

Validity of Test: English Lecturers perceptions of Test Validity in Intensive Reading Course' Final Exam at Universitas Prima Indonesia in academic year 2022/2023

M Meisuri¹, Kresna Ningsi Manik²

¹Universitas Negeri Medan and meisuri@yahoo.com

²Universitas Prima Indonesia and kresnaningsihmanik@unprimdn.ac.id

ABSTRACT

This study was inspired from the researcher' effort in order to enrich knowledge related to language testing' component, one of component is validity and also curiosity to investigate the validity of Intensive Reading Final Test given by English lecturers in University of Prima Indonesia (UNPRI) in academic year 2022/2023. The research addresses two key questions: the lecturers' perception of test validity in Intensive Reading Final Test in English Education Study Program of Teacher Training and Education Faculty of Universitas Prima Indonesia (UNPRI) and the reason of English lecturers in English Education Study Program of Teacher Training and Education Faculty of Universitas Prima Indonesia (UNPRI) Medan have such perceptions. This study applies qualitative research design in which involving 10 English lecturers at University of Prima Indonesia (UNPRI) as the respondents of this research, analyzing the perception of English lecturers related to validity of test in Intensive Reading Subject Final Exam. Findings shows that lecturers show their positive response to test validity with the information as follows: ALWAYS carry out an analysis of the question items before carrying out the Final Test" is in the percentage of 80%, "FREQUENTLY carry out crosschecks between basic competencies and indicators determined in the questions" is in the percentage of 100 %, "ALWAYS (60 %) and FREQUENTLY (20%) Carry out reliability/consistency measurements (reliability)", "ALWAYS (60%) Providing feedback on test results given to students. Lecturers take such action in preparing final semester exam questions for Intensive reading courses are: Analyzing test items is crucial in creating effective assessments, Feedback on tests is crucial for students' learning and development, crosschecking between indicators and test items to make sure that educational tests are valid and reliable, the reliability of question items critical for producing accurate, consistent, and trustworthy assessment results. It supports the meaningful interpretation of scores and allows for valid comparisons across individuals or groups. A way to make a high-test validity, it can be done in some ways such as compile the test grid before writing and selecting items to be tested, and making or selecting items on the test based on standard competence, basic competence, learning objectives that have been formulated in the syllabus

Keywords: Language Testing, Validity, Reliability, Evaluation, Assessment

1. INTRODUCTION

Assessment plays a pivotal role in the teaching and learning process, serving as a means for educators to gauge the effectiveness of their methods and curricula, as well as to evaluate students' grasp of the material. Indeed, the link between evaluation and teaching is inseparable, as understanding the needs of students and identifying successful teaching strategies relies heavily on the assessment process. Assessments are essential for discerning students' proficiency in a particular subject, and both subjective and objective tests can be employed for this purpose. The overarching aim of education is to comprehend the intricacies of the learning process, which is comprised of three integral elements: input, transformation, and output—all of which necessitate evaluation.

To gauge students' progress in the classroom, teachers commonly administer two types of tests: formative tests and summative tests. Formative tests are conducted earlier in the semester, preceding the summative tests held at the semester's end. Through these assessments, instructors can measure students' achievements and assess the extent to which instructional objectives have

been met. Crafting effective tests, however, is a challenging task for educators, given the criteria of validity, reliability, and practicality that must be satisfied. As Douglas Brown asserts, a well-designed test should adhere to these axiomatic criteria, ensuring that it can be administered within specified constraints, is dependable, and accurately measures its intended outcomes. This implies that a good test must fulfill multiple interrelated criteria when being created by teachers.

This study focuses on exploring lecturers' perspectives regarding the Test Validity of the Final Exam in the Intensive Reading subject. The importance lies in ensuring that tests crafted by lecturers align with the syllabus and effectively measure the intended content. The absence of validity undermines confidence in drawing inferences and conclusions from the test results. The significance of validity is evident in its role, emphasizing that tests must accurately measure a student's competence or ability. Topic "validity" is actually a very familiar topic in its correlation to test, and particularly in connection with this study, so validity discussed namely validity on language testing. Validity is very essential component in a language test as it repairs measuring instruments by inspecting the items on the measuring instrument and discarding, repairing, or replacing items that are not good or do not meet the requirements.

