

A Study of the Ambiguous Aspects of Language Testing

Akihiko Higuchi
Misaki Arima

Abstract

This present study deals with the language testing of Japanese junior high school students with an emphasis on the ambiguous aspects within the testing framework. This study collected teacher-designed English achievement tests from fifteen Japanese teachers of English at ten junior high schools in Kagoshima prefecture. They were closely analyzed by the authors. The analysis discovered two major problems in the tests. First, due to an ambiguous evaluation criteria, there was little consensus and mutual understanding among English teachers as to what to test in terms of evaluation. In other words, individual teachers had different views as to what kind of tasks should be utilized in order to measure the students' English proficiency with respect to each evaluation criteria. Some of the typical questions were written in the medium of 'word order questions', 'aural dialogue questions' and 'written dialogue questions' etc. Second, the study also determined that the Japanese teachers of English lacked a fundamental knowledge of language testing. These two factors were the major problems in language testing found in this study. A number of pedagogical implications and limitations are also discussed later in this study.

1 Introduction and rationale

It has been discussed by several researchers that the validity of English achievement tests and assessment are considerably questionable. It is also claimed that their primary functions and purposes are misconstrued by language teachers. For example, Negishi and Wakabayashi (1997) collected and analysed English achievement tests from Japanese junior and senior high schools. Their works revealed a number of flaws in the tests designed by Japanese teachers of English; as is often the case, such tests can be found in our schools. This results in a situation where most achievement tests fail to reflect the students' English proficiency. Under these circumstances, the test results and evaluation are highly problematic.

Ohtomo (1995) summarizes the nature of criterion-referenced test in terms of its purpose, developing procedures, descriptor of objectives, generalization of test performance and the information it produces. As his study points out, few of the tests have these characteristics. In other words, the teachers' interests are not in what the students can do in terms of actual English proficiency, but rather their interests lie in comparing one student's performance with that of other students.

Though playing a significant role in instruction procedure (Shizuka, 2002), studies in language testing and assessment are still limited (Negishi, 2003). Therefore, this study attempts to reveal

the actual situation of language testing in Japanese junior high schools in light of the works of Negishi and Wakabayashi (ibid).

2 Historical overview

This section reviews a historical overview of English language testing and the nature of evaluation for Japanese junior high school students. According to Kajita (1994), in the early 1970s, the evaluation theory proposed by Bloom et al. (1971) gave significant attention to researchers in Japan. Bloom (ibid) supported formative assessment, mastery learning, and suggested the taxonomy of educational objectives. Mastery learning was originally proposed by Carroll (1963 quoted from Bloom, ibid) which claimed that the majority of the students should achieve the goals of subjects, if given adequate time and appropriate instruction to the needs of each students. According to Bloom (ibid) the taxonomy of educational objectives was the educational evaluation method which classified educational objectives into three domains; cognitive domain, affective domain and psychomotor domain. In other words, educational objectives were categorized within the hierarchical framework from simple to complex in terms of students' intellect, emotion and behavior. Above all, the taxonomy considerably contributed to the analytical evaluation which is still applied in a Japanese context in language assessment (Furukawa, 2015).

When the national curriculum was revised in 1980, MEXT put the emphasis on the evaluation of emotional traits such as interest and positive attitude toward learning for the first time.

However, these traits were criticized among teaching professionals at first on account that they could not be measured objectively (Kajita, ibid). After that, MEXT launched a guideline for analytical evaluation that was composed of four evaluation aspects; (1) interest, willingness and positive attitude toward learning, (2) the ability to think and make a decision, (3) the ability to express learners' view, and skills to be acquired in the learning, and (4) the fundamental knowledge and ability to understand the learning task.

As a result, students' English proficiency had been evaluated in terms of 'listening', 'speaking', 'reading', 'writing' along with 'interest and positive attitude toward communicating in English' (Tanaka, 2005).

In addition, this revision introduced 'domain-referenced assessment' (Otomo, ibid) which aimed to analytically evaluate students' performance on each subject. Although the domain-referenced assessment was based on the objectives of each subject, it was distinguished from criterion-referenced assessment which evaluated what the students are able to do in terms of the evaluation criteria. By its nature, rather than criterion-referenced assessment, domain-referenced assessment was characterized as norm-referenced assessment, in which individual students' performance was assessed with reference to that of other students.

