Jurnal Ekonomi
Volume 13, Number 02, 2024, DOI 10.54209/ekonomi.v13i02
ESSN 2721-9879 (Online)
https://ejournal.seaninstitute.or.id/index.php/Ekonomi

Validity and Reliability of the Maritime English Seafarer Proficiency Exam

Adesvan Gumay[1], Siwi Woro Herningsih[2], Evi Sukma Viamita[3]
[1,2,3]Politeknik Pelayaran Banten

Article Info
Keywords:
Validity of Maritime English Seafarer Skill Exam,
Reliability of Maritime English Seafarer Skill Test

ABSTRACT
The purpose of this study is to determine the Validity and Reliability of the Maritime English Seafarer Expertise Exam conducted by PUKP 02 Banten and Surrounding Areas where the exam is held at the Banten shipping Polytechnic campus and questions are given in the form of multiple choice. This study is categorized as a descriptive analysis because it is intended to describe the level of difficulty, discriminating power, validity and reliability of the Maritime English test for level III cadets at the Sailing School for the 2019/2020 school year by analyzing the validity and reliability of each item. The results of this study concluded: 1) The validity of the seafarer expertise exam questions majoring in nautics there are 43 question items or equivalent to 71.66% of the total items are said to be valid, so the test has good validity because the value of the correlation / coefficient relationship is greater (>) than table (Rtable) = 0.1071 for a significance level of 5%. For the Engineering Department there are 41 question items or equivalent to 68.33% of the total items are said to be valid, so the test has good validity because the value of the correlation / coefficient relationship is greater (>) than the table (Rtable) = 0.2483 for a significance level of 5%. 2) The reliability of the nautical department test device was found to be 0.5968 while for the engineering department of 0.6789 the figure was in the interval 0.40 to 0.70 with moderate interpretation. Thus it can be concluded that the Maritime English sailor skill test has sufficient reliability.

Corresponding Author:
Adesvan Gumay
Politeknik Pelayaran Banten
vangoemay@gmail.com

INTRODUCTION

As the largest maritime country in the world, Indonesia should be more 'king' in the maritime world. Just imagine, we have vast ocean waters stretching, connecting many islands. We also have a lot of shallow waters that greatly affect the availability of natural resources, and of course they are very important for the lives of our people.

It has become a common dream for cadets of commercial shipping schools in Indonesia to be able to work and gain experience on foreign ships. This happens because of the very significant difference in income between seafarers working on Indonesian ships and seafarers working on ships of other countries.

What is the main shortage of Indonesian seafarers? Their English skills are still very shallow. In fact, to work on foreign ships, mastering Maritime English is a must, moreover, they are prospective officers on board. Where English is the main language of instruction in

Jurnal Ekonomi
Volume 13, Number 02, 2024, DOI 10.54209/ekonomi.v13i02
ESSN 2721-9879 (Online)
https://ejournal.seaninstitute.or.id/index.php/Ekonomi

communicating on board. Even for shipping in Indonesia, the use of English is an obligation. English used in a maritime context is Maritime English.

Mastery of Maritime English and less qualified makes Indonesian sailors unable to compete in the international market. Indonesian sailors are currently still far behind sailors from the Philippines, because they have good Maritime English. Today the Philippines is the ruler of the world market share with sailors working in almost all countries that have shipping companies.

Especially in shipping schools, the measurement of the achievement of marine competence is carried out in an exam entitled seafarer expertise test (UKP). The exam is conducted by an institution outside the educational institution. The institution is the Board of Seafarer Expertise Examiners (DPKP). Where in its implementation, DPKP has representatives in every region in Indonesia, namely the Seafarer Expertise Examination Committee (PUKP). In Indonesia there are 13 PUKP spread throughout Indonesia. One of them is PUKP 02 Banten Region which houses shipping high schools on parts of Java and Sumatra.

