

REVIEW

Open Access



Test review of Iranian university entrance exam: English Konkur examination

Ali Khodi^{1,2*} , Sayyed Mohammad Alavi¹ and Hossein Karami¹

* Correspondence: Ali.khodi@ut.ac.ir

¹University of Tehran, Tehran, Iran

²University of Neyshabur, Neyshabur, Iran

Abstract

The present paper appraises a standardized test, the entrance exam of Iranian universities, known as “Konkur” that is administered annually as a means of gaining admission to higher education in Iran. This norm-referenced test is administered for students majoring in mathematics, experimental sciences, and humanities whose scores along their weighted GPAs in the last 3 years of high school are used as indicators of students’ rank. Based on the rank achieved, they would find the opportunity to select the highly regarded university for their education. Due to the importance of such a high-stake test which may bring about social and long-time consequences for the participants, the present paper tries to evaluate the test and its psychometrics aspects. It is ostensible that the exam provides a limited situation for measuring the participants “knowledge of language” rather than their “knowledge about language.” Therefore, the dimensionality and validity of the test are debatable. Thus, the present review tries to characterize Konkur examination and discusses the rooms for untouched aspects for the betterment of its quality.

Keywords: Evaluation, Iranian university English exam, Konkur examination, Test review, Validity

General description

The Iranian National University Exam, known as the UEE, is called Konkur examination in Iran. It is probably the changed form of the French term “concours,” which refers to the process of sourcing, screening, and selecting people for different purposes. UEE is administered in three majors including foreign languages. This is a large scale high-stake standardized test of the English language initially administered in 2002 for the entire country (Razmjo, 2006). The test administration is done through the cooperation between the National Organization of Educational Testing of Iran (NOET) and major universities of each city throughout the county. This is a very competitive exam for those who wish to enter public tuition-free universities (Hosseini, 2007). It is a comprehensive test, which integrates the assessment of reading comprehension ability, vocabulary knowledge, the knowledge of language functions, the ability of answering cloze tests as well as knowledge of grammar and structure of English into one multiple choice test. This test makes inferences about the English ability of the participants. All topics addressed in the test could be found in the books introduces as the main

sources of instruction in the school years in Iran. The test is considered to be a norm-referenced test that is designed to evaluate the students' ability. Mainly, the examinees are native speakers of Farsi who only sit for the test in the standard exam centers, within and outside of the country, approved by the NOET.

At the time that the test was administered for the first time, the main focus of language instruction was on enabling students in terms of reading comprehension ability because they were supposed to be able to cope with the demands of reading technical texts at university. Thus, we can see that the focus of this test was given to reading comprehension, structure of the language, grammar, and vocabulary (Farhady, 1985). After 18 years the structure of the test is remained almost intact. Recently, the course books are reviewed and revised and we believe that a new test format is required to assess the communicative ability of students rather than purely memorized grammar and vocabulary.

You can find the test pamphlet that is distributed among the participants in four codes (A, B, C, D). All pamphlets, even in different codes, present the same questions, the same sub-sections, and the same sequence of sub-sections; the only difference in these codes is either the sequence of items or the sequences of the options for items from one person to another. That is, although the examinees answer the same questions, they receive the questions in different orders. The time of administration is 105 mins. All items are dichotomously scored. A correction for guessing is also applied whereby, three incorrect answers would remove one of the correct answers. The number of the questions is 70 presented in sub-sections, namely, grammar (10 items), vocabulary (15 items), sentence structure (5 items), language functions (10 items), cloze test (15 items), and reading comprehension presented as three separate texts (15 items). Statistically speaking, we can see that the test content is not distributed among different section and skills equally; it is as following: about 27.15% for structure and grammar, 34.28% for vocabulary, and 38.57 for reading comprehension (Razmjo, 2006).

Test formats

Based on the content of the books presented to the students at high schools, the test is designed and accommodates six sections as following:

Grammar section

This section includes 10 questions asking about English grammar. The questions are presented in the form of incomplete sentences that should be completed by the option that is to be selected by the students; the options could be phrases, words, prepositions, or verbs. In terms of the sequences of question, there is no pre-determined rule and they are sequenced randomly for each participant rather than on a regular basis that is the same for all participants. Sometimes two grammatical rules are mixed in one question that seems to be very complex for the students to detect the idea behind the question and find the correct answer.