After the national curriculum was revised in 1989, the evaluation aspects of English learning have been considerably changed. At that time, MEXT revised the evaluation aspects in order to correspond to the major elements which underlay the students' autonomous leaning. Yamagiwa (1993) summarized these elements as 'the attitude of proactive learning', 'the ability to think and make a decision', and 'the ability to express learners' view'. In light of these major elements, the existing evaluation aspects were improved in part as follows: (1) interest, willingness and positive attitude toward leaning, (2) the ability to think, make a decision, and express learners' view, (3) skills to be required during the learning, and (4) fundamental knowledge and ability to understand.

As a result, the students' English proficiency has been evaluated in terms of (1) 'interest, willingness and positive attitude toward communicating in English', (2) 'the ability to express learners' view in English', (3) 'the ability to understand English', and (4) 'the ability to understand English linguistic knowledge and different cultures'. In this way, speaking and writing had been categorized as productive skills, and listening and reading categorized as receptive skills.

The national curriculum revised in 1998 took over the existing evaluation aspects and put the emphasis on the improvement of students' achievements. Accordingly, norm-referenced assessment called 'domain-referenced assessment' was apparently shifted to criterion-referenced assessment in 2000, although the actual situation in the assessment has not been changed (Ohtomo, *ibid*).

The introduction of evaluation aspects seemed to be successful in that the students' English proficiency was assessed in analytical terms (Watanabe, 2011, Wakabayashi, 1992). However, these twofold evaluation aspects; productive skills and receptive skills were called into question. Wakabayashi (*ibid*) points out that listening comprehension and reading comprehension cannot be explained in the same manner although these two skills are classified into the same evaluation aspect. For instance, the evaluation aspect of productive skills is not enough to explain a student who is fluent in speaking, but requires some improvement in writing.

As we have seen, in light of the historical overview, it is obvious that the evaluation aspects of students' English proficiency were developed regardless of English education (Negishi, 2005). Ohtomo (*ibid*) points out that adopting Bloom's (*ibid*) taxonomy to English education is highly problematic since it was not foreign language education, but general education the taxonomy took account of. Indeed, this evaluation method caused some problems in the evaluation of English language education.

3 Present study

With the works of Negishi and Wakabayashi (1997) in mind, English achievement tests were collected and closely analyzed by the authors. They were designed by fifteen Japanese teachers at

ten junior high schools in Kagoshima prefecture. After the analysis, it became clear that there were two major problems. Due to the obscure evaluation criteria, this study revealed that there was little consensus and mutual understandings among English teachers as to what to test for on each individual evaluation aspect. In other words, individual teachers had different views on what kind of tasks should be utilized in order to measure their students' English proficiency with respect to each individual evaluation aspect. This kind of problem was highlighted in the structure of the word order questions, aural dialogue questions and the written dialogue questions.

More specifically, fourteen among fifteen teachers comprised word order questions which apparently measured the same aspects of the students' English proficiency. All of the questions were decontextualized, they were also accompanied by a Japanese translation. However, nine teachers intended to measure grammatical knowledge with these word order questions, whereas five teachers intended to measure writing proficiency (See Appendix I).

The disagreement among teachers on the evaluation aspects was also found in dialogue questions in aural and written dialogues. In fact, thirteen among fifteen teachers included aural dialogue questions. Two teachers among them asked students to choose an appropriate response after listening to the dialogues. What is at issue here is that one teacher regarded these questions as a speaking test, while the other teacher regarded them as a listening test (See Appendix II).

In the same manner, thirteen teachers among fifteen included questions which required the students to read dialogues, though the tasks that followed were slightly different. For example, some of them asked the students to fill blanks along the written conversation and others asked them to answer questions in order to test if they understood the contents. Nine teachers considered these kind of questions as being used to measure the students' reading comprehension skills, whereas eight teachers used these questions as a way to measure their 'speaking proficiency' (See Appendix III).

These problems are considered to be caused by the ambiguous evaluation criteria as Negishi (2005) put it. For example, according to the evaluation criteria by the National Institute for Educational Policy Research (henceforth, NIER) (2011), 'writing proficiency' has to be evaluated in terms of 'productive skills' and 'grammatical knowledge'. Nevertheless, an evaluation aspect of grammatical knowledge exists itself as 'the ability to understand English linguistic knowledge and different cultures'. Since grammatical knowledge relates to all the skill areas such as speaking, listening, reading and writing (NIER, *ibid*), little agreement has been reached concerning the distinction between these evaluation aspects among teachers.