The Seafarer Expertise Exam is divided into two stages, namely the CBA (Computer Based Assesment) Exam and the Comprehensive Test or Practical Exam. In the CBA exam, cadets are required to answer questions consisting of several subjects where each subject consists of 30 (thirty) multiple-choice questions. And one of the subjects tested was Maritime English. This study focuses on different sides of the problem to examine quality standards, to measure the breadth of validity and reliability of the Maritime English Seafarer Expertise Exam.

Some forms of questions in the test are subjective questions in the form of open and limited essay questions, and objective questions in the form of multiple-choice, true or false questions (true and false) and arranged questions (arrangement questions) All of these forms require extensive experience. It is a common belief that the more questions varied, the more reliable and valid tests there are. As teachers we must consider all types of questions during the preparation of the test in order to produce good test standards. In addition, by using evaluation, teachers will understand the abilities and progress of their students.

Educational evaluation is needed to measure how successful the teaching and learning process has been experienced by students. A valid evaluation must have phases, using instruments called tests. This is very important for teachers so that the ideal test must have validity, reliability, and usefulness, Gronlund (in Burhan Nurgiyantoro, 2016: 168) stated that "essentially test kits must have the qualities of validity, reliability, and usability".

As an assessment tool in a formal education, seafarer skill test questions are prepared by the Seafarer Expertise Examination Council (DPKP) based on questions made by teachers in each shipping school. The test consists of 30 multiple-choice items with four choices. From the results of the Maritime English test, it has not reflected or has not met the good test qualifications, especially the quality of these items which are sometimes very difficult to understand. Based on the above statement, the author is interested in analyzing

Jurnal Ekonomi
Volume 13, Number 02, 2024, DOI 10.54209/ekonomi.v13i02
ESSN 2721-9879 (Online)
https://ejournal.seaninstitute.or.id/index.php/Ekonomi

the items of the Maritime English test given to cadets in the implementation of the seafarer expertise exam.

METHODS

The research was conducted at the Shipping High School which conducts the Seafarer Expertise Exam at PUKP 02 Banten and has received approval from the Directorate General of Sea Transportation regarding the implementation of its education. The schools include; SMK-P Jakarta, SMK-P Lusiana Tangerang, SMK-P Djadajat, SMK-P Malahayati Jakarta, SMK-P Buana Bahari Cirebon, SMK-P Pembangunan Jakarta, and SMK-P Jakarta Raya. The author takes a bank of UKP questions in Maritime English and other data needed for analysis.

This study is categorized as a descriptive analysis because it is intended to describe the level of difficulty, discriminating power, effectiveness of deception, validity and reliability of the Maritime English test for level III cadets at the Sailing School for the 2019/2020 school year by analyzing the validity and reliability of each item. According to Gay (2010: 276*) "Descriptive research sound is very simple-just ask some people some questions and count responses-but there is considerably more to it than just asking questions and reporting answers. A set of basic step should guide descriptive research studies"*. The steps in descriptive research are 1. Identify the topic or problem; 2. Choose suitable samples; 3. Collect valid and reliable data; 4. Analyze and conclude the results.

This study is also considered a quantitative study because the author used some numerical data that was analyzed statistically. Kumar 2011 argues that there are three types of qualitative research namely; cross-sectional studies, before-and-after studies, and longitudinal studies. This research is also included in before- and after studies where researchers analyze an object that is applied repeatedly to the same sample. Analysis was carried out on these two results.

The population of this study is cadets of level III shipping schools who have received approval from the Ministry of Transportation for the 2019/2020 academic year. Researchers determined this population because level III cadets are considered to have completed all learning processes so it is expected that the selected sample even though they are from different schools but have almost the same intellectual abilities as each other.

The average total number of cadets in each school ranges from 15-100 people consisting of two different majors, namely the Nautka and Engineering Departments. Where these two majors get learning Maritime English subjects. And in the implementation of the seafarer skill exam, there are also these subjects.

