

Learning- and Growth-Oriented Assessment in EFL

Ken NORIZUKI

Educational assessment may be either end-oriented or development-oriented (summative or formative assessment). Of these two important functions of assessment, only the former tends to be emphasized among teaching practitioners. End-oriented assessment is often no more than the product of an end-of-course test. The results are most likely group-referenced, whether the test is norm-referenced or criterion-referenced, and final grades are awarded on the basis of relative standing of the scores or the magnitude of the scores defined as levels of attainment. Little or no useful diagnostic information is provided to learners other than categorical grades at the end of instructional programs. In the context of tertiary EFL education, learners who are relatively more proficient may attain good grades with minimum efforts and those who are less proficient may fail to do so in spite of their efforts. This does not seem to be fair.

This article reports on how these problems could be mitigated, referring to assessment tasks that the writer has employed and samples of dialogue journals that he has collected recently. It is argued that assessment should have a positive role in fostering learners' language learning and enhancing their ability and motivation to use the target language in a meaningful context.

1. Types of Assessment

1.1 Summative and formative assessment

Assessing students' learning can be either end-oriented or development-oriented (Weir and Roberts 1994). These two purposes of assessment are labeled as summative and formative assessment. The former type of assessment concerns the general level of a group or individual learners' course-end achievement. Teachers/Evaluators (hereafter TEs) are usually content if the students' general level of achievement is as high or higher than they expected as a result of their instructional effort, or higher than the control group's status as tested in an experimental research design. The second type of assessment is designed to cover the whole teaching/learning process in a given instructional program. The purpose of this assessment is to help foster the students' learning and make teachers aware of their instructional problems.

While both types of assessment play essential roles in formal education, the

latter is often neglected or scarcely recognized, especially in tertiary instructional programs. Some teachers engaged in classroom assessment (TEs) are not committed to anything more than grading students' levels of final attainment measured by a single one-hour test. Students are mostly concerned about grade levels, or just whether they pass or fail, but rarely about the process or quality of learning. Those students who have done well in regular class work or assignments may not gain high scores on a single semester-end test and their daily efforts may not ultimately be rewarded with good grades. On the other hand, some test-wise students are able to achieve good or passing grades, just by focusing their attention on the final exam. Is this state of affairs fair?

1.2 Norm-referenced and criterion-referenced assessment

In the context of secondary school education, norm-referenced assessment was once regarded as an official method of dividing students into different grades. This method is an effective means of discriminating among or rank-ordering students. A potential problem with using this method in classroom assessment is that the performance of an individual student is assessed in relation to that of all other students in the class. This means that students in a relatively competitive class or school have more difficulty gaining good grades than those in a less competitive class or school. The current Cumulative Guidance Record, published by the Ministry of Education, Culture, Sports, Science and Technology, dictates a complete shift from norm-referenced to criterion-referenced assessment in lower secondary school evaluation schemes. It is expected that this shift to criterion-referenced assessment will prevail in upper secondary education as well (Matsuzawa, 2002).

Meanwhile in tertiary education, norm-referenced assessment is usually not a preferred option except for placement, selection and proficiency decisions. Criterion-referenced assessment is desirable, but is often hard to achieve. If instructional materials or textbooks are rather demanding, it tends to be more difficult to assess students purely in the light of the level of achievement. If the materials are not very demanding relative to students' pre-course ability levels, students may obtain good grades fairly easily.

While both norm-referenced and criterion-referenced types of assessment have distinct roles in tertiary or any other formal education, the latter is more suited to the purpose of assessing the individual students' learning status. A subcategory or a variant of the latter type is individual-referenced assessment (Matsuzawa, 2002). In the last-mentioned framework, the students should be assessed in terms of the amount of learning or growth they show. Thus, less proficient students will not automatically be disadvantaged in a sense that they can learn and show signs of

growth just as well as more advanced students.

2. Methods of assessment

2.1 Testing

Tests are still the most frequently used means of EFL assessment. They are variously called exams, tests and quizzes, depending on “their scope, importance, function, and the care with which they have been prepared”. (Genesee and Upshur, 1996: 140). There are a wide variety of different testing formats, including multiple-choice, fill-in, short-answer, performance, just to name a few. All these test types can be divided into two basic categories: selected response and constructed response (Brown and Hudson, 2002).

Selected-response test items are those that require students to choose the correct answer from among a set of options. These items are considered easy to score objectively. Well-designed items of this kind require a relatively short time for students to complete. They are, however, often difficult for the test writer to create. They can be quite daunting for students of limited proficiency if incorrect options, or distracters are confusingly difficult. And they may not suit the purpose of assessing writing and speaking skills as they require no language production on the part of the students.

