The Speaking Test of Korea's NEATs 2 & 3: Its Usability for University Admission Qualifications

WonKey LEE

(Seoul National University of Education)
The Speaking Test of Korea's NEATs 2 & 3: Its Usability for University Admission Qualifications

WonKey LEE
Seoul National University of Education

Abstract
Korea's experiment of using an internet-based speaking test of English for university admission qualifications is unprecedented and hence still controversial. In this paper, how to test and rate test-takers' speaking performance of English by internet is briefly discussed, and the possibilities and problems of using this speaking test for university admission qualifications are discussed.

1. Introduction

The human society keeps changing. No one can stop the human society changing. In this ever-changing society, a great deal of new knowledge is born every day, and also a great deal of old knowledge becomes obsolete. In this knowledge & information-based society, all kinds of knowledge and information are everywhere. It is thus crucially important to learn where the knowledge and information that "I" need for my purpose actually exist. That is, "know-where" is as important as, or even more important than, "know-how," in that know-where oftentimes comes before know-how.

It is often said there are two different views of knowledge: Objectivism and Constructivism. In objectivism, knowledge is viewed as objective entity, existing outside the individuals, and individuals should learn it by their efforts. On the other hand, in constructivism, the value of a piece of particular knowledge is constructed by individuals themselves, not given by others. In the modern society,

---

1 President of Korea English Language Testing Association, and Professor of Seoul National University of Education in Korea. (wkl@snue.ac.kr)
the constructivism becomes more significant, reducing the objectivism to be less valid.

As science and technology, traffic and communication means are making startling progress, interaction and communication opportunities between people have increased dramatically. The opportunities for using English in the form of oral interaction have strikingly increased, and those increased opportunities demand a lifelong endeavor to learn English of even the grassroots.

In this context, the conventional ways of teaching and testing of English have turned out to be ineffective in making students competent English speakers. English language teaching (hereafter ELT), especially in an EFL environment, is rarely efficient and effective because the ELT in EFL contexts is usually fraught with a variety of unfavorable conditions. In other words, TENOR (situation), and HCLO (structure) are most common. Here TENOR stands for 'Teaching English for No Obvious Reasons', and HCLO stands for 'High Cost, Low Outcome.'

Despite the unfavorable circumstances, to reform this kind of TENOR situation and HCLO structure of ELT more efficiently and economically, a variety of communicative teaching methods are now being actively employed in Korea. Together with these efforts for improved ELT, the washback effect of testing is seriously experimented now in Korea.

There are some exemplary cases by which teaching was improved or reformed by the effect of testing. Here are some examples;

First, it is reported that the Oral Bagurt Examination Project improved ELT of Israel by implementing a speaking test to schools (Gefen & Vorhaus 1989). Second, the Nationwide English Listening Test Through Radio Broadcast for Middle and High Schoolers in Korea (started in 1983) is a classic case. When the results of the National Listening test started to be used as one of the key information for high school entrance in 1986, teachers, students, parents and administrators suddenly and seriously took great interest in this listening test, and since then listening of English started to be actively taught at schools as well as at private institutes. Third, the College Scholastic Ability Test (CSAT:
大学修学能力試験) is also an exemplary case for positive washback effects. The CSAT is currently the most high-stakes exam for high schoolers, of which results are being used as university entrance qualification data. This paper-and-pencil test includes a listening test section. Since the implementation of this test, the ways of school teaching are divided into two: the one before the CSAT, and the one after the CSAT. It is generally evaluated that school teaching after the CSAT faced a fundamental change in the teaching methods in the positive direction.

2. Introducing the NEAT

The Korean government is now determined to reform the conventional education which has been evaluated as not satisfactorily functioning for the future knowledge & information-based society. As one of those efforts, the current Lee Myungbak government (2008-2013) revised the national curriculum which is the backbone foundation for school education in Korea. As a result, teaching hours of primary English were increased, teacher education, both pre-service and in-service, was much intensified, and a national test of English for high schoolers is to be introduced, and more.

