Effects of Distractor Characteristics on the Use of Elimination Strategies by Japanese EFL Learners in Vocabulary Tests

Yuko HOSHINO
Graduate School, University of Tsukuba / The Japan Society for the Promotion of Science

Abstract

The aim of this study was to explore whether the use of elimination strategies differed with the change of distractors in multiple-choice vocabulary tests in sentential context. The distractors were classified into two frequency levels and into three different types, and the participants thought aloud while answering the test. Participants' elimination strategies were coded into four types, and the results showed that frequency of elimination uses did not differ with any change of distractors. In other words, the uses of elimination strategies were constant across any types of items this study adopted. However, there was a critical difference between types of elimination (i.e., informed elimination and random elimination) regarding their roles on the test performance. When participants adopted random elimination, the results of the test were totally different from the overall test performance, but this was not the case in terms of informed elimination. The findings implied that random elimination has a strong influence on the construct of tests; hence, it is necessary to create tests that reduce the use of random elimination as far as possible.

1. Introduction

Elimination is a specific test-taking strategy for multiple-choice (henceforth, MC) tests, and it is almost inevitable that it will be used when answering such tests. Cohen and Upton (2006; 2007) who investigated the use of test-taking strategies found that their participants used elimination in 80% of all cases when they took TOEFL reading sections. However, this peculiarity of MC tests leads to the criticism that test-takers can reach correct answers without a clear understanding of these answers when they rely on elimination too much. Moreover, because elimination is regarded as related to guessing, it is often classified as test-wiseness, which hinders test validity (e.g., Allan, 1992; Davis, Brown, Elder, Hill, Lumley, & McNamara, 1999; Diamond & Evans, 1972; Slakter, Koehler, & Hampton, 1970).

However, it is only random elimination that might be directly related to guessing, whereas other types of elimination might not. According to Bachman and Palmer (1996) who distinguished between random guessing and informed guessing, the former occurs when a choice is made randomly; that is, if there are four choices, each choice has a 25% possibility of being selected. On
the other hand, people who use informed guessing do not select one choice in a completely random fashion but narrow down the possible answers by using the knowledge that is intended to be measured. Hence, unlike random guessing, informed guessing is more or less relevant to the construct of the test, and what we should focus on is random guessing. Nevertheless, no researchers have subcategorized guessing strategies or elimination strategies except for Cohen and Upton (2006; 2007) who revealed that the use of random elimination was much less frequent than that of elimination based on linguistic knowledge.

When we move our focus back to elimination strategies in general, researchers so far think that the use of elimination has some relationship with the types of distractors. For example, MC item-writing guidelines suggest creating plausible distractors in order to attract test-takers (Haladyna & Downing, 1989; Haladyna, Downing & Rodriguez, 2002). If elimination strategies are not supposed to be used at all, any distractors will play the same role and we need not be concerned about what kinds of distractor we prepare. However, research findings have shown that distractors affect test performance. For example, tests with distractors that were semantically related to the correct option were revealed to create more difficulty than tests with unrelated distractors, other conditions being equal (e.g., Greidanus & Nienhuis, 2001; Jenkins, Matlock, & Slocum, 1989; Nagy, Herman, & Anderson, 1985). This is one proof that test-takers tend to rely on elimination, at least to some extent, because unrelated distractors are supposed to be eliminated more easily than semantically-related distractors and lead to higher proportions correct. Therefore, there seems to be some relationship between types of distractor and the use of elimination strategy, which in turn might affect test performance. However, few researchers have shed light on these topics; hence, the current study aimed to elucidate them using MC vocabulary tests in sentential context.

