

## **All that glitters is not gold: Curriculum alignment and improving students' test scores**

Mohammad Reza Ghorbani, UPM, Malaysia

Arshad bin Abd Samad, UPM, Malaysia

Mohd Sahandri Gani bin Hamzah, UPM, Malaysia

Nooreen bte Noordin, UPM, Malaysia

Due to the centralized educational system in Iran, high school principals have directed English teachers to raise their students' achievement test scores so that they will look good on their school report card. Test scores provide the only benchmark against which the students' progress at school can be measured. Although according to the regulations teachers have to add the oral and written exam scores, and record their combination divided by two on the final report card, they usually rate their students' oral ability based on their performance on the written exam. In this study the report card average score of 47 students in grades one, two, and three at the Iranian senior high school in Kuala Lumpur was compared with that of the newly developed oral and written exam. The report card average score of the three classes was 16.37 out of 20 in comparison to the recent average score of 11.58 out of 20 which is based on the combination of the scores gained from the newly developed oral and written exam. The difference between the report card average and the recent exam average was 4.79 scores, which seems to be due to the exclusion of the oral exam scores by some teachers. The information gained from the focus group interview revealed that most teachers rated their students based on their written performance and neglected the oral exams. This may in part reflect a problem with the instruction in which the teachers teach to the test.

**Keywords:** Curriculum alignment; Test score improvement; EFL classroom; Language testing; Backwash effect.

### **1. Introduction**

The term "curriculum alignment" is usually associated with positive washback. Curriculum alignment implies that the curriculum is modified according to test results in order to improve the quality of education

(Andrews, 1994; Linn, 1983; Madaus, 1988; Shepard, 1990, 1991, and 1993). English (1992) defines curriculum alignment as the relationship between the curriculum content and the assessment tools or the relationship between what is taught and what is tested.

Due to the centralized educational system in Iran, high school principals have directed English teachers to raise their students' achievement test scores so that they will look good on their school report card. Test scores provide the only benchmark against which the students' progress at school can be measured. This paper reports that, although according to the regulations teachers have to add the oral and written exam scores, and record their combination divided by two on the final report card, they usually rate their students' oral ability based on their performance on the written exam. The paper will reflect a problem with the instruction in which the teachers teach to the test.

## **2. Background**

As Brown (2002, p. 14) states: "Washback becomes negative when there is a mismatch between the content (e.g., the material/ abilities being taught) and the test". In fact, much time of the class is always spent on materials that appear on the test. Sometimes, students like and prefer to learn English communicatively but the test they have to undergo is multiple-choice and discrete-point. Therefore, they have to focus on smaller parts of language at the expense of integrated skills. Since there is no correlation between test and curriculum objectives, this washback effect of the test on teaching and learning is negative.

The underlying theme of curriculum alignment is teacher involvement and attention to students' needs and diversity. Curriculum alignment, an extremely controversial issue due to concerns of its effectiveness, has had its supporters as well as critics. Some researchers (Wraga, 1999; Strong, Silver & Perini, 2001) maintain that curriculum becomes ineffective when it is aligned to invalid or biased assessments. Glatthorn (1999), on the other hand, argues that it can be an effective means of helping classroom teachers enable students to perform well on high-stakes tests if it is used in good professionalism.

Accountability is a means by which individuals take responsibility for their actions to assure others that there are some safeguards in place to encourage

good practices and to prevent bad practices or abuses. Wraga (1999) considers proper curriculum alignment as an effective tool for managing teachers. English (1992) also believes that accountability involves an estimate of the adherence of what is taught to what was supposed to be taught.

Teaching to the test, in its extreme form, means cheating or giving students actual questions from a secure version of a test. Commonly, it means direct preparation for a particular test. For example, teachers teach their students how to fill in answer sheets and focus their instruction on limited skills and exercises most likely to appear on the test (Kober, 2002).