The efficacy of a test, whether it is deemed good or not, is contingent on the lecturer's skill in test construction, particularly in the cognitive domain, a test is expected to provide precise information to assess students' performance on learned material. At times, tests are created without due consideration for their content, leading to a lack of validity in measuring students' achievement levels. This issue is compounded by the prevalent use of traditional assessment methods by many lecturers in higher education. Regarding to topic of test validity, here some relevant studies conducted before as reference for the betterment of this study completion.

Validity of Test Instruments [1], published under License by IOP Publishing Ltd, Journal of Physics: Conference Series, Volume 1364, 2018 1st Workshop on Engineering, Education, Applied Sciences, and Technology 10 August 2018, Surabaya, Indonesia. This study discusses the definition of validity, types of validity, testing test validity, testing test validity rationally, construct validity, content validity, testing test validity empirically, forecast validity, comparative validity, question validity, and validity testing calculations. items of measurement results. The results of the research from the distribution of questionnaires to 20 (twenty) respondents were tested for validity using the quantitative data analysis method, and it was discovered that 7 (seven) items were valid and 3 (three) were invalid.

An Analysis on The Validity Test Question with The Exam Content Outline in First Semester for The Xi Grade Student of SMA N 3, [2]. This study is intended to examine the validity of the final test of class XI senior high school students in the first semester. Validity itself is significant in testing. Some teachers in schools still have not reviewed the validity of the test. Research has shown that fact remains unhealthy. To get the results, the researchers used data from analyzing school documents. To achieve this, it is hoped that it can help teachers better structure tests.

An Analysis of The Validity of English Test Made by The Teacher in SMA Negeri 1 [3]. The purpose of this study is to determine whether the English test administered by the teacher is content valid. The researcher compares the English test to the syllabus, lesson plan, and material taught to determine the content validity. This study employs a qualitative descriptive method, with document analysis in the form of a mid-semester test and a teacher interview. This study gathered information

from an English teacher who teaches seventh grade at SMP Negeri 1 Bontomarannu. The results obtained after analyzing the teacher's test, a total of 15 questions were created by the teacher who teaches in seventh grade in the mid-semester test. There are 13 questions that are classified based on their content validity, namely in accordance with the syllabus, lesson plan, and material covered. While there are two questions that are classified as not conforming to the content validity.

Based on those studies above, there has not any study investigating the lecturers' perception towards final test validity in intensive reading subject and how and why the lecturers as test as testers, evaluators and implementers of learning on campus implement some strategies/approaches in construction test in Intensive Reading Final Exam.

To understand the lecturers' perceptions of test validity in the Fourth Semester Intensive Reading subject at the University of Prima Indonesia (UNPRI) during the academic year 2022/2023, this research is conducted. The researcher, being a permanent lecturer, aims to leverage this position to facilitate data collection and gather necessary information to enhance the quality of the study. Based on the background above, the researcher formulates the research question as follow:

- 1) How is the lecturers' perception of test validity in Intensive Reading Final Test in English Education Study Program of Teacher Training and Education Faculty of Universitas Prima Indonesia (UNPRI) Medan?
- 2) Why do English lecturers in English Education Study Program of Teacher Training and Education Faculty of Universitas Prima Indonesia (UNPRI) Medan have such perceptions?

Referring the problem statement above, the objectives of the research are formulated as follow: 1). To describe lecturers' perception of test validity in Intensive Reading Final Test in English Education Study Program of Teacher Training and Education Faculty of Universitas Prima Indonesia (UNPRI) Medan. 2). To describe the reason of English lecturers in English Education Study Program of Teacher Training and Education Faculty of Universitas Prima Indonesia (UNPRI) Medan have such perceptions. This research is expected to give useful information about the test validity in Intensive Reading Subject Final exam items in English Education Study Program of Teacher Training and Education Faculty at Universitas Prima Indonesia (UNPRI) Medan in academic year 2022/2023.

Based on the information provided, the lecturers of the English Education Study Program of the Teacher Training and Education Faculty of Universitas Prima Indonesia (UNPRI) Medan can create a good test to evaluate the students, as well as make a concerted effort to measure the students' achievement level and the end of evaluation. So, lecturers truly understand whether or not they are successful in the teaching-learning process. Furthermore, the research might be used as a resource for another researcher who is interested in doing a comparable case study.