The author took 100% of the total number of level III cadets in each school as a sample because the population was less than 1000 subjects, this is based on the opinion of Gay and Airasian (2000: 134) argued *"For smaller populations, say, N = 100 or fewer, there is little point in sampling; survey the entire population. If the population size is around 500 (give or take 100), 50% should be sampled."*

Jurnal Ekonomi
Volume 13, Number 02, 2024, DOI 10.54209/ekonomi.v13i02
ESSN 2721-9879 (Online)
https://ejournal.seaninstitute.or.id/index.php/Ekonomi

The author uses random purposive sampling technique. According to Gay and Airasian (2000:139) *"Random purposive sampling (with small sample): selecting by random means* participants *who were purposively selected and who are too numerious to include all the study*. From these opinions, it can be concluded that random purposive sampling is one sampling technique where researchers determine sampling by setting special criteria that are in accordance with the research objectives and are taken part of the pupulation that meets the criteria randomly. So it is expected to answer research problems. In this study, every cadet in the population in each school will be selected who will take part in UKP, because it is considered that these cadets have completed all the learning processes in class.

Validity analysis

Like other assessments, correlations in test validity have predictors and criteria predictors of test validity are test item scores, while criteria are the total scores of that test. If variable I is pure discrete data or dichotomous data, while the second variable is in the form of continuous data (total test item score), the appropriate technique used in finding correlation between variables I and II is the biserial point correlation technique, where the number of correlation indices is symbolized rpbis (Sudijono, 2003: 185).

Biserial point correlation (rpbis) can be used to see phenomena in respondents' answer patterns, where a large and positive coefficient value will indicate that students can answer the question item well, while small business points indicate that the question item cannot be answered well by the respondent. *Henryson (197l) suggests that the r_{pbis} tells more about the predictive validity of the total test than does the biserial r, in that it tends to favor items of average difficulty.* It is further recommended that r_{is} an assessment of the combined relationship of item criteria and difficulty level.

According to Sudijono (2003: 187-189) there are several steps in analyzing the validity of test items:

a. Tabulate and count test items from numbers 1 to 30 in item analysis tabulation format.
b. Find the average of the sum of scores, Mt, using the formula:
$$Mt = \frac{\sum Xt}{N}$$
c. Find the number of standard deviations, SD_t, using the formula:
$$SD_t = \sqrt{\frac{\sum X_t^2}{N} - (\frac{\sqrt{X_t}}{N})^2}$$
d. Find or count Mp for a test item from 1 to 30
$$M_p = \frac{\text{Jumlah skor peserta tes yang menjawab dengan benar}}{\text{Skor peserta tes yang menjawab dengan benar}}$$
e. Calculates the rpu correlation coefficient of test items from 1 to 30, by using the formula:
$$r_{pbis} = \frac{M_p - M_t}{SD_t}\sqrt{\frac{p}{q}}$$

Jurnal Ekonomi
Volume 13, Number 02, 2024, DOI 10.54209/ekonomi.v13i02
ESSN 2721-9879 (Online)
https://ejournal.seaninstitute.or.id/index.php/Ekonomi

Where:

RPBIS = The biserial correlation points are the strength of the correlation between variable 1 and variable 2, which in this case is seen as the validity coefficient.

MP = The average score that the test taker has for a correctly answered test item.

Mt = average score of the total score.

Tsp = standard deviation of total score.

P = The proportion/number of test takers who answered the analyzed test items was incorrect

q = The proportion/number of test takers who answered the analyzed test items correctly.

In the interpretation of this provision db of (N-nr) is used, =86-2 = 84 sudijono, (2003: 190). 60 degrees of freedom are then checked with a table of "r" values of moment products. So the results are as follows:

1. at 5% significance level (r_t) = 0.213
2. poda 1% significance level (r_t) = 0.278

If the value (f_{pbis}) of the result correlation coefficient is greater (>) than the table of values (r_t) = 0. 213 to 5% level, the results obtained are significant, this means that the test items are considered valid. If the value (r_{pbis}) of the correlation coefficient is smaller(<) than the tablevalue (r_t) = 0.213 for the 5% level, then that level obtained is insignificant. This means that the test items are invalid.