As Vallette (1994) pointed out, washback is particularly strong in situations where the students' performance on a test affects their future careers and lives. In such cases, teachers often feel obliged to teach for the test, especially if their effectiveness as a teacher is in part evaluated by their students' performance. Many people think that holding teachers accountable for students' achievement will result in better education. They assume that the best data about students' improvement come from achievement tests. Although such scores are undoubtedly useful for accountability purposes, educators recognize that they have some limitations.

Many English teachers presuppose that teaching to the test is an acceptable practice, whereas there is much documentation to indicate that it is not. Although alignment of curriculum to the broad objectives of achievement tests is logical, improving students' test scores doesn't mean that their capabilities to use the language orally have improved. Because raising the test score is so often the single most important indicator of school improvement, teaching comes to resemble testing more and more. Schools send messages to their teachers about the importance of test-curriculum alignment and teachers design instruction with such alignment in mind. These effects are evident in Iranian high schools. Unfortunately, only test score improvements are taken to signal school improvement in Iran. Test-curriculum alignment and teaching to the test have influenced the meaning of scores.

The focus on educational accountability has increased pressure to raise test scores in Iranian high schools. To see: 1) what criteria English teachers use to evaluate their students' achievement, 2) to what extent the students' report

card scores indicate their achievement of the oral activities in the textbook, and 3) to what extent teachers teach and test towards the final test format, this study compared the English language scores of the Iranian senior high school students who had registered for the educational year 2006-2007 at the Iranian high school in Kuala Lumpur, Malaysia with their performance on a face to face oral exam. Based on the scores from their performance on the newly developed written and oral tests, on the one hand, and their previous report card scores, on the other hand, six students were purposively selected for the focus group interview. Before proceeding to the research method, a brief description of the Iranian educational system, English education at high schools in Iran, and Iranian schools abroad is provided in the following text to help the reader understand the research context.

### 3. Research Context

Iran's educational structure is in operation with the following duration and age ranges:

<b>Level</b>	<b>Duration</b>	<b>Age</b>
1) Primary school	5 years	6-11
2) Junior high school	3 years	11-13
3) Senior high school	3 years	13-16
4) Pre-university	1 year	16-17
5) University (BA)	4 years	18-21

The school system is under the jurisdiction of the Ministry of Education. In addition to schools, this Ministry also has responsibility for some teacher training and some technical institutes.

Students in grade three at junior and senior high schools will not graduate or be promoted unless they pass a test at the provincial or national level. From time to time the exam for the first and second grade students is held at the provincial or national level. Teachers are being held accountable for ensuring that their students meet academic standards and pass the tests. Principals are being held accountable for raising average test scores each year at the school, district, and provincial levels. Schools with rising test scores are often praised and receive rewards, while those with falling scores are severely criticized. In such a pressure-cooker environment, many teachers and principals feel they have to use all legitimate means to improve student' test scores.

In Iran, English is taught as a foreign language (EFL) and is practiced within a context-restricted environment, in which the determiners of language learning phenomenon depend on classroom activities, determined by the textbook and classroom teacher. English teaching starts from the first grade of grade of junior high school. All high schools follow the curriculum standards. The Ministry of Education compiles, develops and publishes textbooks and teaching materials, for nationwide public and private high schools inside and outside the country. Most high school English teachers implement the grammar translation teaching method in their classrooms to meet the expectations of the national curriculum and exams whose format and structure are prescribed by the Ministry of Education.

The Iranian schools around the world have been established by the Islamic Republic of Iran so that Iranian children living abroad will be encouraged to follow their studies at Iranian schools. These schools are usually less populated than the schools inside the country. Although the teaching and learning situation is different from one country to another, the Iranian educational system, rules, standards and textbooks are prescribed by the Ministry of education for Iranian high schools abroad, too. The Ministry of Education sends specialized and efficient education manpower to these schools.

#### **4. METHOD**

##### **4.1 Sample and Location of the Study**

"It is not always possible to use probability sampling in educational research. Instead, a researcher can use non-probability sampling. In non-probability sampling, the researcher selects individuals because they are available, convenient, and represent some characteristic the investigator seeks to study" (Creswell, 2005, p. 149). In this study, all high school students who were studying at the Iranian high school in Kuala Lumpur in 2006-2007 were studied. There were 47 students at the senior high school level, 17 students in grade one, 16 students in grade two, and 14 students in grade three. They came from different parts of Iran. They were, therefore, a microcosm of the whole population, that is, they represented Iranian senior high school student population.