In this paper, the last one, the national test of English, will be focused on. An initial serious discussion about introducing a national English test was made in 1997, when Korea was experiencing a big economic turmoil caused by the shortage of the foreign currency reserves. There were lots of outcries voiced high about the outflow of foreign currencies, which was considered vice at that time. The target of criticism was the considerable amount of foreign currencies for a variety of international tests of English Korean test-takers had to pay.

With this kind of a bit nationalistic emotion and future needs in mind, the then-government decided to introduce a new national test of English to hopefully replace some major international tests of English such as the TOEFL or TOEIC.

The first model drafted in 1998 was a five-level one which ranges from grade schoolers to high schoolers. But in 2008 it was reformed into a 3-level test for high schoolers and university students, and renamed the National English
Ability Test (NEAT). The NEAT Level 1 is meant for university students, and NEATs 2 & 3 are meant for high schoolers. To help cut and reduce the parents' financial burdens on private tutoring, middle and grade schoolers are excluded from the NEAT.

The NEAT is a brainchild of the current government's new English education policy: Public School English Education Reinforcement Policy (PSEERP). It was announced that a key purpose of the NEAT was to activate the teaching of speaking and writing of English at schools, which has not actually been conducted despite its compelling necessity. Teaching of the speaking and writing of English at schools had not been active. Even though the national curriculum has encouraged teaching of all four skills in a balanced way, but it was not materialized for many practical reasons, such as;

- few societal needs for speaking and writing ability so far
- English teachers' low competency and insufficient professional training
- university entrance-related national tests did not include speaking and writing tests mainly for reliability and practicality reasons

The current College Scholastic Ability Test of which scores are mainly used as one of the important determiners for university admission, has been mostly focused on listening and reading of English. No wonder that the CSAT has been often criticized. It is actually limited in improving students' communicative ability. One of the main reasons for this limitation is, it is often argued, that it does not include speaking and writing tests.

To rectify these limitations, a new test of English which includes speaking and writing tests has been called for for a long time.

1) Needs for a large-scale test
To make this new national test have a substantial impact on ELT, it is attempted to replace the current College Scholastic Ability Test with the NEATs 2 & 3. To use
this test result as an important element of university entrance qualification data, a large scale administration is a must. The total number of university entrance candidates is approximately 600,000 plus every year. A large scale test is needed to accommodate these 600,000 plus candidates.

2) Needs for an internet-based test
To assess a speaking ability in a large scale, there seemed no other better way than using computers. So it was decided that the best possible option was an internet-based test, because fortunately internet infrastructures are quite well established in Korea, and also the internet-based testing was considered a future-oriented method.

If speaking ability is tested by means of computer, the interactiveness of speaking cannot effectively be assessed. Only unidirectional speaking, ie. oral statement which is only part of interactive oral communication, becomes the target of assessment. This means the (construct) validity of the speaking test is considerably dented. Despite this limitation, the internet-based test method seems inevitable in administering a large scale speaking test to 600,000 plus test-takers. In brief, if a large scale test of English speaking is conducted despite some theoretical and practical limitations, the washback effect of the large-scale national test will definitely lead schools to teach the speaking of English, and in the long run, students’ English speaking proficiency will go up generally. In sum, Internet-based speaking tests of English may inevitably reduce the validity of the speaking tests for its inherent lack of interactiveness, but it seems reasonable to opt for the internet-based speaking test of English, even if it is not very best option. There seems to be a strong trust in the washback effect.

3. The Nature of NEAT 2 and 3
The NEATs 2 and 3 are based on the national curriculum, thus the two NEATs are basically an achievement test. However, if we consider the nature of the English language, we can assume that, when students achieve the objectives of
the national curriculum, they will gain proficiency in English to the extent that they achieve the objectives of the curriculum. Thus the NEATs 2 and 3 can be viewed as an achievement test as well as a proficiency test, that is, the nature of the NEATs 2 & 3 may be defined as kind of a prochievement test.

