TEST-WISENESS ITEM ANALYSIS ON SMAN 1 KEDIRI’S SELECTION TESTS IN THE ENGLISH SUBJECT

Rica Naza Qimala
Email: qimala.za@gmail.com
State University of Malang

ABSTRACT: This study aims/aimed to analyze test-wiseness in three sets of SMAN 1 Kediri’s selection tests,. The main focuses of this study are to analyze types of test-wiseness appearance/which appeared in the selection tests, the frequency of their appearance, and also the item analysis of items with the test-wiseness. This study employs/employed a quantitative research design. The findings show that there are five types of test-wiseness found in the selection test, and the total frequency of their appearance is 6.3%. Item analysis of item with test-wiseness shows that the existence of test- wiseness in multiple choice items reduces the quality of the items.

Keywords: test-wiseness, selection test, item analysis

In the needs of learning English/In order to receive students with good ability in English, many favorite schools usually have selection tests to screen potential test-takers who are expected to be able to join the programs that the schools have. Those programs usually emphasize students who are using English actively in the classroom. Furthermore, those favorite schools usually have full supports by the regulation to administer the selection tests. One of the supports can be seen from Kediri’s Mayor Regulation in the year of 2012 which states that the selection test of the first grade in senior high schools which implement RSBI/, - RSBI/Acceleration/Inclusive, uses the following steps to screen the new students: an administration selection, an academic selection, a psychological test, an English ability test, and SKHUN scores.

In addition, test itself is basically a set of language tasks with measurement properties to know the specific skill with purposes to assess students’ achievement, selection, diagnosis, or combination of it (Hill, 1995). The selection
test is administered before students enter a university or a high school. This test is intended to select and screen potential students who have the criteria which are needed by the school related to the school’s programs. Furthermore, this test is also expected to know and predict future performance, in which the students will carry out when they enter certain education institutes. Therefore, there will be a minimum score as a standard whether the students pass or fail (Sulistyo, 2002).

As we know, common selection tests use multiple-choice tests because of their advantages. Based on Seaman’s (2003) statement, using multiple-choice tests makes it possible to test many topic areas than what people can do in an essay. Not only that, it is also efficient to be marked. In addition, because the assessment is able to be carried out objectively, the reliability of multiple-choice tests is assured (Sulistyo and Rahmajanti, 2003). Furthermore, in line with the needs to know the potential of the students, the questions should be answered based on the students’ knowledge and intelligence.

Nevertheless, some questions in the selection test might be easily answered by the students because the questions contain test-wiseness. Test-wiseness is a subject’s capacity to utilize characteristics and formats of the test and/or test-taking situation to receive a high score (Milman, Bishop, and Ebel as cited by Rogers and Bateson 1991:160). The questions which have test-wiseness on the multiple-choice items might be easier to be answered because the stem or the distractor of the multiple-choice items give the students cues to choose the right answer even though the students do not have any related knowledge about the materials being tested (Diamond and Evans, 1972). Then, the question will not be functional to know the students’ intelligence yet mislead to test the students’ strategy in test taking (Holton et al., 2008).

Since the function of a selection test is to collect the accurate data of real potential students, education institutes need a good instrument to get the accurate data needed. Therefore, the instrument is usually in the form of a set of tests which is really good, at least in terms of having high reliability and validity. This is because the high reliability of the test will give a consistent score of the test-takers, and the high validity will give the accurate score of the test-takers.
The test validity is the issue in which the researcher is interested to investigate. This is because the valid test should give the real score of a particular skill, in this case English ability in terms of reading comprehension. Yet, the test-wiseness which may exist in the test, will mislead the score of English ability into the test-taking strategy in the sense of cues accidentally given by the test. The cues themselves have a tendency to make the test-takers more easily guess the right answer without knowing the knowledge of the material being tested. Even though the test-wiseness is rarely considered, this factor actually has a significant influence on the test validity.

In addition, if test-takers possess both materials being tested and test-wiseness knowledge, their score can be greater than the score of test-takers who know little about test-wiseness. Then, one should be careful to interpret the meaning of test scores if the test score can be influenced by test-wiseness because it can result a potential validity problem (Rogers and Bateson, 1991). Therefore, the test should be clear from test-wiseness items to give the equal chance to all test-takers to get a realistic score which reflects their ability. The purpose of this study is/ was thus to analyze test-wiseness in three sets of SMAN 1 Kediri’s selection tests. Three main focuses of this study is/were to analyze types of test-wiseness appearance in the selection tests, the frequency of their appearance, and also the item analysis of items with the test-wiseness.