Constructive-response test items are those that require students to produce language by writing or speaking. These test items are relatively easy to construct. Well-designed items of this type are considered a more valid measure of writing and speaking skills. If the input and expected output are straightforward to interpret for both examinees and examiners, the items can be used effectively and meaningfully for assessing reading and listening skills. However, this item format usually allows for varying degrees of correct or acceptable responses, which makes it more difficult to score. If the answer is retrievable from the test input, the student may copy the relevant part without fully understanding the intent of the question or the meaning of the chosen response itself. Items requiring the students to produce translations in their L1s with regard to specific or overall parts of the test input can be a way to check the amount and quality of the students’ comprehension. However, the items are obviously not amenable to assessing more productive skills and can be counterproductive for students learning receptive skills as well, discouraging them from trying to understand the target language input the way it is written or spoken. Another difficulty is that this method works only if the TE understands all of the students’ L1s (e.g., Japanese, Chinese, Vietnamese, Indonesian, etc.) in the group. For the writer of this article, the last requirement alone substantially precludes the possibility of using this method.

In summary, both selective-response and productive-response item formats have advantages and disadvantages. EFL TEs should employ a variety of response formats wisely after taking into account the purpose, context, or other important practical considerations involved in their assessment.

2.2 Non-test methods

In recent assessment literature, a variety of non-test methods (e.g., observation, portfolios, conferences, journals, questionnaires and interviews) have been collectively called alternative assessments (Alderson and Beretta, 1992; Weir and Roberts, 1994; Genesee and Upshur, 1996; Brown, 1998; 2001; Ekbatani and Pierson, 1999). A common misinterpretation that could stem from this term is that these methods are distinct from and incompatible with test-based methods. Of course this is not true. Brown and Hudson (2002: 78) have reservations about the use of this misleading term and call these methods personal-response items instead. These items are defined as those that allow for the responses and the ways the tasks are accomplished to be quite different for each student. Brown and Hudson's term may not cover researcher- and program-centered observation techniques and the word "item" does not fit most of the methods in this category well. However, the term is a simple and yet reasonable reflection of a sound assessment model that all assessments, in spite of all the difference in structures, should ultimately be based on the examinees' responses.

Unlike testing methods, personal-response, non-test methods require no prescribed criteria for correct or acceptable responses. This allows TEs to assess an individual student's learning process more flexibly. A major problem is how to assess rather unstructured responses in an objective and fair manner.

3. Objectives of assessment

In norm-referenced assessment, educational objectives may not be clearly identified. In criterion-referenced assessment, setting up educational objectives is a starting point and one of the most crucial steps in the process of conducting continuous and target-oriented assessment. Objectives can take many different forms. Brown and Hudson (2002) categorize a wide range of objectives into three major domains: instructional objectives; performance objectives and experiential objectives. A slightly modified and somewhat simplified version of their definition can be a useful frame of reference in the present discussion.

Instructional objectives are narrowly-focused, short-term specific objectives. They can be organized around structural items (e.g., present and past tenses), functional items (e.g., introductions and greetings) and vocabulary items (e.g.,

household and classroom objects).

Performance objectives concern the students' ability to use the target language to accomplish something in a real-life situation. For example, in a business English course, the ability to communicate with business clients through e-mail may be included as part of its performance objectives. In tertiary EFL education, these objectives are ideally oriented towards students' academic or other prominent interests, defined partially in the context of language for specific purposes, or specific skills that individual students wish to acquire (see Douglas, 2000; 2001; Davies, 2001; Cumming, 2001).

Experiential objectives relate to what the TE wants the students to experience in the process of their learning. In lower secondary education, for example, these objectives can be inferred from one of the four assessment criteria specified in the Cumulative Guidance Record, according to which students' interest in the target language and its culture and their motivation to learn are to be assessed. It is probable that tertiary education TEs can accommodate these objectives even more flexibly. One way is to encourage students to report on their autonomous learning process through portfolios, journals or/and diaries. The use of journals in tertiary education will be exemplified in the subsequent section.

4. Assessment in progress

4-1 Selective-response methods

The following three sections will report on some assessment tasks employed by the writer himself or the Faculty of Communications and Informatics (FCI), Shizuoka Sangyo University (SSU), with which the writer is affiliated.

Selective-response items, as defined by Brown and Hudson (2002), are frequently used in EFL textbooks. A well-written item of this type offers an effective clue to understanding a rather demanding language input and may also give learners an enjoyable form of problem-solving activity. However, such items can be more frustrating than constructive-response items for some students if they are too tricky or undercontextualized. The following is an item of this category found in a textbook that the writer has used recently, which exemplifies this problem.

Example 1

A: Where do you think I should put this picture?

B: If I were you, I'd put it either next to the door or between the windows.

A: Well, I think beside that picture is better.

Question: Where is the picture put?

- (A) Next to the door
- (B) Between the windows
- (C) Beside the other picture
- (D) Above the other picture

Guessing the end product of a suspended or interrupted dialogue is pragmatically sensible, but it does not seem very fair to assume that Person A's final remark will be their ultimate decision without receiving Person B's following, 'unwritten' response. This does not mean to criticize a particular textbook that otherwise offers a lot of high-quality exercises and useful expressions, but to indicate that TEs need to be aware of potential problems that selective-response items may inadvertently create.