Vocabulary

In this part, 15 questions are presented in the form of incomplete sentences. The students should select the best option for the completion of the sentence meaning. Among

the options provided for the participants the correct answer was instructed previously within the classroom context, however, the distracts may be new words for them. The part of speech of the options may differ across questions but it is tried to be the same among the options of each single item in order to prevent the possibility of random guessing on the part of the participants.

Sentence structure

In this part, there are 5 questions; each option of the question presents a sentence and the participants should select the option in which there is no grammatical mistake based on the stem of the question. Mainly, the sentences are long in the form of compound and complex sentences and the mistake could be presented in any component of the sentence.

Language functions

In this part, several conversations (mainly does not exceed 3) are written and it is composed of 10 questions. The participants should complete the conversations with the best answers from the options. The correct answer should serve as a complement for the conduction of the function that is happening between two sides of the conversation.

Cloze test

In this part, participants should read a passage in which there are 15 blanks (mainly occurs at a regular number distance, for instance every ten words) and select the option which completes the sentence best. Since the blanks are presented in one text, misunderstanding or failure in finding the correct answer of one blank may lead or mislead the students to select the proper option for the next blanks.

Reading section

Each test has three reading comprehension texts whose length ranges from 350 to 500 words covering a wide range of topics such as academic, scientific, and social issues. For each text, there are 5 multiple choice items asking about the content of the text, meaning of the vocabulary, and sentence interpretations.

Test qualities and psychometrics aspects

Reliability

Due to the importance of the test consequences, the Konkur examination constructors should try their best to meet all the necessary conditions for the test reliability. The quality and number of the items stress that the objectivity of measurement is seriously considered resigned to the fact that a sufficient number of items (N=70) were presented all in the multiple choice format and were assessed through machine-scoring which is a reliable scoring procedure (Roberts, Altenberg, & Hunter, 2020). We believe that the major concerns for reliability are the imbalanced number of items in each sub-section, the equal weight for the selection of wrong option in different subsection, and the interference of the skills in sections such as cloze test or grammar. The analyses that we run on internal consistency of the test shows that the level of reliability is not equal among different subsections partly due to the unequal number of items included in different sections and it

ranges for the lowest part belonging to grammar to the highest level which belongs to sentence structure section.

Generalizability and dependability of findings

Another concern in terms of reliability is the issue of dependability of findings and generalizability of the results. How is this possible to be sure that the outcome of the test is the real performance of the participants? Conceptually, dependability means how much the results of a test show the intended level of the construct we wanted to measure. The use of neutral texts and sentences in this test shows that the developers were aware of the issue and they tried to prevent the existence of any potential bias in the function of items. About the generalizability of the results, we should be aware that the participants were all Iranian students, and the content was taken from the books in high-school; therefore, the findings could be generalized to similar contexts of the country of administration rather than an international level. Khodi (2020) ran a generalizability analysis on a sample of 5000 examinees and reported that 86% of the total variance can be explained by individuals, which is a high degree of reliability of the test. Since, in addition to the gender of the participants, their major was different; they examined the potential contribution of major to the performance of students. It is reported that the interaction of individuals' fields of study and the overlapping questions in the test sections caused an error of about 1.5%. It shows that the national entrance examination does not have a bias against any group of participants with different educational backgrounds.

Validity

As an academic test that is designed to assess the English level of test takers, it should enjoy some certain qualities the most important of which is believed to be validity. Based on Messick (1989) and Bachman (1990) validity accommodates a wider range of concepts including construct validity, content representativeness, and criterion-related validity. For the present test, validity means measuring what the test is supposed to measure while we believe that the social aspect of validity (Chalhoub, 2016) should be added to this old definition. We believe that no test could be considered valid outside of the specific use and context it is designed for (Messick, 1989). Thus, in proceeding some points related to the validity of the test are mentioned and explained.

We can see in the nature and structure of the question in the test that in spite of the construct definition of language proficiency, Konkur designers found it difficult to fully operationalize it due to the constraints and considerations of other test qualities to be manifested practically. For instance, speaking, writing and listening were not accommodated in the Konkur examination on account of the vast regional differences of participants due to accessibility issues to proper instruction. The exclusion of these skills is due to the the subjectivity in scoring these productive skills which may would pose some concerns in the matter of validity and reliability.

