



The Validation Of Reading Comprehension Test By Employing Rasch Model (A Study Case At English Department Of Papua University)

Imelda Mallipa

Universitas Papua

Email: i.mallipa@unipa.ac.id

Abstract: This research investigated the validity of Reading Comprehension test that was taken from Longman complete course for the TOEFL test, Preparation for computer and paper test written by Debora Philips (2001), page 528 until page 537. Thirty-five undergraduate students of English Education major took part in this study. The result of the test are dichotomously scored and analysed by employing Rasch model with the application of winsteps software. The findings revealed that all items in Reading Comprehension Test were valid based on the criteria proposed by Boone et al. (2014). The test information function and the level of students' ability was low. To get the optimal of test information function, the test should be administered to the students in medium ability. It is suggested to use this test to improve the reading comprehension of students with low level proficiency.

Key words: Test Validation, Reading comprehension, Rasch Model

Abstrak: Penelitian ini bertujuan untuk menyelidiki validitas soal "*Reading Comprehension*" dalam buku persiapan TOEFL halaman 528- 537 yang di tulis oleh Debora Philips pada tahun 2001. Ada 35 mahasiswa program sarjana strata satu (S1) pada Jurusan Pendidikan Bahasa Inggris di Universitas Papua yang mengambil bagian dalam penelitian ini. Data dikotomi dari hasil test dikelolah dengan pemodelan Rasch melalui penggunaan program lunak komputer (*software*) winsteps. Hasil yang diperoleh menunjukkan bahwa semua butir soal dalam test "*Reading Comprehension*" memenuhi kriteria valid berdasarkan kriteria (Boone et al.(2014). Informasi yang didapatkan dari pengukuran cukup rendah karena level abilitas mahasiswa rendah. Untuk menghasilkan informasi yang optimal, test ini seharusnya diujikan pada mahasiswa dengan abilitas sedang. Penelitian berikutnya disarankan untuk menggunakan test ini untuk meningkatkan abilitas mahasiswa dengan kemampuan rendah.

Kata Kunci: validiatas soal, kemampuan membaca, Pemodelan Rasch

A. INTRODUCTION

Assessing reading comprehension involves complex process. Therefore, the validation of the test is needed as a logical evidence to support the interpretation of students' score on test and as references in gathering data about the information of students' performance on reading

comprehension test. It can give information whether the test items test what should be tested or not. In teaching and learning process, the validation of test can give the accurate interpretation. If a test is valid as indicator of reading comprehension, then it can be used to improve reading skills/abilities of students with lower proficiency levels.

Since the TOEFL is used as a standard test to measure the level of English proficiency level of students including reading comprehension, many teachers of English reading comprehension especially teachers in higher education apply the test in their classroom without check its validity. It is believed that the test has been developed well by the expert. Though teacher may tend to use that existing test, they are still expected to generate evidence from their classroom.

The kind of reading comprehension test in TOEFL is Multiple-choice (MC) form. This kind of test has been questioned on its validity and relevancy in the age of communicative language assessment (Huhta & Randel, 1996:94). It might be difficult for teachers to make the interpretation about students' comprehension on reading by using students' score on MC test without the validity of the test. This present study tried to find the evidence of the test validation from the context of my classroom.

The students in my classroom had learned reading strategies and skills formally for two years. They were given reading comprehension test that was taken from the TOEFL Preparation book. The result of test was analyzed by using simple statistic. It showed that almost all of students had low score in comprehending the Reading text. Only 1 out of 35 students who got high score and 5 students who were in medium, . To come into the interpretation about the ability of students to read and understand the English Reading text, the evidence was required to show that the measure was accurate.

Table 1: Reading Comprehension Category

Category	Score	Frequency
Very High	41 - 50	0
High	31 - 40	1
Medium	21 - 30	5
Low	11 - 20	23
Very Low	1 - 10	6
Sum		35

To validate the test, Rasch Model measurement was applied. It was believed that Rasch Model Measurement could give complete information about the quality of each test item and the ability of students. Since validation is about building strong argument as foundation in explaining what students are able and are not able

to do, the Rasch model measurement provides sufficient informain.

Based on the above problems in reading comprehension, the present research intended to answer the following questions:

1. Is the Reading Comprehension test valid measure based on Rasch Model?"
2. What is the students' level of Reading ability based on Rasch Model?

