

Overlooked Motivational Strategies: Question Types and Test Scoring Methods

Tomoyuki Kawashima

Gunma University

Contact: tkawashima@gunma-u.ac.jp

Scoring tests is such a routine task that teachers seldom consider whether or not this process could have pedagogical applications. This paper reports on action research which explored the possibility of using question types and test scoring methods as motivational strategies. Two types of tests, cloze and translation tests, were given to first-year senior high school students in high-stakes achievement tests, and scoring methods which differed from the conventional exact-answer scoring method were adopted. In order to examine whether the question types and scoring methods motivated students to study English, a questionnaire survey was administered to 118 students. Student reactions were analyzed in accordance with their English proficiency levels. Data analysis suggested that the translation test and its scoring method contributed to enhancing motivation of students with lower proficiency levels.

Introduction

Tests are a double-edged sword for students. While good results can generate high motivation, poor results can demotivate students. Therefore, it is worthwhile for teachers to take into account how good test results may influence student motivation. However, scoring tests is a routine task for teachers, and teachers often repeat conventional question types and scoring methods without giving a second thought to potential benefits that tests can provide. To explore the potential benefits of using tests as a motivational strategy, an action research project was developed by a group of high school teachers. The new question types required greater effort for preparation on student side, while the new scoring methods allowed students to obtain good scores with relative ease which they would not have achieved in usual tests. To examine a) whether students found the new question types and scoring methods motivating to study English, and b) whether English proficiency levels exerted any effects on student reactions, a questionnaire survey was conducted.

Importance of Enhancing Self-Confidence for Motivation

Unsatisfactory performances in tests can demotivate learners. Sakai and Kikuchi's (2009) investigation into 656 Japanese senior high school students revealed that the test score is the second most demotivating factor for low motivated students. Furthermore, in a 2004 study conducted by Falout and Maruyama, 38.9% of low proficiency and 31.7% of high proficiency university students attributed demotivation to disappointment in performance, and it was the leading demotivating factor in both groups. Literature shows that Japanese learners are more inclined to attribute the cause of their demotivation to internal factors such as reduced self-confidence (Tsuchiya, 2006). This finding suggests that self-confidence is a crucial factor for Japanese learners to maintain their motivation to study English. Indeed, self-confidence influences achievement (Clement, Dornyei & Noels, 1994; Matsuda & Gobel, 2004). Expectancy-value theories explain that motivation is produced by two factors: an individual's expectancy of success in a given task and the value the individual attaches to success in that task (Dornyei, 1998, p. 119). In other words, the more likely they feel they can achieve the goal and the more valuable the goal is for them, the higher and more positive their motivation to accomplish the task will be. Linnenbrink and Pintrich (2003) argue that learners will attempt tasks slightly beyond their skill level if they overestimate their current proficiency level. An interesting finding is reported that students get higher scores on tests if the test begins

with easier test items (Jinks & Lorschach, 2003). These studies indicate that students' perceived competence and prospect of completing the task will influence their achievement.

As has been argued above, one area where we can see the influence of students' perception on their achievement is testing. The positive or negative influence of testing on teaching and learning is called washback (Bailey, 1996). Alderson and Wall (1993) propose 15 possible effects as Washback Hypotheses, and some of them refer to the potential washback on learner's motivation. For instance, "a test will influence how learners learn", and "a test will influence attitudes to the content, method, etc. of ... learning." (p. 120) On the basis of these hypotheses, this research project was designed and it attempted to examine whether question types and scoring methods would have beneficial washback on student motivation.

Method

Participants

This research was conducted at a public senior high school in central Japan. A total of 238 students were enrolled in the first year, and three teachers taught the same course to these students. Students were divided into six mixed ability classes at random. Data was collected from half of the students (N=118). They were in three classes which the author was in charge of. All the students were bound for university and their English proficiency levels were higher than the national average. Students scored 59.6 on average in a national standardized proficiency test administered by Benesse Corporation in which the mean Standard Score (SS) is always fixed at 50. The SSs ranged from the lowest 39.5 to the highest 77.2, and the standard deviation (SD) was 7.6. To examine the effect of proficiency levels, participants were divided into three proficiency groups based on their SSs. Upper Proficiency Group (UP) consisted of 35 students who earned SSs higher than 63.4, or the group's mean SS (59.6) plus the group's half SD (3.8). Likewise, Lower Proficiency Group (LP) was made up of 42 students whose SSs were below 55.8, or the group's mean SS (59.6) minus the group's half SD (3.8). The remaining 41 students whose SSs were between 55.8 and 63.4 were classified into Middle Proficiency Group.