As to the sampling procedure for the focus group interview, based on the scores from their performance on the newly developed written and oral tests, on the one hand, and their previous report card scores, on the other hand, six students were purposively selected. Two students, whose oral and written scores had the most difference and their report card scores were more than the combination of the newly developed written and oral tests, were chosen from each class. Three of the selected students were boys and three of them were girls, that is, one boy and one girl were selected from each class. The participants were volunteers and willing to discuss the topic without force.

Krueger (1994) suggests four to five participants in each focus group interview because small groups supply more opportunity for participants to share ideas and provide for interaction among interviewees. Creswell (2005) suggests four to six participants and believes that "Focus groups are advantageous when the interaction among interviewees will likely yield the best information and when interviewees are similar to and cooperative with each other" (p.215). The group size should be small enough to allow each participant to share insights, but large enough to provide diversity of information. In this study, a group of six participants was interviewed.

#### **4.2. Materials and instrument**

Three of the current English Language Teaching textbooks *English Book 1, 2, and 3* (Birjandi, 2006, 2005, and 2003) that are locally designed to cater for and respond to the English language needs of Iranian students in grade one, two and three at senior high schools were taken to serve as the corpus of the present study. Three achievement tests based on the content of the textbooks were used to measure students' written performance. The one-on-one oral exam was also based on the content of the textbooks as recommended by the Ministry of Education (See Appendix A).

Before conducting the semi-structured focus group interview, an interview guide was developed (See Appendix C) to direct the conversation toward the desired topics and issues. It helped the researcher know what to ask about, in what sequence, and how to pose his questions.

The semi-structured focus group interview was employed in this study to obtain rich information to help interpret the meaning of the data which were collected from the students' scores on the report card and the newly developed oral and written exam. The questions for the focus group

interview were developed based upon the recommendations from Krueger (1994). Two high school students helped pilot the focus group interview guide in order to see whether the questions and the process on the guide were appropriate. Based upon the result of the pilot test, the final focus group interview guide was developed.

### **4.3. Procedures**

This research was done in two phases. In the first phase, three written achievement tests which had already been developed and validated based on the regulations and circulations of the Iranian Ministry of Education were used to measure students' written performance in grade 1, 2 and 3 at the Iranian senior high school in Kuala Lumpur in the beginning of the Iranian 2006-2007 school year. Students' performance was measured through a group-administered test. The students in the 3 classes were tested orally one by one and assessed by three raters. The written and oral exam included items based on the recommendations and guidelines circulated by the Ministry of Education (See Appendix A).

During the oral test administration, which took approximately 5 minutes for each student, the examiner asked the student a series of questions in English that were part of the testing materials. The student's reply was compared to a suggested answer and scored on the answer sheet as either correct or incorrect. The total score was used to assign students to an instructional level for conversation classes, but could also be expressed as a grade equivalent to be used for this study.

Promotion through the Iranian education system is based on middle- and end-of-term examinations. At schools, system of grading is based on a 0-20 scale. An average scale of at least 10 is required for promotion. Grading for the English subject is based on a 0-40 scale divided by 2 in which the outcome is the same as grading in the other subjects. So what appears on the students' report card is their score out of 20 (5 scores from the oral test + 15 scores from the written test). The third grade students' final exam scores is recorded only based on their written exam performance.

A comparison of the scores from the students' report card which is the combination of their oral and written exams rated by their previous teachers and the scores from the two newly administered oral and written tests was made through. The score of each student in grade one, two, and three on each

exam has been tabulated in tables 1, 2, and 3 (See Appendix B). After determining the average of each grade, the average of the whole students' scores in the three grades was also computed, tabulated, and compared as follows.

