CHAPTER II

REVIEW OF RELATED LITERATURE

This chapter is going to discuss theories used in this research. Previous studies is going to be presented too in

A. Review of Related Literature

In this sub-chapter, the researcher is going to describe some theories and definition used in term of analyzing predictive validity of "selection test" of FLDI of Nurul Jadid.

1. Assessment, Test, and Evaluation

Before going to testing in language learning, we need to distinguish test, evaluation, and assessment.

a. Assessment

Assessment can be any kind of responses, questions, answers, and/or activities done by learners in ongoing or in unspecific time after or before teaching-learning process.⁹

⁹ H. Douglas Brown, *Language Assessment - Principles and Classroom Practices* (New York: Pearson Education ESL, 2004), 4.

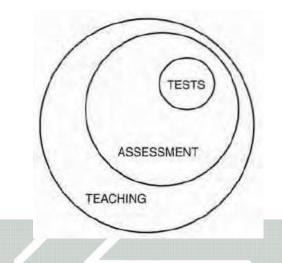


Figure 2.1 Teaching, Assessment, Test

In addition, assessment can be an informal or formal assessment, formative or summative assessment, and norm-referenced or criterion-referenced test.¹⁰

1) Formal – Informal Assessment

A well planned assessment, having certain measurement rubrics, and recording its result is included in a formal type of assessment. Any test, when it has been prepared or conducted must be a formal assessment because it has met the criteria mentioned. However, formal assessment does not only take form of test, it can be drama project, portfolio project or other assessment that does not meet a demand of test' criteria.

Informal assessment, of course, has opposite criteria showed by formal assessment. It can be any response such answers, comments,

¹⁰ Ibid, 5.

arguments, which came from the students; it also does not have clear measurement rubrics, and usually unplanned.

2) Formative – Summative Assessment

Formative assessment is scoring done in an ongoing teachinglearning process with forming students' comprehension, skill and knowledge of language as purpose of the scoring activity. Other purpose of this assessment is determining students' development in teaching-learning process. By doing this assessment, teacher can observe students' comprehension in ongoing teaching-learning process, decide whether his/her teaching methods are appropriate or not, and change the methods used if necessary. Teacher or administrator can apply formative assessment periodically (beginning, middle, or end) or continuously in ongoing teachinglearning. In this case, informal assessment is a form of formative assessment because formative assessment will not record the result of the assessment but rather "repair" methods used and lead students to the objective of the course.

An achievement or any other test usually conducted to summarize all results of learning process done by students is included in summative assessment. This assessment's purpose is to gather information for further analysis in order to determine students' comprehension and rank them according their achievement. The one with good result will be recommended to proceed to higher class or school level. But in the other way, if the result is not approaching indicators or goals set by teacher, so the student will stay in the same class or do remedial.

3) Norm-referenced – Criterion-referenced Test

Norm-referenced test is kind of test that the result of it is written in numerical and/or percentile rank record with placing students in rank order according its score as its purpose and having no feedback from the test-administrator. This test, usually take form of Scholastic Aptitude Test (SAT) or Test of English as Foreign Language (TOEFL).¹¹

In a same line, criterion-referenced test is also a test that recording the test-takers' result in numerical record with feedback from test-administrator (teacher). Here, the teacher needs extra effort and time to deliver or explain many things about the feedback given to the test-takers (students).

b. Test

According to Brown, test is method to accurately measure capability, comprehension, or qualification of a person in certain scope and domain.¹² A test purpose is to know who test-takers are, what they can do, how far they can do it, what benefit they have for the tester, and of course in a desired domain.

¹¹ H. Douglas Brown, *Language Assessment - Principles and Classroom Practices* (New York: Pearson Education ESL, 2004), 7.

¹² Ibid, 3.

According to McNamara, based on the purpose of test, there are two types of test, achievement test and proficiency test:¹³

1) Achievement Test

Achievement test connects with teaching-learning processes in the past which conducted in order to know test-taker's comprehension about the materials taught in the course and also to determine whether the objectives or goals or indicators set by the institution had been reached by test-taker. Final examination is an example of this test, because it examines test-taker's comprehension about material taught in the previous course.

Even though this test type is interested in testing test-taker's comprehension got from the past, but this test is often used as reference in determining test-taker's future.