Reliability analysis

The author determines the reliability of the direct test. There are two formulas proposed to carry out item analysis called K-R_{20} and K-R2l. In this study, the authors are better off using the K-R20 formula. According to Gay (2001:245) "*K-R20 is applicable to test whose items are scored dichotomously (0 or 1); thus, it useful with test items that are scored as true/false or right/wrong*. From the theory above states that the K-R_{20} formula is the right formula to use, this is because in this study the questions are assessed in the form of multiple choice and the dichotomy used 0 for wrong answers and 1 for correct answers then the author thinks that the K-R_{20} formula is the most appropriate choice. According to Sudijono (2003: 252) there are steps taken to determine the reliability of test items as follows:

a. Calculates multiple-choice test items from numbers 1 to 60 in item analysis tabulation format.
b. Find the total variant (St2) using the formula:

$$S_t^2 = \frac{\sum {X_t}^2 - \frac{(\sum X_t)^2}{N}}{N}$$

c. Make calculations to determine reliability by using formulas:

$$r_{xx} = \left(\frac{n}{(n-1)}\right)\left(\frac{S_t^2 - \sum pq}{S_t^2}\right)$$

Jurnal Ekonomi
Volume 13, Number 02, 2024, DOI 10.54209/ekonomi.v13i02
ESSN 2721-9879 (Online)
https://ejournal.seaninstitute.or.id/index.php/Ekonomi

Where:
Rxx = Test reliability coefficient
n = Number of test items available in the test
1 = constant angka
St2= total varian
p = Proportion/number of test takers who answered the analyzed items correctly
q = Proportion/number of test takers who answered the analyzed items incorrectly or Q = 1-P
∑pq = Sum of multiplications p by q (Ga, 2001:245)

Furthermore, Purwanto (1997: 139) states that the interpretation of the correlation coefficient (r) number is usually based on the following benchmark:

Table 1 Interpretation of the correlation coefficient number (r)

Interval	Interpretasi
0,00 – 0,20	Very Low
0,20 – 0,40	Low
0,40 – 0,70	Keep
0,70 – 0,90	Tall
0,90 – 1,00	Very High

Difficulty level

In these cases, the item should not be too easy or too difficult ; There must be a balance between the two. If the test is too easy, the learner will not be able to answer the test. On the other hand, if this test is too difficult, the cadets will be frustrated because they don't know how to answer. Analyzing the difficulty level will provide information about the difficulty level for each item. The difficulty level of the item is indicated by the number of cadets who answered the item correctly. Item difficulty index shows how easy or difficult the item is According to Surapranata (2004: 12) calculating the level of difficulty can use the following formula

$$P = \frac{B}{JS}$$

Where:
P = difficulty index
B = number of test takers in the total group who answered correctly
JS = total number of test takers

To find out the level of difficulty, the author uses the reference given by Sumarna Surapranata (2004: 21).

Table 2 Interpretation of difficulty levels

P	Interpretasi
P<0.30	Difficult
0.30<P<0.70	Keep
P>0.70	Easy

Jurnal Ekonomi
Volume 13, Number 02, 2024, DOI 10.54209/ekonomi.v13i02
ESSN 2721-9879 (Online)
https://ejournal.seaninstitute.or.id/index.php/Ekonomi

Differentiating power

Sebuah item yang bagus harus dapat membedakan antara para taruna yang lebih cerdas dengan yang kurang. *According to J. stanley Ahman and Marvin D. Glock (1967: 187), "The discrimination power of test its ability to differentiate between pupils who have achieved well (upper group) and those who have achieved poorly (the lower gro up)."*

From Ahman and Glock's statements, it can be concluded that if the test items are answered correctly by the upper group and answered incorrectly by the lower group of the test it is good because it can distinguish between the two. Sudijiono (2003: 389) states that to calculate the index of differentiating items using the following formula:

$$DP = \frac{P_a - P_b}{J}$$

Where:

