



Writing English Language Test Items as a Learning Device: A Principle of Habit Formation Rules



I Nengah Astawa ^a, Nengah Dwi Handayani ^b, Ida Bagus Nyoman Mantra ^c, I Ketut Wardana ^d

Article history: Received 20 July 2017, Accepted in revised form 26 November 2017, Approved 18 December 2017

Available online 28 December 2017

Correspondence Author ^a

Abstract



Keywords

*Measurement;
Reliability;
Test;
Validity;
Writing English;*

The present study aims to uncover how valid and reliable the multiple-choice test items as learning device are and their effect on permanent language habit formation. In this pretest-posttest one group pre-experiment study, we observed and analyzed the English test constructed by junior high school teachers. The Pretest on constructing the test items was administered to fifteen English teachers and some weaknesses were revealed. These can be seen from the total score of 855 and the mean score was 57.00. Treatment was given by ensuring them the rules of writing the test based on a principle of habit formation. The competence of teachers in constructing appropriate test items increased to higher quality indicated by scores of 79.33. It means the treatment has significantly influenced the ability of English teachers in constructing appropriate tests. The tests were tried out to 40 students of Junior High School and the result showed that 63 test items were valid and 37 was invalid. The fifty-valid test was administered to the students of different Junior High School and the result showed that the mean score was 72.53. 21 students from 40 students passed the minimum rank and 19 students still needed improving. The highest score was 95 and the lowest one was 26. The medium was 78 and the mode was 53. After measuring the reliability level, the standard deviation ranged 0.50 - 1.00, meaning that the test was appropriate to be applied as measuring device. The valid test was then reconstructed to be raw material of software module.

^a Mahasaraswati University Denpasar, Bali, Indonesia (nengah_astawa@yahoo.com)

^b Mahasaraswati University Denpasar, Bali, Indonesia (ndwihandayani@gmail.com)

^c Mahasaraswati University Denpasar, Bali, Indonesia (bagusmantra@hotmail.com)

^d Mahasaraswati University Denpasar, Bali, Indonesia (wardanak3tut@yahoo.co.id)

*e-ISSN : 2550-7001, p-ISSN : 2550-701X© Copyright 2017. The
Author. SS Journals Published by Universidad Técnica de Manabí.
This is an open-access article under the CC BY-SA 4.0 license
(<https://creativecommons.org/licenses/by-sa/4.0/>)
All rights reserved.*

Contents

Abstract	135
1. Introduction.....	136
2. Research Method.....	137
3. Results and Analysis	137
4. Conclusion	140
Acknowledgements.....	140
References.....	141
Biography of Authors	144

1. Introduction

Education assessment must have been conducted periodically and holistically based on the objective of the study, learning experience and appropriate rules of test construction as a learning device. In a simple term, the test is a method or a way used to measure knowledge, ability, and performance on a certain domain.¹⁶ Therefore, the test must be functioned appropriate either as measurement or learning device.¹⁷ Through test distribution to the students, teachers can acknowledge the level of students' competence mastery, help the students solve or map their problems, find the learning difficulties, diagnose the strength and the weaknesses of teaching method and motivate the students to improve learning achievement.⁷ From this few, whatever the final result is, teachers can act towards the effectiveness of teaching methods and give the feedback, either by remedial program or enrichment. However, in schools, due to the factors of time efficiency, the teachers used the previous English test, copied them from a test bank, and distributed inappropriate test.

There are numerous types of English test written by teachers and test makers to measure the language comprehension and the language skills¹¹, but some of them do not realize that the test is treated as the representation of learning expression. However, the test items that the local teachers construct, have not affected directly by the students' language ability as a communication tool. The assessment is a way to know language ability, language skills, and language behavior based on learning experience and the target in real life communication, so the tests must be simply constructed objective, transparent, and fair.⁵ The weaknesses of the test constructed by local English teachers can result in the false output of learning. If the test is not valid so the result cannot be trusted as an individual representation of learning experience. For example, the test written does not measure what it is supposed to measure.^{18,19,20} The test does not match with the purpose of language function and set the communicative language formation. The language task was given usually does not mirror the real use of language for daily life, such as inviting friends for meals, sending a friend short message by mobile phone, writing letters or asking someone to take pictures. The students' language ability is only measured by multiple choice test items that can lead them to guess, not to understand. This type of test is only measuring the cognitive domain.