2. LITERATURE REVIEW

Evaluation is a one of step that cannot be separated in teaching learning process. Evaluation is the process of describing and refining information that is useful for preparing alternatives [4]. Evaluation can complement the meaning of tests and measurements and can also mean outside of between. Evaluation results can provide answers to professionals, reliable both in quantitative and qualitative data. According to [5] as quoted from Djiwandono, "evaluation is a process collect information about the teaching learning process

as a basic to make a decision". Evaluation is an important activity for the teacher and students. Although the focus in the evaluation phase is on the student's self evaluation, teacher ore also engaged in evaluation activities. In order to know how well the result of teaching and learning process, a teacher must evaluate it. By evaluation the teachers can collect information or can have picture describing how well the teaching learning activity succeeded.

In summary, the researcher concludes that evaluation is a crucial final step in the teaching learning process. Lecturers utilize evaluation to gather information about the learning process and student achievements, aiding them in deciding the success of the learning process. Analyzing the results is essential for a lecturer to fully understand the teaching and learning process. Evaluation allows instructors to gather facts and form opinions on the success of the teaching and learning activity. Tests, one of the evaluation instruments, have definitions proposed by specialists based on distinctive perspectives and thinking methods as follows:

A test is a way to assess someone's skill or understanding in a specific subject. It involves a procedural approach with the objective of measuring a person's ability, knowledge, or competence within a certain domain. Brown emphasizes that testing can reveal people's knowledge. According to J. B. Heaton, a test is aligned with the teaching that has occurred, and sometimes teaching is heavily influenced by the test [6]. Andrew Harrison views tests as necessary but often seen as an unpleasant imposition from outside the classroom, serving to establish standards but using up valuable class time.

In essence, a test can be defined as a critical evaluation procedure, a method for determining the presence, quality, or truth of something. It typically consists of a series of questions, problems, or physical responses designed to assess knowledge, intelligence, or ability.

Fathony, as quoted by Thoha, describes tests as measurement tools organized by questions, commands, and directions to elicit responses or answers in line with the given instructions. In conclusion, a test is a systematic and objective method of measuring a person's ability, organized by questions, commands, and directions for data collection.

Types of Tests

There are numerous types of tests used to assess student performance. In his book, [7] described four types of tests. They are as follows:

Placement Test

The purpose of a placement test is to sort new students into teaching groups so that they can begin a course at roughly the same level as the other students in the class. These tests grammar and vocabulary knowledge, as well as students' productive and receptive skills, are typically based on syllabuses and materials that students will follow and use once their level has been determined. A placement test typically includes a sampling of curriculum material (that is, it has content validity) and thus provides an indication of the point at which the student will find a level or class to be neither too easy nor too difficult, but appropriately challenging.

Diagnostic Test

A diagnostic test is intended to identify a specific aspect of a language. A diagnostic test in pronunciation may be used to determine which phonological features of English are difficult for a learner and should thus be included in a curriculum. While placement tests are intended to demonstrate how well a student's English is in relation to a previously agreed-upon system of levels, diagnostic tests can be used to reveal learner difficulties, gaps in knowledge, and skill deficiencies during a course. Diagnostic tests and placement tests serve different purposes. Diagnostic tests are used to motivate students to complete remedial work, whereas placement tests are used to group them.

Achievement Test

An achievement test is directly related to classroom lessons, units, or even the entire curriculum. Achievement tests are limited to specific curriculum material covered in a specific time frame and are given after a course has covered the objectives in question. Achievement tests can serve as indicators of features that a student needs to work on in the feature, but their primary role is to determine course objective acquisition at the end of a period of instruction. In an educational system, achievement test scores are frequently used to determine the level of instruction for which a student is prepared. High achievement levels typically indicate mastery of grade-level material as well as readiness for advanced instruction.

Proficiency Test

Proficiency test is to assess the student's ability to apply what he was learnt in actual situations. An important element in proficiency testing is to assess in some way the student's ability to repair breakdowns in communication, by asking for a repetition or an explanation, for example, or by apologizing and rephrasing what he has just tried to say.

