carried out by Ramezaney (2014) who showed that pre-university students, repeatedly asked their instructors to explain the

questions included in the university entrance exam in their class activities. This was because they believed that the university entrance matters more than their English class. The teachers also held the same views, signaling the overt zest of the majority of the teachers to prepare the learners for the university entrance exam.

In line with the present findings, Alqahtani (2021) held that "exams might influence teaching mechanisms, attitudes, motivation, and content assessments"(p.22). They discovered that the tests negatively affect "the education settings' teaching mechanisms, teaching staff attitudes and motivation, and the content of assessments" (p.22).

Also, states that the modifications associated with testing need to be carefully made and piloted; therefore, the operationalization of new testing systems occurs only if positive changes are made in the educational system. From this new vantage point, it could be concluded that, though Iranian National Curriculum was an appreciated educational movement, turning a blind eye to the need for structural and contextual modification in the Iranian university entrance exam has undermined the positive effects of the curriculum to a great extent. As shown above, the actual practices of teaching, testing, and learning at Iranian high schools in Iran were in sharp contrast with expectations of the curriculum mainly as a result of the washback effects of NUEE.

Despite the belief held by Messick (1996) that those language tests which feature better design and contents are more likely to generate positive washback on students' learning, many other studies such as those conducted by Cheng and Curtis (2004), Cheng (2005) and (Winke, 2011) have shown that merely modifying language assessments does not necessarily lead to the realization of the intended educational outcomes and any change and reform needs to be carried out inside the whole education network. It seems that in the Iranian context, the reverse has been the case because educational systems and curriculum have undergone a significant change but testing and evaluation are seen as separated from the system and therefore, the university entrance exam, at least its English module, is remained intact.

CONCLUSION

This study was an attempt to develop and validate a model of the washback effects of the National University Entrance Exam on English language education at Iranian high schools. To answer the first research question, the preliminary qualitative data were used to explore the components of washback effects on diverse dimentions of English language education at Iran's highschools, namely: perception and attitude, educational process, policy making, emotional and concequential factors, assessment, and social and cultural factors. Accordingly, a tentative scale was developed , and based on the gathered quantitative data, the answer to the second question was provided based on the exploratory and confirmatory SEM analyses. The results led to the loading of 58 items onto five components of the educational procedure, attitudes, and perceptions, educational policy-making, emotional and consequential factors, and social and cultural factors. The findings revealed that the explored model enjoys a good index fit.

Contrary to the present results, Pakzad and Salehi (2018) concluded that MA UEE does not have a considerable effect on the classroom activities of teachers.

However, Booth (2012) has found that the influences of washback are different based on the status or level of the stakes of a test. Rahman, Ibna Seraj, Kamrul Hasan, Namaziandost, and Tilwani (2021) also found that secondary school certificate English examinations on English language practice in Bangladesh had strong negative washback effects on English practices due to factors including lack of congruence between the objectives of the curriculum and the test format, respndents' negative attitudes towards the test, pressure from parents and schools to assign good grades in the final exams which explain the highly exam-oriented approach.

The current research, in addition to showing consideration for the methodological weakness of previous model makings by employing a scientific and theoretical approach to model making, managed to cater for this important dimension of the effect and incorporated the social and cultural effect of the NUEE into the model.

References

Ahmad, S., & Rao, C. (2012). Examination washback effect: Syllabus, teaching methodology and the learners' communicative competence. *Journal of Education Practice, 3*(15), 173-183.

Alderson, J. C. (2004). Foreword. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research context and methods*: Lawrence Erlbaum Associates Publishers.

Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language testing, 13*(3), 280-297. doi:https://doi.org/10.1177/026553229601300304

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

Alqahtani, F. (2021). The impact of language testing washback in promoting teaching and learning processes: A Theoretical Review. *English Language Teaching, 14*(7). doi:https://doi.org/10.5539/elt.v14n7p2

Andrews, S. (2004). Washback and curriculum innovation. In L. E. Cheng, Y. E. Watanabe, & A. E. Curtis (Eds.), *Washback in language testing* (pp. 59-72): Routledge.

Andrews, S., Fullilove, J., & Wong, Y. J. S. (2002). Targeting washback—a case-study. *30*(2), 207-223.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1): Oxford University Press.

Bailey, K. M. (1996). Working for washback: a review of the washback concept in language testing. *13*(3), 257-279. doi:https://doi.org/10.1177/026553229601300303

Baird, J.-A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: fields apart? *Assessment in Education: Principles, Policy & Practice, 24*(3), 317-350. doi:https://doi.org/10.1080/0969594X.2017.1319337

Booth, D. (2012). *Exploring the washback of the TOEIC in South Korea: A sociocultural perspective on student test activity*. ResearchSpace@ Auckland,

Chapman, D. W., & Snyder, C. W. (2000). Can high stakes national testing improve instruction: reexamining conventional wisdom. *International Journal of Educational Development, 20*(6), 457-474. doi:https://doi.org/10.1016/S0738-0593(00)00020-1

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

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*: Routledge.

Dong, M. (2020). Structural relationship between learners' perceptions of a test, learning practices, and learning outcomes: A study on the washback mechanism of a high-stakes test. *Studies in Educational Evaluation, 64*, 100824. doi:https://doi.org/10.1016/j.stueduc.2019.100824

Elder, C., & Wigglesworth, G. (1996). Perspectives on the testing cycle: setting the scene. *Australian Review of Applied Linguistics. Series S*(13), 1-12.

Elford, G. W. (2002). *Beyond standardized testing: Better information for school accountability and management*: Scarecrow Press.

