THE ANALYSIS OF LISTENING TEST ITEMS ON NATIONAL EXAMINATION TRY-OUT FOR 12th GRADERS OF SMK 3 SEMARANG

Ninuk Krismanti
STKIP PGRI BANJARMASIN
ninukkrismanti@gmail.com

Abstract

National examination burdens students as test-takers. Thus, before facing the real exam, it is common for students to have try-out examination administered by either their schools or the Department of Education of each region in Indonesia. As English is one compulsory subject tested in the national examination, it is also tested in the try-out exam. Listening section always becomes part of the test along with other skills of English. This paper is intended to analyze listening test items on the try-out test for the 12th graders of SMK 3 Semarang by analyzing the Item Facility (IF) and Item Discrimination (ID). The purpose of the analysis is to reveal whether or not the listening test items have met the standard of a good test to be considered valid. From this study, it is found that of 14 test items being investigated, half of them belong to easy level category, and none of them belongs to the difficult level. It is also revealed that regarding discrimination power, one of the 14 items tested has no discrimination power, and only one item has powerful discrimination. The writer proposes several alternative changes to improve the test items so that the listening test items in the try-out examination have similar level to the ones in the national examination

Key words: Item Facility, Item Discrimination, Listening, Try Out, National Examination.

INTRODUCTION

Ideally, national examination is conducted as an indicator of the success of national education in Indonesia. The conducting of national examination which requires certain passing grade standard has given fright and burden to students as test-takers. Thus, intensive preparations are commonly taken before the national examination. One of the preparations taken is by giving try-out to the test-takers. The try-out can be managed either by each school independently or by the Department of Education of each province regionally. The test given in try-out is made as similar as possible to real test given in the national examination. Thus, the validity of the try-out test needs to be assessed in order to give an accurate simulation for the national examination taken by the students later on.

As one of compulsory subjects tested in the national examination, English test is also given in the try-out. Similar to the real test, the try-out of English is also divided into four parts to assess students' ability in listening, speaking, reading and writing. This paper is intended to assess the listening section of an English test try-out for vocational students in Semarang by using Item Facility (IF) and Item Discrimination (ID) analysis.

REVIEW OF RELATED THEORIES

Validity, according to McCowan and McCowan (1999:3), means the extent to which a test measures what it is supposed to measure. One way to test the validity of the test is by assessing the test items. Test items an

Brown (2004:58-61) divided three item indices that can be use to accept, discard, or revise items: item facility (IF), item discrimination (ID), and distracters efficiency. However, since the writer does not have sufficient data, the test cannot be analyzed by using distracters efficiency category. Thus, this paper only focuses on IF and ID analysis. The explanation of the use of IF and ID for analyzing particular items are described as the following:

1. Item facility (IF)

   Item facility (IF) is defined by Brown (2004:58) as the extent to which an item is easy or difficult for the targeted group of test-takers. Henning (1987:10) states, "A test should never be selected or developed without due consideration of the abilities and other characteristics of the intended examinees. Violation of this principle is most commonly observed in the setting of examinations that are too difficult or too easy for the examinees." From the statement, it can be inferred that a test will be considered violating the criteria of a good test if there are too many easy or difficult questions. Such test will make the identification of low-ability and high-ability students difficult to conduct. Therefore, assessing IF of the test is important to test the validity of the items given so that the test given to the test-takers are in a balanced proportion.

Formula to calculate item facility (IF):

$$IF = \frac{\text{\# of Ss answering the item correctly}}{\text{Total \# of Ss responding to that item}}$$

Descriptions of the IF result:

Range (%)	Level of Difficulty
0 – 14.99	Very difficult
15 – 85	Medium
85.01 – 100	Very easy

Source: Brown (2004: 59)

2. Item discrimination (ID)

   Brown (2004:59) defines item discrimination (ID) as the extent to which an item differentiates between high and low ability test-takers. An item on which high ability students and low ability students have equal score indicates poor ID because it cannot discriminate the two groups. On the other hand, an item that is answered correctly by most of high–ability students and incorrectly by most of the low-ability ones has good power of discrimination.