In the other hand, achievement test cannot be a predictor in predicting test-taker's future performance if the syllabus of the course does not imply clearly that the test can do prediction.

2) Proficiency Test

This test is the opposite of achievement test which focuses on the past time; proficiency test focuses on the future of test-taker. This test type's purpose is to predict the future-like performance of test-taker without considering his/her educational background.

¹³ Tim Mcnamara, Language Testing (Oxford: Oxford University Press, 2000), 6.

In the practice of proficiency test, it has criterion that must be reached by test-taker so that his/her capabilities can be considered in standard level set and having bright future-like performance which needed by institution holding the proficiency test.

One of examples of this test is driving test to get driving license. In this test, test-taker needs to join several subtests such as traffic sign test, driving test itself, and driving theory test. Each subtest has its own criterion to be fulfilled by test-taker which determining whether the test-taker is feasible to have driving license.

c. Evaluation

Cross stated that evaluation is a determining process whether the objective/s of a course or teaching-learning process have been reached or achieved.¹⁴ Evaluation, according to Kusuma, must be systematic and continue in order to describe student's capability.¹⁵

Teacher often misplaces evaluation in specific time (beginning of the course, middle, and end), and the result of this misplacement is lack of information received from the misplaced evaluation. The lack of information that teacher got causes a massive misused methods which triggering mislead student. Kusuma stated that it is batter to

¹⁴ A Cross, *Home Economic Evaluation* (Colombus Ohio: A Bell & Howell Company, 2013), 5.

¹⁵ Mochtar Kusuma, *Evaluasi Pendidikan, Pengantar, Kompetensi dan Implementasi*, (Yogyakarta: Parama Ilmu, 2016), 3.

evaluate students or method used every single day by making and schedule of the evaluation systematically.¹⁶

In teaching, teacher must be aware of students' capability in comprehending materials taught in class, because each individual has different comprehension level. To know student's comprehension, teacher must evaluate his/her development since the beginning until the end of the course. The purpose of evaluation, according to Republic Indonesia's law number 20 year 2003 about national education system chapter 157 verse (1), to control the quality of national education as accountability form of caretaker of education to students, institutions, and education program.¹⁷

After those explanations above about test, assessment, and evaluation, we can draw a conclusion that each of them has similarity and dissimilarity. Test is a planned assessment which usually conducted in the end of a course or before based on its purpose, for example: final examination or admission test. Assessment is any planned and unplanned question or instruction for students to know his/her comprehension, including test. Evaluation is a systematic process to determine student's comprehension in the beginning, ongoing, or at the end of a course or teaching-learning process. So, the similarity of those terms is that the

¹⁶ *Ibid*, 3

¹⁷ Republic Indonesia's law about national education system, no.20 year 2003.

purpose of conducting those (test, assessment, evaluation) in determining the method used and student's comprehension.

The dissimilarity between them is about the procedure applied in conducting each of the three terms: test usually conducted before or after teaching-learning process, assessment knows no specific time in term of applying it (can be planned or unplanned), and evaluation also knows no time (can do evaluation in any moment) but need to be planned.

2. Language Testing

After knowing a brief explanation of test, evaluation, and assessment, we can guess what language testing is, and how important it is. Language testing is, of course, a test to measure test-takers' capability, comprehension, or performance in certain language. As skill, language demands good speaking, writing, listening and reading of a speaker, so language testing provides instruments to measure and grade speakers' ability in acquiring certain language.

This testing may take form of proficiency test, final examination, admission test and et cetera; tests known by worldwide as credible language testing to measure speakers' comprehension are IELTS and TOEFL, actually many kinds of good test but not as famous as those.

As a test, language testing means to measure test-takers ability in certain language; and followed by high scale of validity as one of important point in making good test.

Before going to the next sub-chapter, the researcher think that it would be necessary to sum up the explanation above about assessment, test, evaluation and language testing. As explained above, assessment, test, and evaluation has similarity and dissimilarity. The purpose of all three terms is the same, examining students' comprehension and teaching methods used, and the dissimilarity is about procedure used in applying the three terms. Language testing, in the same line also examines students' comprehension in learning or mastering language and also methods used in teaching language. The researcher assumes that all three terms might involve and influence language testing. In other words, language testing may be in form test, assessment, or evaluation; and covered those three terms.