B. LITERATURE REVIEW

1. Test Validation and Reading Comprehension

Test validation is the process of generating evidence to justify the interpretaion of abilities from the test scores, Weir (2005). The evidence provides information about how the result of score reflects the accurate representation of students's level of knowledge or skills. It includes the data relating to construct-based, content-based and criterion-related validities, together with the reliabilities of test.

Construct validity is the investigation of whether a test measures the construct adequately. In this process, the test is analysed to ensure that it can measure a consruct in individuals through the using of statistical analysis as an adequete basis for determining the existence or non-existece of irrelevant variables. For instance, the presence of writing abilty in the test for students's proficiency in reading.

The test for assesing the reading comprehension of students that addresses the component of skills and strategies should reflect the purpose of reading in a reading process. Students may activate different levels of strategis and skills to cope with different reading purposes. When in-depth comprehension is the requirement of the test, then the students are expected to find information in text by reading carefully. it is diffrent from reading test for very specific reading goals that requires students to scan the text or read quickly through the text.

The good reading test items are able to distict the students with different proficiency. Ghafournia & Afghari (2013) found that the students with different level of reading proficiency use different strategies in comprehending reading text and answer reading questions. Those who have high level of reading proficiency link the main idea of reading text with the questions, find the main ideas of reading text through scan and skim, comprehend the

reading text without translating word for word, and guess the meaning of unknown through using the contextual clues, while the low level of proficiency group the words with similar meaning, visualize new words, group the words with similar pronunciation, write the main ideas of reading texts in key sentences, and read passages several times. The result of their study also shows that the students with more linguistic knowledge used comprehending test-taking strategies more frequently to comprehend the reading text and answer the questions

The ability of students to answer the reading questions is affected by vocabulary size. Ibrahim, Sarudin, and Muhamad (2016) who studied the relationship between vocabulary size and reading comprehension of ESL learners at a public university in Malaysia found that there was statistically significant difference between score in reading comprehension and vocabulary test. To be a good comprehension in a reading text, the students should have levels of vocabulary familiarity more than 75% (Bonk, 2000), or master the most frequent 2,000 words (Schmitt, 2000 & Laufer 1992). Another factor that gives high contribution to comprehending reading text is prior knowledge (Ozuru et al., 2009; Shaprio, 2004; Carrel and Wise, 1998; & Chen 2008). The students with high prior knowledge have better performance on reading test than those with low prior knowledge (Abdelaal & Sase, 2014).

2. Rasch Model

The Rasch model is model-based measurement in which trait level estimate depends on both the persons' responses and on the properties of the items that were administered (Embretson & Reise, 2000:3). This model shows the response pattern consistency for items and also for persons. Therefore, the using of this model in validating the test can give accurate picture of the difficulty & the dimensionality of test items and underlying the ability of students. It can show the consistency of an item in assessing students with different ability.

Rasch Model was developed based on the probability of person's response (students) to give correct answer on an item. The students who have higher ability than another should have greater probability of solving any item of the type in question, and similarly, one item being more difficult than another means that for any person the probability of solving the second item correctly is greater one (Rasch 1960:117). It

means that if item 1 is the question with the highest level of difficulty in test, then the students with lower ability are not able to answer the question correctly. If a student with lower ability gives correct answer, then the student only tries to answer by guessing.

The result of persons' responses and the properties of the items is shown in variable maps. The maps can show both the position of items and students along the line of the variable. It is easy to identify both the students with high low level and it is easy to recognize the difficulty level of items. The students that have the same level of proficiency and the items with the same level of difficulty will be at the same position.

C. METHODS

Participants

Thirty-five undergraduate students of English Education major took part in this study. They were at the end of the fourth semester at Universitas Papua. They have received formal teaching about reading skills and strategies for two years. The students would be assigned to TOEFL and IELTS preparation as one subject included in curriculum.

Materials

The materials were taken from Longman complete course for the TOEFL test, Preparation for computer and paper test written by Debora Philips (2001), page 528 until page 537. The test contained the items that required students to answer the questions about main idea, the organization of ideas, stated and unstated details, pronouns referents, implied details, transitions questions, vocabulary (structural, word parts, context), and the specific information.

Procedure

The test was administered to 35 undergraduate students of English Education. There were 50 questions that all were in multiple-choice format. The test took time for 100 minutes. The result of the test are dichotomously scored and analysed by employing winsteps software version 3.73.

