Englisia NOVEMBER 2017
Vol. 5, No. 1, 17-28

CORRELATION BETWEEN ABILITY TO RECOGNIZE SENTENCE ERRORS AND ABILITY TO PRODUCE GRAMMATICALLY CORRECT UTTERANCES

Masrizal
University of Southampton, England
masrizal@unsyiah.ac.id

ABSTRACT

This article summarizes and reports an empirical study investigating students' ability in recognizing grammatical errors and producing grammatically correct sentences. 38 university students were involved in a set of grammar tasks which were specifically created to measure their ability to both identify errors and avoid them in language productions. The main purpose of the study is to prove whether their ability to pinpoint errors within sentences resembles their ability in producing grammatically correct sentences using the same features. The study also measures the appropriateness of the test items in order to see how it affects students' performance. Final test data collected from the students in two different groups reveal that their ability to recognize sentence errors has positive correlation to their ability to produce correct sentences. The correlation figure among the more proficient students (group 2) is relatively larger, indicating that the amount of knowledge on relevant features positively influences, to a certain extent, the quality of language production and responses.

Keywords: *error recognition; sentence production*

INTRODUCTION

The ability to spot grammatical errors in sentences is a very important skill required from L2 students. In an academic setting, this ability shows learner's proficiency of a particular language, both in passive (receptive) and active (productive) skills (Read, 2015). In a passive context, an L2 learner is required to be able to recognize errors and decide whether a sentence, or parts of it, has fulfilled

necessary grammatical requirements. On the other hand, this skill is necessary when the learner is required to supply a part of a phrase or sentence into either written or oral production. It is essentially required to assure that the produced utterances comply with basic language requirements.

This particular study has particularly looked at students' ability in recognizing English language errors and supplying correct parts into sentences. Thirty eight undergraduate university students have been involved in a set of grammar tests which took place in two different classes. The main purpose of this study was to look at whether participants' ability in recognizing sentence errors correlates positively with their ability in producing correct sentence structures. In addition, it would finally try to evaluate the appropriateness of the test items by using two different measures, *difficulty index* and *discrimination index*.

Test Specifications

Table 1:Test Specification	
Purpose of the instrument	This test was designed to assess test-takers' ability in recognizing English sentence errors and supplying correct parts in the similar context. It would also predict whether their ability to recognize sentence errors resembles their ability to produce correct form in the same context.
Construct or domain that will be measured	English grammar knowledge was assessed in this test.
Length of the test	Thirty minutes.
Context in which the instrument is to be used	This instrument was used in an English medium education. In this case, it is used to assess university students in the department of English Education.
Characteristic of intended participants	Participants are university students from the Faculty of Education and Teacher Training, majoring English Education. Two groups of participants took part in the test, one of which being in the third semester while the

	other is in the fifth semester.
Conditions and procedure of administering the instrument	The test was administered in two sample classes by an assigned lecturer. Test sheets were manually distributed and participants had been required to complete the test within the allocated time.
Procedures of scoring	For the multiple choice questions in part A, each correct answer was given one point. Incorrect and unanswered questions were marked '0'. For part B, the marking procedure is still the same. However, spelling was checked before deciding whether an answer was correct or wrong. If the word was misspelled, but lead to a correct answer, it would be regarded as correct.
Intended level of difficulty	This test is designed for intermediate to lower advanced level of English grammar ability.
Reporting of the results	Correlation between skills in each part, item difficulty, and discrimination index.

DISCUSSION

How construct validity is ensured and checked

In order to establish construct validity for this test, every endeavour has been done to prevent the presence of two main threats to construct validity identified by Messick (1989), construct under-representation and construct-irrelevant variance. As further discussed by Zheng and De Jong (2011), a number of ways could be alternative solutions in order to prevent the presence of the threats. To avoid construct-under representation, the tasks had been ensured to have sufficient coverage of target language situations, especially in regards to situational and interactional authenticity (Bachman & Palmer, 1996). However, since this only assesses grammar knowledge, the efforts have been done to assure that all the questions given are relevant to the test takers background knowledge.