A solution to this problem is to tell the students informally that the last speaker in the suspended dialogue (Person A) has a final say in the matter, and the other speaker (Person B) will eventually follow his/her opinion. This solution would not be justified in a more formal test or exam setting, however.

A similar problem was found in the process of making English entrance tests. The writer recently examined an FCI English entrance test administered in February 2002, and found that no less than 12 selective-response items out of 25 underwent varying degrees of modification before they were finalized. Example 2-1 was replaced by Example 2-2 for a reason similar to that discussed above for example 1, as there is no way of knowing whether person B would respond in the positive or negative.

Example 2-1 (before revision, conversation gap-filling: Item 12)

A: You look really hot.

B: Yes, I came on my bicycle.

A: Would you like some orange juice?

B: (12)

- 1 I hate orange juice.
- 2 Yes, I'm really hungry.
- 3 Yes, please. I'm really thirsty.
- 4 No, thank you. I'm not thirsty.

Example 2-2 (after revision, Item 12)

A: You look really hot.

B: Yes, I came on my bicycle.

A: Would you like some orange juice?

B: Yes, please. (12)

- 1 I hate orange juice.
- 2 I'm really hungry.
- 3 I'm really thirsty.
- 4 I'm not thirsty.

When the earlier (V1) and finalized (V2) versions of the whole test were post-hoc analyzed using two randomly selected samples of SSU students, all the V2 test quality statistics in the analysis (e.g., mean score, item discrimination, NR reliability, CR dependability, etc.) outperformed the V1 counterpart (for more details, see Norizuki, 2002). While this finding is encouraging for the entrance test makers and demonstrated that their hard effort was worthwhile, EFL TEs must recognize that all the selective-response items cannot be subjected to such intensive scrutiny. In short, language learning cannot be measured nor assessed by selective-response items alone.

4-2 Constructive-response items

The writer of this article used to depend heavily on selective-response items in quizzes and tests used for classroom assessment. Now constructive-response formats are almost always used instead in his quizzes and dominate approximately nine tenths of the total items in his term-end or end-of-course achievement tests. (In some of his recent term- and course-end tests, all items were made in constructive-response formats, reflecting the aims of the classes, as shown in example 7.) The rationale for this change is reflected in the writer's current view that learners need to practice producing certain forms of target language in order to improve their ability to understand the language input better and to act on it more meaningfully. Examples 3 to 6 refer to actual items recently used in his classroom quizzes. (The concept of a quiz is not usually discussed nor defined in language assessment literature. A rare exception is Cohen (1994).)

Example 3

1. Daniel learned about (c) blossom viewing. (花見)
2. He is learning Chinese (c). (漢字)
3. They (漢字) are very complicated, so he is often (c).
4. Kenji is planning to study (a) next year. (留学する)
5. Kenji reads an English newspaper (a) almost every day. (声を出して)
6. Daniel thinks that Kenji should (c) with English-speaking people as often as he can.
7. Kenji says, "I look (f) to having our first English lesson." (待ち望む)
8. The buttons are in an (e).
9. Did you see the memo from (p)? (人事部)
10. Just (d) the extension number. (電話をする)
11. Your computer will (b). (ピーと鳴る)
12. I'm sorry, you must have the (w) number. (間違い)
13. The woman will try to (r) him by e-mail. (連絡をする)
14. Employees need to fill out the (i) form. (保険)
15. Each department's fax number is (p) above the fax machine. (掲示される)

Example 4

Type 1:

Students read pages 34 and 35 and the special column known as the Yellow Panel on page 39 of their textbook and then answer the questions in English or Japanese. Short answers are acceptable.

pp.34-35

1. When did Henry Ford become an apprentice at a machine shop?
2. How long did he work as a travelling repairman for a firm that made portable steam engines?
3. Which car inspired William H. Murphy to set up the Detroit Automobile Company? (どの車に感動して会社を設立?)
4. Why did Ford fail twice to put his car into production quickly enough? (なぜ2回失敗?)
5. Where did Henry Ford build the Quadricycle? (p.35 YP どこで作った?)
10. Where can shares and stocks in the company be sold and bought? (p.39 YP)

Type 2:

The students read pages 38 and 39 and fill in the blanks in the summary passage.

pp.38-39

In the early years of motoring, cars were luxury items and expensive to buy.
6() made a big profit on each car. Henry Ford believed that the car should be
7() to everyone as a means of personal transport. He was determined to sell cars
to the 8() market. After Malcolmson left the company, the company started
building 9(a) cars for ordinary people.

Example 5

The students are to answer the questions they will hear by filling in the blanks and then copying the complete sentences or phrases below onto the separate answer sheet.