In order to solve this problem, Matsuzawa (2002) provided English teachers a clue for interpreting these evaluation criteria. According to his study, students' grammatical knowledge has to be tested in terms of declarative knowledge and procedural knowledge. The terms declarative

and procedural knowledge originated in cognitive psychology. Declarative knowledge is a very early form of knowledge (DeKeyser, 2001), the kind of knowledge a novice foreign language learner acquires.

On the other hand, implicit knowledge is tacit and intuitive knowledge which has been completely internalized by the proficient learner (Ellis, 2009). Therefore, Matsuzawa (ibid) suggests that grammatical knowledge be treated as ‘the ability to understand English linguistic knowledge and different cultures’, and that it should be exclusively treated as declarative knowledge. It stands to reason that procedural knowledge which is required in language use should be treated as ‘the productive knowledge’ and ‘receptive knowledge’. Matsuzawa (ibid) also claims that these evaluation aspects originally aimed to evaluate students’ English proficiency in terms of language use and usage since too much emphasis has been placed on the latter. However, the situation remains unchanged, since the evaluation aspects are not understood well enough by the English teachers. As a first step toward the better understanding, it is time to reconsider such ambiguous evaluation aspects.

As for the second major problem, this study revealed a limited discussion on the test validity among the English teachers. Because of this, a number of test items failed to show the students’ English proficiency. For example, nine teachers among fifteen intended to test the students’ reading comprehension utilizing texts students had already read during English classes. With a text they have already learned, students do not have to read and understand it; they are simply asked to recall its contents. Hughes (2003) and Negishi and Wakabayashi (1993) point out that most achievement tests include this kind of ‘pseudo-reading comprehension tests’. In order to provide the evidence that the students have acquired reading comprehension skills, it is necessary that teachers understand what the reading comprehension skills are. Conversely, objectives of reading comprehension were abstractly construed by the teachers (Ohtomo, ibid.).

In another case, ten teachers among fifteen intended to measure students’ ‘speaking proficiency’ by utilizing paper-and-pencil tests, this despite the fact that researchers assert them as less valid (Negishi and Wakabayashi, ibid, McNamara, 2000, Shizuka, 2002). These tests may show if the students can write in a given situation or fill the blank in a written dialogue, however, there is no evidence that they can speak in an actual conversation. Such tests can produce a negative backwash effect, therefore they should not be included (Negishi, 1993).

Furthermore, many test items relied more on skills other than those they intended to measure. Such problems were found in the listening and comprehension questions. For example, seven teachers tested students’ listening comprehension by utilizing test items which required the students to write appropriate responses in English to questions in the listening materials (See Appendix IV). In this case, students who failed to answer them are regarded as lacking in their listening

comprehension skills regardless of whether or not they understood the questions. Negishi and Wakabayashi, (ibid) suggest that in questions of this type, it is almost impossible to ascertain where a student's deficiencies lie in terms of comprehension or writing proficiency. Because of this, reading questions should be independent from other questions in the other three language skills (Hughes, ibid). In the same way, five teachers evaluated the students' writing skills in their reading questions (See Appendix V). The ability required to properly elicit the students' receptive skills need to be well understood by teachers.

4 Conclusion and limitation

In summary, this study found that two major problems existed in achievement tests for Japanese junior high school students. One was the disagreement in interpreting the evaluation aspects of tests. This was evident in the way the tests were constructed. The main cause of this problem was found in the ambiguity of the test evaluation aspects which had been developed without carefully considering what is required for accurate English language testing.

In conclusion, the evaluation aspects should be reconsidered and supported by pedagogical theory, elaborations and achievement targets in language learning. Learners' English proficiency are tested over five skill; listening, reading, writing, spoken interaction and spoken production in CEFR-J (Tono, et al.,2014). In addition, concrete guidelines in testing and evaluation in conjunction with good teaching methods, are needed for English teachers.

The other problem was validity in the teacher-designed tests. It was triggered by the lack of fundamental knowledge in developing tests among teachers. There were limited discussions as to what English skills need to be acquired by the students. There were also limited discussions as to what skills and knowledge language teachers need to acquire in order to be able to create effective tests that effectively evaluate each skill. For these reasons, a number of test items failed to show the students' English proficiency and what the students could do by using English. Hirata (2003) insisted that validity in the evaluation method and testing procedures should be prerequisites for criterion referenced assessment.