METHOD

This study is/ was conducted using quantitative research as research design. Furthermore, the frequency of test-wiseness’ appearance in three sets of English subject test of SMAN 1 Kediri’s selection tests and the item analysis are/were counted statistically.

The main focus on/of this study is/was to analyze selection tests in SMAN 1 Kediri. The selection tests were made by an independent institution, LPMP, for the sake of RSBI classes in the year of 2010, 2011, and 2012. Since the selection tests for RSBI classes were just held three times, the researcher used the entire selection tests. Table 1 below shows the total items in each year.
To analyse items in the selection tests, the writer re-administered the selection tests in three of 9th grade classes of SMPN 3 Kediri. There are 9 classes with 320 students as a total. The three classes were randomly chosen as representative samples of each year of the test administration. The selection test of year 2010 was administered in 9C, the year 2011 was administered in 9E, and the rest was administered in 9D. In addition, two classes consist of 33 students while the other class consists of 31 students.

After the data and the test-wiseness principles were ready, the researcher began to analyze the test-items. To analyze the test-items, the writer looked for cues in the test-items which were suitable with the test-wiseness principles.

The first category of test-wiseness is internal cues. In this category, the cues usually lie in the stem and/or the alternatives of the multiple-choice items. Six types of test-wiseness in this category are presented as follows. Grammatical Help means there is a grammar cue, in the form of specific determiner, verb agreement, or else, put incidentally or accidentally by the test developer so that the test-takers are able to guess the key to the question. Then, association means there is a similar word, phrase, sound stated in the stem which is also stated in the key. Moreover, this cue will lead the test-takers to the correct answer because of the similarity between the stem with one of the alternatives. Next, length has the correct answer longer and more specific than the distractors. Then, overlapping distractors means the distractors are almost similar with the correct answer. In this cue, synonym or antonym can be examples of overlapping distractors that exist in the test items. Then, absurd and specific alternatives cue implies the existence of the distractors which are obviously wrong and the correct answer which is obviously correct. This kind of cue makes the test-takers easily guess which one is the key. The last cue is spelling which represents spelling errors in the distractors
so that the test-takers are able to guess the key just by catching which spelling is correct and suitable.

Then, the second category of test-wiseness is external cues. Different from the internal cues, the external cues deal with how the test developer develops the multiple-choice items, such as Give-Away, Order of Answer, and Number of Alternatives, which are explained as follow. Give-away means the developer incidentally put the key of the question in the stem of other questions so that the test-takers easily lead to the key just by reading other questions. Besides, the order of answer sometimes makes the students get high scores. It will likely happen if the test developer mostly puts the key of the questions in the same places. Then, if the test-takers realize it, the test-takers will gain a high score easier. Last but not least, the number of alternatives is able to influence on how the test-takers gain a good score. It works with a condition of a stem only has three options of alternatives so that the test-takers have 33% chance to guess the correct answer in each alternative. Moreover, that will be better if the alternatives consist of four options instead of three. The intention is to reduce the chance to guess the correct answer from 33% become 25% to each alternative.

Then, after analyzing the whole sets of selection tests, it was continued to the next step that is counting the frequency of test-wiseness that is in the whole sets. In addition, there were four frequencies which were counted: the frequency of test- wiseness in year of 2010, year of 2011, year of 2012, and the frequency of test- wiseness which appeared in three sets of selection tests. As a secondary data, the item analysis was also checked using a computer program application named ITEMAN (Item and Test Analysis) version 3.00. Using the ITEMAN, the researcher reveals item analysis which consists of item validity, difficulty index, item discrimination, distractor effectiveness, and reliability of the selection tests.

**FINDINGS AND DISCUSSION**

**Findings**

Based on the theories which this study uses as a base, it shows that several types of test-wiseness cues exist in the selection tests. However, there are some types of test-wiseness which no longer exist in the multiple-choice items. As seen
in table 2 below, those types are spelling cues, give away cues, order of answer cues, and also number of alternatives cues. Here, eleven items with test-wiseness which were found are presented as follows.