In this study, three kinds of distractors were adopted following Morimoto (2008a; 2008b): paradigmatically-related (semantically-related) distractors, syntagmatically-related distractors, and control distractors which had neither a semantic nor a syntagmatic relationship (unrelated). The results of her studies indicated two points: (a) test-takers were attracted more to syntagmatically-related distractors than to unrelated distractors, and (b) the test with unrelated distractors was easier than the test with syntagmatically-related distractors. Although her studies did not show consistent results regarding paradigmatically-related distractors, it would be interesting to include this type of distractor because past studies that used word-in-isolation tests agreed that tests with semantically-related distractors were more difficult than tests with unrelated distractors (Greidanus & Nienhuis, 2001; Jenkins, Matlock, & Slocum, 1989; Nagy, Herman, & Anderson, 1985). In addition to the three types of distractor above, the current study prepared two frequency levels for choices because frequency is one of the most dominant factors that affect performance in vocabulary tests (e.g., Schmitt et al., 2001). By considering these distractor characteristics, the following research questions were addressed:
RQ1: Does the use of elimination strategies differ with the change of distractor characteristics?
RQ2: Does the performance of test-takers who use elimination strategies differ with the change of distractor characteristics?

2. Method

Twenty university students with a variety of majors participated in this study. They had been learning English for at least six years at the time of the experiment, and most of them had correctly answered at least 70% of vocabulary items in the STEP Eiken’s 2nd grade test. They answered an MC vocabulary test in sentential context, which is a gap-filling format with four choices (one correct option and three distractors). The test contained 39 items in total, and they had distractors with different characteristics: (a) distractors in 13 items were paradigmatically-related words that had a paradigmatic relationship with the word in the correct answer; (b) distractors in another 13 items were syntagmatically-related words that had a syntagmatic relationship with one of the words in the context; and (c) distractors in the remaining 13 items were unrelated words that did not have the relationships stated above. The selection of words in the distractors was the same as Morimoto (2008a; 2008b) who used WordNet 3.0 (Princeton University Cognitive Science Laboratory, 2006) for the selection of paradigmatically-related distractors and adopted the loglog score from the British National Corpus to select syntagmatically-related distractors. All of the distractors in an item had the same word class as the correct alternative, as well as the same frequency level in the JACET 8000 word list (JACET, 2003). Nine items out of 13 each belonged to frequency level 1 in JACET 8000 (level 1 items), and the remaining four items each belonged to level 2 (level 2 items). After all the distractors were selected, three English native speakers checked the tests, and confirmed that only the correct answers could fit into the context. Furthermore, the results of a pilot study indicated that all types of distractors attracted lower-level participants more than upper-level participants; hence, the distractors in this study were regarded as functioning well. Examples of the items can be seen in the Appendix.

Participants answered the vocabulary test in sentential context by using the thinking aloud technique, but before that they were instructed on how to think aloud. This instruction, based on Cohen and Upton (2006), included an explanation of the purpose of this study and how to think aloud, and also involved practicing thinking aloud. After confirming that they were sure of what to do, they took the MC vocabulary test. They were allowed to spend as much time as they wanted, and all of them finished the test within 77 minutes. They gave prior consent to participation in this study and to having their voices recorded.

As for coding of the transcribed think-aloud protocol, the framework of test-taking strategies by Cohen and Upton (2006) was adopted. Their scheme includes many categories; however, this study focused only on elimination strategies. There were seven strategies: (a) selects
options through elimination of other option(s) as unreasonable based on background knowledge; (b) selects options through elimination of other option(s) as unreasonable based on paragraph/overall passage meaning; (c) selects options through elimination of other option(s) as similar or overlapping and not as comprehensive; (d) selects options through their discourse structure; (e) discards option(s) based on background knowledge; (f) discards option(s) based on vocabulary, sentence, paragraph, or passage overall meaning as well as discourse structure; and (g) uses the process of elimination (i.e., selecting an option even though it is not understood, out of a vague sense that the other options could not be correct). Among these strategies, (a) and (e) overlapped so they were integrated into one strategy. Moreover, (b) and (d) were not used because this study adopted sentential contexts; hence, paragraphs and discourse structures never appeared. Consequently, the remaining strategies can be summarized as follows: (1) elimination of other options because they were similar or overlapping and not as comprehensive; (2) elimination based on background knowledge; (3) elimination based on vocabulary or sentence, and (4) elimination out of a vague sense of incorrectness (related to random guessing). First, two raters coded 10% of the data and then reached 79.94% agreement. After disagreement was resolved by discussion, the author alone coded the remaining data.