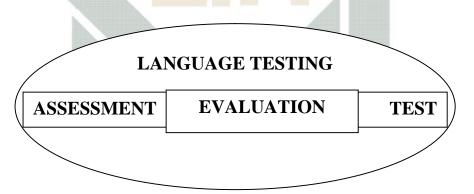


Figure 2.2 Language testing (evaluation, assessment, test)

3. Validity

One of important things in a research, evaluation or testing, and data measurement is validity. Joppe stated that:

"Validity determines whether the research truly measures that which it was intended to measure or how truthful the research results are. In other words, does the research instrument allow you to hit "the bull's eye" of your research object? Researchers generally determine validity by asking a series of questions, and will often look for the answers in the research of others."¹⁸ McNamara also stated, in his book *Language Testing*, about

validity:

"The purpose of validation in language testing is to ensure defensibility and fairness of interpretations based on test performance.... If no validation procedures are available, there is potential for unfairness and injustice."¹⁹

In other words, research is valid if it truly examines the object of the research and provides results from it. For example, a measurement tool, let's call it as speedometer, is valid when measures the speed of a car; and no longer valid when it measures the speed of snail. Even though the speedometer measures speed, it will be invalid if the speed is too slow (in this example is a snail's speed) or too fast. A test is in same condition as instrument of measurement mentioned, it can be classified as an invalid

¹⁸ Marion Joppe, "The Research Process" *University of Guelph-School of Hospitality, Food and Tourist Management* (<u>https://www.uoguelph.ca/hftm/research-process</u>, accessed on April 25, 2016)

¹⁹ Tim McNamara, *Language Testing* (Oxford: Oxford University Press, 2000), 48.

test when a kid takes the test that tended to be taken by adult; even it is possible for the kid to finish it perfectly.

In addition, Messick stated that validity of a test is all about suitability, significance and utility of the result of the test itself,²⁰ and others viewed validity as the interpretation's level of a test score.²¹ From this additional perspective, validity is all about the qualification of interpreted test score or research result with the support of evidence and theory.

In study of language testing, validity has many categories that help teacher, institution or any individual to make and hold fine and fair test. Brown categorized validity into five parts²²: a) content validity, b) construct validity, c) criterion validity, d) consequential validity, and e) face validity.

a. Content Validity

This validity concerns about positive relation between the content of a test and the test's purpose which means that the test's content must exactly examines and produces results that fulfilling the purpose of the test. For example, if a test wants to know the speed of football player's

²⁰ Samuel Messick – Ruth Linn. (Ed.), *Educational Measurement: Validity* (New York: McMillan, 1989), 13-103.

²¹ American Educational Research Association, American Psychological Association, and National Council on Measurment in Education, *Standards for Educational and Psychological test* (Washington, DC: American Educational Research Association, 1999) 9.

²² H. Douglas Brown, *Language Assessment - Principles and Classroom Practices* (New York: Pearson Education ESL, 2004), 22.

shot, the right examination of this kind of purpose is to ask the player to shoot a ball toward goal and measure the speed of his shot.

b. Criterion Validity

Every test must have a minimum score as standard result whether the test-takers pass the test or not; and to recognize that they truly achieve the standard of a test purpose, criterion validity comes to measure the test-takers' achievement by comparing the current result with other result of similar test and purpose to support the evidence that they deserve it. For example, an English admission test's minimum score is 80, test-takers must reach the score, and if they did, the examiner will compare the result with other result of similar test, purpose and standard to the current test to prove that the test-takers have truly have ability in reaching the minimum score or standard.

According to Hughes, criterion validity is used when there are two or more tests which executed simultaneously (one test is main test and others are assistance test).²³ Word "simultaneously" means that the tests are held in the same day and at the same time (even not exactly the same time). Brown divided criterion validity into two parts²⁴:

²³ Arthur Hughes, *Testing for Language Teacher* (Cambridge: Cambridge University Press, 2003),
27.

²⁴ H. Douglas Brown, *Language Assessment - Principles and Classroom Practices* (New York: Pearson Education ESL, 2004), 24.

1) Concurrent Validity

This validity comes to measure the result a test by comparing other similar test's result that has similar purpose with the first test in order to provide evidence and clarify the first's result.