<table>
<thead>
<tr>
<th>Table 2 Types of Test-Wiseness in the Selection Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Types</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>Internal Cues</td>
</tr>
<tr>
<td>Grammatical Help</td>
</tr>
<tr>
<td>Association</td>
</tr>
<tr>
<td>Length</td>
</tr>
<tr>
<td>Overlapping Distractors</td>
</tr>
<tr>
<td>Absurd-Specific Alternatives</td>
</tr>
<tr>
<td>Spelling</td>
</tr>
<tr>
<td>External Cues</td>
</tr>
<tr>
<td>Give-away</td>
</tr>
<tr>
<td>Order of answer</td>
</tr>
<tr>
<td>Number of Alternatives</td>
</tr>
</tbody>
</table>

Let us start from the item with grammatical help cue. Item 6 below is taken from the selection test in the year of 2010. As seen in below, the underlined word, finger, in the stem shows that there is a cue lies in this item.

6. Each finger puppet needs … finger to play.
   a. one
   b. two
   c. three
   d. four

The cue, which exists in Item 6, leads the test-takers to choose option a as the answer directly. This is because the singular form of finger in the stem should follow a singular number. Therefore, an option which has one as the answer should be the correct answer.

Then, an item with the association cue which exists in Item 5 of the selection test year of 2012 is also presented in the following passage.

5. Why do engineers think the body of tunas, the mackerels, and billfish are nearly ideal from the point of view?
   a. It is covered with a slick.
   b. Their body make it possible to move smoothly and slippery.
   c. The eyes lie flush with the body and do not protrude at all.
   d. The fins are stiff, smooth, and narrow, qualities that also help cut drag.

As seen in Item 5, there is a repetition in the underlined phrase in the stem, the body of tunas, the mackerels, and billfish, which transforms into the underlined phrase in the alternative b, their body. Moreover, this is a type the
association cue, where there is a repetition of a word or phrase both in the stem and in the correct answer. Therefore, it is possible for the test-takers to directly choose b as their answer although they do not know what the question is about.

Next, three items with length cue will be presented below. The first item being discussed is taken from Item 43 on the selection test year of 2011. The following item is the stem and alternatives of Item 43.

43. The rules are issued to …
   a. keep the mountain green and free from pollution.
   b. protect the visitors from wild animals.
   c. get rid of the dangerous animals.
   d. get money as much as possible.

The context of Item 43 above is about hiking rules of Pangandaran Mountain. As we can see, the phrase in option a is the longest one. In addition, it is also quite specific for the alternative a to answer the stem in Item 43. Therefore, there is the tendency that the test-takers will easily choose the alternative a as an answer. Besides Item 43, the length cue is also found in Item 99 of the selection test in the year of 2011.

99. What happened when people do the overfishing?
   a. A lot of fish will breed again.
   b. The fish population will increase.
   c. The sea will be full of fish population.
   d. The other creatures will lose their food and in danger of extinction.

Item 99 has the length cue because the key answer of this test item obviously lies in option d. Not only because the length of the option which is the longest one, but also the underlined sentence above, the other creatures will lose their food and in danger of extinction, is explicitly seen as the correct answer. Therefore, with this kind of alternatives, there is a big possibility for the test-takers to choose option d as the correct answer.

Next, the last item with length cue is found in the Item 9 of the selection test in the year of 2012. The stem and the alternatives are provided below.

9. From the letter we learn that …
   a. Lila has been to Srau beach many times before.
   b. Lila and Poppy spent their holidays together to Srau beach.
   c. Lila was so disappointed spending the holiday at the beach as it rained.
   d. Lila was so amazed at what she found at the beach as it was the first for her to visit the place.
The context in Item 9 above is about a recount text in the form of a letter from Lila to her best friend, Poppy, telling that she has a great holiday in Srau beach. As we read the alternatives, people commonly will notice that alternative d is the longest and the most specific one. Even if the test-takers do not catch the context of the letter, there is a tendency they will choose the alternative d as the answer.

Then, after presenting items with the grammatical help, the association, and the length cues, an item of the overlapping distractors cue is also presented below. The first item that follows is taken from Item 39 of the selection test in the year of 2011.

39. The conflict of the story is found in paragraph …
   a. one and two.
   b. two and three.
   c. two.
   d. one.

The conflict in the story mentioned in Item 39 lies in the paragraph two. However, there are three alternatives in this item which use paragraph two as the alternatives. This kind of alternatives not only will tend to be confusing even for the high intelligence test-takers, but also will be arguable, later on. Therefore, this item has overlapping distractors cue.

The next item with overlapping distractors cue is taken from Item 19 of the selection test year of 2012.

| Good morning, |
| I’m just dropping in to say, |
| “I hope you have a wonderful day!” |
| May your morning be sunny side up, |
| Your afternoon bright too. |
| May the whole day bring good news to you. |

19. Who is the text probably intended to?
   a. A boss.
   b. A teacher.
   c. A good friend.
   d. A business counterpart.