D. RESULT AND ANALYSIS

Rasch model with the application of winsteps software version 3.73 was used to analyse the data to check the validity and reliability of the test and the Reading

Comprehension of students. The using of Rasch Model showed the relationship of students' response and the items while the reading ability showed the level of students' Reading proficiency. The validity of test as instrument is shown in Table 1, the validity of each item is shown in Table 2, and the level of students' ability in Reading Comprehension test is shown in Graph 1 and Table 3

The mean item difficulty was set at 0 by definition and the mean examinee ability was -0,87. It means that the reading comprehension of the students in this test was lower than the difficulty level of test. The person reability and Item reliability was 0.68 and 0.83 indicating a good reability for Raeding comprehension test. Item separation was 2.20 that means the test items can be categorized into 3 level of difficulty: difficult, moderate, and easy.

Table 1. The Validity of Test

SUMMARY OF 35 MEASURED Person

	TOTAL		MEASURE	MODEL	INFIT		OUTFIT	
	SCORE	COUNT		ERROR	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	17.3	50.0	- .87	.34	1.00	-.1	1.02	.0
S.D.	5.3	.0	.61	.03	.17	1.2	.36	1.1
MAX.	32.0	50.0	.68	.46	1.41	2.4	1.89	2.1
MIN.	6.0	50.0	-2.40	.32	.76	-1.8	.61	-1.4
REAL RMSE	.35	TRUE SD	.50	SEPARATION	1.41	Person RELIABILITY	.67	
MODEL RMSE	.34	TRUE SD	.50	SEPARATION	1.48	Person RELIABILITY	.69	
S.E. OF Person MEAN = .10								

SUMMARY OF 50 MEASURED Item

	TOTAL		MEASURE	MODEL	INFIT		OUTFIT	
	SCORE	COUNT		ERROR	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	12.1	35.0	.00	.44	1.00	.0	1.02	.0
S.D.	6.6	.0	1.16	.17	.09	.6	.23	.7
MAX.	30.0	35.0	2.82	1.02	1.18	1.0	1.89	1.8
MIN.	1.0	35.0	-2.79	.35	.84	-1.3	.27	-1.4
REAL RMSE	.48	TRUE SD	1.06	SEPARATION 2.20	Item RELIABILITY	.83		
MODEL RMSE	.47	TRUE SD	1.06	SEPARATION	2.25	Item RELIABILITY	.83	
S.E. OF Item MEAN = .17								

The items of Reading Comprehension Test were valid based on the criteria proposed by Boone et al. (2014). The items was valid should meet at least one of the three criteria. The criteria are as following:

- a. The value of Outfit Mean Square (MNSQ) in the range of $0.5 < MNSQ < 1.5$
- b. The value of Outfit Z-Standard (ZSTD) in the range of $-2.0 < ZSTD < 2.0$

c. The value of Point Measure Correlation (Pt. Mean Corr) in range of \pm <Pt Measure<0.85.

Based on the criteria, it was found that all of items (50 items) in reading comprehension test were valid. It means that all the items in this test

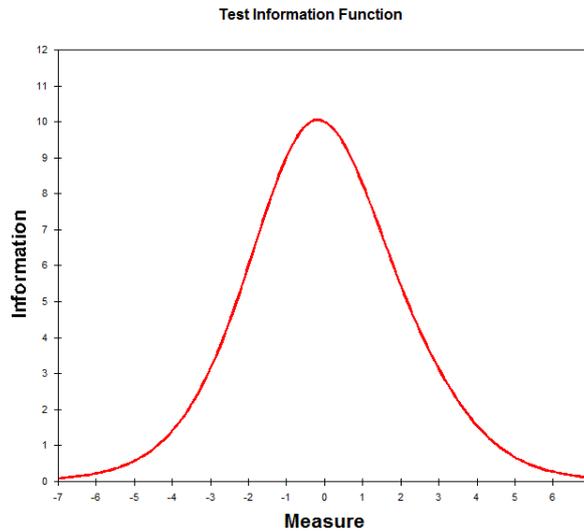
played the function to measure what it intended to be measured. There were 11 items (item 33, 41, 3, 28, 30, 6, 45, 36, 8, 22, 38) that met all the criteria, while there were two items that fitted only the value of Outfit Z-Standard (ZSTD) and others fitted both the value of (MNSQ) and the value (ZSTD).