As for the limitations of this present study, the number of the data was limited. For the subsequent study, therefore, the validity of the evaluation aspects for English education should be investigated including a quantitative study. In so doing, the actual situation in language assessment and testing will be more effectively evaluated for Japanese junior high school students. To achieve these aims, further elaboration will surely be necessary.

References

- Bloom, B. et al. (1971) *Handbook on Formative and Summative Evaluation of Student Learning*. McGraw-Hill, Inc.
- DeKeyser, R. (2001) 'Automaticity and automatization' In Robinson, P. *Cognition and Second Language Instruction*. Cambridge : CUP
- Ellis, R. (2009) 'Implicit and Explicit Learning, Knowledge and Instruction' In Ellis, R. et al. *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching*. Bristol : Multilingual Matters
- Furukawa, O. 'B.S. Bloom riron no nihon ni okeru juyo to hatten' *Konan university kyosyoku kyoiku center nenpo kenkyu hokokusyo* file:///C:/Users/higuchi/Downloads/K01534.pdf
- Hirata, Y. (2003) *Chugakko eigoka zettaihyoka no hoho to jissai*. In Hirata et al. Tokyo : Meiji Toshō
- Hughes, A. (2003) *Testing for Language Teachers 2nd Edition* CUP
- Kajita, E. (1994) *Kyoiku ni okeru hyoka no riron I gakuryokukan hyokakan no tankan*. Tokyo : Kaneko Shobo
- McNamara, T. (2000) *Language Testing* Oxford : OUP
- Matstuzawa, S. (2002) *Eigo kyoshi no tame no atarashii hyokaho*. Tokyo: Taishukan Shoten
- Negishi, M., Wakabayashi, S. (1997) *Musekinin na tesuto ga ochikobore wo tsukuru*. Tokyo: Taishukan Shoten
- Negishi, M. (2005) 'Donatta? kantenbetsu zettaihyoka' *Eigo kyoiku*. Tokyo: Taishukan Shoten
- Negishi, M. (1993) *Tesuto no tsukurikata*. Tokyo : Kenkyu Sha
- Negishi, M. (2003) 'Zettai hyoka wo kangaeta tayo na hyoka hoho no kaihatsu' *Chugakko eigoka zettaihyoka no hoho to jissai*. In Hirata, et al. Tokyo : Meiji Toshō
- NIER (2011) *Hyokakijun no sakusei hyokahohoto no kuhukaizen notameno sankosiryō*
http://www.nier.go.jp/kaihatsu/hyouka/chuu/10_chu_gaikokugo.pdf
- Ohtomo, K. (1975) 'Eigokyoiku no nakano hyoka-Criterion-referenced Test-' Jinbun Kenkyu. Kanagawa : Kanagawa daigaku jinbun gakkai
- Ohtomo, K. (1995) 'Daigaku koko ni okeru eigo no nyushi de kangaeru koto' *kyoiku*. Tokyo: Taishukan Shoten
- Shizuka, T. (2002) *Eigotesutosakusei no tatsujin manyuaru*. Tokyo : Taishukan Shoten
- Tanaka, M. (2005) *Korekarano eigogakuryokuhyoka no arikata*
- Tono, Y. et al. (2013) *The CEFR handbook A resource book for using CAN-DO descriptors for English language teaching*. Tokyo : Taishukan Shoten
- Wakabayashi, S. (1992) 'Eigo kyoiku nisshi' *kyoiku*. Tokyo: Taishukan Shoten
- Watnabe, Y. (2011) 'Eigogakuryokuhyokaron' In Ishikawa, S. *Language Testing and Assessment*. Tokyo : Taishukan Shoten
- Yamagiwa, T. (1993) *Gakuryokukan no tenkan*. In Yamagiwa, T. et al. Tokyo : Kaneko Shobo

Appendix I

4 【書く力】 26 点

1. 次の日本語に合うように、【 】内の英語を並び替えなさい。ただし、英文の決まりに従って一文書くこと。(3点×4=12点)

(1) もし晴れば、私たちはソフトボールをするつもりです。

【clear / , / we'll / if / softball / play / it's / .】

(2) 一週間は、七日あります。

【seven / there / days / a week / are / in / .】

(3) 私は、ハワイに行きたい。

【Hawaii / go / I / to / want / to / .】

<言語に関する知識> 5 次の問いにそれぞれ答えなさい。

日本語の内容と合う英文になるように()内の語を並べ替え、2番目と4番目に来るものを順に記号で答えなさい。選択肢にコンマは無いので気をつけること。

なお、文頭に来る語も小文字で書いてあります。(1点×5)

(7) 父が新しい自転車を買ってくれるんだ。

(ア a new bike イ is going to ウ me エ my father オ buy) .