In regards to construct-irrelevant variance, it has been ensured that no test-takers were advantaged or disadvantaged by the test as a result of their personal background. Everyone speaks the same first language and is learning English as a foreign language. Everyone shares the same topical knowledge and, therefore, the probability of providing correct answers are purely dependent on their own personal knowledge regardless of any non-academic background everyone shares (Kuncel & Sackett, 2014).

Impact of consequences of the test on stakeholders

This test is expected to give an overview about students' strengths and weaknesses in analysing English sentence errors. Often, non-native English speakers tend to have better ability in recognising errors from pre-produced sentences, while at the same time they struggle to produce such utterances on their own. From this mini test, which obviously covers limited features of English grammar, it is expected that their weaknesses can be revealed so that further adjustments can be made in regards to teaching materials, classroom test design, and lesson coverage.

Design of assessment tasks and scoring system

As previously mentioned, the test instrument consists of two separate but related parts. Part A and B assess analytical and productive skill respectively. Further details and how the questions in both parts are connected to each other will be given in the following table.

Part A The following sentence **parts** have been supplied incorrectly. Participants have to **identify** which one is incorrect in each particular sentence.	Questions	Part B The basic form of the following sentence **parts** have been provided. Participant need to **insert/supply** them into the gap by using correct forms.
S-V agreement (incorrect verb form)	Q1	**S-V agreement.** (supply correct verb form)
S-V agreement (incorrect copula verb)	Q2	**S-V agreement.** (supply correct copula)
S-V agreement (incorrect copula verb)	Q3	**S-V agreement.** (supply correct copula)
S-V agreement(incorrect passive verb form)	Q4	**S-V agreement.** (supply auxiliary in passive form)
S-V agreement (incorrect verb form)	Q5	**S-V agreement.** (supply correct verb form)

Inverted subject and verb(incorrect copula)	Q6	**Inverted subject and verb**(supply correct copula)
S-V agreementwith *either … or* (incorrect copula verb)	Q7	**S-V agreement**with *either … or* (supply correct copula verb)
Parallelism in object/complement (*to+inf*form is not parallel)	Q8	**Parallelism in object/complement** (supply parallel *-ing*form)
Aux+V(*modal + inf,* incorrect infinitive)	Q9	**Aux+V**(*modal + inf,* supply correct infinitive)
Aux+V(*aux had + past participle,* incorrect pp)	Q10	**Aux+V**(*aux has + past participle,* supply correct 'be' form)
Correlative conjunction*not only…but* (incorrect pair)	Q11	**Correlative conjunction***not only…but* (supply correct pair)
Word form	Q12	**Word form**
Plural & singularnoun using *amount vs. number*(incorrect reference)	Q13	**Plural & singularnoun** using *amount vs. number*(supply correct reference)
Pronoun (incorrect pronoun)	Q14	**Pronoun** (supply correct pronoun)

Administration of assessment tasks

The test took place in two grammar classes at Syiah Kuala University, Indonesia. These classes were chosen due to the availability of access to the targeted participants. Special authorisation had initially been granted and the class lecturer, who happens to be a colleague of mine, had initially expressed her willingness to distribute the test materials as well as to administer the test herself. Prior to the test date, a research assistant has been hired to prepare the test materials and handed them to the lecturer.

Thirty eight students coming from grammar 1 and grammar 3 classes (further labelled as group 1 and 2 respectively) participated in the test. Technically, the students from grammar 3 class are considered to be more proficient in English grammar, while those from the other group are mainly starters or at pre-intermediate level at the most possible. Therefore, I expected to see better results produced by the students from grammar 3 group due to having higher proficiency.

As mentioned elsewhere in this paper, this test consists of two different parts. The first part contain multiple choice items about recognising sentence errors, while the other is a kind of filling the gap questions in which test takers are required to productively supply correct grammatical forms into the gap. Each item consists of three answer choices, except in number 11 to 14 of part B. There are still debates over whether a multiple choice test item should contain fewer or more options (Lee & Winke, 2013). In this test, the options are kept to minimum so that test-takers' needs for *testwiseness* to succeed can be minimized (Rogers & Harley, 1999).The instruction for the test has been provided as clear as possible in order to assure that the test was completed within the time provided. It is expected that their level of proficiency in these two tasks can be distinguished after completing the test.

