Akdeniz Language Studies Conference 2012

The backwash effect of the test items in the achievement exams in preparatory classes

Turan Paker[*]

Department of English Language Teaching, *Faculty of Education, Pamukkale University, Denizli 20070 Turkey*

Abstract

The purpose of the study is to give insights about the four language skills and language use assessed in achievement tests and the possible backwash effect of the items on both learning/teaching processes in English preparatory classes at the tertiary level. For this purpose, some samples of achievement tests from 13 institutions have been collected and analyzed in terms of test items used to assess four language skills, language use and vocabulary, and their potential backwash effect. The results have revealed that reading skill and subskills, language use and vocabulary knowledge are assessed in the achievement tests by all institutions. However, listening, writing and speaking skills are assessed by 70 % of the institutions. In addition, 15% of them also assess translation in their achievement tests. All in all, four language skills and their subskills are tested to some extent in almost all institutions. This is a very effective approach to create needs for the learners to focus on four language skills and to get them ready for their academic life as the tests items/tasks have tendency to assess performance rather than language knowledge solely.

1. Introduction

Assessment is an indispensable part of teaching/learning process. Although it has been carried out as pen and pencil tests traditionally, alternative assessment types are in our agenda in performance-based testing era (Bailey, 1998; Bachman & Palmer, 1996; Brown, 2004; Brown & Hudson, 2002; Genesee & Upshur, 1996; Hughes, 2003; Shohamy, 2001; Weir, 1990). No matter what we teach in the classroom, in

───────────

[*] Corresponding author. Turan Paker, Tel.:+90 535 929 1923; fax: +90 258 296 1200
*E-mail address:*tpaker@gmail.com

fact, our test items create the needs for our learners to master the knowledge, skill or performance; especially in countries like Turkey, tests play a dominant role in determining the future of students' lives. Hence, we have the unavoidable phenomenon in this process; the backwash or washback effect. For the sake of consistency, we will use "backwash effect" throughout this article. Backwash effect is simply defined as "the effect of each test item on teacher's teaching and learner's learning in terms of positive and negative aspects (Alderson & Wall, 1993; Bachman, 1990; Brown, 2004; Brown & Hudson, 2002; Cheng, 2005; Hughes, 2003; Weir, 1990). According to Brown & Hudson (2002), test items should be directly related to language teaching/learning process as they represent the components of language curriculum at pre-defined level. As long as the test items are parallel with the objectives of the syllabus/curriculum, they will have potential positive backwash effects on the learners; otherwise, they will influence their learning in a negative way.

As language test types, the most common ones are proficiency, placement, diagnostic and achievement tests. Among the others, the achievement tests are naturally curriculum based and assess the objectives of the curriculum (Bailey, 1998; Brown, 2004; Brown & Hudson, 2002; Cheng, Watanabe & Curtis, 2004; Hughes, 2003; Weir, 1990). Hughes (2003, p.13) points out that the purpose of achievement tests is "to establish how successful individual students, groups of students, or the courses themselves have been in achieving objectives. Hence, we have the problems of reliability and validity; content validity, construct validity and criterion validity. A test is considered to be valid as it can ensure the representation of the instructional objectives (Brown, 2004; Brown & Hudson, 2002; Hughes, 2003; Messick, 1996; Weir, 1990). According to Hughes (2003), such representation will be reflected in test items as proper sample of skills, subskills, structure and vocabulary. On the other hand, in order to improve the reliability of the tests, a number of strategies are suggested by the scholars (Bailey, 1998; Brown, 2004; Brown & Hudson, 2002; Hughes, 2003).

From time to time, there is a confusion of items and tasks used in achievement and proficiency tests. According to Hughes (2003, p.11), the proficiency test is based on "a specification of what candidates have to be able to do in the language in order to be considered proficient." Learners are expected to be proficient after a certain level. On the other hand, achievement tests assess a point in the process of learning a language. It gives us a general picture of the students at a certain level such as A1, A2, B1, B2, C1 or C2 (The ALTE Framework). Hence, the achievement tests are more specific and are based on the level studied. For that reason, language testers in SLA or FLA contexts should take into account the second language acquisition process in testing (Bachman & Cohen, 1998). They should keep in mind that the learners will go through a learning process, and the test items should not discourage them to learn new skills or knowledge. The test items should be designed in such a way that they should be constructive rather than destructive. We know that each learner goes through interlanguage period (Selinker, 1972) in their learning process. The implication of this period on testing is that our students will inevitably make two types of mistakes/errors in the exams: L1 transfer and L2 development mistakes/errors, which are unavoidable (Bachman & Cohen, 1998). However, in the tests, test designers are so behaviouristic that they expect learners to carry out the items/tasks in the exam without mistakes. It is an important inconsistency from the point of second language acquisition process.