The overlapping distractors cue exists in the Item 19 because the distractors are too plausible so that the distractors are confusing even for the high intelligence test-takers. Even though this kind of test item does not lead the test-
takers into the correct answer, this is a part of test-wiseness cue since it is debatable because the alternatives seem to be all correct.

Other items with overlapping distractors cue are also found in Item 27 and Item 29 in the year of 2012’s selection test. Here, Item 27 is firstly presented.

27. From the caution above we are informed that …
   a. the floor is being cleaned.
   b. anyone should keep the floor clean.
   c. if one is not careful, he/she may slip and fall.
   d. people are supposed to walk like what is shown in the caution.

In this item, distractor b is overlapped with the key answer, which is option a. At a glance, the test-takers will choose option c, since it is the explicit answer of Item 27. However, if we examine more, the answer is actually a since we are being informed, not being warned. This kind of alternatives will appear to be confusing and too plausible, even for the high intelligence test-takers.

29. What does the first paragraph tell us about?
   a. How to plant Aloe Vera.
   b. The size of Aloe Vera plant.
   c. The shape of Aloe Vera leaves.
   d. What people can use Aloe Vera.

Item 29 above also has overlapping distractor cue since there are two alternatives which are likely to be the correct answers. Since those two alternatives are too plausible, the test-takers will find difficulty recognizing the correct answer.

Two last items with the test-wiseness cue that exists in the test items are in the Item 2 and Item 96 of the selection test in the year of 2011. The following passage is the explanation of items with the absurd-specific alternatives cue which are found in the selection tests.

**THANK YOU FOR NOT FEEDING THE ANIMAL**

2. The notice means …
   a. you may give some food to the animal.
   b. you’re not allowed to feed the animal.
   c. you’re allowed to feed the animal.
   d. you can feed the animal.
Based on the alternatives of Item 2, the option $b$, which has been underlined, has a big possibility to lead the test-takers into the correct answer since it is the only negative option, you’re not allowed to feed the animal. The other options have affirmative sentences which are quite absurd and it makes the option $b$ become a specific answer to be chosen.

Then, the last item that follows is an item with test-wiseness cue is Item 96 of selection test in the year of 2011.

96. What do the countries bordering the North Sea use it for?
   They use it as …
   a. a dustbin.
   b. a marine park.
   c. a place for surfing.
   d. a game for preservation.

The item above possesses absurd-specific alternatives since there is only option $a$ being talked in the text. Therefore, the test-takers have a tendency to directly choose option $a$ since others alternatives are not stated in the text at all. Moreover, there were no external cues, which consist of give away, order of answer, and number of alternatives, found in the selection test. Each item in the selection test did not give away answers to other items because different reading topics are presented in every four or five number. Moreover, the order of answer cue was also not found in the selection test because the correct answers are spread quite evenly in each option of alternatives.

In addition, since there are four alternatives in each item, the number of alternatives cue was not found as well. This is because with four options in alternatives, the chance of test-takers blindly guessing the correct answer is only 25% of each item. Therefore, the fair options have been presented in the selection test.

### Table 3 Frequency of Test-Wiseness in Each Year

<table>
<thead>
<tr>
<th>Year of the Selection Test</th>
<th>Test-Wiseness</th>
<th>Total Item</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>f</td>
<td>%</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2011</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>2012</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>Total*</td>
<td>11</td>
<td>6.3</td>
</tr>
</tbody>
</table>

*) Total of year 2010-2012
The types had been known, the frequency of test-wiseness in each year could be counted. As can be seen in Table 3, out of 25 items in year 2010, there was one test-wiseness cue found. The frequency of test wiseness which appears in selection test year of 2010 is 4%. Moreover, with 100 items in the selection test year of 2012, there were five test-wiseness cues found. That makes the frequency of the test-wiseness that appears equal to 5%. Furthermore, the selection test in year of 2012 has five test-wiseness items out of 50 multiple-choice items where the frequency of the test-wiseness is 10%.

As a total, there were 11 test-wiseness cues found in the selection tests in the year 2010 up to 2012. Moreover, with 175 multiple-choice items as a total, the frequency of test-wiseness appearing in the selection tests is 6.3%. Hence, the frequency of test-wiseness appearing is indeed varied. Yet it is clearly seen that there were test-wiseness that lies in each year of the selection tests.