DB = Discriminating power (index of differing items)

PA = number of test takers in the upper group who answered correctly

PB = number of test takers in the lower group who answered correctly

J = total number of test takers in the group

To find out the level of Discriminating Power or the ability to reduce which cadets are capable and which cadets are not capable, the author uses the reference of Anas Sudijono (1996: 218)

Table 3 Interpretation of the index of distinguishing items

Differentiating Index	Interpretasi
Less than 0.20	Less
0,20-0,40	Satisfactory
0,40-0,70	Good
0,70-1,00	Very good
Below 0.00	Very Lacking

If an item has a negative distinguishing power index (below 0.00), it must be omitted. If an item has a differentiating power index from 0.00 to 0.20, it must be revised. On the other hand, if the item has a differentiating power index of more than 0. 20 items must be retained. That means if the items have satisfactory, good and very good discriminating power they are accepted, but if the items have less index, the items must be revised.

Qualitative research also requires validity *and reliability as* well as quantitative research. In this regard, Fraenkel and Wallen (2005: 150) suggest that validity *refers* to the accuracy, meaningfulness and usefulness of conclusions drawn by researchers based on the data collected. While reliability is the power of data that can describe the authenticity and real consistency of any existing data based on time, place and situation. Based on the opinion of Gay, Mills and Ariasian (2009), validity is the degree to which qualitative data is measured accurately as it should be and reliability is the degree to which qualitative data is measured in a consistent manner as it should be.

From the results of the data validity trial, several question items were declared invalid. For the question of the Nautical department In the interpretation of this provision db of (N-nr) is used, =43-2 = 41. Then it is checked with a table of "r" values of the moment

Jurnal Ekonomi
Volume 13, Number 02, 2024, DOI 10.54209/ekonomi.v13i02
ESSN 2721-9879 (Online)
https://ejournal.seaninstitute.or.id/index.php/Ekonomi

product. So the result is 0.3008 at 5% significance level. After comparing the R value of pbis obtained from 75 items of questions tested, 64 questions were declared valid. Of the 64 questions, 60 questions will be used as research instruments.

For Engineering major questions In the interpretation of this provision db of (N-nr) is used, =26-2 = 24. Then it is checked with a table of "r" values of the moment product. So the result is 0.3882 at 5% significance level. After comparing the R value of pbis obtained from 75 items of questions praised, 63 questions were declared valid. Of the 63 questions, 60 questions will be used as research instruments.

The results of the reliability test found that both question packages for both nautical and engineering majors were at a "high" level. For the initial testing of the nautical department, a correlation coefficient value of 0.70009 was obtained while the final test result was 0.8289. For the initial test of the engineering department, a correlation coefficient value was obtained which was 0.7556 while the final test result was 0.6706

Credibility

Credibility is the determination of qualitative research results that are credible or trustworthy from the perspective of participants in the research. The steps taken to gain credibility are by: (1) *member-checking* , namely member numbers or codes on the data studied based on sub-focus, then making a form to check whether data that has not been analyzed is for analysis, (2) *discussions with colleagues*, by asking for help to evaluate the data analysis that has been done by researchers. The discussion was conducted to check whether all data had been analyzed (3) *The determination of the researcher,* through how to read repeatedly and recheck all the data that had been analyzed and match it with the recapitulation.

Transferability

Transparency is the degree of power of qualitative research that can be generalized or transferred to other contexts or settings. The transferability of a qualitative perspective is the responsibility of a researcher in generalizing. A researcher can increase the transferability of a research result by describing the context of the study with assumptions that are central to the study. In this study, transferability is obtained by describing the context in respondent data on research findings so as to get detailed research findings. Thus, research findings can be subject to study for further research.