Table 2. Validity of Items

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL		INFIT		OUTFIT		PT-MEASURE		EXACT MATCH		Item
				S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	OBS%	EXP%		
23	5	35	1.04	.49	1.18	.6	1.89	1.8	A-.23	.20	85.7	85.7	S23	
26	2	35	2.09	.74	1.08	.3	1.53	.8	B-.11	.14	94.3	94.3	S26	
12	2	35	2.09	.74	1.07	.3	1.41	.7	C-.07	.14	94.3	94.3	S12	
21	6	35	.82	.46	1.16	.6	1.38	1.0	D-.11	.21	82.9	82.8	S21	
48	8	35	.44	.41	1.05	.3	1.36	1.3	E .07	.23	80.0	77.4	S48	
27	6	35	.82	.46	1.15	.6	1.22	.7	F-.04	.21	82.9	82.8	S27	
39	1	35	2.82	1.02	1.03	.4	1.17	.5	G .00	.10	97.1	97.1	S39	
11	24	35	-1.71	.38	1.16	1.0	1.15	.8	H .02	.26	62.9	70.0	S11	
43	6	35	.82	.46	1.13	.5	1.14	.5	I .00	.21	82.9	82.8	S43	
44	7	35	.62	.43	1.14	.6	1.12	.5	J .02	.22	77.1	80.0	S44	
50	12	35	-.17	.37	1.14	1.0	1.13	.8	K .05	.26	62.9	67.4	S50	
40	5	35	1.04	.49	1.08	.3	1.11	.4	L .07	.20	85.7	85.7	S40	
5	10	35	.12	.39	1.11	.7	1.10	.5	M .08	.25	74.3	72.2	S5	
20	13	35	-.30	.36	1.10	.9	1.10	.7	N .11	.27	60.0	65.3	S20	
10	15	35	-.56	.36	1.06	.7	1.10	.8	O .16	.27	60.0	62.2	S10	
31	12	35	-.17	.37	.98	-.1	1.09	.5	P .26	.26	74.3	67.4	S31	
34	19	35	-1.06	.35	1.07	.8	1.07	.7	Q .17	.28	51.4	61.9	S34	
9	8	35	.44	.41	1.07	.4	1.03	.2	R .14	.23	74.3	77.4	S9	
15	7	35	.62	.43	.97	.0	1.06	.3	S .23	.22	82.9	80.0	S15	
29	20	35	-1.18	.36	1.06	.7	1.06	.5	T .18	.28	60.0	62.9	S29	
47	1	35	2.82	1.02	1.03	.3	1.05	.4	U .03	.10	97.1	97.1	S47	
18	12	35	-.17	.37	1.02	.2	1.05	.4	V .22	.26	68.6	67.4	S18	
42	12	35	-.17	.37	1.04	.3	1.00	.1	W .22	.26	62.9	67.4	S42	
7	25	35	-1.86	.39	1.04	.3	1.02	.2	X .21	.26	65.7	72.4	S7	
13	18	35	-.93	.35	.98	-.2	1.03	.3	Y .28	.28	68.6	60.9	S13	
38	15	35	-.56	.36	1.01	.2	1.01	.1	Y .25	.27	60.0	62.2	S38	
25	8	35	.44	.41	1.00	.1	.95	-.1	X .25	.23	74.3	77.4	S25	
14	17	35	-.81	.35	.99	-.1	.97	-.3	W .30	.28	62.9	60.7	S14	
2	19	35	-1.06	.35	.97	-.3	.96	-.3	V .33	.28	62.9	61.9	S2	
1	18	35	-.93	.35	.97	-.3	.96	-.3	U .33	.28	74.3	60.9	S1	
24	15	35	-.56	.36	.95	-.5	.97	-.2	T .34	.27	71.4	62.2	S24	
49	19	35	-1.06	.35	.97	-.3	.95	-.4	S .33	.28	62.9	61.9	S49	
16	15	35	-.56	.36	.96	-.4	.92	-.6	R .36	.27	65.7	62.2	S16	
17	15	35	-.56	.36	.94	-.6	.95	-.4	Q .36	.27	71.4	62.2	S17	
32	8	35	.44	.41	.95	-.2	.86	-.4	P .35	.23	74.3	77.4	S32	
37	17	35	-.81	.35	.94	-.6	.92	-.7	O .37	.28	57.1	60.7	S37	
19	12	35	-.17	.37	.94	-.4	.88	-.6	N .38	.26	68.6	67.4	S19	
35	5	35	1.04	.49	.94	-.1	.84	-.2	M .31	.20	85.7	85.7	S35	
46	7	35	.62	.43	.94	-.2	.85	-.4	L .35	.22	77.1	80.0	S46	
22	14	35	-.43	.36	.91	-.8	.89	-.8	K .41	.27	74.3	63.5	S22	
8	20	35	-1.18	.36	.91	-.8	.89	-.9	J .42	.28	71.4	62.9	S8	
36	16	35	-.68	.35	.91	-1.0	.90	-.9	I .42	.27	71.4	61.3	S36	
45	7	35	.62	.43	.91	-.3	.79	-.6	H .40	.22	82.9	80.0	S45	
6	12	35	-.17	.37	.88	-.9	.82	-1.0	G .47	.26	68.6	67.4	S6	
30	15	35	-.56	.36	.87	-1.3	.84	-1.4	F .49	.27	65.7	62.2	S30	
28	20	35	-1.18	.36	.87	-1.3	.84	-1.3	E .49	.28	71.4	62.9	S28	
3	30	35	-2.79	.50	.87	-.3	.72	-.6	D .45	.21	85.7	85.7	S3	
4	12	35	-.17	.37	.87	-1.0	.82	-1.1	C .49	.26	74.3	67.4	S4	
41	1	35	2.82	1.02	.86	-.2	.27	-.5	B .44	.10	97.1	97.1	S41	
33	11	35	-.03	.38	.84	-1.1	.78	-1.2	A .53	.26	77.1	69.7	S33	
MEAN	12.1	35.0	.00	.44	1.00	.0	1.02	.0			73.9	72.8		
S.D.	6.6	.0	1.16	.17	.09	.6	.23	.7			11.1	11.3		