Table 2.1 A framework to considered for each of the four categories

Category	Content	Purpose	Considerations
Placement	General reference forward to future learning	Grouping	Speed of results variety of tests interview
Diagnostic	Detailed reference back to classwork	Motivation remedial work	Short-term objectives new example of the material taught
Achievement	General reference back to course	Certification comparison with others at the same stage	Decision about sampling similar material to that taught in new contexts
Proficiency	Specific purposes reference forward to particular applications of language acquired	Evidence of ability to use language in practical situations	Definition of operational needs authenticity context strategies for coping.

The Characteristics of a Good Test Validity

The validity of a test is the extent to which it measures only what it is supposed to measure. The test must aim to provide an accurate measure of the skill being measured. Its meaning refers to the extent to which the outcomes of an evaluation procedure serve the specific purposes for which they are intended. For instance, if a test is intended to assess oral comprehension, it should not be used to assess another skill, such as reading comprehension. Brown divides validation in three types of validation which important in teaching learning proses includes: content validity, face validity and construct validity.

Content Validity

According to Andrew Harrison, content validity is concerned with what goes into the test. The content of the test should be decided by considering the purposes of the assessment, and then drawn up as a list known as a content specification. The content specification is important because it ensures as far as possible that the test reflects all the areas to be assessed in suitable proportions and also because it represents a balanced sample, without bias towards the kinds of item which are easiest to write or towards the test material which happens to be available.

According to [8] content validity is the validity of the assessment instrument in terms of the content of teaching material covered by the assessment instrument. A way to make a high content validity test can be done in a way; compile the test grid before writing and selecting items to be tested, and making or selecting items on the test based on standard competence, basic competence, learning objectives that have been formulated. A test is said to have content validity if a test actually samples the class of conditions, that is, the universe of subject matter about conclusions are to be drawn.

The test actually involves the test in a sample of the behavior that is being measured. What is the importance of content validity? First, the greater test's content validity, the more likely it is to be an accurate measured of what it is supposed to measure. Secondly, such a test is likely to have a harmful backwash effect. Furthermore, to conclude that content validity is deal with is the content of test item to be tested is reflected of the curriculum or not.

Face Validity

An important face of consequential validity is the extent to which student view the assessment as fair, relevant, and useful for improving learning or what is popularly known as face validity. Face validity refers to the degree to which a test looks rights, and appears to measure the knowledge or abilities it claims to measure, based on the subjective judgment of the examinees who take it, the administrative personnel who decide on its use, and other psychometrically unsophisticated observers.

Construct validity

A test, part of test, or a resting technique is said to have construct validity if it can be demonstrated that it measures just the ability which it supposed to measure. A construct is any theory, hypothesis, or model that attempts to explain observed phenomena in our universe of perceptions. For most of the tests that you administer as a classroom teacher, a formal construct validation procedure may seem a daunting prospect. Construct validity is a major issue in validating large-scale standardized tests of proficiency.

Reliability

A trustworthy test is consistent and dependable. If you give a test to some students or a group of students on two different occasions, the results should be similar. The consistency of a test's reliability. It would be pointless to try to time people's waits with a piece of elastic. A tape measure that stays the same length all the time is required so that one person's waist is known to be eighty-one centimeters and another's waist is known to be ninety-one centimeters.

Reliability measures estimate how much variation we can expect under various conditions. There are several methods for estimating test reliability. The three fundamental methods and the information each provides.

The first method is the test-retest method, which involves administering the same test to the same group twice with any time interval between tests ranging from several minutes to several years. The second method is equivalent to method, which is also known as alternate forms or parallel forms. This method involves administering two equivalent forms of a test to the same group during the same testing session. The final method is the internal consistency method, which requires only one test administration. Because this method does not require two tests, memory and practice effects are not involved. The consistency of scoring by two or more scorers is referred to as scorer reliability. One would not expect high scorer reliability if very subjective techniques were used in test scoring.

3. METHODS

In this research, the researcher employs descriptive qualitative research. Qualitative refers to get the understanding deeply about the way things are, why they are that way and how participant perceive them. The form of descriptive qualitative research is used to analysis the data. It is descriptive because the objectives of this study are to finding the information as many as possible [9]. The researcher distributes questionnaire to the respondents, displays the questionnaire results in Microsoft Excel from each respondent along with question points and answer choices.