In this research, we aimed at finding answers to the following research questions:

Which skills and subskills are tested in the achievement tests in various schools of foreign languages in Turkey?

How is the distribution of the type of items/tasks/subskills?

What are their potential backwash effects?

2. Methodology

In this study, the participants were 13 Schools of Foreign Languages in various universities in Turkey, 10 of which were in state and 3 of which were in private universities. The data were collected from the achievement tests used in these institutions. All achievement test documents were taken from the testing office of each school, and they have been analyzed in terms of test items used to assess four language skills, language use and vocabulary, and their potential backwash effect. The data have also been analyzed in line with second language acquisition process and the competencies expected in those institutions. As the data are based on written documents, the test items and procedures followed for speaking tests have been ignored as they have different organizations and procedures for the purpose of testing.

3. Results and Discussion

The data about the achievement tests reveal that the reading skill, language use and vocabulary knowledge are tested 100% in all institutions. On the other hand, the listening and writing skills are tested 70%, the speaking skill is tested 60%, and only in some schools, translation is tested 15% (see Table 1). It seems that most of the institutions cover four language skills in their achievement tests, and as a consequence of which students will have to study and focus on all skills. Thus, these tests will provide positive backwash effect to help them get ready for their future academic life in which the medium of instruction is in English.

Table1. The distribution of skills in achievement tests

Skills: Speaking	Listening	Language Use	Reading	Vocabulary	Writing	Translation
60%	70%	100%	100%	100%	70%	15%

When the subskills tested are analyzed, our findings reveal that different subskills are tested in the form of various test items. In Listening and Reading as receptive skills, some subskills such as scanning, skimming, information transfer, dictation (word/phrase level), note taking (guided), referencing, inferencing, deducing the meaning of new words from context have been tested (see Table 2). They all prepare students for their academic life and their implication is positive from the point of backwash effect. However, the weight of each skill in achievement tests differs from one institution to the other. For listening skill, the range is between 10-25 % and for reading skill, it is between 15-30%.

Table 2. The distribution of listening and reading skills, subskills and tasks (item types) in the achievement tests

Skills	The weight of the skill in the tests	Subskills	Item Type
Listening	10-25%	-Scanning -Information transfer -Dictation/note taking (word/phrase level) -Note taking (guided)	-Multiple Choice -Fill in the blanks (sentence/cloze test) -Fill in a table -Matching

Readin g	15-30%	-Skimming -Scanning -Information transfer -Referencing -Inferencing -Deducing the meaning of new words from context	-Multiple Choice -True/False -Open-ended items -Matching -Fill out a form/a table -Sequencing sentences to make a summary -Sequencing the paragraphs -Sentence completion -Choose the irrelevant statement in a paragraph -Place the appropriate sentence in a paragraph -Reaction to the given situation -Find the paraphrased statement -Cloze test with MC -Fill in the blanks from the word box -Complete the dialogue with MC -Put sentences into a paragraph

The test items to test the subskills of receptive skills are various. Listening subskills are tested mostly by multiple choice, fill in the blanks (sentence/cloze test), matching, True/False, and filling in a table /chart (dictation/guided note taking) (see Table 2). On the other hand, reading subskills are tested with more various test items/tasks compared to listening subskills. The most common test items/tasks are multiple choice, true/false, open-ended items, matching, filling out a form/a table, sequencing sentences to make a summary, sequencing the paragraphs, sentence completion, choosing the irrelevant statement in a paragraph, placing the appropriate sentence in a paragraph, reacting to the given situation, finding the paraphrased statement, cloze test with multiple choice, filling in the blanks from the word box, completeing the dialogue with multiple choice, and reordering scrambled sentences into a paragraph (see Table 2). All these subskills show that the students are tested on certain academic subskills in order to see how much they can achieve them. They all provide positive backwash effect on the part of students, and we believe that they prepare the learners for their academic life.

The test items to test the subskills of writing skill are various, and the weight of writing in achievement tests is between 15-25%. Writing subskills are tested mostly by various modes of writing such as description, comparison/contrast, cause and effect, problem solution, argumentative type either at paragraph or essay level. In testing writing skill, some test items are communicative and contextual such as writing a paragraph on a given topic(s), writing an e-mail, writing a letter of advice, writing an application letter, writing an essay on a given topic(s), writing a story about last holiday, making an outline of the given text (see Table 3). We believe that these items reinforce real life tasks and have positive backwash effect on the learners.