1. Two () and (数) colas.
2. () colas.
3. (数) yen.
4. They are ().
5. They have some interesting ().

Example 6

The students explain one word chosen from those written in five word cards drawn from an envelope in English. They are encouraged to write as much as they can and in as much detail as possible on the separate answer sheet. (There are twenty cards altogether. Among the words written on the cards are 'elephant', 'season', 'Australia' and 'gesture'. This quiz is based on a game activity in which they are to guess the words from a teammate's drawings or verbal hints.)

Example 3 shows the most frequently used type of quiz items in the writer's classes. The students are to listen to the oral input from the audio tape, and to fill in the blanks with words used in the input. The Japanese clues are provided to facilitate the students' comprehension and their attention to the target expressions in a rather complex input. The main advantage of this format is the ease of preparing items. Another important advantage, which should not be overlooked, is the ease of covering a wide range of key expressions. A disadvantage is that students tend to focus on specified words alone but not the way the words are used in context. A quiz of this format takes about 10 to 15 minutes to complete.

Example 4 illustrates two types of quiz items used together in the second-year reading class. Type 1 items are all in the form of *wh*-questions. Japanese clues are

provided to alleviate ambiguity and to avoid common misinterpretations that can be made by the students. The students are allowed to answer either in English or Japanese and also in short forms. The second type is a summary of a three-page-long text. The students can focus on key expressions, but since the sentence structures are largely modified and simplified, they are not likely to complete the blanks fully without understanding the organic link between the original text and the summary. A quiz of this type is quite challenging, but not too difficult if the proper amount of time is accorded. Twenty to thirty minutes should be allowed for the students concerned.

Example 5 forms a part of a single quiz coupled with items illustrated in example 6. This format is designed to prompt students to produce a full sentence or meaningful unit of utterance in response to the stimulus question. Although there is very little creativity allowed for in the students' production, it is hoped that students will be more conscious of their target language use through their controlled output than in the word-only item.

Example 6 illustrates an attempt to incorporate a kind of supplementary fun activity into the broader assessment framework. A simple, well-designed game or game-like activity is entertaining for the majority of the students, and is a useful way to create a cooperative and harmonious mood among the students. There remains the question of what it is aimed at in the process of language learning and assessment. The purpose of using the particular set of questions listed in example 6 for classroom assessment was largely symbolic, reminding the students that the game was purposefully introduced in the course of the language program. Rather than clinging to the formal assessment of narrowly-focused instructional objectives and lowering the confidence of students of limited proficiency, assessment should have an added role in arousing students' natural interest in learning to use a foreign language. This affective role enables students to recognize that the quiz is not fearful nor stressful, but can give them an enjoyable learning experience. (Brown (1998) provides a practical collection of informal assessment endorsed by teaching practitioners around the world.)

These quizzes can be considered as a variety of class progress tests. They serve not only as a part of final summative grading, but also as a teaching device "to stimulate learning and to reinforce what has been taught." (Heaton 1988: 171)

Example 7 illustrates a recent end-of-course test for the first-year mandatory EFL class. (The name of the course is Reading Skills I. But the intended nature of the course is not limited to reading, but rather the integration of the four skills.) No selective-response items were included in the test as the major aims of this class were intensive reading aloud and writing (or copying) of passages in the target

Example 7: An end-of-course test for the mandatory EFL class

リーディング・スキルズ I A/B 定期試験 (July 30, 2002)

学籍番号 () 氏名 ()

【1】 テープを1回聴き、1~10の空所に入る単語や語句を書きなさい。

L1

1. No one () the Internet, but anyone can use it. (所有する)
2. With only a computer and a phone line, you can send and (r) letters in minutes. (受け取る)

L3

3. About the end of the thirteenth century, Japan was first (i) to Europe in a book. (紹介された)
4. Many Japanese words come from Portuguese words. One such (e) is *paz*. (例)

L4

5. It's 8:40. Tina is running (a) the platform. (...に沿って)
6. They have to (w) () the next train. (...を待つ)

L5

7. Sue was (h) by her uncle. (雇われた)
8. The hats on the wall are (m) () banana leaves. (...で作られている)

L6 (Emi と Mrs. Smith の会話)

9. Everyone in Ms. Hill's class makes a speech, and it's Emi's (t) next week. (番)
10. Students in Mrs. Smith's class can (c) any topic they like. (選ぶ)

【2】 テープを繰り返し聴き、11~18の質問に英語で答えなさい。

L2

11. What did Pascal do in 1642?
12. What did Mr. Shima do in the United States about 30 years ago?

L4

13. When did zen come to Japan?
14. How long has the Great Buddha in Kamakura been outside?

L8

15. What have students learned through debate?
16. What does the teacher hope her students will always do?

L12

17. How did many white people treat black people in the early 1960s?
13. After the 381-day boycott, what did King and his followers win?