Confirmability

Confirmability and objectivity are the results of the research obtained confirmed to others. In the context of this study. Confirmability is carried out by conducting expert tests who have competence and expertise in the same field of study

RESULTS AND DISCUSSION

Validity of the Maritime English Seafarer Skills Exam

Overall, the validity of the Maritime English Seaman Expertise Test questions majoring in Nautical Affairs is good. Of the total 60 questions tested, 43 questions were valid. In other words, 71.66% of the total questions were found to be valid while the remaining 28.33% or 17 questions were declared invalid. However, for the level of difficulty of the questions,

Jurnal Ekonomi
Volume 13, Number 02, 2024, DOI 10.54209/ekonomi.v13i02
ESSN 2721-9879 (Online)
https://ejournal.seaninstitute.or.id/index.php/Ekonomi

the majority of questions are still too difficult for students, namely 36 questions or 60% of the total while the remaining 24 questions or 40% of the questions are in the medium category and none of the questions are included in the easy category. In terms of student distinguishing ability, 53.34% or 32 questions are considered good, while 28 or 46.66% are considered less.

Overall, the validity of the Maritime English Seafarer Expertise Test questions of the Engineering Department is good. Of the total 60 questions tested, 41 questions were valid. In other words, 68.33% of the total questions were found valid while the remaining 31.67% or 19 questions were declared invalid. However, for the difficulty level of the questions, the majority of questions are still too difficult for students, namely 36 questions or 50% of the total while the remaining 20 questions or 40% of the questions are in the medium category and 10 questions or 16.67% are in the easy category. In terms of the ability to distinguish students, 56.66% or 34 questions are considered good, while 26 or 44.33% are considered less.

Table 4. Recapitulation of Question Validity

	SUM		PRESENTATION (%)	
	NAUTICAL	TEKNIKA	NAUTICAL	TEKNIKA
VALID	43	41	71,6	68,3
TDK VALID	17	19	28,4	37,7
SUM	60	60	100	100

Table 5. Recapitulation of the Difficulty of the question

	SUM		PRESENTATION (%)	
	NAUTICS	TECHNIQUE	NAUTICS	TECHNIQUE
DIFFICULT	36	30	60	50
KEEP	24	20	40	33,3
EASY	0	10	0	16,4
SUM	60	60	100	100

Reliability of Maritime English Seafarer Proficiency Test

A test is reliable if it is consistent and reliable. Reliability refers to the consistency of test scores. Also, it refers to the scope of the test that produces consistent results if different markers mark it.

The reliability of the Maritime English Seafarer Skill Test questions is calculated using the KR-20 formula. The calculation is done manually with Excel. The number of question items is 60. Giving interpretation of the test reliability coefficient (r using the benchmark in table 1, if r11 falls into the medium, high and very high categories then the question is declared reliable. However, if r11 falls into the low and very low categories, then the learning outcome test is declared unreliable.

The correlation coefficient value found in the question package of both nautical and engineering majors is at the "medium" level at 0.5968 for nautical majors and 0.67789 for

Jurnal Ekonomi
Volume 13, Number 02, 2024, DOI 10.54209/ekonomi.v13i02
ESSN 2721-9879 (Online)
https://ejournal.seaninstitute.or.id/index.php/Ekonomi

engineering majors located in the interval 0.40-0.70. Thus, it can be concluded that the Maritime English test tested has sufficient reliability.