To achieve the objective of the study, the teachers should be capable of mapping the students' competence targeted which breaks down in learning syllabus into standard competence and basic competence.²¹ Therefore, to measure the objective of the study planned, teachers should be able to construct as good test as possible. The good test must be attractive, educated, valid and reliable. The test may not contain any trap because this can harm the students. Whatever the test is, it must be able to dig back the students' knowledge in such a way.²² In understanding the test successfully, the students should gain three kinds of knowledge: linguistic knowledge, prior knowledge and textual.²³ Weaknesses of the test that teachers made must be overcome by giving the English teachers comprehension about the essential purpose of the test, features of a good test, and significance of the learning outcomes. Respecting the essential factors of language assessment, it is considered important

to rely on the process of language teaching and learning on assessment as it helps monitor students' educational progress and evaluates the quality of school systems. It is necessary to measure students' achievement by administering tests. Therefore, based on the phenomenon, this study was conducted through experiment and expose facto research design on analysis of constructing English test items based on the principle of habit formation rules. This is due to the urgency of test used to measure learning experience related to the objective of English learning as a foreign language that covers three domains, namely affective, cognitive and psychomotor.²⁴

This communicative test was then constructed in the frameworks of communicative domains by integrating those domains into every step of test construction.²⁵ The objective of the study is to 1) finding out the ability of the local English teachers in constructing the test, 2) analyzing the significance of treatment given to the teachers in constructing a test, and 3) describing the effect of their test towards students' test results. There was a lot of studies conducted and published on the topic of assessment, but yet there are few study concerns with the treatment to teachers in constructing a good test based on the principle of habit formation rules. The previous study on language testing was conducted.² Most of the result and the discussion of those study reported the quality of test items made by teachers, student's errors in understanding test instruction and the features of a good test as a measurement device. In contrary, this study may further concern with the treatment given to English teachers through the principle of habit formation rules and find out the output of the test administered.⁷ From this view, it can be hyphotized that the treatment of principle of habit formation rules can influence the teachers' ability in constructing the test significantly and the test items are considered appropriate as a measurement device.

2. Research Method

Based on the problem and treatment, the present study employed an experimental and exposes facto research design. The one group pretest-posttest design was chosen as the method to analyze the data and the later applied to describe the validity, reliability, acceptability of the test. One group pretest-posttest design is an experiment research design that is conducted by one group only without a control group.²⁰ The steps were started with pretest on teachers' ability in writing English test, a workshop about the principle of writing the test based on principles of habit formation rules and administered posttest to measure the effect of treatment. The stages of application of habit formation rules as basic principle of constructing the tests were mainly carried out in six steps of sequence, namely: (1) analyzing the relevant test models used by the teachers in junior high schools, (2) planning competence and the standard to be measured, (3) gathering local English teachers in Mengwi district in workshop, (4) testing the teachers of constructing the test based on the principle, and 5) conducting trial out teachers' test to the students and 6) judging the quality of the test as appropriate measurement device.

3. Results and Analysis

Referring to research design, one group pretest-posttest design, the data were taken from 15 English teachers from 10 junior high schools (SMP) in Mengwi district. The pretest that consisted of assignment to make good test for final examination was administered to know the preexisting ability of teachers in constructing appropriate test, such as Multiple-choice test items/error-recognition multiple-choice items, true-false test item, filling the blanks, the measurement of the teacher's ability was judged by using test construction rubrics.⁸ From the data analysis, it was known that the teachers still made a lot of mistakes in constructing both objective and subjective test. The test was not constructed based on rules of writing a good test. The mean score that those teachers obtained was 57.00. Then the treatment was conducted to them in a workshop that applied the method of writing good tests. For one week. The reviewers were the experts in language testing. The stimulation of the workshop was fulfilled with writing their own test and presented in the training program. The post-test was administered to know the effectiveness of the treatment. The ability of teachers in constructing English test was getting better, either in objective test or in a subjective test. The

Astawa, I., Handayani, N., Mantra, I., & Wardana, I. (2017). Writing English Language Test Items as a Learning Device: A Principle of Habit Formation Rules. International Journal Of Social Sciences And Humanities (IJSSH), 1(3), 135-144. doi:10.21744/ijssh.v1i3.67

evidence was the mean score was increasing into 79.33. The influence of the treatment was then analyzed by using t-test. The summary of the teachers' achievement in constructing English test can be seen in the following table.