2) Predictive Validity

Usually, a test that should be measured by this validity has purpose to eliminate test-takers and pick some who reached the standard determined by an institution when they wants to study or work in it. Hughes stated this validity is used to predict student or worker candidates' performance and achievement.²⁵

The difference between predictive validity and concurrent validity is the purpose of a test made. The purpose of a test with concurrent validity tends to see how high the achievement of a test-taker, and then provide evidence from another test to support the result of taken test. Whereas, the purpose of a test with predictive validity tends to predict test-takers performance in future time.

As the main topic of this research, predictive validity will be discussed in further explanation in different sub-chapter.

c. Construct Validity

²⁵ Arthur Hughes, *Testing for Language Teacher* (Cambridge: Cambridge University Press, 2003),
27.

Before going to the definition of this validity which is also called by construct-related evidence, it must be clear what construct is. Construct, according Brown, is a complex theory, hypothesis or model of bigger idea that explains phenomena in conception domain.²⁶

Construct validity is a validation measuring the theories or topics used in an assessment that are connected to specific language construct. For example, an institution wants to have English written test that requiring test-takers to write an opinion for about two thousand words. Several considerations in the test that help scoring analysis: word spelling, vocabularies used, grammatical accuracy, and idea of the text.

Usually, construct validity is used to measure large scale test such as TOEFL and IELTS. Fraenkel added content and criterion validity as part of construct validity since the scale of the test measured is large.²⁷

d. Consequential Validity

This part of validity is considered to watch over the consequences that may occur in a before and after test, and an ongoing test. The impact can appear in form of test-takers preparation, their socioeconomic aspect, and the way they learn and perform after the test²⁸. For example, McNamara stated that an appearance of a test may trigger

²⁶ H. Douglas Brown, *Language Assessment - Principles and Classroom Practices* (New York: Pearson Education ESL, 2004), 25.

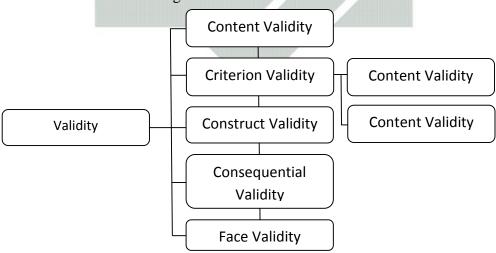
²⁷ Jack R. Fraenkel, et.al., *How to Design and Evaluate Research in Education* (New York: McGraw-Hill Education, 2011)

²⁸ H. Douglas Brown, *Language Assessment - Principles and Classroom Practices* (New York: Pearson Education ESL, 2004), 26.

at least a course that provides its student with information how to pass the test; some test-takers can afford to pay in order to join the course, but the rest probably cannot.²⁹ Other example is the impact of the test on test-takers' preparation that takes more their time for study than socialize with their neighbors or friends.

e. Face Validity

Face validity, the last part of validity, can be described as a testtakers' view on the test.³⁰ This validity is a subjective matter of the testtakers that not even expert can judge that the test has high face validity or not. Let me put it like this, for example, a test may have low face validity because the test-takers think that the test is misplaced. They believe that there is a better test that more appropriate to be tested because they were not prepared for the test, and they thought that it must me something else.



²⁹ McNamara, *Language Testing* (Oxford: Oxford University Press, 2000), 54.

³⁰ H. Douglas Brown, *Language Assessment - Principles and Classroom Practices* (New York: Pearson Education ESL, 2004), 26.

Figure 2.2 Brown's five validities

In short, validity is the pin point of any research or testing (in this case language testing) to make the research acceptable in the scholars' point of view. In language testing, validity has five types of validation: content validity, criterion validity, construct validity, consequential validity, and face validity. All of those types are connected each other even construct validity only used in large scale of proficiency test as measurement.

4. Predictive Validity

Predictive validity is a validation procedure taken to show effectiveness of a test in predicting future-like performance of people in certain activity. As explained in the first chapter, collage entrance test, driving test, and other admission batteries including language proficiency test are tests or assessments that need predictive validity approach to help officer or institutions to choose or pick person needed.

Ary et al stated that predictive validity is a relationship between a score measured and score from a criterion.³¹ Also, the researcher has mentioned above that predictive validity is part of criterion validity that measure prediction of test in indicating test-takers future performance. As part of validity, this predictive validity has big role in delineating institutions

³¹ Donald Ary, et al., *Introdcution To Research In Education*, (USA: Wadsworth, Cengage Learning, 2010), 229.

future success, in small scale, test-takers' delineation in performing their abilities to work or study as the institution demanded.