Scoring of performances and analysis of results

The scoring for this test has been done as simple as possible. Each correct answer is worth one point, while the incorrect or unanswered items are not given any score. The answers, along with the score obtained by all participants were then calculated and summarized in relevant tables to be further analyzed appropriately.

Evaluation of participants' performance

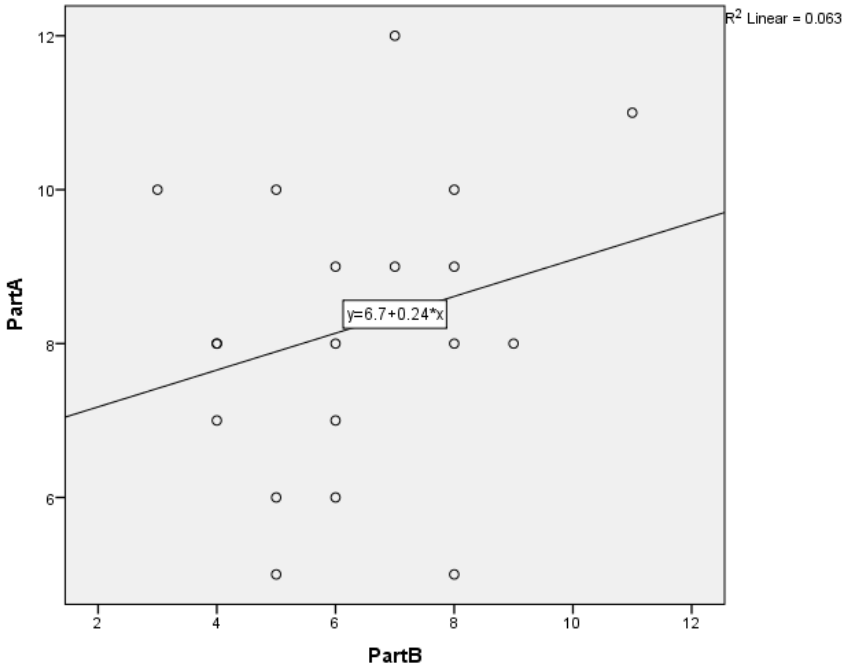
Group 1

After running Pearson's correlation test on SPSS, it is clearly seen that the Pearson's r coefficient for this particular group is 0.251. This means that there is a positive, but small, correlation between the questions in part A and B. However, at Sig. (2-tailed) value of 0.299, which is greater than 0.05, we can determine that there is no statistically significant correlation between the two variables. These lead us to conclude that a better score in one part of the test, i.e. Part A, might probably have a little contribution to the increase in the other, i.e. Part B, or vice versa.

Correlations

		PartA	PartB
PartA	Pearson Correlation	1	.251
	Sig. (2-tailed)		.299
	N	19	19
PartB	Pearson Correlation	.251	1
	Sig. (2-tailed)	.299	
	N	19	19

In addition, the distribution of test results by each participant can be seen in the following scatterplot chart.



Considering this above correlation value, there is still possibility that the results of the test are not fully representative to the actual students' proficiency. With this type of results, the chances that some answers come from guessing are big, especially if we look at particular results of individual questions. Average correct answer achieved by the whole group is 8.2 out of 14 in part A, while in part B there are only 6.3. This simply shows that part A seems to be easier for them considering that they do not have to produce their own form of answer. In question number 10, for example, Pearson's correlation coefficient -0.368 at 0.121 level of significance proofs that this particular question in one part of the test correlate negatively with its relevant pair in the other. Therefore, we cannot confidently confirm whether a student who is good in one part of the test would perform equally in the other.