For the NEATs, six pilot tests were conducted for the last three years (2009-2011) and the basic format of the test has been determined. The most distinctive feature of the NEATs 2 & 3 is that they include speaking and writing tests which the CSAT does not include. Listening and reading tests of the NEATs are almost the same as the current CSAT in format and method of testing. Listening and reading tests are given in the multiple choice question type. Another characteristic of the NEATS is that grammar is not explicitly tested in order to give more weight on the fluency aspect of communication in English.

A distinction is officially made between these two separate tests: NEAT-2 and NEAT-3. The NEAT-2 is basically a CALP test; a higher level basic academic English-oriented test. This is meant for the test-takers who want to specialize in English-related, or English-using subjects in university.

On the other hand, the NEAT-3 is a BICS test; a lower level practical English-focused test. This is meant for the test-takers who want to find their ways in areas which are less English-related when they are admitted into university.

These two tests are designed to report their test results in four grades: A, B, C and F. Here 'F' means a failure, and thus the tests have only three grades. This crude grading of the test results is originally meant not to encourage a tough competition among students, and not to arouse an excessive level of test-anxiety. This 3-grade reporting approach may cast some doubt about the usefulness of the test results. Obviously if the test results of only three grades are used as a university entrance qualification data, discrimination between test-takers cannot be effectively made, resulting in the test itself being powerless.

Considering the intense competition in university entrance in Korea, the inability to discriminate test-takers will definitely make universities employ a more discriminating device other than the NEATs, unwantedly causing out-of-
school private tutoring to thrive, which has been restrained by the government with a maximum effort.

4. The Speaking Tests of NEATs 2 & 3

Two essential elements of speaking tests are elicitation and marking. In other words, how to elicit speech from test-takers and how to mark objectively are the two key ingredients of speaking tests.

1) Elicitation

To conduct a speaking test, first of all a ratable amount of speech sample should be elicited from the test-takers. The objectives and contents of the NEATs 2 & 3 speaking test are based on the national curriculum, and the nature of speaking asked of test-takers is of unilateral speaking, which is not interactional one. This is mainly because the NEATs are given via Internet through the medium of computer. The current level of computer technology, despite its dazzling advancement, is not sufficiently satisfying in presenting speaking test items which successfully incorporate unpredictability, interactionality of natural communication. So the internet-based speaking test of English may inherently have a limitation in that it can only test non-interactional oral statement about the given prompt materials, or questions.

For item writing, the item writer accreditation system is employed; Secondary school teachers voluntarily apply to be recruited as prospective writers, and 60-hour off-line item writing training is given by the specially-mobilized training team. Only the successful completers of the training program are qualified and accreditated as official item writers.

The plans of securing NEAT item writers are; 80 item writers were accreditated in 2010, 240 item writers in 2011, and 720 item writers will be accreditated in 2012 accumulatively.

There are two ways of item writing in the NEAT:
(1) the "stay-together" item writing (合宿制)
Qualified item writers are gathered in a place and stay shut off from all outside contacts until item writing is completed. Fees for items and for staying away from home are paid.

(2) the "invite & screen" item writing (公募制)
Qualified item writers enter test items, which will be screened by the test item screening committee. Fees for items chosen are paid.

2) Rating
For rating of speaking and writing, there are two types of rubrics used:
One is a general rubric, and the other is a task-specific rubric. A general rubric is the assessing criteria for general or overall ability regardless of particular items. A task-specific rubric is assessing criteria meant for each particular item.

The NEATs are a national curriculum-based achievement test, so each test item has its own task-specific rubric. It is because it is not easy to check if the test-takers have achieved the national curriculum objectives with the general rubric.

For rating, a rater accreditation system is also used, which is similar to the item writer accreditation system. School teachers who want to work as a rater can apply to recruited, and get trained both on-line and off-line in rating, and qualified and accredited.

The NEATs employ a multiple rater system: Two raters are employed for each test-taker. In the case that there is a big discrepancy between those two raters, a third person who is more experienced gets involved and determines the final grade.