【3】 テープを繰り返し聴き、19~26の空所に入る単語 (各2語) を書きなさい。

L7

19. Mika が理解できなかった理由 : People in London speak English with all (k) of (a) .
20. Mika が感じたこと : For her, (e) English was more difficult than (w) English.
21. Ted と Mika が翌日参加したもの : They went on a (s) (t) of London by bus.

L6

22. 話しをするとき : Half of all American sometimes (f) (n) .
23. 彼らが心配なこと : They worry about giving (w) answers and making a poor (i) .
24. 学校で言われること : You must (s) looking at things in a (n) way.

L8

25. 動物園禁止側の意見 : We should (p) animals in their natural (s) .
26. 動物園許可側の意見 : (W) animals are a part of our earth, and we should not (d) them.

【4】 テープを繰り返し聴き、全文を解答用紙に書きなさい。わからないところがあっても想像して、筋の通る話にすることが望めます。(L11)

Some ar' is eat (). Other ar' is eat anir' , that eat ().
 In () 20' ars, many () of anir' have died .
 Pe' have kille' animals for (). They hp' also killed ? als
 their () . P' se have aler' () anim' . omes
 when they' houses ar' . If ar' is can't fin' , () to
 live, ther' out.
 (is killing w' animals, tr') becor' , and r'
 ar') . Birds' eat () . I can't (healthy egg

language.

Items 1 to 10 in Section 1 take the form of a quiz format shown in example 3. The audio language input was given only once for these items. Meanwhile, for the rest of the items in sections 2, 3 and 4, the students were allowed to listen to the input as often as they wanted within a specified time range (90 minutes minus the time spent for section 1) using the memory function of the self-learning language laboratory system.

Section 2 items are a variant of item formats in examples 4 (Type 1) and 5, tapping the students' ability to comprehend the text delivered in audio input, and to answer in English (in full or short forms) to the stimulus questions printed on the test paper.

Items in section 3 are a variant of those in section 1. The Japanese clues are provided as a signpost for the missing expressions in the expected responses.

In section 4, students have to reproduce the whole text mutilated with diagonal lines with the aid of audio input. They are encouraged to create meaningful alternatives if they cannot identify some missing parts.

These four types of items in the course-end test were pre-specified before the test so that the students would not feel anxious about what they would see in the test. In addition, all the item types were more or less familiar to the students as they were frequently used in daily classroom assessment and learning activities. For the learners, the final test is actually a series of quizzes and learning tasks combined together. The course-end test can justifiably be a summary or gist of daily continuous assessment. It should not be the one and only means of assessment to be taken into account in awarding a final grade.

One important missing element in the writer's present assessment framework is a formal means of assessing speaking. In the mandatory EFL class mentioned above, two short read-aloud tests were conducted on an individual basis during the regular class session. These were extremely time-consuming, and students have to wait long for their turns while doing some in-class assignments. The tests reveal something about the students' ability in pronunciation and fluency, but virtually nothing about their communicative ability. Ideally, a more formal, comprehensive means of oral assessment should be devised. The issue needs to be examined in a future study.

4-3. Personal-response items

In a typical tertiary EFL class that the writer is in charge of, the students' language ability levels vary enormously. If summative assessment for final grading is based entirely on cognitive language ability measures, whether they are tests,

quizzes or a combination of the two, those students with higher pre-course proficiency will have a notable advantage. In other words, more proficient students may obtain good grades without much effort whereas less proficient students may fail to do so in spite of their efforts. Is this fair?

In order to tackle this problem and to make assessment more learning- and growth-oriented, the writer of this article introduced dialogue journals in his listening class in the 2000/2001 academic year and in the reading (general EFL) class at the outset of the 2002/2003 academic year (for a discussion of listening journals, see Hasegawa and Norizuki, 2001). Table 1 provides a set of guidelines for writing reading journals adopted in the current fall/winter semester.

Table 1: Guidelines for Writing Reading Journals

| |
|--|
| *Students are asked to keep journals on what they've studied at home or elsewhere outside of the class in terms of the following points. They are to submit the journals five or six times on specified dates during a semester. |
| 1) Materials used |
| a) classroom textbook: numbers & names of units and pages studied |
| b) Internet sites or other materials used: names of sites, articles, etc. with attached photocopies or printouts if possible |
| 2) Method (be as specific as possible) |
| 3) Things they find difficult to understand or their questions to the teacher |
| 4) Comments (effects of learning, things they find interesting, etc.) |

The following 9 excerpts are from journals written or drawn by different students enrolled in the writer's listening and reading classes in the 2002/2003 year.

Excerpts 1 and 2 show two of the best journals delineating the student's process of learning listening, reading and writing skills. In excerpt 1, the student followed a similar set of pre-specified guidelines. She was particularly adept at elucidating important linguistic and communication problems, referring to actual problems and her own episodes of language use. In each of her entries, she tried to include some good questions to create opportunities for communication with the teacher. In excerpt 2, the student followed the guidelines described above. She adopted a highly intensive, well-organized approach to read-aloud and write-a-lot exercises, starting with intact passages followed by three levels of mutated passages, and eventually learning to read and write passages without looking at them. She diagnosed her major problem as lack of vocabulary knowledge, and provided self-devised vocabulary quizzes on a regular and frequent basis.