**Table 4.**

*Comparison of the average score of each grade on the written and oral test*

Grade	MEANS			
	Report Card (out of 20)	Written (out of 30)	Oral (out of 10)	Written+Oral (out of 20)
1	17.29	25.60	5.07	15.33
2	17.18	26.77	5.85	16.31
3	14.64	22.12	4.08	13.10
Total	49.11	74.49	15	34.74
Mean	16.37	24.83	5	14.91
SD	1.49	2.41	0.88	1.64

In the second phase, based on the scores from their performance on the newly developed written and oral tests, on the one hand, and their previous report card scores, on the other hand, six students were selected purposively for a focus group interview. Two students, whose oral and written scores had the most difference and their report card scores were more than the combination of the newly developed written and oral tests, were chosen from each class.

The following suggestions and recommendations which were made by Krueger (1994), Ary, et al. (2002), and Creswell (2005) were followed to gather data through the interview.

- 1) A quiet, suitable place was located for conducting the interview.
- 2) Consent was obtained from all of the participants before the interviews.
- 3) A focus group discussion guide, including the questioning route, moderator's guide, and discussion outline, will be developed in advance to provide the direction for group discussion.
- 4) The moderator took the keynotes and the assistant moderator took the detailed notes throughout the discussion, including notes on participants' body language.
- 5) All of the interview session was tape recorded in order to avoid

missing the interviewees' comments.

- 6) The participants were thanked in a courteous manner after concluding the interview.

The note-based content analysis was used to analyze the focus group interview. The raw data for the note-based content analysis relied primarily on 1) filed notes which were based on observations and comments in the interview, 2) a debriefing session, and 3) summary comments at the conclusion of each interview. The use of tape was primarily to verify specific quotes and to translate the oral summary at the conclusion of the interview. The analysis was based upon the systematic note-based analysis adapted from Krueger (1994) with some modifications.

After the interview, the draft report for each interview was sent to each of the interviewed students to check whether the content was valid. A peer expert with a background in research was asked to review the frequency of counting and data interpretation in order to check the researcher's accuracy in this analysis.

## **5. Results and Discussion**

Table 4 displays the average scores of students in grade one, two, and three. The performance of students on the written tests is much better than their performance in the oral exam. The report card average score of 47 students in the three classes was 16.37 out of 20 in comparison to their recent average score of 11.58 out of 20 which is based on the combination of the scores gained from the newly developed oral and written exam. The difference between the report card average and the recent exam average was 4.79 scores. In fact, this reduction is because of the addition of the oral exam score and then dividing the sum of the written and oral exam scores by two.

The fact that scores on the written tests are much higher than oral scores seems to be logical because it implies that the teachers rated their students based on their written performance and neglected or didn't hold any oral exams. Although according to the regulations they have to add the oral and written exam scores, and record their combination divided by two on the final report card, they rate their students' oral ability based on their performance on the written exam.

As it can be seen in tables 1, 2, 3 (See Appendix B), and 4, it is clear that the oral skills of students are considerably less developed than their written skills. The average of the oral exams in all grades is less than the written exams. The combination of the recent oral and written exam scores is less than the report card scores which is supposed to be the combination of previous written and oral exam scores. The tables indicate that if the oral exam scores are neglected and students are rated based on their performance on the written test, there will be little difference between the previous and recent marks. But if the combination of the recent oral and written exam scores is compared with the previous exam scores, the recent scores in all cases, due to the oral exam, are less than the previous ones.

As a result, the low rates in the oral exam may in part reflect a problem with the instruction in which the teacher teaches to the test. The oral parts of the textbooks are not tested at the provincial or national level. That is why teachers narrow down the content of the textbooks to what shows up on the test. "The only reasonable, direct inference you can make from a test score is the degree to which a student knows the content that the test samples" (Mehrens, 1984, p. 10). Students' lack of ability to use the language orally indicates that test-curriculum alignment and teaching to the test are distorting instruction in Iran.

Out of the six students who took part in the focus group interview, four of them stated that they were not tested orally, one of them said that he had been tested orally but only her written exam score was reported on the report card, and one of them stated that she was tested orally and her report card score was a combination of the written and oral exam.