Some other items/tasks in writing are related to the discourse and organization of a paragraph such as making an outline before writing, putting the cohesive words in appropriate place and writing a topic sentence/a conclusion sentence in a text. These items are good to reinforce some basics about the organization of either a paragraph or an essay. We believe that they have positive backwash effect for the learners to acquire and produce the organization of a paragraph or an essay. On the other hand, the others are completely mechanical items such as rewriting some sentences by using pronouns instead of nouns,

rewriting some sentences (paraphrasing at sentence level), finding and correcting the misspelled words and writing 10 sentences (comparison/contrast) by using the pictures of two related objects. These items reinforce grammatical features rather than communicative value of the task. They lead students to feel perfectionist in using the language and inhibit them from using their ideas freely and creatively. Students have to monitor themselves (Krashen & Terrell, 1983) more than they can make use of creative thoughts and ideas. In such cases, students focus on the form rather than meaning, and they cannot come up with creative writing. Moreover, Turkish students have more tendencies to do so due to the training they have gone through in both primary and secondary schools.

Table 3. The distribution of writing skills, subskills and tasks (item types) in the achievement tests

Skills	The weight of the skill in the tests	Subskills	Item Type
Writing	15-25%	-Description -Comparison/contrast -Cause and effect -Problem solution -Argumentative	-Write a paragraph -Write an e-mail -Write a letter of advice -Write an application letter -Write an essay -Write a story about your last holiday -Re-write some sentences by using pronoun instead of nouns -Re-write some sentences (paraphrase) -Find and Correct the misspelled words -Put the cohesive words in appropriate place -Write a topic sentence/ conclusion sentence in a text -Make an outline of the given text -Write 10 sentences (comparison/contrast)
Translation	4-8%	-From Turkish to English -From English to Turkish (sentence level)	- Multiple Choice

Testing translation has been used in the form of multiple choice type in only a few institutions. Although it is not common to test translation at lower levels, some institutions test translation at sentence level at intermediate and upper levels. The weight of translation in the achievement tests of those institutions is between 4-8%. Translation could be part of the achievement tests if it is part of the curriculum/syllabus. But the test items should be given in a context. However, testing it through only sentence level as multiple choice type is not appropriate simply because it does not represent real life. In real life, students do not translate in the format of a multiple choice, rather, they directly translate a piece of text such as a chunk, phrase, statement, thought, or a factual information from target to native language or vice-verse. Moreover, the sentences to be translated in these tests are not given in a context. They are

just isolated sentences, and they have no discourse and contextual clues so that students can make use of while translating them.

In terms of testing vocabulary, most of the institutions test vocabulary as a separate section although some of them integrate it into reading skill, and the weight of vocabulary test is between 10-20% in different institutions. The test items to test vocabulary are mostly contextual in a cloze format such as deducing the meaning of new words from context. However, some test items are isolated sentences in multiple choice format such as finding synonym/antonym of a given word in a sentence. Furthermore, there are test items to test the structure of vocabulary items such as using the appropriate form of the words such as noun, adjective, adverb or verb forms (see Table 4). On the other hand, those who test vocabulary as part of language skills embed some items in reading, listening, writing or speaking skills such as deducing the meaning of an underlined word in the text or finding the meaning of reference words/phrases or they assess as part of their scale as in writing and speaking sections (Brown, 2004; Hughes, 2003).

The test items to test language use range from very mechanical to contextual and communicative items. All the institutions test language use in a separate section in the achievement tests, and they give weight between 20-50 %. Our analysis is based on whether the test items reinforce functional or structural use. Our results reveal that most of the items are mechanical type such as multiple choice, writing appropriate form of verbs in a cloze test, filling in the blanks with appropriate form of verbs, nouns, adjectives, etc., completing isolated sentences, making sentences, rewriting some sentences (transformation drill), answering open ended items with full sentences (short answers are not evaluated), and finding the mistake and correcting it (see Table 4). However, there are a few functional test items such as completing the dialogue, making statements based on a given situation, and matching statements with appropriate pictures/statements. It is a pity that most of the items are prepared to test structural use, and they are quite mechanical. Testing offices should come up with more communicative ways to test language use. Otherwise, mechanical drills do not help our learners much to digest the information, acquire and use it functionally in a communicative situation (Krashen & Terrell, 1983), simply because they do not prepare our learners for the real life; neither academic nor non-academic.