The level of students' ability in reading test can be seen in Graph 1. X-axis showed the level of students' ability in answering reading text questions and Y-axis showed the accumulative proportion. It is shown that both the level students ability and the test information function was low. To get the optimal test information function, the test should be administered to the students in medium ability, Sumintono & Widhiarso (2015).

Graph 1. Test information Function

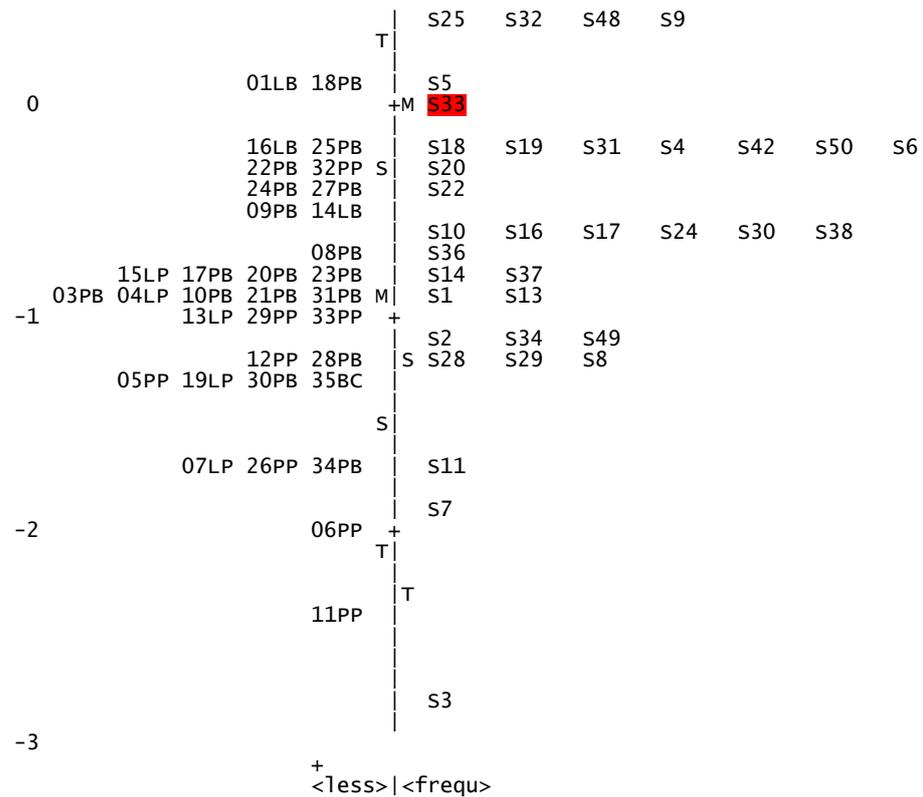


Students' ability (right) and items' difficulty (right) side is located in logit on the item-person map (Table.3). The students who were located above the 0.0 logit were the students with the high ability and the students below 0 were the students with the low ability. Most students were located below 0.0. Only three students (02PB, 01LB, 18PB) out of thirty-five students who were positioned between logit 0.0 – 1.0. In the other hand, the items (item 33) that located at the 0 point on the item-person map is the mean of the item difficulty estimates. The