This research focused on the lecturers in English Education Study Program of Teacher Training and Education Faculty of Universitas Prima Indonesia (UNPRI) Medan in academic year 2022/2023. The researcher choses 10 lecturers as the subject of the research. The location of the research taken place at Universitas Prima Indonesia (UNPRI). This research focused in English Education Study Program of Universitas Prima Indonesia (UNPRI) in academic year 2022/2023.

The location of Universitas Prima Indonesia (UNPRI) is on Jalan Sampul, No.3, Sei Putih Bar., Kec. Medan Petisah, Kota Medan, Sumatera Utara 20118. The researchers initially employed a perception survey consisting of seven questions (refer to Table 4.1) to gather data from all participants. The collected data will be thoroughly examined and discussed in the subsequent sections on Data Analysis, Findings, and Discussion. Subsequently, the researchers conducted interviews with lecturers using an open-ended script (refer to Table 4.2). The responses from these interviews were transcribed and subjected to analysis. In summary, data collection involved distributing questionnaire scripts to both lecturers and students for completion and interviewing English lecturers as per the data requirements.

The researcher analyzed the data from interviews with lecturers using a systematic approach, including transcribing recorded information, categorizing data based on test validity, and

verifying accuracy through participant confirmation. Conclusions were drawn to understand why lecturers chose these analytical steps. The research encompassed various stages such as data collection, data reduction, and data display. Through analysis, the reasons behind lecturers' choices in constructing test items became evident. The researcher demonstrated that the valuable contributions of lecturers' perceptions and insights played a crucial role in constructing a well-designed test. The findings highlighted how these contributions positively influenced the overall quality of the final test.

4. RESULTS AND DISCUSSION

The data in this research is the perception of lecturers of final test validity in the intensive reading subject was collected through observation, questionnaire' distribution and interview session. The data was analyzed by identifying and classifying by using theory as data analysis' contemporary category. The data was presented by the researchers as an orderly collection of information that allows for conclusion-making and action-taking [10]. The final action was making conclusions, which was accomplished by deciding what the data meant or by identifying patterns or regularities after reading the matrices or displays.

Lecturers' Perception of Final Test Validity in Intensive Reading Subject

The total number of respondents is 10 lecturers from English Education Study Program. After distributing the questionnaire to the respondents, then it is got the result as follows:

1) Carry out item analysis of questions before the implementation of the National Final Examination

The Most dominant result is "SELALU melakukan analisis item butir soal sebelum pelaksanaan UAS (ALWAYS carry out an analysis of the question items before carrying out the Final Test" is in the percentage of 80% and in the same percentage, followed by "SERING Melakukan crosscheck antara kompetensi dasar yang terdapat dalam silabus dengan item tes". "SELALU melakukan analisis item butir soal sebelum pelaksanaan UAS" (ALWAYS carry out an analysis of the question items before carrying out the Final Test" is in the percentage of 80%. This shows that before constructing the test questions for Intensive Reading Final Test, the lecturer always carried out an analysis of the question items.

2) Carrying out crosschecks between indicators and test items

Next, "SERING Melakukan crosscheck antara kompetensi dasar dan indikator yang ditentukan terhadap soal" is in the percentage of 100 %, this show that the lecturer is very attentive to verify or compare test items with basic competency and indicators in the syllabus to ensure accuracy of the test.

3) Measure the reliability/consistency of the question items

"Melakukan Pengukuran keandalan/keajegan (reliability)", Selalu 60 %, sering 20 % and jarang 20%. This shows that lectures are more which are always attentive to carry out reliability/consistency measurements (reliability) of the test items before the Final Test execution than lecturers who often and rarely do the same thing.

4) Measuring the differentiating power of a question (the ability of a question to differentiate between students who are smart (high ability) and students who are less intelligent (low ability)

“Mengukur daya beda soal (mengukur kemampuan siswa antara yang pintar dengan yang kurang pintar = jarang 60%, sering 20 %, tidak pernah 20%. Here, Lecturers do not really focus on the ability of a question to differentiate between students who are smart (high ability) and students who are less smart (low ability) in the test.