Table 4. The distribution of language use and vocabulary, and tasks (item types) in the achievement tests

Skills	The weight of the skill in the tests	Subskills	Item Type
Language Use	20-50%	-Structural use -Functional use	-Multiple Choice -Write appropriate form of verbs in a cloze test -Fill in the blanks -Complete the dialogue -Sentence completion -Make statements -Make sentences -Odd one out -Matching -Open-ended items -Re-write the sentence

			(transformation) -Find the mistake and correct it
Vocabulary	10-20%	-Deducing the meaning of new words from context -Finding synonym/ antonym of a given word in context -Using the appropriate form of the words	-Cloze test with MC -Fill in the blanks with the words from the word box -Matching -Multiple Choice

In achievement tests, it should be kept in mind that what we test is not the end product, and our learners are somewhere in the middle of acquisition/learning process. They have not digested some structural knowledge yet; however, they can recognize or even use them haphazardly. According to Ellis (1985), you can't change the route but you can change the rate. Thus, we should accept initially that they will make mistakes such as subject-verb agreement, preposition, spelling, plural, and conjugation when they are required to carry out certain tasks/performances. The items/tasks prepared for the purpose of assessment should enable the testers to tolerate such mistakes rather than deduce some marks due to such kind of deficiency on learners' performance. Otherwise, the backwash effect of this type of test is that "you know the language if you know the grammar." In such cases, the learners spend most of their time studying and mastering grammatical knowledge, and they are considered successful learners provided that they can achieve well in such 'discrete item' tests (Oller, 1979). As a result, the learners know about the language but they cannot communicate through that language. According to Newmark & Reibel, (1968 cited in Johnson, 1981), this kind of grammatical competence is 'necessary but not sufficient.' Johnson (1981) points out the fact that this kind of teaching produces 'structurally competent' students who are often communicatively incompetent and able perhaps to form correct sentences to describe the daily habits of a character in their textbook but unable to transfer this knowledge to talk about themselves in a real-life setting.

4. Conclusion

This study has attempted to describe what achievement tests have covered in different preparatory programs at the tertiary level, the type of test items/tasks used and their implications on backwash effect. Our results reveal that four language skills and their subskills are tested to some extent in almost all institutions. This is a very effective approach to create needs for the learners to focus on four language skills and to get them to prepare for their academic life. The tests have tendency to assess performance rather than language knowledge solely. Moreover, achievement tests are usually designed in terms of such a slogan, "test what you teach (Hughes, 2003)." Therefore, as long as four language skills are thought, they should be tested because any skill which is not tested is simply ignored by the students. Although the weight of each skill changes from one institution to the other, it is very encouraging to see all four skills to be tested from the backwash effect point of view. However, all tests are still perfectionist, in most of which there is only one correct answer. Learners are not allowed to make any mistakes such as subject-verb agreement, article, preposition, spelling, word choice, etc. Some test items even require full answers. We cannot see any answer variability and user flexibility to respond to the test items/tasks except the writing sections. Our suggestion is that testing offices should produce items/tasks

which are more contextual and communicative in which learners will feel flexible to respond to carry out the required tasks without too much anxiety to make mistakes is in real life. The achievement tests should be designed on skill based rather than knowledge based test items/tasks.

References

Alderson, J. C. & Wall, D. (1993). Does washback exist? *Applied Linguistics,* 14, 115-129.
Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press.
Bachman, L. F. & Palmer, A.S. (1996). *Language testing in practice.* Oxford: Oxford University Press.
Bachman, L. F. & Cohen, A.D. (1998). *Interfaces between second language acquisition and language testing research.* Cambridge: Cambridge University Press.
Bailey, K. M. (1998). *Learning about language assessment: dilemmas, decisions, and directions.* London: Heinle & Heinle.
Brown, H. D. (2004). *Language assessment: principles and classroom practices.* London: Longman.
Brown, J. D. & Hudson, T. (2002). *Criterion-referenced language testing.* Cambridge: Cambridge University Press.
Cheng, L. (2005). *Changing Language Teaching Through Language Testing: A Washback Study.* Cambridge: Cambridge University Press.
Cheng, L, Watanabe, Y. & Curtis, A. (Eds.) (2004). *Washback in language testing: Research contexts and methods.* London: Lawrance Erlbaum Associates.
Ellis, R. (1985). *Understanding second language acquisition.* Oxford: Oxford University Press.

Genesee, F. & Upshur, J.A. (1996) *Classroom-based evaluation in second language education.* Cambridge: Cambridge University Press.

Hughes, A. (2003). *Testing for language teachers.* Cambridge: Cambridge University Press.

Johnson, K. (1981). Some background, some key terms and some definitions. In K. Johnson & K. Morrow (eds.). *Communication in the classroom.* London: Longman.

Krashen, S. & Terrell, T. (1983). *The natural approach: Language acquisition in the classroom.* Oxford: Pergamon Press.

Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13* (3).

Oller, J. W. Jr. (1979). *Language tests at school.* London: Longman.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics,* 10, 29-230.

Shohamy, E. (2001). *The power of tests: A critical perspective of the uses of language tests*. London: Longman.

The ALTE Framework, http://www.teachingenglish.org.uk/sites/teacheng/files/framework_english.pdf Retrieved on 07.05.2012.

Weir, C. J. (1990). *Communicative language testing.* London: Prentice Hall.