Three students said that their teachers skipped the oral activities in the book and asked the students to learn them only for the written exam. Three of the interviewed students stated that they just repeated oral activities after their teachers but their teachers didn't put much emphasis on them in comparison to the writing and reading exercises.

All of the interviewed students believed that they were weak at using oral activities of the textbook and their report card scores could be considered only to indicate their written performance. Only one of them stated that there was an oral exam, the score of which was added to the written exam score

and the combination of the two divided by two was reported on the report card.

A girl student (M. A.) commented that she agreed with her teacher to teach to the test because her parents didn't care about learning and only the exam score was important for them. She said:

Within two or three years we are supposed to sit for the university entrance exam which is only held in a written form and affects our future career. Our parents expect us to pass the exam, so we expect our teacher to teach to the test.

Another student (N. M.) said:

Our teachers and we know that it's very important to improve our oral skills but the final evaluation system at the provincial or national level makes us spend more time on those parts of the book that are most likely to show up on the final exam.

Two of the students commented that their written performance on the exams at the provincial and national level was supposed to be the main benchmark to evaluate their achievement and their teachers' success in teaching. They believed that teachers should not be blamed for teaching to the test because they were doing the right thing to protect themselves and their students against the external pressures, such as parents, principals, etc.

Depending on how it is done, teaching to the test can be either productive or counterproductive. It can alter the interpretation of test scores because it involves teaching specific content. Therefore, the direct inferences that can be reasonably drawn from test scores are weak and general in nature, and inaccurate if instruction is limited to the actual objectives sampled in the test or, worse yet, to the actual questions on the test. The temptation exists to teach too closely to the test and teachers are pressured to do so. In fact, the outcome is high scores on a paper and pencil test at the expense of listening and speaking abilities.

The ultimate purpose of any test is to improve teaching and learning. Some forms of test preparation are very far from this ideal, while others move closer toward it. As a general guideline, a test preparation practice is inappropriate if it raises test scores without also improving students' mastery

of the broader subject being tested. Higher test scores do not necessarily mean more learning.

## **6. Conclusion**

As the results show, it seems that English teachers neglect the oral exam and rate their students based on their performance on the written exam. Because they are not accountable for the oral exam and students are not tested orally at the provincial or national level, they don't pay much attention to it. They narrow down the content of the textbooks to what shows up on the test. In a nutshell, they do so because of the washback effect of the test.

Teaching to the test stifles creativity and encourages cheating. Schools that don't do well on tests forfeit thousands of Iranian Rials in reward money. One indicator of a teacher's overall evaluation is based on test scores, a significant motivator for teaching to the test. This measure is one of main factors in evaluating teachers which will force them to spend inordinate amounts of time teaching to the test and teaching test-taking skills.

In Iran, school principals and teachers can pay a heavy price if they do not forge a strong link between what is taught in classrooms and what is tested at the provincial or national level. Doing well on the provincial or national tests can mean generous cash rewards.

Instead of using test results to penalize schools, tests should become stronger tools for improving schools. Teaching to the test is exactly the right thing to do as long as the test is measuring what students are supposed to learn.

Chances are the teachers and principals are honest people who would not use dishonest tactics to raise scores on tests. But when the stakes are raised on testing, the incentives to cheat grow much larger. Cheating is inexcusable. This is an area where the Ministry of Education needs to be extremely vigilant.

When inappropriate or questionable forms of teaching to the test do occur, it's usually the teachers who are blamed. The real fault for inappropriate forms of "teaching to the test" lies not so much with teachers, but with policymakers who have created accountability systems centered on higher test scores, with little regard for how these scores are attained.