Table 1
The summary of teachers' achievement in constructing English test

Measurement	Pretest	Post-test
Score	855	1.190
Mean	57.00	79.33
Dx	22.33	
t-observed	4.36	

Table 2
Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	pretest	57.0000	15	13.73213	3.54562
	Posttest	79.3333	15	4.46681	1.15333

The result of descriptive table paired sample statistic each of which N was 15 obtained the mean of pretest: 57.00 and posttest were 79.33. In general, the data analysis of spreading of post-test can be said that the standard deviation of pretest (Std = 13.73213) was nearly to std of post-test (Std=4.46681). Based on the data above the t-observed of df 14 was 2.03 with a significance level of 0.05. It showed that t-observed 4.36 was higher than t-table: 2.03. It can be concluded that the treatment of workshop of the principle of habit formation rules has a significant influence on teachers' ability in constructing a good test.

Table 3
Paired Samples Test

		<i>Paired Differences</i>			<i>95% Confidence Interval of the Difference</i>		t	df	<i>Sig. (2-tailed)</i>
		<i>Mean</i>	<i>Std. Deviation</i>	<i>Std. Error Mean</i>	<i>Lower</i>	<i>Upper</i>			
<i>Pair 1</i>	<i>pretest</i>	-22.33333	10.31411	2.66309	-28.04510	-16.62157	-8.386	14	0.000
	<i>Posttest</i>								

Based on analysis of t-test on pair sampling by using software SPSS 17 for Windows showed a value of $p = 0.000$ ($t = -8.386$, $df = 14$) $< \alpha = 0.05$. From the calculation, t-observed was 4.36 and bigger than t-table, 2.03, therefore the H_0 hypothesis was rejected and H_1 (Hypothesis Alternative) was accepted. It means the treatment has significantly influenced the ability of English teachers in constructing appropriate tests. Instead of the test, the study also used secondary data, the type of data that supported the analysis through administering the questionnaire. There were 10 questions asked of the teachers about their opinion, feeling and psychological effect of the given treatment and their test construction skill. The answers were

then calculated and described by the compared percentage. It showed that 47% of the teachers felt very satisfied, 44% of them were satisfied, 6% were satisfied enough. And 3% unsatisfied. Therefore, it can be concluded that teachers had a positive response towards workshop that influenced their comprehension and skill in constructing a good test based on principles of habit formation rules. Herewith, the chart that shows the psychological effect of the given treatment.