(8) カナダでは野球をしますか。

(ア played イ in ウ is エ Canada オ baseball)?

(9) あの公園をきれいにしよう。

(ア make イ park ウ let's エ that オ clean).

Appendix II

1 聞き取りテスト 英語は2回ずつ放送します。メモをとってもかまいません。

- 2 これから、Yuki と Bob との対話を放送します。この対話では、Yuki の最後の発言がチャイムの音で消されています。対話の流れに合うように、このチャイムの部分に入る表現として最も適当なものを、下のア～エの中から一つ選び、その記号を書きなさい。

ア Yes. I have many things to do.

イ Yes. But why?

ウ Yes. Let's go.

エ Yes. Where is it?

2 【話す力】 19 点

1. 対話を聞いて、チャイムの部分に適する応答をア～ウから 1 つ選び、その記号を書きなさい。

(各2点) 6点

(1) ア Yes, it's easy for me to get up early.

イ Yes, it's difficult for me to get up early.

ウ No, it's easy for me to get up early.

(2) ア Yes, I'll take them.

イ Yes, please. I want a card.

ウ No. I want a different one.

(3) ア You're welcome.

イ Certainly. Here you are.

ウ Sure. How about these cups?

Appendix III

【読むこと】

8 次の文章をよく読んで、後の問いに答えなさい。

Ms. Wood: Yuki, Where did you work?
 Yuki: I worked at a nursery school near the soccer stadium.
 Ms. Wood: Really? What did you do there?
 Yuki: I sang songs and played with the children.
 Ms. Wood: Did you enjoy the experience?
 Yuki: I had a wonderful time there. The children liked me almost at once.
 Ms. Wood: Do you like to play with children?
 Yuki: Yes, very much. I would like to work at a nursery school ①() () ().
 I like to take care of children.

(1) 下線部①は、「将来」という意味の言葉が入る。3語で答えよ。

(2) 次の質問に答えよ。

ア 由紀が職業体験でしたことは何か。具体的に日本語で2書け。

2 【話すこと】

次の英文を読んで、後の問いに答えなさい。

Satomi: Tom, we are going to go to Awaji Island this Saturday. Do you want to come with us?
 Tom: Island? (①) will we get there? By ferry?
 Satomi: No, we will go by car. There is a big bridge to Awaji Island. Also, there are no trains on the island.
 Tom: That sounds interesting. ② .
 Satomi: The fish there is very good in all seasons, so we often go to our favorite sushi restaurant. Now it's fall, so you can enjoy ③ (have) the fish of the season.
 Tom: Wow! What fish is good now?
 Saba and sawara are good in Awaji now. There are many other seasonal foods for each season in Japan.(④)example, *kuri*, *matsutake*, and *kaki* are popular in fall.

1 ①, ④の()に適する語を、それぞれ1語で答えなさい。

2 ②に適する文を選び、記号で答えなさい。

3 ③の()内の語を適する形に変えなさい。

Appendix IV

<聞くこと>

問5 これから英語で質問をします。それに対する適切な答えを4語以上の英文で書きなさい。
 英文を書く時間は30秒です。

Appendix V

5 【読むこと】 18点

1 次の文章を読んで、下の問いに指定された語数の英語で答えなさい。(2点×2)

Today, many Japanese are working abroad as volunteers. Among them is Yamamoto Toshiharu, a doctor and photographer.

When he was in elementary school, he visited Africa with his father. He was shocked to see a lot of sick people who were suffering from hunger and illnesses. Since then, he has been interested in international cooperation. He has visited a lot of developing countries and helped many poor people. He also started a program to train people who want to work as doctors or nurses in those countries.

When he visits developing countries, Mr. Yamamoto always meets children. He asks them to draw something which is important to them.

Here is a picture drawn by Maruao, a 13-year-old boy in Tuvalu. This picture shows that it is important to stop climate change. If not, Tuvalu may go under the sea some day in the future.

Look at the next picture. It was drawn by Jatariuc, 15, in Romania. This picture shows the flag of her country. She says, "The most important thing to me is my country."

(1) How did he feel when he saw people who were suffering from hunger and illnesses? (6語)

(2) What is the most important thing to Jatariuc? (4語)