Overall, we believe that the construct of academic English was operationalized as the reading, grammar, and vocabulary skills that are critical to success of a first year student at university, or potentially it is the impact of needing such skills which at universities that has led to a test with such a format. It seems that the Konkur examination does not enjoy

a full representation of construct validity as there is a wide gap between the intended curriculum and the test.

Factor structure of the test and test dimensionality

It is not clearly stated that the construct of measurement, that is language proficiency, is defined as a unidimensional or multidimensional construct. The form and content of the test accentuate that there are several dimensions for the test, but on the other hand the sum-up procedure of scoring shows that no weighted score is dedicated to these dimensions and all are taken into account similarly. Even the difficulty of items does not contribute to the calculation of the final score of the participants. It means that answering a very difficult question would bring the same score as answering an easy question will bring. We suggest that for such a high-stake test with major social and life-long consequences the application of weighted scores and item difficulty level to the scoring procedure because it would increase the quality and dependability of the results. In the scoring procedure, in addition to what we stated, we can find another major concern. In spite of the fact that there is a wide range of item response theory models such as bifactor, higher-order or unidimensional models are taken as the basic framework of analysis, unfortunately in the scoring procedure of the test we could find no sign of using these models in validation procedure of the finding. In an independent study we made a comparison of these models and checked if the nature of language in this test is multidimensional or unidimensional. We found that the factor structure of language proficiency is best explained through the testlet model rather than being measured through the bifactor model (Alavi, Karami, & Khodi, [in press](#)).

Impact and washback

It is believed that “testing is never a neutral process and always has consequences” (Stobart, 2003, p. 140). Evaluation of washback is a complex and multi-dimensional act that does not exist naturally and is taken as the aftereffect of teachers, educators, or other factors’ contribution in the test-taking procedure (Alderson & Wall, 1993; Bailey, 1996; Cheng & Falvey, 2000; Spratt, 2005). For the present test in particular, the washback effect occurs due to the fact that its structure is in practice a centralized, measurement-driven system whose orientation is bound to the teachers-dominated classes, textbooks and testing impact (Ghorbani & Neissari, 2015). Although the great emphasis for washback is suggested for communicative-oriented methods, and it is stressed that tests should explicitly be designed to bring positive washback (Cheng, Watanabe, & Curtis, 2004), apparently in Iran this occurs for a reading-oriented test. Therefore, very few Iranian students finish high school with the ability to speak English effectively in spite of mastering the prescribed textbooks (Farhady, Jafarpoor, & Birjandi, 1994), and English instruction in most of the Iranian academic situations seems to be ineffective and impractical (Hosseini, 2007).

In Iran, it was found that that the EEU negatively and implicitly influences English teachers for instruction of the content and format of the test (Salehi & Yunus, 2012) and regarding the UEE format and importance, students potentially spend more time on grammatical structures, vocabulary, and reading than writing, pronunciation, speaking, and listening exercises (Farhady et al., 1994; Ghorbani, 2012). The ultimate

objectives of the EFL program stay to be outlined that has led to different repercussions over the various periods of the educational program including assessment programs. The large scales and high s-stake test named Konkur functions alike an agreement among the instructors for deciding about the material of instruction, and its negative impacts known as washback and the implementation of the actual curriculum fails (Jahangard, 2007). This exam is extremely important not only to students and their parents but also to the larger society and the whole society is affected by its impacts. For instances, the traffic limitations are changes in the administration day and many parents wait outside the administration centers until to the end of the exam. As this importance accentuates, many non-governmental institutes have started to present the simulated exams, supplementary classes, and books for students to achieve the utmost skill for taking the test, a procedure that starts even starts almost 3 years before the test. Teachers are also influenced by the Konkur and try to adapt their instruction with the hidden curriculum and students' needs and even preferences.