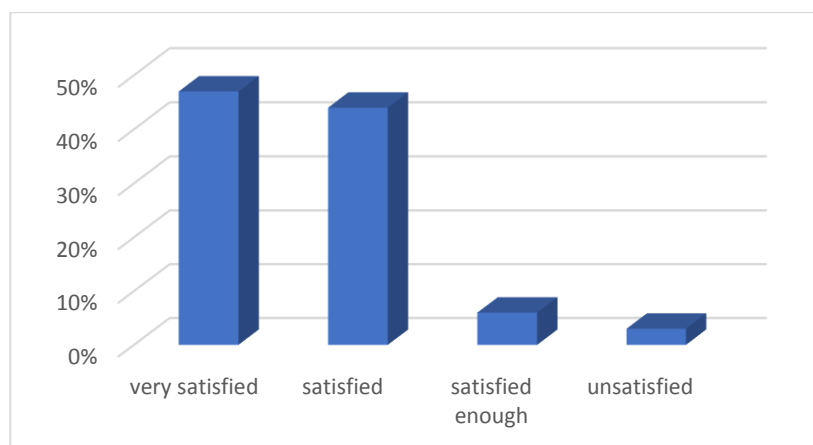


Chart 1. The psychological effect of the given treatment

The next step of the study was analyzing the teachers' work on the multiple-choice test items. They were supposed to work together to construct 100 test items tried out in SMP2 Mengwi. This was conducted to know how valid and reliable the test they made. Valid means that the test can measure what is supposed to be measured and nothing else (Heaton, 1988). Students' learning result was taken from a test that informed the teachers about their language performance. The construction of the test was based on a table of test specification containing an indicator of learning, types of test, and key answer. The test items were arranged in order from the simplest to the most difficult ones. The test was written based on the rules with six features and administered to the students by using random sampling with a lottery system. The students involved in this study were 40 students consisting of 15 males and 25 females. The students did the test in 90 minutes and it was ensured that no one was cheating. The students' answers were checked, scored and tabulated. To know the quality of the test, meaning how valid and reliable the tests were, there were some methods used in this study. The steps were 1) ranking their scores orderly from the highest to the lowest one; 2) dividing students into two groups, the upper (1-20) and the lower group (21-40); 3) counting the correct and incorrect answer of each number in every group; 4) formulating the level of facility index (FV) or level of difficulties and the easiness of each item (CU+CL/N). The range can be transcribed that the value that was less than 0.30 was rejected because the item was too difficult for the students, the items were accepted when the result ranged between 0.30-0.70, while the tests were rejected if the score was more than 0.70 which was considered too easy. The other part of test item selection was discrimination index (DV) used to analyze to what extent each test item could discriminate the able group and unable group. The formulation was CU-CL/n. The criteria of DV ranged 0.00- 0.20 was very poor discrimination, 0.21- 0.30 was poor, 0.31- 0.70 was very good and 0.00 - 0.10 was excellent (Heaton, 1988, Brown, 2004, Fulcher and Davidson, 2007).

From test item analysis, it can be informed that of 100 test items which were tried out to 40 students, there were 63 test items considered valid and 37 items were invalid. The invalidity was caused by some facts that the options were easy to guess, some were too easy and the others were too difficult. It was decided to use 50 of 63 valid test items to be administered to other 40 students of the different school, the seventh grade of SMPN 3 Mengwi. Based on the result of the test, the mean score of students' achievement was 72.53. Of the 40 students, 21 students were classified very well, and the rest or 19 students were considered sufficient. The highest score was 95 and the lowest one was 26. The median was 78 and the mode was 53. To know the level of reliability, it was used a formula called Kr 21 (Heaton, 1988). The range of the reliability is between 0.50 and 1.00. This test reliability score was 0.96 meaning that the test fell into high reliability and the test was appropriate to be used as a measurement device.

4. Conclusion

In conclusion, assessing the competence level of the students' ability has progressed a long way since the beginning learning process. A measurement device, a test administered to acknowledge the level of students' competence mastery, help the students solve or map their problems, find the learning difficulties, diagnose the strength and the weaknesses of teaching method and motivate the students to improve learning achievement. Therefore the tests constructed by 15 teachers after the treatment based on the principle of habit formation rules were considered valid and reliable. A well-ruled test construction makes that the purpose of communicative language testing is achievable. Therefore, based on the phenomenon, this study was conducted through experiment and expose facto research design on analysis of constructing English test items based on the principle of habit formation rules. This is due to the urgency of test used to measure learning experience related to the objective of English learning as a foreign language that covers three domains, namely affective, cognitive and psychomotor. The test was able to dig the students' closest environment knowledge. The test matches the purpose of language function with the setting of communicative language formation. The language task given mirrored the real use of language for daily life, such as inviting friends for meals, sending a friend short message by mobile phone, writing letters or asking someone to take pictures with various language expressions.

Acknowledgements

The authors would like to express their profound gratitude to all educators, teachers, and researchers for their help in conducting this study. Deep sincere gratitude is also dedicated to those who have endeavored their support, contribution, valuable input, and suggestion during the writing of this paper. Sincere gratitude also goes to the editors who have reviewed and approved this article to be published.