Context of Study

The entrance examinations to university have a dominant influence on the teaching and testing at high schools. Nishino (2011) revealed that the entrance examinations to university are the second major reason, after the class size, why teachers are hesitant to introduce communicative language teaching. For teachers at high ranked schools, preparing students for entrance examinations is the ultimate goal to strive for. They believe giving classroom tasks and question types for tests which are similar to those in university entrance examinations is the most efficient. Kikuchi (2006) looked into the difficulty of the reading passages for entrance examinations of 20 prestigious universities in Japan in 1994 and in 2004. In spite of the Ministry's repeated request for a shift toward communication-oriented teaching at senior high schools, Kikuchi found that reading passages for entrance examinations remained very difficult and that translation tasks were still frequently used. Under these circumstances, it is fairly common among high school teachers to continue giving translation tasks for in-house tests.

This research was undertaken when teachers were in desperate need of a change. The school had a long history of sending graduates to prominent universities, but its popularity was on a decline. Teachers were under greater pressure to raise student proficiency levels. After deliberate discussion, they decided to set recitation of the textbook text as the main goal of their lessons and adapt the test accordingly. They were aware of the significant benefits of recitation in enhancing English proficiency levels on the basis of their own learning and teaching experiences. Teachers also decided to change question types to the tasks to reproduce texts from the textbook. Their choice was unusual, but the teachers made the drastic decision under a strong conviction that it is the test that has the greatest impact on students. College-bound high school students take many tests. With regards to high-stakes tests, they usually take in-house achievement tests (5 times a year), in-house proficiency tests (3~4 times a year), and external proficiency tests by private companies (3~6 times a year). In-house achievement tests have important consequences for students because their scores from these tests form a principal component of their semester grade. In addition, the transcript of grades are submitted to university when students apply for admission. On the other hand, in the case of external proficiency tests, it is customary for test-takers to indicate the names of

universities they want to attend in registration. By doing so, test-takers can learn their relative positions in comparison with their seniors at the same school or with their competitors outside. This information is crucial for students as well as for teachers in choosing the universities they will apply to.

Question Types

In most high school achievement tests in which translation tasks are common, students write more Japanese than English. Consequently, they are often used to remembering a Japanese translation of the text before tests. When teachers decided to give new types of questions in tests, they thought it necessary to give guidance to students and emphasize the importance of reproducing the English precisely. One question type was a cloze test where students filled the gaps in the texts with the exact word(s) which were in the texts from a textbook. Teachers chose missing words intentionally not only from the parts of the text which included new words and phrases but also from the other parts in order to remind students of the importance to remember the whole text. The sentences with blanks were provided either in English or in Japanese, and students always had to answer by using the missing English words from the original English text. The other question type was a translation test. The translation test required students to translate whole or part of a Japanese sentence into English using expressions from the original text. The formats of the tests remained the same throughout the year.

Scoring Methods

The number of answer blanks in the cloze test varied from 14 to 17 across test administrations, but the total number of scores which were given to each cloze test was always kept at 11. Hence, the scores were compressed. Table 1 shows the scores students earned using both the conventional and new methods in the case of a cloze test with 17 blanks. The scoring method was intended to be lenient to students with lower proficiency levels. Even if they answered as few as three blanks correctly, they still got four points. On the other hand, it set the ceiling for students in the higher proficiency range. Students who gave 17 perfect answers and those who answered 14 questions correctly, both received the maximal points of 11.

Table 1: Scoring Method for the Cloze Test

Numbers of Correct Answers	0	1-2	3-4	5-6	7-9	10-13	14-17
Scores Earned with Conventional Scoring Method	0	1-2	3-4	5-6	7-9	10-13	14-17
Scores Earned with New Scoring Method	0	2	4	6	8	10	11

In the case of the translation test, two points were always allocated with each question. The new scoring method differed from the conventional method in that one point was subtracted only when two errors were made. Therefore, both students whose translation was flawless and those who made one error received two points. Moreover, one point was given even when three errors were made (Table 2).

Table 2: Scoring Method for the Translation Test

Numbers of Errors	4	3	2	1	0
Scores Earned if Marked Conventionally	0	0	0	1	2
Scores Earned if Marked with New Method	0	1	1	2	2

These unconventional scoring methods were adopted in order to reduce the severity of the test and to encourage students. If marked conventionally, points were always deducted for an error. However, the new scoring methods had a wider range of tolerance to errors. This was because the level of requirements became higher with the new question types. Any deviance from the original text in the textbook was counted as an error. For instance, even a missing article was

regarded as an error. For this reason, the teachers thought it necessary to tolerate a minimum number of errors and to encourage students to try without being afraid of making errors.