<table>
<thead>
<tr>
<th>No</th>
<th>Item of Difficulty</th>
<th>Item of Discrimination</th>
<th>Item Validity</th>
<th>Effectiveness of Distracters (number of distracters that work)</th>
</tr>
</thead>
<tbody>
<tr>
<td>6-2010</td>
<td>Very easy</td>
<td>Very good</td>
<td>Sufficient</td>
<td>1</td>
</tr>
<tr>
<td>2-2011</td>
<td>Very easy</td>
<td>Very bad</td>
<td>Not valid</td>
<td>0</td>
</tr>
<tr>
<td>39-2011</td>
<td>Easy</td>
<td>Very bad</td>
<td>Not valid</td>
<td>1</td>
</tr>
<tr>
<td>43-2011</td>
<td>Easy</td>
<td>Very good</td>
<td>Sufficient</td>
<td>1</td>
</tr>
<tr>
<td>96-2011</td>
<td>Very easy</td>
<td>Very bad</td>
<td>Not valid</td>
<td>0</td>
</tr>
<tr>
<td>99-2011</td>
<td>Very easy</td>
<td>Very good</td>
<td>Sufficient</td>
<td>1</td>
</tr>
<tr>
<td>5-2012</td>
<td>Easy</td>
<td>Good</td>
<td>Low</td>
<td>1</td>
</tr>
<tr>
<td>9-2012</td>
<td>Easy</td>
<td>Very good</td>
<td>Sufficient</td>
<td>1</td>
</tr>
<tr>
<td>19-2012</td>
<td>Easy</td>
<td>Very bad</td>
<td>Not valid</td>
<td>1</td>
</tr>
<tr>
<td>27-2012</td>
<td>Very difficult</td>
<td>Very bad</td>
<td>Not valid</td>
<td>1</td>
</tr>
<tr>
<td>29-2012</td>
<td>Easy</td>
<td>Very good</td>
<td>High</td>
<td>1</td>
</tr>
</tbody>
</table>

Moreover, Table 4 above is the summary of item analysis on item with test-wiseness. Generally, the difficulty index of those items is in the very easy and easy categories, yet there is one item which is very difficult. Next, as we have seen in Table 4, out of eleven items, five items have very bad ability to discriminate between test-takers who master the competence being tested and who do not. On the other hand, there is one item which is good and five items which are very good on discriminating the competence of the test-takers.

Besides, there were five items which were not valid, there were two items in low and high level, yet the rest of the items had sufficient level of item validity. Then, out of eleven items, there were two items which did not have workingdistractors.
In addition, 9 items possessed only one working distractor. The distractors appeared to be not plausible so that the test-takers could get the key more easily. Therefore, the findings show that the item difficulty of items with test-wiseness is in the easy and very easy level. In terms of item discrimination, five items can describe very well, yet the others cannot describe the competence of test-takers.

**Discussion**

Based on the theories which this study as a base, it was shown that several types of test-wiseness cues exist in the selection tests. However, there are some types of test-wiseness which no longer exist in the multiple-choice items. Those types are *spelling cues, give away cues, order of answer cues*, and also *number of alternatives cues*.

Since the proportion of correct items which possesses test-wiseness is quite high, it seems there is a tendency that the junior high school students are able to recognize and take advantage of test-wiseness as well as students in the age of elementary school and senior high school do (Millman, 1965 as cited by Diamond & Evans, 1972 Diamonds & Evans, 1972). Hence, in line with Rogers & Bateson’s (1991) conclusion there is also a possibility that the good score which the test takers get is the result of both their ability in recognizing and taking advantages of test-wiseness which lies on the multiple-choice items.

Even though test-wiseness is rarely considered in creating multiple-choice items, it does exist in the selection tests. However, the types of test-wiseness which appear in the test-items seem to be reduced. This phenomenon is perhaps influenced by the carefulness of the test developer in developing test items. As a result, even though the term ‘test-wiseness’ is rarely heard, the test developer has avoided it indirectly. It may be a result of many books and online articles which discuss creating good multiple-choice items in which the distractors seem to be plausible for those who do not have competence, and also not overlapping between each distractor. Or else, the small portion of test-wiseness existence is perhaps caused by many revisions before the selection tests were administered. This is in line with the weaknesses of multiple-choice items which is aforementioned in this study that creating multiple-choice items consumes much
time since it needs revision over revision to make it better (Salkind, 2013; Anonym, 2013).