How to find whether the predictive validity of a test is high or not? to know it, we have to use correlation analysis. In examining the prediction of a test, Kusuma in his book, *Evaluasi Pendidikan, Pengantar, Kompetensi dan Implementasi*, stated that there are two important technical terms: predictor and criterion.³² Predictor is a test that its predictive validity is being measured; meanwhile, criterion is predicted performance of test-taker by the test that indicating success in a learning process, and usually takes some period time to be reached. Donald Ary et al. also stated that criterion only available in future time.³³

It must be kept in researcher, teacher, and institution minds that creating criterion is little bit tricky, because they must consider base rate of the test-taker. Base rate is test-taker's ability used to reach the criterion made; the higher test-taker's base rate means that the criterion is too easy, just the opposite, the lower test-taker's base rate means that the criterion is too hard to reach.³⁴

³² Mochtar Kusuma, *Evaluasi Pendidikan, Pengantar, Kompetensi dan Implementasi*, (Yogyakarta: Parama Ilmu, 2016), 50.

³³ Donald Ary, et al., *Introdcution To Research In Education*, (USA: Wadsworth, Cengage Learning, 2010), 229.

³⁴ Mochtar Kusuma, Evaluasi Pendidikan, Pengantar, Kompetensi dan Implementasi, (Yogyakarta: Parama Ilmu, 2016), 51

Kusuma also stated a procedure in validating predictive validity of a test:³⁵

- Make test item(s) which is appropriate with test-maker's goal
- Determine subject of the pilot study
- Indentify criterion to be reached
- Wait for criterion variable to appear
- Achieve the criterion
- Correlate two scores from the test and the criterion

So, the first step is to make our own test or use other test to be used as predictor; off course the test must be valid and reliable. The second step is to choose subject to become the population of this validation. The criteria of the subject must be clear so that the examiner can choose it wisely. The third step is creating criterion to be reached by the subject; this criterion also must be valid and reliable. The fourth is waiting for the criterion's appearance. In assessing predictive validity, we must wait for the criterion to appear because, unlike concurrent validity, it would take some time for the subject to reach the criterion as described by Ary, et al:³⁶

³⁵ Ibid, 52

 ³⁶ Donald Ary, et al., *Introduction To Research In Education*, (USA: Wadsworth, Cengage Learning, 2010), 229.

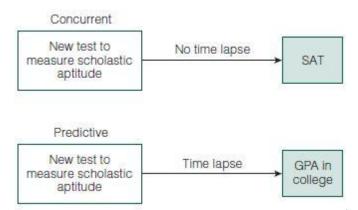


Figure 2.3 Description of Criterion Validity by Donald Ary et al

The fifth step is that the subject must reach the criterion. Example for the fourth and fifth step: if the examiner assessing predictive validity of an admission test (predictor) of a collage, he must wait for his subject (student in this case) to study for, at least, a semester, finish the examination of the semester (criteria or indicator) and wait for the exam's score (criterion). The last step is correlating the admission test and the examination of the semester. If the correlation between the two scores is high this means that the test has high predictive validity.³⁷ But, need to know that this study will pass several steps because the pertinent institution (FLDI) had conducted selection test (predictor) and first semester final examination (criterion). So, the researcher will directly correlate those variables (predictor and criterion).

There is one more thing to be put in caution, as Kusuma stated that general principle of test also is applied in examining predictive validity of

³⁷ Mochtar Kusuma, *Evaluasi Pendidikan, Pengantar, Kompetensi dan Implementasi*, (Yogyakarta: Parama Ilmu, 2016), 53.

a test: there is no test has perfect prediction, so the score of a test is also imperfect.³⁸

B. Review of Previous Studies

A research, in order to be accepted must have foundations underlying it; and one of the foundations is using previous studies for this research. Knowing previous study finding, we can understand what had been done and undone yet that can help us in doing next research.

In this case of predictive validity analysis or investigation, many institutions, groups, or individuals had conducted predictive validity analysis on admission or placement tests to evaluate how predictive the test was, and it will be briefly described in following sub-chapters along with differences of this research compared with previous studies mentioned below.