Data Collection and Analysis

After participants took two achievement tests with the new question types and scoring methods, a questionnaire was administered. It had two statements: “The cloze test and its scoring method motivate me to study English,” and “The translation test and its scoring method motivate me to study English.” Participants responded by choosing an answer closest to their ideas from 1 (strongly disagree) to 6 (strongly agree). They were also asked to write the reasons for their choices. In addition, the impact of the new scoring methods for each proficiency group was examined by comparing the scores with those obtained when marked conventionally. The answer sheets for the second achievement test of all the participants were copied and partially marked again with the conventional scoring methods. The achievement test was always based on two chapters in the textbook. Due to time limitations, the cloze and translation tests on the first chapter were marked again for comparison. Participants were not informed of their score differences between conventional and new scoring methods when they answered the questionnaire. Hence participant evaluations were based on their own experiences of the new methods.

Results and Discussion

The analysis of the questionnaire showed that the means of the ratings for all the 118 participants were 4.55 for the translation test and 4.21 for the cloze test. This suggests participants found the question types and marking methods for both cloze and translation tests more motivating. It also indicates that the translation test was more favorably received by participants. Table 3 presents the results of further analysis into the difference among proficiency groups. It indicates that the mean scores of three proficiency groups centered around 4.2 for the cloze test, whereas participants in LP found the translation test more highly motivating. However, when one-way ANOVA was run on the ratings, no statistical difference was detected between groups in either question type.

Table 3: Reaction to Question Types and Scoring Methods

Proficiency	Upper Group (N=35)	Middle Group (N=41)	Lower Group (N=42)
Cloze Test	4.20	4.24	4.20
Translation Test	4.40	4.48	4.76

The analysis of qualitative data revealed that participants positively responded to the cloze test and its scoring method. A total of 54 comments were collected, and 68.5% were in favor. They referred to the ease of getting higher scores, e.g., “I can get full marks if I work hard.” Owing to greater possibilities of obtaining high scores, participants became more willing to prepare for the tests. For instance, “If I memorize the English text, I can get higher scores. So I feel I can work hard next time too.” In addition, the leniency of the new scoring method reduced stress during the test preparation. One participant wrote “As a few mistakes do not affect my scores, I can relax and prepare for the test.” On the other hand, two reasons for dissatisfaction with the new method emerged from the 17 negative comments. More than half of the comments referred to fairness. For example, “It is not fair that the scores are the same even if the number of correct answers differs.” Furthermore, three comments mentioned negative washback. They stated that it might encourage students to slacken their efforts in memorizing words, and that they might end up memorizing the minimal number of expressions. However, this concern did not seem to be a widely held one, for there were many positive comments referring to the eagerness to learn target sentences more precisely. For example, “The cloze test makes me feel like memorizing the English texts perfectly.” “In order to get more points, I pursue higher accuracy of my knowledge.”

Regarding the translation test, 51 comments were gathered, of which only four comments (7.8%) were negative. Nearly half of the positive comments were about the sense of achievement. Typical comments were as follows: “This way, it is easier for efforts to bear fruit,” and “The harder we work, the more scores we can get.” Some students mentioned, “The translation test demands that we should study the details, not just the outline of the text. So we need to check our answers

carefully for that.” The other half of the comments were about adequacy of the level of requirement. Any deviation from the original text was counted as an error. For this reason, the lenient scoring method mitigated the harshness of the test and it helped maintain participants’ motivation to study hard. They stated, “It is very difficult to remember the use of articles. That’s why, the moderately low level of requirements in scoring makes me feel like giving it a try.” On the contrary, a small number of comments (N=4) showed that they were apprehensive that some students would make light of small mistakes. Overall, it could be summarized that except for a few, participants favorably received the question type and scoring method of the translation test.

The differences in scores by two scoring methods lent support to the finding that the translation test and its new scoring method motivated LP participants. The cloze test used for this investigation had 31 gaps and full marks were 22 points. If marked conventionally, the full marks would have become 31 points, meaning a nine-point increase. Table 4 shows the percentage of participants who would have been affected by the gain or loss if marked with the conventional scoring method. As far as UP is concerned, nearly two thirds of the participants would have gained 6 to 9 points with the change in scoring methods. On the other hand, the scores of 22% of the participants in Middle Proficiency Group would have been one or two points lower. The situation would have been even more unfavorable for LP. As much as 59% of the participants would have lost one or two points. These results underpinned the legitimacy of the teachers’ conviction that the new scoring method would have been more advantageous for Middle and Lower Proficiency Groups.