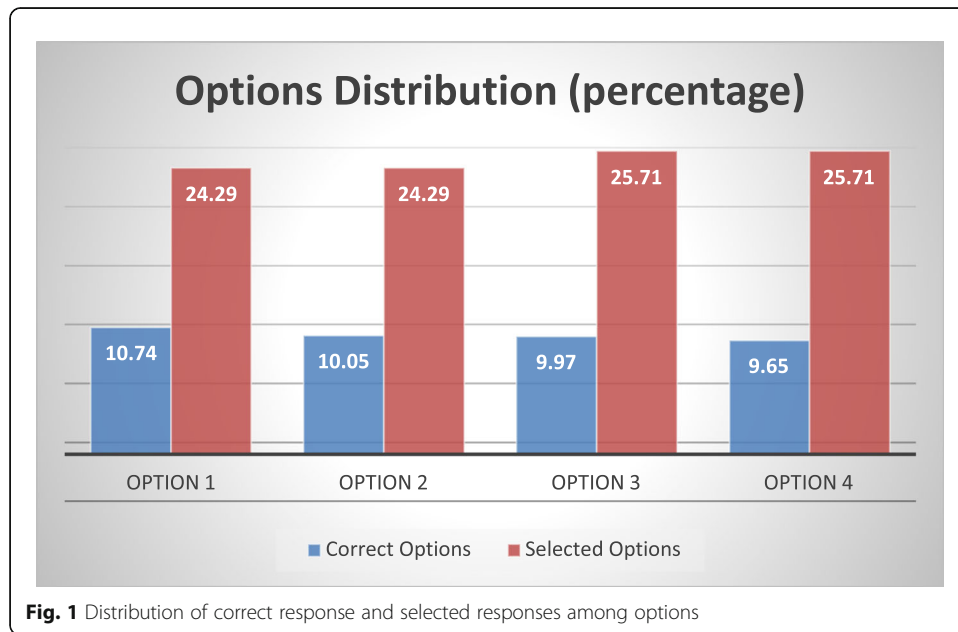
Clarity of the test

The information that is distributed by test administration board aims at providing transparency the test configuration, test time, and format. In this particular case, the Konkur examination, such information is always published, disambiguated, and clarified. It is believed that the content presented in the textbooks would suffice mainly the needs of some students and for achieving higher ranks in the Konkur and having a good command of English; a supplementary sources are needed which fundamentally differ among the students of different schools and regions and this hidden syllabus is never determined and there is no consensus about it (Salehi & Zamanian, 2012).

Conclusion

UEE enjoys a well-situated presentation of the content instructed and addresses the curriculum in a comprehensive manner. Although the washback looks somehow deconstructive, it could be reestablished with regards to the new course books and the priorities that have been newly accepted and changed. Challenges remain, inevitably, in terms of overlooking listening, writing, and speaking could be resolved possibly through addition of the new contents to the instructional curriculum and new test format. The construct irrelevant variance also may exist but could be suppressed by application of weighted scores and adding new variables such as applying students' educational background. The remaining concerns should be resolved once a consensus achieved about the definition of the constructed that is assessed basing in the test and new technological tools could be used in the assessment procedure. One major reason that some participants fail in the exam is the incompatibility of the nature of the items of the test with the item format, for instance, assessing the communicative ability of the participants through filling the blank of a conversation. The overall analysis of UEE depicts that although the context of the test and the content are in accordance, the nature is not a good indicator of participants real language ability due to lack of some important language skills in the test. Drawing on the social and life-time consequences of the test in Iranian society, it is promising that a well-deserved number of research studies evaluate UEE and provide suggestion for its betterment.

Appendix 1



Appendix 2

Table 1 Construct reliability and validity

Dimension	Cronbach's alpha	rho_A	Composite reliability	Average variance extracted (AVE)
General factor	0.833	0.883	0.853	0.104
Cloze test	0.540	0.650	0.673	0.161
Function	0.507	0.562	0.684	0.196
Reading	0.594	0.669	0.711	0.167
Grammar	0.346	0.536	0.517	0.173
Structure	0.168	0.315	0.503	0.248
Vocab	0.584	0.672	0.687	0.166