1. Previous Studies

Mary Kerstjens and Caryn Nery had been conducted an analysis on IELTS scores and students performance in academic domain.³⁹ This research's purpose was to know how high IELTS's prediction on students from non English country who learned in Australia in different major of studies by using Pearson correlation product moment correlating their IELTS scores and first semester academic performance in form of grade

³⁸ Ibid, 50.

³⁹ Mary Kerstjens - Caryn Nery, "Predictive Validity in the IELTS Test: A Study of the Relationship Between IELTS Scores and Students' Academic Performance", *IELTS Research Reports* Vol. 3, 2000, 85.

point average (GPA). They also used questionnaire for the students and interview for academic staff in technical and further education (TAFE).

The result of their research showed that correlation between IELTS scores and GPA of the students was positive even though from the research IELTS was not a significant predictor for academic performance since only reading skill that proven as critical skill involved in academic performance. The students and staff's responses to questionnaire and interview stated that they agreed that reading skill was most influence material in academic performance; and they considered higher IELTS scores as aid in helping students' learning process. The staff also added that many factors influence students academic performance whether inside or outside education domain.

Another study investigated IELTS as significant predictor also conducted by Patricia Dooey and Rhonda Oliver. They sought whether IELTS could be a credible predictor for college-students' success on academic performance in Curtin University Australia remembering that this international test was one of admission test predicated as "must-pass" test to study in English-speaking country, in this case is Australia.⁴⁰ College-students participated in this investigation were joining three different majors, business, engineering, and science.

⁴⁰ Patricia Dooey – Rhonda Oliver "An Investigation into the Predictive Validity of the IELTS TEST as an Indicator of Future Academic Success", *Prospect* Vol. 17 No. 1, April, 2002, 36.

This research which based on correlation method stated that IELTS (still) was insignificant indicator of academic success of the students. It was caused by the evidences in Dooey and Oliver research showing only a subtest, reading, that could be a significant predictor for all students from three differences majors. The result was the same with research conducted by Kerstjens and Nery which showed that IELTS was insignificant predictor for college-students success.⁴¹

Besides IELTS, TOEFL is also considered international English proficiency test and, usually, in the same time, and admission test also. In this case of predictive validity analysis, TOEFL had been investigated by Zhang Yan. He had conducted a research to know whether the test had high predictive validity or not on students' first term's GPA joining international exchange students between UBC (University of British Columbia) and Ritsumeikan University by using regression as analysis tool.⁴² He involved five variables: writing scores, speaking scores, gender, total TOEFL scores, and TOEFL sectional scores.

The predictive validity found in the research was medium as Zhang Yan elaborated his finding according to the variables.⁴³ Unsurprisingly,

⁴¹ Mary Kerstjens - Caryn Nery, "Predictive Validity in the IELTS Test: A Study of the Relationship between IELTS Scores and Students' Academic Performance", *IELTS Research Reports* Vol. 3, 2000, 105.

 ⁴² Zhang Yan, master thesis: "Predictive Validity of TOEFL Scores on First Term's GPA as the Criterion for International Exchange Students" (Vancouver: University of British Columbia, 1995), 21.

total TOEFL scores as predictor on the students' GPA was in mediocre or medium level. The result of this first variable found because TOEFL was not the only factor affected the students' academic performance such as dormitory or class environment. Sectional scores was in small level as indicator for the GPA since only section II (writing and grammar knowledge) showed to be in medium level because this skill was considered to be a useful skill in doing task since many assessments were made in form of productive skill (writen); the others two, section I (listening comprehension) and section III (reading skill) were small and negligible. Last, as predictor on the students' GPA, speaking scores was minor in predicting the GPA, writing scores had small prediction on the students' GPA.

There was a simple and temporary explanation about a confusing finding about two results of two writing's assessments, why writing knowledge's score in section II was medium and an independent writing score was small. Zhang Yan explained that in section II, it was writing knowledge not writing skill, remembering Japanese students' intermediate skill in writing, "they might need more basic writing knowledge" Zhang Yan stated,⁴⁴ and gender was a medium predictor on the students' GPA since widely known that non-language and academic domain could whether disturb or help students' success.

⁴⁴Zhang Yan, master thesis: "Predictive Validity of TOEFL Scores on First Term's GPA as the Criterion for International Exchange Students" (Vancouver: University of British Columbia, 1995), 71.