Formula to calculate ID:

$$ID = \frac{\text{High group \# correct} - \text{low group \# correct}}{\frac{1}{2} \times \text{Total of your 2 comparison groups}}$$

Descriptions of the ID result:

Range	Level of discrimination
≥ 0	No discrimination
0.01-0.99	Moderate discrimination
≤ 1	High discrimination

Source: Brown (2004:59-60)

METHOD

This study is carried out under a mixed methods approach. According to Creswell (2003:20), the data of a mixed methods approach can be both in forms of numerical and textual data. Data in this paper are taken from listening section of English national examination try-out in SMK 4 Semarang. Test-takers of the try-out are 77 students of 12th grade of the school who are scheduled to follow national examination at the same year as the try-out is given. The students are divided into two groups: those who got package A (39 students) and those who got package B (38 students). There are 14 listening items for each package. This paper is only aimed to analyze 14 items in package A.

The data are analyzed using item facility (IF) and item discrimination (ID) calculations to determine whether or not corrections of the items are needed. The result of IF and ID calculations are used along with test-making guidelines prepared by the Ministry of National Education to revise the try-out.

FINDINGS AND DISCUSSION

Item number	Correct Answer	%	Item Facility (IF)	Item Discrimination (ID)
1	30	77	.77	.74
2	38	97	.97	.32
3	12	31	.31	-.11
4	37	95	.95	.42
5	36	92	.92	.63
6	20	51	.51	.32
7	39	100	1	.32
8	24	62	.62	1.05
9	33	85	.85	.63
10	37	95	.95	.42
11	31	79	.79	.74
12	35	90	.90	.63
13	35	90	.90	.53
14	26	67	.67	.63

Table 1 IF and ID distributions per question item

A. Item Facility (IF) Analysis

Range (%)	Level of Difficulty	Number of items
0 – 14.99	Very difficult	
15 – 85	Medium	1, 3, 6, 8, 9,11, 14
85.01 – 100	Very easy	2, 4, 5, 7, 10, 12, 13

Table 2 Distribution of listening test difficulty

Based on the table above, it can be seen that there are no questions with difficult level occur in the test. On the other hand, there are too many questions are in easy level category. This unbalanced proportion does not fit the criterion of a good test which should have all three levels of difficulties (from easy to difficult) or balanced distribution. Thus, some questions from easy-level difficulty category should be changed into difficult-level questions.

In examples below we change item no. 4, 5, and 7 which are in easy-level category into difficult-level category:

Item no. 4

Test-making guidelines:

> *Menentukan respon yang tepat terhadap pernyataan/pertanyaan lisan yang berisi pemberian saran/ pendapat.*
> Determining appropriate responses for oral statements/questions on giving advices/suggestions.

Original version:

(4)   M   : I can't do this assignment. It's hard to understand the sentences.

> A. You had better take a rest.
> B. Why don't you consult your dictionary?
> C. If I were you, I would take it.

Explanation of problem and solvency:

The distracters A and C are not closely related to the key assignment – understanding the sentences. Good distracters should either have similar sound or have close lexical meaning to the key. Thus, for question number (4), the distracters should be changed into the following:

> A. You had better take a look at your book
> C. If I were you, I would consult my note

Item no. 5

Test-making guidelines:

> *Menentukan respons yang tepat terhadap pernyataan / pertanyaan lisan yang mengungkapkan sebuah undangan atau penawaran.*
> Determining appropriate responses for oral statements/ questions on invitation or offers.

Original version:

(5)   W   : Will you come to my party next week?

        A. I'd love to, but I have something to do.
        B. Yes, my party will be held next week.
        C. No, there is no party in the week end.

Explanation of problem and solvency:

Distracters B and C are very different from the target answer so that they will not mislead the students. To fix this problem, the distracters can be changed like the following:

        B. Sorry, I think I'll come with my girlfriend.
        C. Glad to hear that.

Item no. 7

Test-making guidelines:

> *Menentukan gambaran umum/informasi tertentu/informasi tersirat dari percakapan lisan singkat tentang kegiatan sehari/hari/kegiatan yang sedang berlangsung.*
> Determining general understanding/ particular information/ implied information from short dialogues on daily activities/ the activities in progress.