In excerpt 3, the student showed how much she understood the words in the blanks each time she listened to the song, and described her specific problems in

detail. She persevered with this laborious approach in each of her six entries.

Excerpt 4 shows a famous fairy tale chosen by the student as part of his extensive reading practice. In his second entry, he chose *Puss in Boots*.

Excerpt 5 displays portions of a student's journal in which he discovered and discussed the unique use of tense and aspects in on-line newspaper articles. He is the only student who is handing in typed journals in the current semester. The teacher (the writer of this article) was pleasantly surprised to find in his first reading journal his ability and willingness to explore his own way of learning English, which was not very noticeable in the first-semester listening class he enrolled in.

Excerpt 6 is a student's handwritten extract of *Aladdin* supplemented with her vocabulary practice and comments. As part of her six-page-long, dense journal entry, the student demonstrated her read-and-write-a-lot approach as a means of exploring and personalizing her favorite story. This journal entry is a showcase example of the ultimate combination of an intensive writing approach and reading for pleasure.

Excerpt 7 shows a part of an article on the latest computer games that the student accessed through the Internet and a portion of his comments and questions about it. With the help of the Japanese translation available, he was able to comprehend a relatively difficult, technical article fairly well and tried to get into the details through communication with the teacher.

Excerpt 8 is a skit writing assignment submitted with the student's listening journal. Her elaborate drawing demonstrates more than any explicit compliments that journal writing can be enjoyable and that it is not necessarily the forced output of a learning activity participated in unwillingly.

Excerpt 9 looks at first sight as if the student had simply copied the lyrics of an English song apparently made by a nonnative writer of English. The student's comments, however, reveal that he actually translated a Japanese pop song into English. He found it particularly difficult to identify the subjects of verbs that are missing or vaguely presupposed in the original Japanese version. Although his translation is far from perfect and a question remains as to the 'translatability' of the original song, his painstaking effort to create English and his own unique language learning experience should be highly commended.

For the on-going reading or general EFL class, the writer is planning to award 10 % of the points of the total summative assessment to journal writing, and maximally 5% more points will be awarded to outstandingly well-written journals.

The aforementioned samples are some of many well-written journals. It is important to notice that they were not necessarily produced by the students with the best English language proficiency or achievement scores. (It was found that the overall ratings for the reading journals collected in the 2002 spring/summer

Excerpt 2: A sample of a student's dialogue journal (2)

- 10月21日(月)
 (1) 教材: テキスト p59, 60 Lesson 2 TRACK 5, 6
 (2) 学習方法: 音読 (各1回)
 余料集 3 パターンを使い 各 TRACK 1 回ずつ書く。(ペ-3に記入)
 (3) TRACK 6 の最後の文がなかなか出てこなくて大変だった。
 (4) 何箇所か書くつもりでスラスラに書いてしまった。
 10月23日(水)
 (1) 教材: テキスト p58 ~ 60 Lesson 2 TRACK 4, 5, 6
 (2) 学習方法: テキストを音読し、各 TRACK 3 回ずつ書く。
 音読 (余料集 3 パターンを使い、間違いやうな単語の発音)
 筆記 (各 TRACK 1 回ずつ書く) (ペ-3に記入)
 (3) なし
 (4) 苦手な単語を書いて覚えた。 → 音読教材を音読して覚えた。
 10月24日(木)
 10月26日(土)
 (1) 教材: テキスト p62 ~ 64 Lesson 3 TRACK 7, 8, 9
 (2) 学習方法: CD のリポート再生 (時間帯があるときは CD をかけた)
 (3) なし
 (4) 単語を聞き取れないところを自然に耳に入らなくて困った。なるべく口に出して CD を聞いて覚えた。
 10月27日(日)
 (1) 教材: テキスト p62, 63 Lesson 3 TRACK 7, 8
 (2) 学習方法: 音読 (各 TRACK 3 回)
 余料集 3 パターンを使い 1 枚ずつ書く。2 回ずつ読む。
 各 TRACK 2 回ずつ書く。(ペ-3に記入)
 (3) TRACK の in this area のところから書く。 (ペ-3に記入)
 (4) 初めに一通り筆記をして、音読してから 1 回筆記したら、1 回目よりスラスラ書けた。
 10月28日(月)
 10月29日(火)
 (1) 教材: テキスト p62 ~ 64 Lesson 3 TRACK 7, 8, 9
 (2) 学習方法: CD をリポート再生し、本を見ながら筆記 + 内容のまとめ + 1 回ずつ単語の発音 (ペ-3, 6に記入)
 (3) あまったペ-3 を使って小テストをしたのですが、16/9 点でスアリスが 2 つありました。父は単語を覚えていないから、と怒りました。
 (4) 内容の整理はできて、単語のスアリスは合格は良かったです。