Group 2

Within this particular group of students, the test seems to reveal slightly different but higher figures. Based on the Pearson correlation coefficient of 0.548, we can see that there is a bigger positive correlation between part A and part B scores of the test. At 0.015 significance level (2-tailed), which is obviously lower than

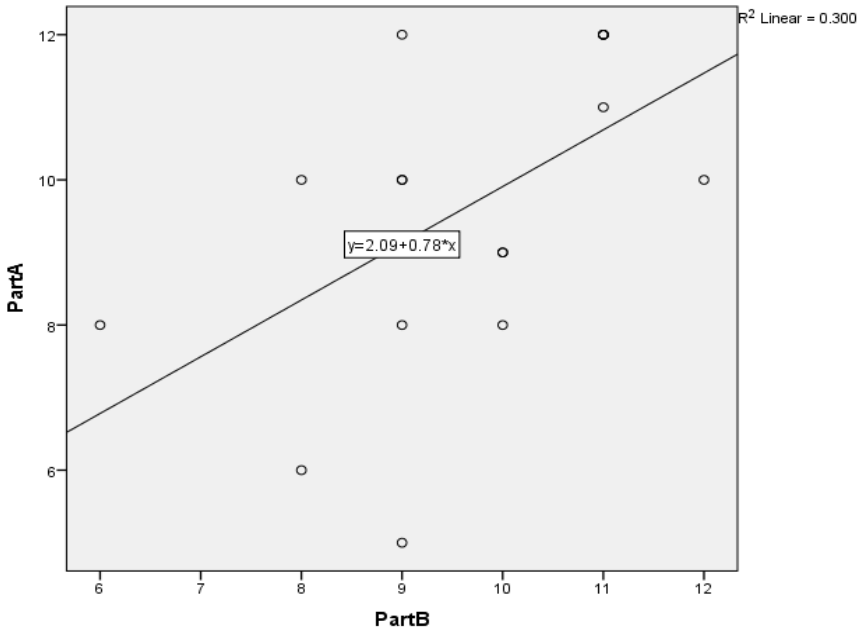
0.05, the correlation between the two parts of the test is significant. Therefore, we can conclude that, among these participants, a good achievement in one part of the test, i.e. Part A, can somehow predict a better accomplishment in the other, i.e. Part B, or vice versa.

Correlations

		PartA	PartB
PartA	Pearson Correlation	1	.548*
	Sig. (2-tailed)		.015
	N	19	19
PartB	Pearson Correlation	.548*	1
	Sig. (2-tailed)	.015	
	N	19	19

*. Correlation is significant at the 0.05 level (2-tailed).

A clear overview on how scores are positively distributed can be seen in the following scatterplot graph.



From the graph, it is confirmed that students' achievement in part A of the test reflects their score in the other part. The group's mean score of correct answer also confirm this, which is 9.5 correct answers for both part A and B. In a simple definition, the number of correct answer they score in one part of the test is not very different from the one in the other part, except for a few students. This is sufficient to tell us that most of the participants in this group have a better proficiency level in

recognizing sentence errors and, at the same time, producing grammatically correct English sentences on their own.

Evaluation of assessment instrument

In order to determine whether the test items are appropriate or not, a set of item analysis is required. This section of the paper will discuss two different measures of item analysis called *difficulty index* and *discrimination index*.The following table will provide an overview and numerical figures regarding these, while further details will be discussed in the subsequent sections.

Item Analysis using Difficulty Index

This measure is used to determine the level of difficulty of the test items. To do this, the proportion of student who answered the test item need to be calculated accurately. This will give information whether a test item is relatively easy or difficult, and if it needs replacing or not. This is done by simply dividing the number of students who choose the correct answer by the total number of students. This formula will reveal the level of difficulty of each item, also known as *p-value*. A general 'rule of thumb' is that an item is relatively easy if the difficulty is more than 0.75, whereas it is more difficult if the difficulty is below 0.25 (FCIT, 2016). Therefore, for example, if an item is answered correctly by 85% of test takers, it would have an item difficulty, or p-value, of 0.85 (Matlock-Hetzel, 1997).