CONCLUSION

Based on the research Analysis of the validity and reliability of the Maritime English seafarer expertise exam for grade XII students in schools whose marine education has been approved by the Directorate of Sea Transportation of the Ministry of Transportation, the following can be concluded: The validity of the sailor skills test questions in the Department of Nautical there are 43 question items or equivalent to 71.67% of the total items are said to be valid, so the test has good validity because the value of the correlation / coefficient relationship is greater (>) than the table (R_{table}) = 0.1071 for a significance level of 5%. Of these questions, there are 18 questions that have a good level of difficulty which are in the medium category. While the remaining 25 fall into the difficult category. There are 23 questions that have good ability in distinguishing students' ability levels, while the remaining 20 have poor abilities. As for the Engineering Department, there are 41 question items or equivalent to 68.33% of the total items said to be valid, so the test has good validity because the value of the correlation result / coefficient relationship is greater (>) than the table (R_{table}) = 0.266 for a significance level of 5%. There are 14 questions that have a good level of difficulty that fall into the medium category, the remaining 20 are in the difficult category and 7 questions are considered too easy for cadets. There are 23 questions that have good ability in distinguishing students' ability levels, while the remaining 20 have poor abilities. The reliability of the test kit for the Nautical department was found to be 0.5968, while from the results of the coefficient calculation, the correlation of the test device for the engineering department was 0.6789, the figure was in the interval 0.40 to 0.70 with moderate interpretation. Thus, from the calculation of the correlation coefficient of the test device, it can be concluded that the Maritime English sailor expertise test has a fairly good reliability.

REFERENCE

Allen, David. (1998). Assessing student learning. New York: Teacher College Press.

Al-Shumaimeri, Y. (1999). An Evaluation of an English language test. KSA: King Saud University, College of Education.

Anderson, P. and Morgan G. (2008). Developing tests and questionnaires for a national assessment of educational achievement. Washington: World Bank.

Arikunto, S. (2013). Research procedure: a practice approach. Jakarta: PT. Rineka Cipta

Bachman. L. F. (1990). Fundamental considerations in language testing. Oxford: Oxford University Press.

Brown, D. (2003). Language assessment: principles and classroom practice. San Francisco: Longman

Brown, D. (2006). Teaching by principles: An interactive approach to language pedagogy. San Francisco: Longman

Creswell, J. (2012). Educational research: Planning, conducting and evaluating quantitative and qualitative research. New York: Pearson Education

Cummins, Jim and Davison, Chris (2007). International handbook of English language teaching. New York: Springer Science + Business Media, LLC.

Festinger , David and friends (2005). Essentials of research design and methodology. Ottawa: John Wiley & Sons, Inc.

Fraenkel, Jack R. and Wallen, Norman E. (2007). How to design and evaluate research in education. New York: McGraw-Hill Companies, Inc.

Fulcher, Glenn and Davidson, Fred (2007). Language testing and assessment. London and New York: Routledge.

Gay L.R. and Airasian, Peter (2000). Educational Research competencies for analysis and application. New Jersey: Prentice-Hall, Inc

Gharbavi , Abdullah and Mousavi, Seyyed Ahmad. (2012) Do Language Proficiency Level Correspond to Language Strategy Adoption?. Khoramshahr Payam e-Noor University.

Graves, K. (2000). Designing language course. Oxford: Oxford University Press.

Haryudin, Acep. (2014). Validity and reliability of English summative tests. Indraprasta University PGRI.

Hawkey, Roger. (2006). Impact theory and practice, studies of the IELTS test ang Progetto lingue 2000. Cambridge: Cambridge University Press.

House, D. J. (2004). Seamanship techniques shipboard and marine operation. Oxford: Elsevier.

Lync, B.K. (2001). Rethinking assessment from a critical perspective. Language testing. 18 (4) 351-372.

Macalister. (2010). Language curriculum design. Oxford: Teacher College Press.

Richards, J.C. (2001). “Reflective teaching in TESOL teacher”. Issues in language teacher Education. Retrieved on 13th of August 2010 from: http//www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED370357.

Robertson and Nunn. (2013). The study of second language acquisition in the Asian contest paperback. The Asian EFL Journal.

Singh, Yogesh Kumar. (2006). Fundamental of Research Methodology and Statistics. New Delhi: New Age International (P) Limited, Publisher

Sudijono, Anas. (2003). Educational epvaluation authors. Jakarta: King Grafindo.

Sugiyono. (2013). Qualitative quantitative research methods and R&D. Bandung: Alfabeta

Wainer and Barun (1988). Test validity. American Education Research.