5. Points to be seriously considered
Despite the necessity and demand of speaking test in the NEAT, there are some points to be seriously considered in the speaking test of the NEATs 2 and 3.
In validity, there is clearly a limitation in that it is only an oral statement test, not an interactive oral communication test. This seems inevitable because a large scale test is needed.

In reliability, the use of a task-specific rating scale, and the rater accreditation system may be of help in securing the inter-rater reliability, but the degree of raters' subjectivity in rating may cause a litigation later on when the national test becomes a high-stakes test by being used as one of the data for university entrance qualifications.

Practicalitywise, there can be some technical problems, but they may be minor, considering the level of Korea's technology. However, the need for a large scale test may cause no-small problems. The test has to accommodate over 600,000 test-takers on the pre-established 1,700 test sites. Currently only about 50,000 test-takers can take the test at one time. Tests should be given about 12 times a year. And the test-takers can take the test twice a year before they apply for a university, there will be 24 times of testing a year. In this case, equating of the 12 different versions of the test can be a big problem. Equating between 12 different versions may be not impossible theoretically, but practically it might be quite daunting, because it is difficult, or almost impossible to keep the same facility, the same discrimination between 12 different versions of a test.

In reporting of the test results, 3 grades may not be very useful for the highly competitive university entrance. Little discrimination between test-takers may cause to increase private sector education, resulting in more burden to students and parents. These points should be carefully taken care of if the NEAT speaking test can be used for university admission qualification data.

6. Conclusion

In this paper, the background of introducing a new national English test, and practical constraints in implementing the NEAT were explored and discussed. The fundamental purpose of introducing the NEAT is to improve school teaching of English, that is, to intensify and activate the teaching of speaking and writing.
(2) **Timing:**
   a) when does the release happen, and
   b) how often is the released spec (itself) revised?

(3) **Audience:**
   a) for whom is the released spec intended, and
   b) do they actually read and make use of it?

(4) **Comparability:** how does the released spec compare to that of other similar tests?

(5) **Outcome:** how is the test’s validity argument affected by spec release? What impact does spec release have on test reform?

(6) **Tradition:** how is spec release different from the long-standing tradition of releasing past tests as sample tests?

These questions are provocative and far-reaching. All of these issues can be addressed productively if test developers and researchers start small; perhaps they could attack one single element of this theory – ‘focus’. For instance, is it (really) a bad idea to tell the wider public what websites are consulted during language test development? Would it really be harmful to let students know that test developers regularly use a website like the thesaurus link, shown above? In fact, maybe releasing that point actually triggers positive washback on instruction and teaching. This paper is offered as a way to stimulate discussion and collaboration amongst language testers in Japan so that the above questions can be explored.

**References**


* Some sample items of the NEATs 2 & 3 are presented below for your information.

**NEATs 2&3 speaking test items (example)**

**(Type 1) Talk on a familiar topic**

1. Do you like reading books? There are many books around us. If there is one book that you like most, please tell us about your favorite book and why. (At least in two sentences)

2. What was the most unique and special gift you've ever had? Please describe it. why was it so special?

3. Tell about your favorite restaurant or the one you have recently visited in more than three full sentences.

4. What do you want to be in the future? Tell us what you want to become when you grow up.

* Your speaking must include;

① what field you are interested in
② what job you are going to get after graduation
1. Read the following passage carefully. What do you think of the writer's argument? You should state your opinion in two to four sentences.

To me, a hero is not only someone who saves a person from a burning building. It could be a parent who helps a child with a difficult homework assignment. It could be a person who stops by to check on you when you're ill. A hero is not just a person who has the courage to take a risk. He is also kind to people in every day situations.

- I disagree with the writer. I think we can call a person a hero when he/she helps us in dangerous situations.