The next previous study mentioned here is a research conducted by Irwan Nuryana and Arief Fahmie about predictive validity of UPCM (Ujian Penerimaan Calon Mahasiswa) at 1999 and 2000 in UII (Universitas Islam Indonesia/Indonesia Islamic University), entitled "Validitas Prediktif Ujian Penerimaan Calon Mahasiswa Universitas Islam Indonesia terhadap Indeks Prestasi Kumulatif Mahasiswa". I included this research because it focused on predictive validity analysis even it had nothing to do with language proficiency. Their basic method in analyzing predictive validity of UPCM on students' GPA in their research was using Pearson correlation product moment as same as the first previous study mentioned; Kurniawan and Fahmie correlated UPCM scores and GPA of students of year 1999 and 2000 whom learned in different major of studies.⁴⁵ From the research, they found that UPCM had become insignificant predictor for students' GPA because there was a subtest in it that had negative correlation with the GPA, the lower students' grade the more GPA they could got.

A research by Renistri Mudela also used predictive validity as main topic for her study.⁴⁶ Her research analyzed the predictive validity of APM

⁴⁵ Irwan Nuryana Kurniawan – Arief Fahmie "Validitas Prediktif Ujian Penerimaan Calon Mahasiswa Universitas Islam Indonesia terhadap Indeks Prestasi Kumulatif Mahasiswa", *Fenomena* Vol. 3 No. 1, Maret 2005, 59.

⁴⁶Renisti Mudela, "Validitas Prediktif Skor Advanced Progressive Matrice (APM) dan Skor Skala Minat Pekerjaan Terhadap Prestasi Belajar Siswa: Studi Deskriptif Korelasional Skor Inteligensi (APM), Skor Skala Minat Terhadap Prestasi Belajar Siswa Kelas X dan XI, academic year 2013/2014" UPI Digital Repository, Indonesia University of Education, (<u>http://repository.upi.edu/13201/</u>, accessed on 23 August, 2016).

(Advanced Progressive Matrice) and SMP (*Skala Minat Pekerjaan* / job proclivity scale) scores towards first and second grade senior high school students' achievement. The population of this research was first and second grade senior high school students with details: 1st grade of SMA Negeri 2 Bandung (342 students), 2nd grade of SMA Negeri 5 Cimahi (344 students), 2nd grade of SMA Negeri 1 Marhagayu (396 students), 2nd grade of MA Negeri 1 Bandung (291 students), 1st grade of MA Persis Katapang (47) and SMK Negeri Katapang (377). Method used in this research was correlational descriptive with document analysis as the data collection technique, since the variables were scores of APM (variable X1), SMP (variable X2), and students' achievement (Y). In analyzing the data, Mudela used Pearson Correlation Product Moment.

After the statistical analysis, she concluded that the relationship among variables showing positive correlation but the significance is weak since variable X1 and X2 (must be) placed in positive row with plus (+) sign but not powerful enough to influence variable Y drastically.

Validitas Prediktif Tes Komptensi Berbahasa Indonesia Menurut Minat Belajar is a research conducted by Elvi Suzanti, the purpose of this research was to assess the predictive validity of Indonesian speaking competence test towards result of Ujian Nasional (UN) or Indonesian national examination section based on junior high school students interest in learning *Bahasa* Indonesia.⁴⁷ The competence test here was formed by two parts, Part A was spelling test, and Part B was reading test.

This research by Suzanti was conducted in state junior high school 51 and 25 Jakarta using survey and ex post facto methods. The sample of this study was 494 respondents from third grade of both schools. Suzanti wanted to know the predictive validity of the competence test by using regression (t) analysis (multiple regressions) for her research which was different with other previous studies in this chapter. In analyzing process, data was divided into two categories based on high or low interest in learning. To indentify high or low interest in learning, Suzanti used regression (t) too.

The result of the analysis showed that the predictive validity of *bahasa* Indonesia competence test was high, and the test could be used as predictor of UN (national examination) in junior high school level. The proof of this claim was that the competence test's predictive validity towards students who had high interest in learning Indonesian (1) and joined the UN was t_{score} 36.36 > t_{table} 01.645; the competence test's predictive validity towards students who had and students with low interest in learning Indonesian (2) and joined UN was t_{score} 02.484 > t_{table} 01.645.