- 11月16日(土)
 TRACK 10
 Do you like traveling? I like it very much. I went to Hokkaido last summer. I visited an Ainu museum. I learned about the Ainu there. Once Hokkaido was the land of the Ainu. The names of places show this. For example, Sapporo, Kushiro, and Asahikawa are from the Ainu language.
 Do you like traveling? I like it very much. I went to Hokkaido last summer. I visited an Ainu museum. I learned about the Ainu there. Once Hokkaido was the land of the Ainu. The names of places show this. For example, Sapporo, Kushiro, and Asahikawa are from the Ainu language.
 TRACK 11
 People in Honshu went into Hokkaido at the end of the Edo period. They changed the Ainu way of life. These changes became official in the Meiji period. Many Ainu customs became illegal. They stopped hunting and fishing, and began farming against their will. They could not use their own language in public.
 People in Honshu went into Hokkaido at the end of the Edo period. They changed the Ainu way of life. These changes became official. They stopped hunting and fishing, and began farming against their will. They could not use their own language in public.
 TRACK 12
 Today some Ainu are trying to keep their traditional ways. Kayano Shigetaru is one of them. He thought, "The life of a people is in its language." So he started Ainu language classes with his friends. This moved many people. Now Ainu teachers give their people language lessons. The language lives again.
 Today some Ainu are trying to keep their traditional ways. Kayano Shigetaru is one of them. He thought, "The life of a people is in its language." So he started Ainu language classes with his friends. This moved many people. Now Ainu teachers give their people language lessons. The language lives again.

| | | | |
|-------------|-------------------|----------|-------------|
| 単語のミニ・テスト | 狩猟 hunting | 7点 / 8点中 | traditional |
| 博物館 museum | 〜に反対 against | | traditional |
| 言語 language | 大家 public | | traditional |
| 時代 period | 伝承 伝承 traditional | | traditional |
| 違法 illegal | | | |

Excerpt 5: A sample of a student's dialogue journal (5)

(3) We are in mood to put with this bull 「モー我慢できない」の見出しだが、英語は、横にしかかかないので日本語より制約があると思う。日本語だと縦の見出しでたりないところを横見出しで補うことができるが、英語ではそれが不可能であるため見出しをできるだけ簡素化し凝縮する必要がある。そのため理解力がある人は、見出しに関する限り単純明解で慣れれば記事よりよみやすいのかもしれないが、とまどうことの1つとして英語には、現在、過去、未来、過去完了、^{*}来完了などさまざまな表現があり悩まされる。

(4) 記事が過去形で書かれていても、見出しは現在形で表現される。新聞雑誌の記事の内容の大半は、ごく最近起きた過去の出来事を報じている。雪印の牛肉すり替え事件も過去のできごとであり、記事の文章は過去形で書かれているが、見出しは、現在形で書いている。現在形の方が読者に訴える力が強く感じるからだろうか。
英語の新聞等の記事は、「見出しでは過去の出来事は現在形で表わす」ということだと学んだ。

英語で読む天声人語 (イントロ版)

発行：朝日新聞社/CLUBA&A
http://www.asahi.com/

I N D E X 2002.01.25(金曜日)

1. We are in no mood to put up with this bull'
2. 「モー我慢できない」

We are in no mood to put up with this bull'

Just imagine such a bizarre scene. Snow Brand Food Co. employees were called to work on their days off to "repack," about 13.8 tons of beef into 663 boxes. I presume they were merely following orders from their superiors, but I must wonder how they felt about performing this unbelievable task.

The latest Snow Brand scam fills me with deep gloom and apprehension because of its very unpleasant implications. For one, it was only a little over a year ago that Snow Brand Milk Products Co., the parent company, caused a serious food poisoning epidemic that severely damaged its brand name. But the latest scam suggests that the Snow Brand group was apparently oblivious to the overriding need to recover its shattered honor at all costs.

Another infuriating implication is that the company tried to take advantage of the mad cow disease scare. For the food industry, this is a grave crisis requiring urgent attention. But Snow Brand not only took advantage of this crisis, it even dared to abuse the government program to buy up potentially contaminated beef with taxpayers' money.

.....

semester class were negatively correlated with TOEIC Bridge total scores, its reading subtest scores and the end-of-course test, measuring -.34, -.36 and -.29 respectively. The ratings for listening journals collected in the same period were moderately correlated with the quiz and the course-of-end test scores, showing .50 and .44 respectively, and yet were virtually uncorrelated with the TOEIC Bridge listening subtest scores (.06.)