3. Results and Discussion

3.1 Preliminary Analysis

The study by Morimoto (2008b) who used similar distractors showed that items with unrelated distractors tended to be easier than those with syntagmatically-related distractors. In order to confirm whether this tendency could be observed in this study, a 2 (frequency: levels 1 and 2) × 3 (types of distractors: paradigmatically-related, syntagmatically-related, and unrelated) two-way ANOVA was conducted. The descriptive statistics are presented in Table 1.

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Distractor</th>
<th>n of Items</th>
<th>Mean</th>
<th>SD</th>
<th>Frequency</th>
<th>Distractor</th>
<th>n of Items</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>Paradigmatic</td>
<td>9</td>
<td>65.56</td>
<td>11.90</td>
<td>Level 2</td>
<td>Paradigmatic</td>
<td>4</td>
<td>45.00</td>
<td>22.36</td>
</tr>
<tr>
<td></td>
<td>Syntagmatic</td>
<td>9</td>
<td>70.00</td>
<td>18.42</td>
<td></td>
<td>Syntagmatic</td>
<td>4</td>
<td>48.75</td>
<td>23.61</td>
</tr>
<tr>
<td></td>
<td>Unrelated</td>
<td>9</td>
<td>74.44</td>
<td>16.56</td>
<td></td>
<td>Unrelated</td>
<td>4</td>
<td>88.75</td>
<td>12.76</td>
</tr>
</tbody>
</table>

The results revealed that there were significant interactions, $F(2, 38) = 13.40, p = .00$, as well as significant main effects of frequency, $F(2, 38) = 9.00, p = .01$, and distractor, $F(2, 38) = 24.83, p = .00$. The subsequent analyses showed that level 1 items were significantly easier than level 2 items, although this was not the case in items with unrelated distractors. Furthermore, items with unrelated distractors were found to be easier than items with paradigmatically- and syntagmatically-related distractors, although no significant difference was detected in level 1.
Despite of some exceptions, the overall results followed the expected pattern: level 1 items were easier than level 2 items, and items with unrelated distractors were easier than those with paradigmatically- and syntagmatically-related distractors.

3.2 Frequency of Elimination Use

In this section, chi-square analyses were conducted to investigate whether the frequency of elimination use differed with the types of distractors. The results are summarized in Table 2 which indicates the use of elimination due to overlapping choices and elimination based on background knowledge were extremely rare compared to elimination based on linguistic knowledge and random elimination. This is because the use of elimination based on background knowledge can be partially controlled when test-writers create items. In order to make tests more valid, that is, in order to make the test measure vocabulary knowledge, the author paid special attention to ensure that testees could not answer items only with background knowledge. Hence, this type of elimination was seldom used. In terms of elimination due to overlapping choices, it is surprising that almost no participants used it, especially in the items with paradigmatically-related distractors, because paradigmatic association means words having similar meanings. That is, although distractors had a semantic relationship with each other in paradigmatically-related distractors, test-takers did not eliminate choices because they overlapped. This is good news for testers since this type of elimination is regarded as test-wiseness (e.g., Allan, 1992). Overall, these two types of elimination strategy were seldom used; therefore, they were not analyzed from this point on.

<table>
<thead>
<tr>
<th>Frequency</th>
<th>Distractor</th>
<th>n of items</th>
<th>Overlapping</th>
<th>Background knowledge</th>
<th>Linguistic knowledge</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>Paradigmatic</td>
<td>9</td>
<td>2</td>
<td>1</td>
<td>76</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Syntagmatic</td>
<td>9</td>
<td>0</td>
<td>3</td>
<td>81</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Unrelated</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td>80</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>27</td>
<td>3</td>
<td>6</td>
<td>237</td>
<td>83</td>
</tr>
<tr>
<td>Level 2</td>
<td>Paradigmatic</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>33</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Syntagmatic</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>37</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Unrelated</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>39</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>12</td>
<td>0</td>
<td>7</td>
<td>109</td>
<td>61</td>
</tr>
</tbody>
</table>

Note. Overlapping means eliminating other options because they were similar or overlapping and not as comprehensive; background knowledge indicates elimination based on background knowledge; linguistic knowledge indicates elimination based on vocabulary or sentence, and random means elimination out of a vague sense of incorrectness (related to random guessing).