2. Listen carefully and state your opinion about the suggested ideas. You should say in full sentences.

- I think the ideas are very nice and I need to start to practice them in my everyday life. I will set my cell phone to beep on the hour and use it as a signal to wear a smile on my face.
(Type 3) Give advice / make a suggestion

1. Your best friend is expected to have a job interview. Referring to the following lists, give him or her some advice to help prepare for the interview for one minute. Be sure to include at least two guidelines below.
   - Look over the job description
   - Anticipate the dialog
   - Practice and review
   - Be comfortable and at ease

   * I think the first step is to find out what the job is like so that you can portray yourself as what the interviewer is looking for. Also, you might want to anticipate interview questions and rehearse them repeatedly because you can't spend time and energy putting your answer together in your head right then. Most of all, don't be nervous and try to be comfortable and natural.

2. One of your best friends likes computer games very much. Whenever he starts a game, he never stops. In the end, he doesn't do his homework or get enough sleep. So he argues with his parents a lot about his game addiction. What would you like to say to him?

3. Imagine there is a fire in your house. The hallway, the only way to escape, is on fire. You have to escape from fire. Describe how you should act in this situation. You should explain in full sentences.
(Type 4) Make up a story

1. An Internet blog site asked its users whether they preferred light or dark backgrounds in blog design. The results were:

How about you? Do you prefer light or dark backgrounds on a blog? Describe the diagram using "majority," "quite a few," and "only." Then say your opinion. You’ll be given one minute to prepare your answer.

The majority like light backgrounds but quite a few readers think it depends upon the blog design. Only one tenth go for dark backgrounds. As for me, I agree with the majority. I think light background is more appropriate for a blog. When I see a blog with dark background, I tend to think the information provided in the blog is unreliable. I don’t know why, but it’s true.
2. Look at the bar chart, which presents the percentage of the subjects that high school students like to study. Compare and contrast each subject's ratio on the basis of the chart, and give us your opinion about your preference.

![Bar Chart](chart.png)

- Students' preferences of the subjects are in the order of Chemistry, Biology, Physics, Astronomy, Composition, Earth Science, English and Math.
- Chemistry is the most favored subject, and Math is the least favored subject.
- Actually I don't like Chemistry very much unlike these students. My favorite subject is Biology.

3. Look at the chart of three students who are of different age, and of different length of living in the USA. The chart shows their reading and listening scores. Describe the chart and draw your conclusion.

<table>
<thead>
<tr>
<th>Student</th>
<th>Current age</th>
<th>Years of Living in the USA</th>
<th>Reading score</th>
<th>Listening score</th>
</tr>
</thead>
<tbody>
<tr>
<td>A(boy)</td>
<td>15</td>
<td>2 years</td>
<td>70</td>
<td>95</td>
</tr>
<tr>
<td>B(girl)</td>
<td>20</td>
<td>5 years</td>
<td>95</td>
<td>90</td>
</tr>
<tr>
<td>C(man)</td>
<td>35</td>
<td>8 years</td>
<td>90</td>
<td>75</td>
</tr>
</tbody>
</table>

**Student A** is now 15 years old. He came to the USA 2 years ago and his reading score is 70, and listening score 95. **Student B** is now 20. She lived in the USA for 5 years, and her reading test score is 95, listening score is 90. **Student C** is now 35. He lived in the USA for 8 years, and his reading score is 90, and listening score is 75. The younger they are, the better they are at listening. Student B's performance is generally good.
(Type 6) Answer a series of questions

1. You will hear four questions about nickname. Answer each question in full sentences for fifteen seconds respectively.

Q1. What is your nickname? If you don’t have one, what is your best friend’s nickname?
Q2. Do you like your nickname? If you don’t have one, do you like your friend’s nickname?
Q3. If you like the nickname, why do you like it? If you don’t, why not?
Q4. Is there any nickname you want people to call by? If there is, what is it? And why do you want it? If there isn’t, why not?

2. You will hear four questions. Answer each question in full sentences for fifteen seconds respectively.

Q1. Which do you prefer more, watching television or going to the movies? Explain why?
Q2. Can you choose the most memorable movie, book, or play you have ever seen? Will you describe it briefly?
Q3. What kind of job do you want to have in the future? Will you explain why you like them?
Q4. What kind of job do you never want to have in the future? Why do you not like them very much?