5) Providing feedback on test results given to students

“Pemberian feedback terhadap hasil tes yang diberikan kepada siswa, sering 60%, jarang 40%. Here, lecturers tend to provide feedback to students after the test results have been carried out. This case functions as an encouragement for students to act to make improvements, corrections for better results. By providing feedback, tests can also function as a motivator for students to learn better.

6) Measuring the level of difficulty of the Intensive Reading questions given to students

“Melakukan Pengukuran tingkat kesulitan soal Intensive reading yang diberikan kepada siswa”, jarang 60 %, sering 20%, and tidak pernah 20 %. This shows that the level of difficulty contained in the questions is rarely paid attention to by lecturers in intensive reading courses when giving questions.

Table 4.1. Persepsi Dosen Terhadap Analisis Butir Soal UAS Mata Kuliah Intensive Reading

PERSEPSI DOSEN TERHADAP ANALISIS BUTIR ITEM SOAL UAS MATA KULIAH INTENSIVE READING

No	ASPEK YANG DITANYAKAN	SELALU (SL)	SERING (SR)	JARANG (JR)	TIDAK PERNAH (TP)
		4	3	2	1
1	Melakukan analisis item butir soal sebelum pelaksanaan UAS				
2	Pemberian feedback terhadap hasil tes yang diberikan kepada siswa				
3	Pengukuran tingkat kesulitan butir soal Intensive reading yang diberikan kepada siswa				
4	Mengukur daya beda soal (kemampuan siswa untuk membedakan antara siswa yang pandai (berkemampuan tinggi) dengan siswa yang kurang pintar (berkemampuan rendah).				
5	Membuat tes dengan mempertimbangkan/melihat kompetensi dasar dalam silabus				
6	Melakukan cross check antara indikator dengan butir soal tes				
7	Pengukuran keandalan/keajegan (reliability) butir soal				

The result of lecturers' perception of Intensive Reading Final Test content validity is presented in chart column as follows:

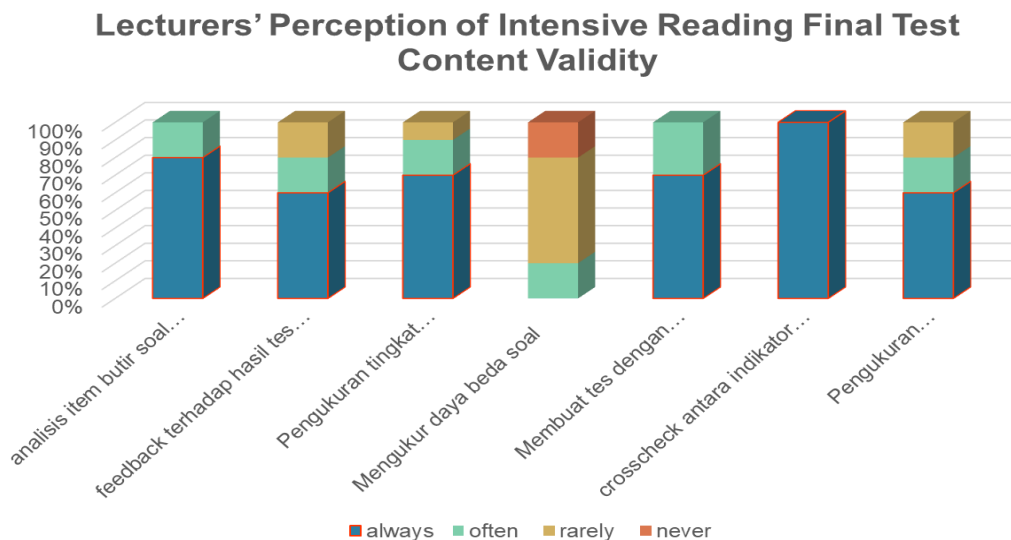


Chart 4.1. Lecturers' Perception of Intensive Reading Final Test Content Validity

The reasons why lecturers take such action in preparing final semester exam questions for Intensive reading courses are:

1. Analyzing test items is crucial in creating effective assessments.

This involves ensuring that the items are valid (measuring what they should), reliable (producing consistent results), and fair. Different types of validity, like content and construct validity, are considered. Reliability looks at score consistency. Difficulty level analysis ensures a mix of easy and hard questions, expressed as the percentage of correct answers. The discrimination index shows how well an item distinguishes between high and low performers. Item-Response Theory (IRT) is a statistical approach used to understand how individuals' responses relate to their abilities.