Appendix 3

Table 2 IRT analysis and dimensionality assessment

Item	Unidimensional		Bifactor		Testlet	
	χ^2	Probability	χ^2	Probability	χ^2	Probability
1	75.29	0.0001	72.59	0.0001	19.72	0.0002
2	24.4	0.0179	25.16	0.0141	13.07	0.0703
3	20.22	0.0095	21.66	0.0056	7.9	0.0952
4	25.42	0.0625	27.12	0.0401	4.1	0.3937
5	60.38	0.0001	70.59	0.0001	9.6	0.0222
6	28.61	0.0118	28.4	0.0079	22	0.0012
7	39.95	0.0051	37.96	0.0089	21.24	0.0194
8	25.34	0.1884	24.99	0.1603	19.52	0.034
9	25.85	0.0039	24.86	0.0031	6.99	0.0304
10	18.48	0.3611	18.76	0.2807	13.26	0.103
11	48.13	0.0001	46.59	0.0001	30.84	0.0001
12	35.76	0.0019	36.25	0.0016	8.06	0.1529
13	15.58	0.3417	15.99	0.249	12.81	0.0766
14	55.6	0.0001	56.17	0.0001	44.71	0.0001
15	88.4	0.0001	80.13	0.0001	20.09	0.0012
16	23.92	0.0208	25.1	0.0051	11.95	0.0629
17	28.48	0.0123	29.32	0.0094	14.43	0.0711
18	82.43	0.0001	69.5	0.0001	13.2	0.0042
19	38.59	0.0012	37.29	0.0011	8.75	0.1876
20	74.45	0.0001	63.45	0.0001	38.5	0.0001
21	18.35	0.1909	18.08	0.1542	7.42	0.2856
22	49.08	0.0003	48.4	0.0004	29.94	0.0002
23	22.47	0.0693	22.48	0.0482	13.92	0.0524
24	25.21	0.0216	26.18	0.0101	9.59	0.1428
25	40.38	0.0001	45.9	0.0001	14.81	0.002
26	59.15	0.0001	63.82	0.0001	22.28	0.0001
27	10.89	0.8167	10.7	0.7744	12.32	0.1954
28	49.23	0.0001	49.62	0.0001	16.99	0.0488
29	25.43	0.1132	25.59	0.082	19	0.0884
30	27.26	0.0985	27.33	0.0728	13.76	0.088
31	27.54	0.0505	28.85	0.0501	13.56	0.1935
32	84.17	0.0001	85.94	0.0001	43.32	0.0001
33	56.35	0.0001	58.48	0.0001	17.26	0.004
34	23.51	0.3166	23.9	0.1994	9.51	0.3933
35	53.4	0.0001	55	0.0001	24.18	0.0005
36	24.43	0.1795	25.3	0.1506	6.51	0.6889
37	41	0.0001	42.57	0.0001	28.3	0.0001
38	18.31	0.0744	18.75	0.0655	8.27	0.3113
39	37.26	0.0007	39.07	0.0001	25.83	0.0005
40	15.35	0.5007	17.44	0.3595	13.13	0.1072
41	65.08	0.0001	65.61	0.0001	33.08	0.0001
42	28.35	0.0568	32.59	0.0186	12.23	0.1407

Table 2 IRT analysis and dimensionality assessment (*Continued*)

Item	Unidimensional		Bifactor		Testlet	
	χ^2	Probability	χ^2	Probability	χ^2	Probability
43	34.78	0.0042	35.21	0.0014	19.38	0.0129
44	15.49	0.4909	15.41	0.4243	5.99	0.4252
45	64.69	0.0001	64.96	0.0001	35.73	0.0001
46	36.11	0.0102	35.69	0.0077	17.03	0.0092
47	25.72	0.058	26.99	0.0192	16.91	0.0179
48	37.82	0.0026	37.71	0.0017	20.04	0.0027
49	18.15	0.1996	18.15	0.1515	11.14	0.084
50	18.47	0.4265	18.55	0.4215	11.75	0.1621
51	25.22	0.1925	26.42	0.0904	14.39	0.072
52	33.87	0.0007	34.41	0.0003	12.38	0.0299
53	14.33	0.5024	14.97	0.382	11.14	0.1323
54	45.02	0.0001	44.56	0.0001	14.89	0.0108
55	33.21	0.0106	36.23	0.0027	9.08	0.3372
56	26.26	0.1569	26.42	0.0902	14.72	0.0986
57	17.07	0.2516	17.51	0.1765	5.56	0.4754
58	45.74	0.0001	45.26	0.0001	20.66	0.0001
59	31.64	0.0111	32.97	0.0029	23.01	0.0001
60	33.27	0.0015	33.32	0.0009	19.62	0.0032
61	29.57	0.0573	29.58	0.0416	19.75	0.0195
62	48.38	0.0004	48.58	0.0002	22	0.0088
63	18.68	0.5443	19.06	0.39	10.26	0.4195
64	13.37	0.5751	13.54	0.4859	16.02	0.0663
65	25.85	0.056	26.62	0.0319	11.34	0.1827
66	22.74	0.1206	21.78	0.1134	14.82	0.0383
67	18.73	0.2254	18.71	0.1319	23.35	0.0003
68	47.3	0.0001	48.86	0.0001	18.57	0.0173
69	21.79	0.2409	22.82	0.1547	13.57	0.0935
70	36.91	0.0022	37.22	0.0012	21.84	0.0094

Abbreviations

UEE: The Iranian National University entrance exam; NOET: The National Organization of Educational Testing ;
EFL: English as foreign language

Acknowledgements

We thank the editor and reviewers.