It must be admitted that there are some students who do not appear to enjoy writing journals. These students may not be happy about the fact that they are being partially assessed by their journals. Some of them may want to write good journals, but simply do not know how. A question can be raised about the fairness of this assessment tool. However, this question is analogous to the fairness of testing varied levels of students on a common metric (for discussions of fairness in testing and assessment, see Davies, 1997; Shohamy, 1997; 2000, Hamp-Lyons, 1997, 2000; 2001, Kunnan, 2000; Ross, 2000; Lynch 2001; Norizuki, 2002). By combining multiple methods of assessment, it is expected that assessment will become fairer and more accountable for the process of language learning and growth on a more individual basis.

It seems that the students in the writer's classes are gradually getting to understand this multi-faceted view of language learning and assessment. As far as journals are concerned, the teacher has more knowledge about individual learners, enabling him to offer both formal and personal feedback to their comments and questions, while an increasing number of students are responding verbally or nonverbally to the feedback. There is a sign of general improvement with the students' ability and motivation to write about their learning process in journals.

5. Conclusion

The concept of assessment has mostly been associated with selection, grouping and grading. These are important functions of assessment, but they are not all what assessment is meant to be. Assessment in the 21st century should exert its active influence on the process of language learning. While fostering a student's self-learning ability and motivation to learn the target language, assessment should help him/her to identify both learning and language problems. Assessment should not be confined in its scope to decontextualized, structural or lexical knowledge of the language. EFL TEs should design their assessment to activate learners' language learning and enhance their ability and eagerness to use the language in a meaningful context.

Acknowledgement

I would like to record my special thanks to nine anonymous students who have kindly allowed me to reproduce their journal samples in this article.

References

- Alderson, J.C. and Beretta, A. editors. 1992: *Evaluating Second Language Education*. Cambridge: CUP.
- Brown, J.D. editor, 1998: *New Ways of Classroom Assessment*. Alexandria, VA: Teachers of English to Speakers of Other Languages, Inc.
- Brown, J.D. 2001: *Using Surveys in Language Programs*. Cambridge: CUP.
- Brown, J.D. and Hudson, T. 2002: *Criterion-referenced Language Testing*. Cambridge: CUP.
- Cohen, A. D. 1994: *Assessing Language Ability in the Classroom (Second edition)*. Boston, MA: Heinle and Heinle.
- Cumming, A. 2001: ESL/EFL instructors' practices for writing assessment: specific purposes or general purposes?
- Davies, A. 1997: Introduction: the limits of ethics in language testing. *Language Testing* 14, 235-241.
- Davies, A. 2001: The logic of testing Language for Specific Purposes. *Language Testing* 18, 133-147.
- Douglas, D. 2000: *Assessing Languages for Specific Purposes*. Cambridge University Press: CUP.
- Douglas, D. 2001: Language for Specific Purposes assessment criteria: where do they come from? *Language Testing* 18, 171-185.
- Genesee, F. and Upshur, J.A. 1996: *Classroom-based Evaluation in Second Language Education*. Cambridge : CUP.
- Ekbatani, G, and Pierson, H. editors, 1999: *Learner-directed Assessment in ESL*. Mahwah, NJ.: Laurence Erlbaum Associates.
- Hamp-Lyons, 1997: Washback, impact and validity: ethical concerns. *Language Testing* 14, 295-303.
- Hamp-Lyons, L. 2000: Fairness in language testing. *Studies in Language Testing* 9, Cambridge: CUP, 30-34.
- Hamp-Lyons, L. 2001: Ethics, fairness(es), and developments in language testing. *Studies in Language Testing* 11, Cambridge: CUP, 222-227.
- Hasegawa, K. and Norizuki, K. (長谷川和則・法月健) 2001: What university students have learned through keeping listening journals. (『リスニング・ジャーナルによる多聴活動のもたらしたもの』), *LET Journal of Central Japan* 12: 56-69.

- Heaton, J.B. 1988: *Writing English Tests (Second Edition)*. London: Longman.
- Kunnan, A.J. 2000: Fairness and justice for all. *Studies in Language Testing 9*, Cambridge: CUP, 1-14.
- Lynch, B. 2001: The ethical potential of alternative language assessment. *Studies in Language Testing 11*, Cambridge: CUP, 228-239.
- Matsuzawa, S. (松沢伸二) 2002: *New ways of Assessment for English Language Teachers*. (「英語教師のための新しい評価法」) Tokyo: Taishuukan (大修館).
- Norizuki, K. 2002: Fairness in institutional English language testing. Paper read at JALT 2002, Shizuoka, November 2002
- Ross, S.J. 2000: Econometric assumptions for edumetric tasks? Some issues of examination ethics and policy. *Journal of Policy Studies 9*, 175-182.
- Shohamy, E. 1997: Testing methods, testing consequences: are they ethical? are they fair? *Language Testing 14*, 340-349.
- Shohamy, E. 2000: Fairness in language testing. *Studies in Language Testing 9*, Cambridge: CUP, 15-19.
- Weir, C. and Roberts, J. 1994: *Evaluation in ELT*. Oxford: Blackwell.