Policymakers and education officials must recognize that they can't keep demanding higher test scores without also paying attention to how their demands affect teaching and learning, in both positive and negative ways. They have a responsibility to help prevent bad or questionable forms of teaching to the test and encourage good practices. They should 1) provide teachers with various kinds of support to help them improve classroom teaching, 2) make sure that provincial and national tests are valid, fair, reliable, and well aligned with standards and curriculum, 3) revise testing systems to include a variety of item formats that measure both basic and more advanced knowledge and skills, 4) design accountability systems that reward good teaching and encourage attention to all important standards and subjects, and use other means to assess progress in standards and subjects not covered by tests, 5) provide teachers with professional development in effective ways to help students master the content in standards, 6) strengthen the handling of testing issues in teachers' pre-service education and professional development, 7) provide students with the extra time and attention they may need to master the academic knowledge and skills contained in standards, which may entail more intensive instruction during the school day, or time after school or during the summer, 8) systematically monitor the intended and unintended effects of testing and accountability, 9) watch out for possible score inflation by tracking trends in tests, questioning miraculous gains in teacher-made tests, and comparing test scores with scores on other assessments and measures of progress.

### ***Acknowledgments***

The writing of this paper was obviously not possible without the personal and practical support of numerous people. Thus I would like to deeply thank the various people who provided me with useful and helpful assistance during my data collection at Iranian school in Kuala Lumpur. I would especially like to thank Dr. Arshad, associate professor Dr. Sahandri, and Dr. Nooreen at the Faculty of Educational Studies, UPM, Malaysia for providing me with nice feedback and invaluable insights.

### ***The Authors***

Mohammad Reza Ghorbani ([mrg872@yahoo.com](mailto:mrg872@yahoo.com)) is TESL Ph.D. candidate at the Faculty of Educational Studies, UPM, Malaysia. He got an AA in TEFL from

Shahid Beheshti Teacher Training College at Mashhad in 1990, a BA in English Literature from Tabriz University in 1995, and an MA in TEFL from Isfahan University in 1999.

Arshad bin Abd Samad ([arshad@educ.upm.edu.my](mailto:arshad@educ.upm.edu.my)) is Deputy Dean of Student & Alumni Affairs and Professor of Educational Studies at UPM, Malaysia. Arshad bin Abd Samad got a B.S from Indiana University, an M.A. from Indiana University, and a Ph.D. from Northern Arizona University.

Mohd Sahandri Gani bin Hamzah ([sahandri@putra.upm.edu.my](mailto:sahandri@putra.upm.edu.my)) is professor of Measurement at UPM, Malaysia.

Nooreen bte Noordin ([nooreen@educ.upm.edu.my](mailto:nooreen@educ.upm.edu.my)) is Professor of Measurement at UPM, Malaysia.

## References

- Andrews, S. (1994). *Washback or washout?* The relationship between examination reform and curriculum innovation. In D. Nunan, V. Berry & R. Berry (Eds.), *Bringing about change in language education* (pp. 67-81). Hong Kong: Department of Curriculum Studies, University of Hong Kong.
- Ary, D., Jacob, L. C., & Razavieh, A. (2002). *Introduction to research in education*. 6th ed. Australia : Wadsworth,.
- Birjandi, P., Soheili, Noruzi, M., & Mahmodi (2006). *English Book 1*. Iranian Textbook Publication and Distribution. Tehran, Iran.
- Birjandi, P., Soheili, Noruzi and Mahmodi (2005). *English Book 2*. Iranian Textbook Publication and Distribution. Tehran, Iran.
- Birjandi, P., Soheili, Noruzi and Mahmodi (2003). *English Book 3*. Iranian Textbook Publication and Distribution. Tehran, Iran.
- Brown, J. D. (2002). *Questions and answers about language testing statistics: Extraneous variables and the washback effect*. Shiken: JALT Testing & Evaluation SIG Newsletter, 6(2), 12-15: <http://www.jalt.org>.
- Creswell, J. W. (2005), *Educational Research – Planning, Conducting and Evaluating Quantitative and Qualitative Research*. Pearson, New Jersey.