Focusing on elimination by linguistic knowledge (hereafter, informed elimination) and random elimination, chi-square analyses were conducted in order to investigate whether the frequency of elimination use differed with the types of distractors. However, there were no significant differences, $\chi^2 (2) = .17, p = .92$ for informed elimination, and $\chi^2 (2) = .44, p = .80$ for
random elimination. These results indicate that the uses of elimination were almost constant across items with different types of distractors. Although level 1 items seemed too easy for university students, they still adopted both informed and random elimination. These results partially support Cohen and Upton (2006; 2007) who indicated that their participants used elimination by knowledge with very high frequency and used random elimination with moderate frequency in vocabulary items, even when the items were fairly easy (average proportion correct was 81%). Another point that supported Cohen and Upton was that informed elimination was used far more frequently than random elimination. However, the results did not show whether the use of elimination affected actual test performance, which was the focus in the next section.

3.3 Effectiveness of Elimination Strategies

In order to see whether informed elimination and random elimination influenced test performance in items with specific distractors, a 2 (frequency) x 3 (types of distractor) ANOVA was conducted, restricted only to cases when either informed or random elimination was used. First, regarding informed elimination, there was a significant main effect of the types of distractors, $F(2, 340) = 9.41, p = .00$, and there was a significant tendency of the main effect of frequency levels, $F(1, 340) = 3.42, p = .07$, whereas interaction was not significant, $F(1, 340) = .79, p = .46$. These outcomes indicated that level 1 items reached higher proportions correct, and items with unrelated distractors were easier than items with paradigmatically- and syntagmatically-related distractors, between which there was no significant difference. This was almost along the same lines as the results in Table 1 in 3.1. Hence, informed elimination did not make items with any type of distractor easier or more difficult than the original test performance; that is, use of informed elimination did not change the overall results of the tests.

The same analysis was conducted for random elimination. The results showed that there was a significant interaction, $F(2, 138) = 3.54, p = .03$, as well as a significant main effect of the types of distractors, $F(2, 138) = 3.64, p = .03$, although the main effect of frequency levels was not significant, $F(1, 138) = .03, p = .87$. The analysis of the simple main effect showed that level 1 items were significantly easier than level 2 items only when distractors had a paradigmatic relationship. These results were completely different from both the results of overall test

Figure 1. Proportions correct in informed elimination.

Figure 2. Proportions correct in random elimination.
performance in 3.1 and those of informed elimination above.

The differences are more apparent in Figures 1 to 3. First, Figure 1 shows the proportions correct of the test when informed elimination was used. As the results of the ANOVA indicated, the success rates of level 1 items were higher than those of level 2 items, and items with unrelated distractors obtained the highest proportions correct. On the other hand, Figure 2, which shows the results of the test when random elimination was adopted, illustrates completely different results. Contrary to the expectation, level 2 items were easier than level 1 items, except for paradigmatically-related distractors. When we compare Figure 2 with the result of the whole test presented in Figure 3, the pattern of test results is totally different, especially in level 1 items when the testees used random elimination. The proportions correct were much lower than those of the whole test, and the graph was v-shaped, while in level 2 items, all of the lines in Figures 1 to 3 move diagonally upwards to the right.

The reason for the different patterns between frequency levels would reflect the difficulty the participants had in explaining their thoughts while they were thinking. In level 2 items, there were some cases in which participants had difficulty reasoning why they selected or eliminated particular choices even when they used linguistic knowledge. Let us consider the examples of protocols and their related items.

<Level 2 item with unrelated distractors>
It is important, however, to ( ) this work from the fiction of, say, B. S. Johnson and Christine Brooke-Rose.

a. bother        b. distinguish*        c. complain        d. satisfy

“I have a vague sense that ‘distinguish’ is the answer. (The interviewer: Why did you choose ‘distinguish’?) There is ‘from’ in the latter part of the sentential context…, I have a vague sense that it is correct. It’s my intuition.”