Authors' contributions

All authors had the same contribution. The author(s) read and approved the final manuscript.

Funding

We received no funding.

Availability of data and materials

The data will be available upon request.

Declaration**Competing interests**

We have no conflict of interest.

Received: 7 February 2021 Accepted: 13 May 2021

Published online: 12 July 2021

References

- Alavi, S.M., Karami, H., & Khodi, A. (in press). Examination of factorial structure of Iranian English Knokur examination: An IRT analysis. *Current Psychology*.
- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied linguistics*, 14(2), 115–129. <https://doi.org/10.1093/applin/14.2.115>.
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language testing*, 13(3), 257–279. <https://doi.org/10.1177/026553229601300303>.
- Chalhoub-Deville, M. (2016). Validity theory: Reform policies, accountability testing, and consequences. *Language Testing*, 33(4), 453–472.
- Cheng, L., & Falvey, P. (2000). What works? The washback effect of a new public examination on teachers' perspectives and behaviours in classroom teaching. *Curriculum Forum*, 9(2), 10–33.
- Cheng, L. E., Watanabe, Y. E., & Curtis, A. E. (2004). *Washback in language testing: Research contexts and methods*. Lawrence Erlbaum Associates Publishers. <https://doi.org/10.4324/9781410609731>.
- Farhady, H. (1985). Evaluating English test in Iranian Konkur examination. *Roshd Amozesh Zaban*, 3(3), 15–17.
- Farhady, H., Jafarpoor, A., & Birjandi, P. (1994). *Testing language skills: From theory to practice*. SAMT Publications.
- Ghorbani, M. R. (2012). Controversy over abolishing Iranian university entrance examination. *Asian Education and Development Studies*, 1(2), 139–152.
- Ghorbani, M. R., & Neissari, M. (2015). Washback effect of the Iranian concours on senior high school students' EFL learning activities. *Iranian Journal of Language Testing*, 5(1), 1–28.
- Hosseini, S. M. H. (2007). ELT in higher education in Iran and India – a critical view. *Language in India*, 7, 1–11 Retrieved October 9, 2008, from <http://www.languageinindia.com/dec2007/eltinindiaandiran.pdf>.
- Jahangard, A. (2007). Evaluation of the EFL materials taught at Iranian high schools. *The Asian EFL Journal*, 9(2), 130–150.
- Khodi, A. (2020). *An appraisal of validity and dimensionality of B.A. Iranian University Entrance Examination (Unpublished doctoral dissertation)*. University of Tehran.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational researcher*, 18(2), 5–11. <https://doi.org/10.3102/0013189X018002005>.
- Razmjo, S. A. (2006). Content analysis of specific questions of the English language test group of the national entrance exam of the country's universities. *Shiraz University Journal of Social Sciences and Humanities*, 1(46), 465–480.
- Roberts, J. A., Altenberg, E. P., & Hunter, M. (2020). Machine-scored syntax: comparison of the CLAN automatic scoring program to manual scoring. *Language, Speech, and Hearing Services in Schools*, 51(2), 479–493. https://doi.org/10.1044/2019_LSHSS-19-00056.
- Salehi, H., & Yunus, M. M. (2012). The washback effect of the Iranian universities entrance exam: Teachers' insights. *GEMA Online® Journal of Language Studies*, 12(2), 609–628.
- Salehi, M., & Zamanian, M. (2012). The use of supplementary materials by Iranian EFL teachers at high schools for talented students. *Journal of Studies in Learning and Teaching English*, 1(2), 167–185.
- Spratt, M. (2005). Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language teaching research*, 9(1), 5–29. <https://doi.org/10.1191/1362168805lr152oa>.
- Stobart, G. (2003). *The impact of assessment: Intended and unintended consequences*. Rutgers.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)