⁴⁷ Suzanti, Elvi, "Validitas Prediktiv Tes Kompetensi Berbahasa Indonesia Menurut Minat Belajar" Badan Pengembangan dan Pembinaan Bahasa Kementrian Pendidikan dan Kebudayaan (<u>http://badanbahasa.kemdikbud.go.id/lamanbahasa/produk/1319</u>, accessed on 23 August, 2016).

Joana Francisca Reni Dwi Astuti had been examined predictive validity of national examination in Indonesia towards students' achievement entitled *Validitas Prediktif Ujian Nasional Terhadap Prestasi Belajar Pada Mahasiswa Universitas Sanata Dharma*.⁴⁸ This research was conducted to find an empirical proof that predictive validity of *Ujian Nasional* (UN) academic year 2004/2005 can predict future-like achievement or grade point average (GPA) of collage students.

Astuti used Pearson Correlation Product Moment as data analysis technique. She correlated not only UN generally but she analyzed all subtest of UN to find relationship between the scores of each section with the students' GPA. The subjects of this study were active students of Faculty of Psychology year 2005 that had joined UN academic year 2004/2005 from first semester to seventh semester (students on IV semester and VII semester for specific). Last, the cumulative scores of UN would be separated according to majors taken by student so see which majors had more potential success if a student is accepted in Faculty of Psychology.

From the research, Astuti found that there was no correlation between UN scores and students' GPA on fourth semester according to the result of statistical analysis (r=0.176, p=0.123), and seventh semester also proofed the same (r=0.188, p=0.099). Here was also the result of each subtest

⁴⁸ Joana Francisca Reni Dwi Astuti, thesis: "Vliditas Prediktif Ujian Nasional Terhadapt PRestasi Belajar Pada Mahasiswa Universitas Sanata Dharma" (Yogyakarta: Universitas Sanata DHarma, 2010), vii.

correlated with GPA, Indonesian (r=0.236, p=0.038), English (r=-0.011, p=0.925), math (r=0.078, p=0.652), and economy (r=0.462, p=0.002). Also, the result of correlation showed that collage students from social studies (*IPS*) major in senior high school had more potential success (r=0.355, p=0.023) than students from mathematical and natural sciences studies (*IPA*) major in senior high school (r=0.048, p=0.783).

The conclusion of her study showed that the predictive validity of national examination (*ujian nasional*/UN) was low and could not be used as one of admission battery for university, collage, or campus.

As last previous study, a study conducted by M Zakaria Adityawarman entitled "*Predictive Validity of the Collage Entrance English Test of UINSA Surabaya toward Students*' Achievement of *Teacher English Education Department (PBI)*" is a research to analyze predictive validity of admission test.⁴⁹ Adityawarman analyzed the predicitive validity of Collage Entrance English test of UINSA (State University of Islamic Studies Sunan Ampel) in predicting students' future performances by correlating its scores with the students' achievement in the intensive course of English Teacher Education Department. This study used Pearson *r* or Pearson product moment correlation; and had been concluded that the Collage Entrance Test had good predictive validity.⁵⁰

⁴⁹ M. Zakaria Adityawarman, *"Predictive Validity of the Collage Entrance English Test of UINSA Surabaya toward Students' Achievement of Teacher English Education Department (PBI)"* (Surabaya: State University of Islamic Studies Sunan Ampel, 2016), 3.

2. This Study

In this thesis, the researcher would like to do predictive validity research which the foundation of the study is clear (as presented in chapter one and two) but the object of this study is different with the previous studies mentioned above. Underlining differences in this analysis is that the object of its scores of high school around 13-17 years old students who decided to learn English in FLDI of Nurul Jadid. The environment of the Islamic-boarding school does help them to learn and practice intensively. Its environment is closed enough for outsider to "disturb" their learning processes, FLDI's rules are strict also, where every member of the institution must speak in English which help their development in mastering the language; it does not like the previous studies that used scores of college-students that took different majors in their universities which language was not their top priority, and non-language factor, environment,⁵¹ might be an obstacle since they learned in different countries.

⁵¹ Zhang Yan, master thesis: "*Predictive Validity of TOEFL Scores on First Term's GPA as the Criterion for International Exchange Students*" (Vancouver: University of British Columbia, 1995), 74.