- English, F. W. (1992). *Deciding What to Teach and Test: Developing, Aligning, and Auditing the Curriculum*. Newbury Park, CA: Corwin Press.
- Glatthorn, A. A. (1999). Curriculum alignment revisited. *Journal of Curriculum and Supervision*, 15(1), 26-34.
- Kober, N. (2002). *Teaching to the Test – The Good, the Bad, and Who’s Responsible*. US: Test Talk for Leaders, Center for Education Policy
- Kruger, R. A. (1994). *Focus groups*. (2nd ed.). California: Sage Publications, Inc.
- Ligon, G. D., & Jones, P. (1982). *Preparing Students for Standardized Testing: One District's Perspective*. Paper presented at the annual meeting of the American Educational Research Association, New York.
- Linn, R. L. (1983). Curricular validity: Convincing the court that it was taught without precluding the possible of measuring it. In G. F. Madaus (Ed.), *The courts, validity, and minimum competency testing* (pp. 115-132). MA: Kluwer-Nijhoff.
- Madaus, G. (1988). The influence of testing on the curriculum. In L. Tanner (Ed.), *Critical issues in curriculum: 87<sup>th</sup> yearbook of the NSSE, Part 1* (pp. 83-121). Chicago: University of Chicago Press.
- Mehrens, W. A. (1984). "National Tests and Local Curriculum: Match or Mismatch?" *Educational Measurement: Issues and Practice*, 3, (3), 9-15.
- Shepard, L. A. (1990). Inflated test score gains: Is it old norms or teaching the test? *Educational Measurement: Issues and Practices*, 9(3), 15-22.
- Shepard, L. A. (1991). Psychometricians' beliefs about learning. *Educational Researcher*, 7, 2-16.
- Shepard, L. A. (1993). The place of testing reform in educational reform: A reply to Cizek. *Educational Researcher*, 22, 10-14.
- Strong, R. W., Silver, H. F., & Perini, M. J. (2001). Making students as important as standards. *Educational Leadership*, 59(3), 56-61.
- Valette, R. M. (1994). Teaching, testing, and assessment: conceptualizing the relationship. In C. R. Hancock (Ed.), *Teaching, testing, and assessment*. Lincolnwood: National Textbook Company.

Wraga, W. G (1999). The educational and political implications of curriculum alignment and standards-based reform. *Journal of Curriculum and Supervision*. 15(1). 4-25.

## Appendix A

The written and oral exam included items based on the following recommendations and guidelines circulated by the Ministry of Education.

### First and Second Grade Guidelines

#### *1. Final written exam (30 scores) which includes:*

#### 1. Spelling

One or two missing letters in 12 words used in sentences will be completed by the student 3 Scores

#### 2. Vocabulary which includes:

Fill in the banks with the words given. 4 scores

There is one extra word. (8 sentences and 9 words)

Fill in the blanks with your own knowledge. (2 sentences) 1 score

Choose the best answer. (4 multiple-choice questions) 2 scores

#### 3. Reading and sentence comprehension

Read the text and answer the following questions. 4. scores

(Completion,

Multiple choice, True/ False, Wh-questions, Yes/ No questions)

The following sentence implies that or according to the sentence . . . 3 scores

The above structure is followed by multiple choice questions

#### 4. Structure

1. Multiple choice items. 2 scores

2. Put the scrambled words in order. 2 scores

3. Complete the incomplete sentences. 2 scores

4. Look at the picture and answer the questions. 2 scores

#### 5. Language functions

Complete the incomplete dialogs. (From the textbook) 3 scores

#### 6. Pronunciation

1. Look at the example and put the words under the right columns. 1 score

2. Which word is different?

***II. Final oral exam (10 scores) which includes:***

1. Reading	
1. Fast reading	1 score
2. Pronunciation and stress	1 score
3. Fluency	2 scores
4. Comprehension	2 scores
2. Retelling and summarizing the reading text	
The student will summarize the text after reading	1.5 scores
3. Language functions	
The student will act out the dialogs and conversations in language functions section	1.5 scores
4. Pronunciation	
The student will pronounce the words which are in the pronunciation section	1 score
Total: 40 scores (30 scores for written test +10 scores for oral test)	

**Third Grade Guidelines**

***III. Final written exam (40 scores) which includes:***

1. Spelling	
One or two missing letters in 12 words used in sentences will be completed by the student	4 Scores
2. Vocabulary which includes:	
Fill in the banks with the words given.	4 scores
There is one extra word. (8 sentences and 9 words)	
Fill in the blanks with your own knowledge. (3 sentences)	3 score
Choose the best answer. (6 multiple-choice questions)	3 scores
3. Reading and sentence comprehension	
Read the text and answer the following questions. (Completion, Multiple choice, True/ False, Wh-questions, Yes/ No questions)	4. scores
The following sentence implies that or according to the sentence....	4 scores