References

1. Prost, A., Colbourn, T., Seward, N., Azad, K., Coomarasamy, A., Copas, A., ... & MacArthur, C. (2013). Women's groups practising participatory learning and action to improve maternal and newborn health in low-resource settings: a systematic review and meta-analysis. *The Lancet*, 381(9879), 1736-1746.
[View in \(Google Scholar\)](#) [\(Scopus\)](#)
2. Astawa, I. N. (2011). PENGAJARAN "ENGLISH FOR GUIDING" BERBASIS PENDEKATAN SOSIOKULTURAL. *Jurnal*, 1(2), 170-178.
[View in \(Google Scholar\)](#)
3. Suparsa, I. N., Mantra, I. B. N., & Widiastuti, I. A. M. S. (2017). Developing Learning Methods of Indonesian as a Foreign Language. *International Journal of Social Sciences and Humanities (IJSSH)*, 1(2), 51-57.
[View in \(Google Scholar\)](#) [\(PDF\)](#)
4. Azwar, S. (1997). Reliability and validity. *Library Student, Yogyakarta*.
[View in \(Google Scholar\)](#)
5. Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Allyn & Bacon.
[View in \(Google Scholar\)](#)
6. Laborda, J. G., Magal-Royo, T., de Siqueira Rocha, J. M., & Álvarez, M. F. (2010). Ergonomics factors in English as a foreign language testing: The case of PLEVALEX. *Computers & Education*, 54(2), 384-391.
[View in \(Google Scholar\)](#) [\(Scopus\)](#)
7. Read, J. (1990). Providing relevant content in an EAP writing test. *English for specific purposes*, 9(2), 109-121.
[View in \(Google Scholar\)](#) [\(Scopus\)](#)
8. García Laborda, J., López Santiago, M., Otero de Juan, N., & Álvarez Álvarez, A. (2014). Communicative Language Testing: Implications for Computer Based Language Testing in French for Specific Purposes. *Online Submission*, 5(5), 971-975.
[View in \(Google Scholar\)](#) [\(PDF\)](#)
9. Maba, W. (2017). The Implementation of Education National Standard in the Instrument of School Accreditation of Bali Province Education Authority. *International Research Journal of Engineering, IT & Scientific Research (IRJEIS)*, 3(4), 1-6.
[View in \(Google Scholar\)](#)
10. Maba, W. (2017). Teacher's Perception on the Implementation of the Assessment Process in 2013 Curriculum. *International Journal of Social Sciences and Humanities (IJSSH)*, 1(2), 1-9.
[View in \(Google Scholar\)](#) [\(PDF\)](#)
11. Maba, W., & Mantra, I. B. N. (2017). An Analysis of Assessment Models Employed by The Indonesian Elementary School Teachers. *International Journal of Social Sciences and Humanities (IJSSH)*, 1(1), 39-45.
[View in \(Google Scholar\)](#) [\(PDF\)](#)





Astawa, I., Handayani, N., Mantra, I., & Wardana, I. (2017). Writing English Language Test Items as a Learning Device: A Principle of Habit Formation Rules. *International Journal Of Social Sciences And Humanities (IJSSH)*, 1(3), 135-144. doi:10.21744/ijssh.v1i3.67

-
12. Mantra, I. B. N. (2016). Promoting Primary School Teachers' Competence through Dynamic Interactive Workshop and Partnership. *Journal of College and University. This is an open access article under the, 2455*, 8028.
[View in \(Google Scholar\)](#) [\(PDF\)](#)
 13. Richards, J. C., & Rodgers, T. S. (2014). *Approaches and methods in language teaching*. Cambridge university press.
[View in \(Google Books\)](#) [\(PDF\)](#)
 14. Richards, Jack C. and Renandya, Willy A. 2002. *Methodology in Language Teaching: An Anthology of Current Practice*. New York: Cambridge University Press.
[View in \(Google Books\)](#)
 15. Yamtinah, S. (2009). Evaluasi Proses dan Hasil Belajar Kimia.
[View in \(Google Scholar\)](#)