<Level 2 item with paradigmatically-related distractors>
As we pass one café, the waiters race out to ( ) tables and chairs inside.

a. attract        b. breathe        c. drag*        d. stretch
“Stretch tables, drag tables, breathe tables, we don’t say ‘breathe tables.’ ‘Breathes, attract tables,’ it’s not…. ‘Stretch table…’ the answer is ‘drag.’ (The interviewer: Why did you choose ‘drag’?) It’s out of a vague sense that it is correct.”

In the first example, the participant seemed to decide the answer based on the collocation between ‘distinguish’ and ‘from,’ and in the second example, the participants checked the fit between the words in choices and ‘tables’ that comes right after the brackets. In both cases, testees seemed to decide or eliminate choices by using linguistic knowledge such as collocation and contextual fit, but both participants said they decided the answers out of vague sense of correctness. That is, they could not clarify the source of their reasons. Hence, it can be assumed that verbalizing reasons was more demanding when the words in the choices were relatively difficult (i.e., frequency level 2 in this case) than when they were fairly easy (i.e., frequency level 1), because there were almost no such cases in level 1 items. These facts indicated that the use of random elimination in level 2 items might not reflect pure random elimination, because it was contaminated by the case when the participants could not explain their thoughts fully. In level 1 items, on the other hand, random elimination was coded only when the participants really had no idea about which choices fit the context. Therefore, regarding random elimination, it would be more valid to interpret the results from level 1 items than from level 2 items.

Integrating the interpretations above, the results showed that using informed elimination did not affect the overall test performance. In other words, the results of the whole test and the results of those who used informed elimination had a similar tendency. However, the results of those who used random elimination were totally different from the results of the overall test; therefore, only random elimination has a malign influence on the validity of the test.

4. Conclusion

This study examined the frequency of use of four elimination strategies employed while answering a vocabulary test in sentential context, and also investigated their influence on test performance. It was found that elimination due to overlapping choices and elimination based on background knowledge were seldom used, whereas elimination based on linguistic knowledge and random elimination were consistently used with items with any type of distractor. Therefore, the answer to RQ1 was that the use of elimination strategies did not differ with distractor characteristics, although some eliminations were more likely to be used than others. In terms of RQ2, the results showed that items with frequency level 1 and with unrelated distractors were identified better when participants used informed elimination, but there were totally different results in random elimination: level 1 items with paradigmatically-related distractors had significantly better scores than the corresponding level 2 items. What is more, random elimination seriously interfered with the test results, especially in level 1 items because these showed a
different tendency from the overall results of the whole test. Hence, the answer to RQ2 was that the test performance differed with distractor characteristics, but the difference was more apparent between use of informed elimination and use of random elimination.

These outcomes imply some important issues in MC tests. First, the use of elimination strategies is almost inevitable. Although we can control elimination based on background knowledge to some extent, participants consistently use informed elimination as well as random elimination in items with any type of distractor. However, the use of informed elimination does not affect the overall test performance, which leads to the second implication, so we do not have to worry too much about it. Random elimination, on the other hand, is a crucial problem because it can change the test results drastically. Hence, what is important for test creators is to make items in which participants are unlikely to use random elimination. Nevertheless, random elimination was used much less than informed elimination; therefore, the results from this study counter the criticism that MC test scores do not reflect testees’ ability because of elimination or guessing. Although it was impossible to control many aspects of distractors such as concreteness and familiarity, the results will contribute toward improving MC testing conditions. If further research uses tests other than vocabulary tests, it will illuminate the features of elimination strategies more fully.

Acknowledgement

I thank Professor Yuji Ushiro for his valuable comments on this paper.

References


JACET (2003). *JACET list of 8000 basic words*. Tokyo: JACET.


**Appendix**

<Level 1 item with paradigmatically-related distractors>
She waved her ( ) in front of her and she said: My daughter wrote the best art book ever written.
a. ability     b. hand*        c. power       d. skill

< Level 1 item with syntagmatically-related distractors>
The bottles you buy today contain rain ( ) which fell up to 80 years ago.
a. forest      b. sky          c. window      d. water*

< Level 1 item with unrelated distractors>
She could ( ) his face quite clearly, as it was lit by the gas street lamp just a few feet behind him.
a. buy         b. develop     c. fall        d. see*