Original version:

(7)   Script:
        W   : Hi John, what are you doing here?
        M   : I am reading a magazine.
        W   : What kind a magazine do you read?
        M   : I usually read a music magazine in my spare time.

        What is John doing?

        A. The boy is listening to music.
        B. **He is reading music magazine.**
        C. The boy is reading a newspaper.
        D. He is spending his spare time.

Explanation of problem and solvency:

There are two problems in this item. The first is the stem which happens to be too easy. The question 'What is John doing?' is exactly the same to the first question in the conversation. The second problem is the distracters. The use of the word 'the boy' is not suitable. Students might not realize that 'the boy' in the sentences is used to replace John. Thus, they will easily think A and C as wrong answer. There are two possibilities of how the question should be given: by changing the stem or changing the distracters.
First alternative:
        Change the stem into: What can we infer from John?
Second alternative:

Change the distracters into:
- A. John is reading a magazine about cooking.
- C. John is reading a music article on newspaper.
- D. He is spending his spare time.

Based on the correction made above, a new distribution of 14 items of listening session of 12th grade vocational high school national examination try-out is as the following:

Range (%)	Level of Difficulty	Number of items
0 – 14.99	Very difficult	4, 5, 7
15 – 85	Medium	1, 3, 6, 8, 9,11, 14
85.01 – 100	Very easy	2, 10, 12, 13

With the distribution above, it is expected that the proportion of the test based on different level of difficulties becomes more balanced.

B.   Item Discrimination (ID) Analysis

Range	Level of discrimination	Number of items
≥ 0	No discrimination	3
0.01-0.99	Moderate discrimination	1,2,4,5,6,7,9,10,11,12,13,14
≤ 1	High discrimination	8

Table 3   Distribution of items discrimination

Based on the table above, it can be seen that item number 3 does not differentiate high and low students. Thus, correction should be made on this item as the following:

Item no 3.

Test Making guidelines:

> *Menentukan pernyataan lisan yang tepat tentang penampilan (ciri-ciri fisik) seseorang/deskripsi benda sesuai dengan gambar.*
> Determining appropriate oral statements on physical appearance (physical characteristics) of someone/ descriptions of objects based on the pictures.

Original version:

- A. He's listening to the speaker.
- B. He is wearing a tie.
- C. He's touching his arms.
- D. He's answering a question.

Problem explanation and solvency:

There is no discrimination in this test item, which means the item is too easy for both the high and low ability students. The problem is the distracters are out of the guideline. All of

the distracters are stating activities while they are supposed to state about physical appearance. Thus, the original distracters need to be changed into following:

A. He has long and curly hair
C. He has short and wavy hair
D. He is wearing a T-Shirt

In terms of ID, other items, but item number 3, do not need any correction because they have shown moderate and powerful discrimination as required in a good test.

CONCLUSION

A good test item should fit certain range of IF and ID. If the range cannot be reached by the item, then improvement for the item is needed. For 14 listening items of the English test try-out that become the subject of assessment in this paper, it is found that (1) 7 of 14 items belong to easy level category while none of the items belong to difficult level category, and (2) 1 of 14 items has no discrimination power while only 1 item has powerful discrimination.

Some alternatives are proposed to improve the proportions of test difficulty distribution. By using IF result as the consideration and the test-making guidelines, the writer changes three of seven items on easy level category to difficult level category. Other than making the changes by using IF, the writer also proposes to change item number 3 based on the ID result. The improved test, however, still need to be tested again in order to assess its effectiveness.

REFERENCES

Brown, H. Douglas. 2004. *Language Assessment: Principles and Classroom Practices*. Pearson Education: NY.

Creswell, John W. 2003. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (Second Edition). Sage Publications, Inc: California.

Henning, Grant. 1987. *A Guide to Language Testing: Development, Evaluation, and Research*. Heinle & Heinle Publishers: China.

McCowan, Richard J. and Sheila C. McCowan. 1999. *Item Analysis for Criterion – Referenced Tests*. CDHS: NY.

http://un.kemdikbud.go.id/files/KISI-KISI-UJIAN-NASIONAL-2016-SMK.pdf