Fox, J. (2005). Revisiting the Storied Landscape of Language Policy Impact Over Time: A Case of Successful Educational Reform. *Curriculum Inquiry, 35*(3), 261-293. doi:https://doi.org/10.1111/j.1467-873X.2005.00329.x

Fullan, M. (2001). *The new meaning of educational change*: Routledge.

Goertz, M., & Duffy, M. (2003). Mapping the Landscape of High-Stakes Testing and Accountability Programs. *Theory Into Practice, 42*(1), 4-11. doi:https://doi.org/10.1207/s15430421tip4201_2

Hayes, B., & Read, J. (2004). IELTS test preparation in New Zealand: Preparing students for the IELTS academic module. In *Washback in language testing* (pp. 119-134): Routledge.

Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.

Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Test-enhanced learning of natural concepts: effects on recognition memory, classification, and metacognition. *Journal of Experimental Psychology: Learning, Memory, Cognition, 36*(6), 1441-1451. doi:http://dx.doi.org/10.1037/a0020636

Linn, R. J. C. f. t. S. o. E. (1992). Educational assessment: Expanded expectations and challenges (Tech. Rep. 351). University of Colorado at Boulder.

Madaus, G. F. (1988). The distortion of teaching and testing: High-stakes testing and instruction. *Peabody Journal of Education, 65*(3), 29-46. doi:https://doi.org/10.1080/0161956880953861 1

McNamara, T. (2000). *Language testing*: Oxford University Press.

Messick, S. (1996). Validity and washback in language testing. *13*(3), 241-256. doi:https://doi.org/10.1177/026553229601300302

Mizutani, S. (2009). *The mechanism of washback on teaching and learning.* ResearchSpace@ Auckland,

Noble, A. J., & Smith, M. L. (1994). Measurement-driven reform: Research on policy, practice, repercussion.

Pakzad, A., & Salehi, H. (2018). Washback effect on TEFL MA exam on Iranian lecturers' classroom activities. *Applied Research on English Language, 7*(1), 43-66. doi:https://doi.org/10.22108/are.2018.106580.1179

Pallant, J. F. (2007). *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS for Windows*. New York: McGraw Hill Open University Press.

Puspitasari, M. (2020). *Investigating the washback effect of the national examination on Indonesian practices: Perceptions of teachers, students and parents of test impact.* University of Glasgow,

Rahman, K., Ibna Seraj, P. M., Kamrul Hasan, M., Namaziandost, E., & Tilwani, S. A. (2021). Washback of assessment on English teaching-learning practice of secondary schools. *Language Testing in Asia, 11*(12). doi:https://doi.org/10.1186/s40468-021-00129-2

Ramezaney, M. (2014). The Washack Effects of University Entrance Exam on Iranian EFL Teachers' Curricular Planning and Instruction Techniques. *Procedia - Social and Behavioral Sciences, 98*, 1508-1517. doi:https://doi.org/10.1016/j.sbspro.2014.03.572

Resnick, L. B., & Resnick, D. P. (1992). Assessing the Thinking Curriculum: New Tools for Educational Reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing Assessments: Alternative Views of Aptitude, Achievement and Instruction* (pp. 37-75). Dordrecht: Springer Netherlands.

Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: washback effect over time. *13*(3), 298-317. doi:https://doi.org/10.1177/026553229601300305

Spratt, M. (2005). Washback and the classroom: the implications for teaching and learning of studies of washback from exams. *9*(1), 5-29. doi:https://doi.org/10.1191/1362168805lr152oa

Tabachnick, B., & Fidell, L. (1996). Using multivariate statistics. Harper Collins.

Tollefson, J. W., & Tsui, A. B. (2003). *Medium of instruction policies: Which agenda? Whose agenda?* : Routledge.

Tsagari, D. (2011). Washback of a high-stakes English exam on teachers' perceptions and practices. *Selected papers on theoretical and applied linguistics, 19*, 431-445.

Wall, D., & Horák, T. (2006). The impact of changes in the TOEFL examination on teaching and learning in Central and Eastern Europe: Phase 1, the baseline study. *ETS Research Report Series, 2006*(1), i-199.

Wang, Y., & Huang, B. H. (2020). Washback of TOEFL preparation courses on students' attitudes and score improvement. *International Journal of Linguistics, 12*(3), 84-104. doi:https://doi.org/10.5296/ijl.v12i3.16940

Watanabe, Y. (1996). Does grammar translation come from the entrance examination? Preliminary findings from classroom-based research. *13*(3), 318-333. doi:https://doi.org/10.1177/026553229601300306

Winke, P. J. T. Q. (2011). Evaluating the validity of a high-stakes ESL test: Why teachers' perceptions matter. *45*(4), 628-660.

Biodata

Farhad Fathinejad is a Ph.D. candidate of TEFL at Islamic Azad University, Central Tehran Branch. He has taught English courses at different universities as well as in-service training courses at the Ministry of Education. He is currently the 'Deputy for Planning and Management of the National Organization for the Development of Exceptional Talents' at the Ministry of Education.
Email: *ffathinejad@yahoo.com*

Behdokht Mall-Amiri is Assistant Professor of Applied Linguistics and a faculty member at Islamic Azad University, Central Tehran Branch. She teaches courses such as, research methodology and teaching language skills at MA and PhD levels. She has published several papers on ELT issues in domestic and international academic journals.
Email: *bmallamiri@gmail.com*

Hamid Marashi is Associate Professor of Applied Linguistics at Islamic Azad University, Central Tehran Branch and Editor-in-Chief of the *Journal of Language and Translation*. He currently teaches graduate and postgraduate courses with his main areas of research interest including innovative teaching practices and learner variables. He has published in international academic journals (including *TESOL Journal* and *Language Learning Journal*) and also presented in international conferences.
Email: *ahmuya@yahoo.com*