Based on item analysis, in terms of difficulty index, it is well said that multiple-choice items with test-wiseness are usually in the category of easy or very easy. It is perhaps caused by the cues that lie in the test-items which make most of the test takers able to answer the items and improve a test score no matter what the content area of a test is (Benson, 1988; Sarnacki, 1979 as cited by Rogers & Bateson, 1991:160). Besides, some items with test-wiseness have ineffective distractors which are not able to discriminate the test-takers well. Hence, the items with test wiseness are usually in the category of easy or very easy, and have ineffective distractors, cannot discriminate the competence of test-takers.

Nevertheless, even though some items which possess test-wiseness have problems in terms of difficulty index, discrimination index, and ineffective distractors, the items analysis shows that the reliability of the selection tests are still in the high level. Overall, it means that the test-wiseness which may affect the condition of the test-takers do not disturb the consistency of the test-takers’ scores (Bachman & Palmer, 1996 as cited by Anonym, 2010). Moreover, the items with test-wiseness are also shown to be not valid. Although Rogers and Bateson’s statement (1991) demonstrated test-wiseness can result a potential validity problem, the result of this study shows that only 45% of the items with test-wiseness have the validity problem (not valid items).

Moreover, based on an item with test-wiseness found in the year of 2010, it is seen that almost all of the test-takers are able to recognize the correct answer. Yet, this item has an ability to discriminate the competence well even though there is only one working distractor in this item. In addition, the item validity, in the year of 2010, shows that the level of validity is in sufficient level. Next, there is a slight increasing frequency on the year of 2011. Out of five items with test-wiseness found in this year, there are three data which have minus value of item discrimination and item validity. Two of those data have high item difficulty; in which all of the test-takers can recognize the correct answer, yet it makes the items unable discriminate the competence because it appears to be too
easy. Another invalid item is likely to happen because the key answer was only chosen by a small proportion of the test takers. On the other hand, two other items which have sufficient level of item validity tend to be able to discriminate the competence of the test-takers.

Then, in the year of 2012, there is a significance increasing frequency of test-wiseness appearance since the frequency rises from 5% to 10%. Although there is an increase, out of five items with test-wiseness, only two items which are not valid. Since those items are not valid, they do not have ability to discriminate the competence of the test-takers. The cause is because one of the items is too difficult and no test-takers are able to recognize the correct answer. Another invalid item happens because the distractors work better than the key answer. On the other hand, three other items are valid, even though the levels of validity are varied from low level to high level. Thus, those items have the ability to discriminate the competence of test-takers.

Even though the frequency of test-wiseness in year of 2012 increased, it has less invalid items than in the year of 2011. Hence, there is the negative relationship between test-wiseness and item difficulty. When the test-wiseness’ frequency is high, the quality of the items are low (very easy or very difficult). The low quality of item difficulty also affects the quality on how an item discriminates the competence of the test-takers. Besides, when the frequency of test-wiseness is high, the item validity will be very low (not valid). Thus, the appearance of test-wiseness is bad for the quality of test items.

CONCLUSION AND SUGGESTION

In this study of analyzing three sets of SMAN 1 Kediri’s selection tests in year of 2010, 2011, and 2012 which were developed by LPMP (Lembaga Penjaminan Mutu Pendidikan) East Java, several conclusions are made. First, out of nine types of test-wiseness items, there are only five types found in the selection test. Those are grammatical help cue, association cue, length cues, overlapping distractors cues, and also absurd-specific alternatives cue. Based on the frequency, it can be concluded that test-wiseness appears only in the small portion (6.3%) of the three sets of selection test. Next, in terms of item analysis, it
shows that there is negative relationship between test-wiseness and item difficulty. Moreover, the appearance of test-wiseness is bad for the quality of test items.

After doing the study in analyzing test-wiseness in terms of its proportion in the selection tests along with types of test-wiseness appearance, and the item validity, the researcher suggests that the test developer as well as English teacher should be more careful in writing and arranging multiple-choice items. Moreover, for the committee of the selection tests, even though the selection test is made by an independent institution, it is better to have some representatives to recheck and examine the multiple-choice items. In addition, the researcher also has a suggestion to the head master of the school to make this study a consideration before hiring other institutions to develop the next selection test for the school. Furthermore, since this study still has limitations, the researcher truly hopes that other researchers are willing to improve the limitation of this study. In this case, getting as many as sources about the latest news of test-wiseness is one of the ways to improve this study. In addition, analyzing multiple-choice items in the wider area or population is also able to enrich the perfectness of this study.

REFERENCES