Based on the figures in the item analysis table, it is clearly seen that the difficulty index of each question from this test varies from 0.05 to 1.00. Students in group 1 seem to struggle more with a number of questions in part B and with very few in part A. Moreover, students in group 2 look more proficient with only one issue in each part of the test. It is also clear that the average difficulty index value is different between the two groups, with group 1 having lower score. According to the p-value, question 11 in part B seems to be the easiest question to both groups.

Question 13 of part A appears to be the most problematic one for both groups. Since index of difficulty measure of both groups are below 0.25, this is considered to be a difficult item to all students of different proficiency level. This suggests that this item needs to be reviewed and replaced if necessary. Another

important point is that students in group 2, which is supposedly a higher performing group, seem to struggle the most in question 1 in part B. Surprisingly, students in group 1 have recorded a completely opposing result, confirming that they seem to perform better in this particular question item. Overall, only a small number of questions are either too easy or too difficult.

Item Analysis using Discrimination Index

This measure is used to know how well an assessment differentiates between high and low scorers. In this regard, we would like to know how often the high-performing test takers would select right answers for each question in comparison to the low-performing ones. If an assessment has a positive discrimination index (which is between 0 and 1), high score participants are expected to choose correct answer for specific questions more often than those with lower total score. On the other hand, an assessment is considered having a negative discrimination index (between -1 and 0) if this happens otherwise (FCIT, 2016). Discrimination index can be determined by subtracting the number of students in lower group who got correct answer from the ones in the upper group, then divided the number of half of the total samples.

According to the measurement results, question 2 in part A has a negative discrimination index in both groups. This means that low performing students are more likely to get this item correct. Considering this, the items need to be carefully analyzed and probably deleted or changed. Apart from this, a number of other questions with negative discrimination index need to be further reviewed, suggesting the replacement of the items or simply re-writing them.

Furthermore, it is interesting to see that four question items given to participants in group 1 have recorded a 0.00 discrimination index. For group 2, there are 3 such questions. This simply means that both the high-performing and low-performing students in each group did not find these questions as being too easy, which indicates that the items are doing a great job to challenge the test-takers. Overall, the fact that most of the items have positive discrimination index has led us to assume that most of these questions are appropriate enough to the students, with a small number of them need to be either revised or discarded.

CONCLUSION

Based the result of the study, a number of conclusions can be reported as in the following.

1. Correlation between assessment parts

The results of Pearson's correlation test proves there is a correlation between students' ability in recognising sentence errors and supplying correct parts of sentences. Based on this result, participants in group 2 produce a higher correlation score, which helps us assume that their ability to recognize errors perfectly matches with their ability to produce correct forms of sentence parts.

2. Item Analysis results

In terms of the test instrument, item analysis through item difficulty and item discrimination index has proven that most of the test items are appropriate. However, some questions, with either too high or too low difficulty index, will need to be reviewed, revised, or even discarded. Likewise, items with negative discrimination index will also need to be treated as such.

REFERENCES

Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*: OUP Oxford.

FCIT. (2016). Classroom Assessment. Retrieved 10 January 2016, from http://fcit.usf.edu/assessment/selected/responsec.html

Kuncel, N. R., & Sackett, P. R. (2014). Resolving the assessment center construct validity problem (as we know it). *Journal of Applied Psychology, 99*(1), 38.

Lee, H., & Winke, P. (2013). The differences among three-, four-, and five-option-item formats in the context of a high-stakes English-language listening test. *Language Testing, 30(1)*, 99-123.

Matlock-Hetzel, S. (1997). Basic Concept in Item and Test Analysis. Retrieved 10 January, 2016, from http://ericae.net/ft/tamu/Espy.htm

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational researcher, 18(2)*, 5-11.

Read, J. (2015). *Assessing English Proficiency for University Study*: Palgrave Macmillan.

Rogers, W. T., & Harley, D. (1999). An empirical comparison of three-and four-choice items and tests: susceptibility to testwiseness and internal consistency reliability. *Educational and Psychological Measurement, 59(2)*, 234-247.

Zheng, Y., & De Jong, J. (2011). *Research note: Establishing construct and concurrent validity of Pearson* Test of English Academic: Pearson Academic Ltd.