items below 0 have less difficult than item 33 and the items above item 33 have more difficult. Any student has 50% probability of succeeding with items located at the same point on the logit scale and have a greater or less than 50% chance of succeeding on item below and above the logit scale. For example, The students with code 02PB have a greater chance to answer the question number 15, 44,, 45, and 46 than the questions number 21, 27, and 43. The student with code 11PP has only 1 question to be answer with a greater chance.

Table 3. The Item-Person map

	<more>		<rare>	
3		+		
			S39	S41 S47
			T	
2		+	S12	S26
			S	
1		+	S23	S35 S40
			S21	S27 S43
	02PB		S15	S44 S45 S46



E. CONCLUSION

The result of the study showed that all items in this reading comprehension test were valid measure and the students' level ability in this research was low. The mean difficulty level of test that was 0.0 logit was greater than the means ability of students that was -0.87 logit. Based on the test information function, the test is suitable to be administered to the student in medium level of reading proficiency. Therefore, the students that took place in this test should be given a specific treatment to develop their reading comprehension.

REFERENCES

- Abdelaal, Noureldin M & Sase, Amal Saleh. 2014. Relationship between Prior Knowledge and Reading Comprehension. *Advances in Language and Literacy Studies; Vol.5 No 6, 131*
<http://dx.doi.org/10.7575/aiac.all.v.5n.6p.125>
- Boone, W.J., Stever, J.R., & Yale, M.S (2014). *Rasch Analysis in The Human Sciences*. Dordrecht: Springer Science, and Business Media.
- Bonk, W. J. 2000. Second Language Lexical Knowledge and Listening Comprehension. *International Journal of Listening*, 14, 14-31.
<http://dx.doi.org/10.1080/10904018.2000.10499033>
- Embretson, S.E., & Reise, S. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Erlbrum
- Carrel, P.L., & Wise, T.E. (1998). The relationship between Prior Knowledge and Topic Interest in Second Language Reading. *Studies in Second Acquisition*. 20(3), 285-309.
- Chen, C. (2008). *The effect of Background Knowledge and previews on Elementry native mandarin-speaking English Language Learner's reading comprehension*
- Ghafournia, Narjes & Afghari, Akbar. 2013. The interaction between Cognitive Test-Taking Strategies, Reading Ability, and Reading Comprehension Test Performance of Iranian EFL Learners. *English Language Teaching; Vol.6, No. 8; 2013, 139-150*.
<http://dx.doi.org/10.5539/elt.v6n8p139>
- Ibrahim, Engku H.E, Sarudin, Isarji, and Muhamad, Aionon J. 2016. The Relationship between Vocabulary Size and Reading Comprehension of ESL Learners. *English*

- Language Teaching*; Vol.9 No. 2; 2016, 116-123. . <http://dx.doi.org/10.5539/elt.v9n2p116>
- Laufer, B. (1992). How Much Lexis is necessary for Reading Comprehension. *Vocabulary and Applied Linguistics*, 126-132. http://dx.doi.org/10.1007/978-1-349-12396-4_12
- Messick, S. 1989. Validity. In R. Linn (Ed.), *Educational Measurement*. New York: Macmillan, pp. 13-103
- Ozuru, Y., Dempsey, K., Mcnamara, D.S. (2009). Prior Knowledge, Reading Skill, and Text Cohesion in the Comprehension of Science texts. *Learning and Instruction*, 19(3), 228-242.
- Shapiro, A.M. (2004). How including prior knowledge as a Subject Variable May Change Outcomes of Learning Research. *American Educational Research Journal*.
- Schmitt, N. (2000). *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.
- Sumintono, Bambang., & Widhiarso, Wahyu.2015. *Aplikasi Pemodelan Rasch pada Assessment Pendidikan*. Cimahi: Trim Komunikata
- Weir, Cyril J. 2005. *Language Testing and Validation: an Evidence-Based Approach*. New York: Palgrave Macmillan.