The above structure is followed by multiple choice questions

Cloze passage 3 scores

4. Structure

1. Multiple choice items. 3 scores

2. Put the scrambled words in order. 2 scores

3. Complete the incomplete sentences. 2 scores

4. Look at the picture and answer the questions. 2 scores

5. Language functions

Complete the incomplete dialogs. (From the textbook) 4 scores

6. Pronunciation

1. Look at the example and put the words under the right 2 score  
columns.

2. Which word is different?

Total: 40 scores

### Appendix B

**Table 1.**

*The scores of the students in grade one at the Iranian senior high school in Kuala Lumpur*

Student	Report Card	Written	Oral	Written+Oral
1	19	27	5.5	16.25
2	20	29.25	6.25	17.75
3	16.5	22.5	4	13.25
4	18.25	25.75	5.25	15.5
5	20	30	9	19.5
6	13.5	20	2.5	11.25
7	14	21	3	12
8	14.75	21	3	12
9	17	25.5	3.5	14.5
10	16.25	22.75	3.75	13.25
11	16	23	3	13
12	19.5	30	9	19.5
13	20	30	9	19.5
14	15.25	24.5	3.5	14
15	18.75	28.5	5.5	17
16	19.5	30	6.5	18.25
17	15.75	24.5	4	14.25
Total	294	435.25	86.25	26.75
Average	17.29	25.60	5.07	15.33
Std	2.22	3.55	2.21	2.81

**Table 2.**

*The scores of the students in grade two at the Iranian senior high school in Kuala Lumpur*

Student	Report Card	Written	Oral	Written+Oral
1	15	25.5	4	14.75
2	19	30	7	18.5
3.	12.75	19.25	3.25	11.25
4	20	30	10	20
5	17.5	28	5.5	16.75
6	20	29.5	8	18.75

7	19	30	8	19
8	16.25	25	5	15
9	14.25	20.75	3.75	12.25
10	18	27.5	6	16.25
11	18.5	29	6	17.5
12	15.75	26.5	4	15.25
13	17	27	4.5	16.75
14	13.5	20.75	3.25	12
15	19	29.5	7	18.25
16	19.5	30	7.5	18.75
Total	275	428.35	92.75	261
Average	17.18	26.76	5.79	16.31
Std	2.35	3.62	1.98	2.69

**Table 3.**

*The scores of the students in grade three at the Iranian senior high school in Kuala Lumpur*

Student	Report Card	Written	Oral	Written+Oral
1	18.25	28.75	7.25	18
2	18	27	6	16.5
3.	15.5	23	3.5	13.25
4	15	23	3.5	13.25
5	12.25	20	3	11.5
6	10.75	17	2.5	9.75
7	20	30	7.5	18.75
8	11	16.75	2.75	9.75
9	17.5	24.5	4	14.25
10	12	20	2.5	11.25
11	11.75	17.5	2.5	10
12	19.5	28	7	17.5
13	13	20.5	3	11.75
14	10.5	13.75	2.25	8
Total	205	309.68	57.25	183.5
Average	14.64	22.12	4.08	13.10
Std	3.45	5.03	1.95	3.44

## **Appendix C**

### **Focus Group Discussion Guide**

- I. Introduction: Welcome and introduce names.
- II. Have the participated students complete the "Participation Information Form" with observed background information.
- III. Students' opinion about high school English teachers' assessment.
  1. What were your previous teacher's criteria for measuring your achievement? Probing questions: such as 1) written, 2) oral, 3) weight of each part.
  2. To what extent do you think your report card score indicate your achievement of the oral activities in the textbook?
  3. To what extent do you think your teacher taught and tested you towards the final test format?
- IV. Summary and conclusion
  1. Is there any other idea/point that you would like to comment or share with us?
- V. Appreciations