 16. Astawa, I. N., Mantra, I. B. N., & Widiastuti, I. A. M. S. (2017). Developing Communicative English Language Tests for Tourism Vocational High School Students. *International Journal of Social Sciences and Humanities (IJSSH)*, 1(2), 58-64.
[View in \(Google Scholar\)](#) [\(PDF\)](#)
 17. Suryasa, I. W., Prayoga, I. G. P. A., & Werdistira, I. W. A. (2017). An Analysis of Students' Motivation Toward English Learning As Second Language Among Students In Pritchard English Academy (PEACE). *International Journal of Social Sciences and Humanities (IJSSH)*, 1(2), 43-50.
[View in \(Google Scholar\)](#) [\(PDF\)](#)
 18. Maba, W., & Mantra, I. B. N. (2017). An Analysis of Assessment Models Employed by The Indonesian Elementary School Teachers. *International Journal of Social Sciences and Humanities (IJSSH)*, 1(1), 39-45.
[View in \(Google Scholar\)](#) [\(PDF\)](#)
 19. Amerta, I. M. S. (2017). The Role of Tourism Stakeholders at Jasri Tourism Village Development, Karangasem Regency. *International Journal of Social Sciences and Humanities (IJSSH)*, 1(2), 20-28.
[View in \(Google Scholar\)](#) [\(PDF\)](#)
 20. Maba, W., Perdata, I. B. K., & Astawa, I. N. (2017). Constructing Assessment Instrument Models for Teacher's Performance, Welfare and Education Quality. *International Journal of Social Sciences and Humanities (IJSSH)*, 1(3), 88-96.
[View in \(Google Scholar\)](#) [\(PDF\)](#)
 21. Basak, A., & Khanna, K. (2017). A Study on the Selection Criteria of Different Hotels of Delhi NCR in Accordance to the HR Policies and Market Trends. *International Journal of Social Sciences and Humanities (IJSSH)*, 1(1), 27-38.
[View in \(Google Scholar\)](#) [\(PDF\)](#)
 22. Amerta, I. M. S. (2017). Community Based Tourism Development. *International Journal of Social Sciences and Humanities (IJSSH)*, 1(3), 97-107.
[View in \(Google Scholar\)](#) [\(PDF\)](#)

-
23. Atwell, N. (1998). *In the middle: New understandings about writing, reading, and learning*. Boynton/Cook Publishers, Inc., A Subsidiary of Reed Elsevier Inc., 361 Hanover Street, Portsmouth, NH 03801-3912.
[View in \(Google Scholar\)](#)
24. Casanave, C. P., & Hubbard, P. (1992). The writing assignments and writing problems of doctoral students: Faculty perceptions, pedagogical issues, and needed research. *English for Specific Purposes*, 11(1), 33-49.
[View in \(Google Scholar\)](#) ([Scopus](#))
25. Flowerdew, J. (1999). Problems in writing for scholarly publication in English: The case of Hong Kong. *Journal of Second Language Writing*, 8(3), 243-264.
[View in \(Google Scholar\)](#) ([Scopus](#))

Biography of Authors

	<p>Drs. I Nengah Astawa, M.Hum is a lecturer in English Education Study program. He is a lecturer of diverse subjects related to English Education. He has conducted extensive researches in various subjects related to English language teaching and literature. His papers have been published nationally and internationally in various reputable journals</p>
	<p>Nengah Dwi Handayani, S/Pd., M.Pd is a lecturer of English who has conducted several studies majoring in English education. Her academic papers have been published in a number of education journals. Her long experiences in teaching English has inspired her to write academic papers and publish them in various journals</p>
	<p>Dr. Ida Bagus Nyoman Mantra, SH., S.Pd., M.Pd is a Doctor in Linguistics from Udayana University and holds a master degree in education from Ganesha Education University. He is a lecturer in English Education at Mahasaraswati University, Denpasar, Bali. He has carried out several studies in language education, oral tradition, and human right laws. His study focus is on investigating functions, values, and meanings of Balinese oral tradition. His articles have been published in several national education and linguistics journals.</p>
	<p>I Ketut Wardana, S.Pd., M.Hum is a lecturer of English majoring in language education and linguistics. He has experienced in teaching students of various grades from children to adults. His academic reviews and articles have been published in various national journals. He has published books in linguistics, English education and other related interesting topics in language education.</p>