Table 4: Score Differences in the Case of Cloze Test

Proficiency	+9	+8	+7	+6	+5	+4	+3	+2	+1	0	-1	-2
Upper Group	25%	13%	8%	17%	4%	8%	0%	8%	4%	0%	8%	4%
Middle Group	0%	3%	9%	6%	13%	6%	0%	9%	3%	28%	19%	3%
Lower Group	0%	0%	0%	0%	4%	17%	0%	8%	4%	8%	46%	13%

Table 5: Score Differences in the Case of Translation Test

Proficiency	0	-1	-2	-3	-4	-5	-6	-7	-8	-9
Upper Group	17%	13%	17%	13%	21%	4%	4%	8%	4%	0%
Middle Group	6%	6%	19%	13%	6%	9%	22%	19%	0%	0%
Lower Group	17%	17%	0%	21%	13%	4%	8%	8%	8%	4%

Table 5 shows how the score would have differed in the case of the translation test with 11 questions. Each question was worth two points, so full marks were 22 points. Regardless of proficiency levels, more than 80% of the participants would have obtained lower scores. However, especially noteworthy was that slightly higher percentage of LP than UP would have been hardly affected. The total score of 34% of LP and 30% of UP would not have changed at all or would have become one point lower. This demonstrates that participants in LP prepared very hard for the translation test.

This action research has limitations as well. As the mean SS (59.6) suggests, participants were in the higher range of English proficiency in comparison with average Japanese high school students. Many of the participants in LP in this study would be categorized into middle proficiency group in other studies. Furthermore, it should be noted that classroom activities were adapted to reinforce the relevance with the new question types as well.

To conclude, the results of quantitative and qualitative data analysis suggest that the new test types and their scoring methods contributed to enhancing students’ motivation to study English. Especially, the translation test and its scoring method were more favorably received by students as proficiency levels went down. The examination of the impact of the new scoring method revealed that answers of as many students in LP as in UP were flawless. This means that the question

type and scoring method contributed to motivating LP participants. They were able to see the goals they should strive for and the process they should go through to attain the goal.

Note: An earlier version of this paper was presented at JACET-Kanto 9th Annual Convention on July 12, 2015.

References

- Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129.
- Bailey, K. M. (1996). Working for washback: A review of the washback concept in language testing. *Language Testing*, 13(3), 257-279.
- Clement, R., Dornyei, Z., & Noels, K. A. (1994). Motivation, self-confidence, and group cohesion in the foreign language classroom. *Language Learning*, 44(3), 417-448.
- Dornyei, Z. (1998). Motivation in second and foreign language learning. *Language Teaching*, 31, 117-135.
- Falout, J., & Maruyama, M. (2004). A comparative study of proficiency and learner demotivation. *The Language Teacher*, 28(8), 3-9.
- Jinks, J., & Lorschach, A. (2003). Introduction: Motivation and self-efficacy belief. *Reading & Writing Quarterly*, 19, 113-118.
- Kikuchi, K. (2006). Revisiting English entrance examinations at Japanese universities after a decade. *JALT Journal*, 28(1), 77-96.
- Linnenbrink, E. A., & Pintrich, P. R. (2003). The role of self-efficacy beliefs in student engagement and learning in the classroom. *Reading & Writing Quarterly*, 19, 119-137.
- Matsuda, S., & Gobel, P. (2004). Anxiety and predictors of performance in the foreign language classroom. *System*, 32, 21-36.
- Nishino, T. (2011). Japanese high school teachers' beliefs and practices regarding communicative language teaching (in Japanese). *JALT Journal*, 32(2), 131-155.
- Sakai, H., & Kikuchi, K. (2009). An analysis of demotivators in the EFL classroom. *System*, 37(1), 57-69.
- Tsuchiya, M. (2006). Profiling of lower achievement English learners at college in terms of demotivating factors. *Annual Review of English Language Education in Japan*, 17, 171-180.

Bio: Tomoyuki Kawashima is Associate Professor at the Graduate School of Health Sciences, Gunma University. Prior to the present job, he taught English to high school students for 25 years. His research interests include pedagogical applications of World Englishes in English language teaching, affective factors in speaking English, and the development of speaking and writing skills.