2. Feedback on tests is crucial for students' learning and development.

It helps them identify strengths and weaknesses, learn from mistakes, and clarify misunderstandings. Positive feedback motivates students, encourages goal setting, and fosters metacognitive thinking. It also initiates a dialogue between students and teachers, guides further study, promotes self-regulation, and transforms assessments into learning opportunities. Additionally, feedback enhances confidence, meets individual needs, and encourages reflective practice for improved critical thinking skills.

3. Crosschecking between indicators and test items

This activity is done to make sure that educational tests are valid and reliable. It helps confirm that test questions accurately measure what they are intended to assess, align with educational goals, and are fair for all individuals. This process also identifies any issues with test items, allowing for improvements and ensuring high-quality assessments. In summary, crosschecking enhances the overall effectiveness and accuracy of educational tests.

4. The reliability of question items

This is critical for producing accurate, consistent, and trustworthy assessment results. It supports the meaningful interpretation of scores and allows for valid comparisons across individuals or groups. Reliable tests are essential in educational, clinical, and research settings where accurate measurement is paramount.

Interview Question' points of Lecturers' Perception toward Test Validity in Intensive Reading subject Final Exam, described as follows:

Table 4.1. Interview Question' points of Lecturers' Perception toward Test Validity in Intensive Reading subject Final Exam

No	Interview Question' points of Lecturers' Perception toward Test Validity in Intensive Reading subject Final Exam
1	In constructing test item, do you think that Analyzing test items is crucial? Why?
2	In constructing test items, why is Crosschecking between indicators, basic competence and test items important?
3	After completion of final test, do you think that giving feedback on tests is crucial for students' learning and development? Why?
4	In your opinion, what is the importance of question items' reliability?
5	Before giving the students exam, do you measuring the differentiating power of a question (the ability of a question to differentiate between students who are smart (high ability) and students who are less intelligent (low ability), why?
6	Before giving the students exam, do you measuring the level of difficulty of the Intensive Reading questions given to students, why?

CONCLUSION

After analyzing and interpreting the acquired data in the previous chapter, this chapter will present the conclusion of the study and the suggestion. This is the last part of the research about analysis of content validity of Intensive Reading final test items in English Education Study Program of Teacher Training and Education Faculty of Universitas Prima Indonesia (UNPRI) Medan in academic year 2022/2023.

A way to make a high-test validity, it can be done in some ways such as compile the test grid before writing and selecting items to be tested, and making or selecting items on the test based on standard competence, basic competence, learning objectives that have been formulated in the syllabus. From the data that is got by researcher, the test validity of Intensive Reading subject in Final Test according to lecturers' perception in English Education study program at Universitas Prima Indonesia (UNPRI) academic year 2022/2023 gains 4 categories (4 questions) as the highest percentage from 7 questionnaire items examined to the lecturers, they are:

1. Analyzing test items is crucial in creating effective assessment 80%,
2. Giving Feedback on tests is crucial for students' learning and development, 60%
3. Crosschecking between indicators and test items, 100%
4. Measuring the reliability of question items, 60%

SUGGESTION

After the researcher obtains the result of the data analysis, the researcher would like to propose some suggestion for the consideration which are significant for the lecturers, teacher, other/future researchers.

1. It is hoped that lecturers of Intensive Reading course, as testers, evaluators and implementers of learning on campus, in preparing the questions, first pay attention to the quality of the question items, so that the questions produced truly meet the measurement

standards. This needs to be done in an effort to improve the quality of the final semester exam questions.

2. It is hoped that this research can be used for similar research by other researchers for conducting more researches about analysis of questions' items on an exam test to pay attention the short comings created by researchers as comparative material and consideration for repairing and improving implementation study.
3. For universities or equivalents, it is expected to provide lecture-training regarding the quality of the questions in particular multiple-choice questions in order that the capabilities of the lecturers in constructing the questions can be further improved.

REFERENCES

- [1] Sudaryono, U. Rahardja, Q. Aini, Y. Isma Graha, and N. Lutfiani, "Validity of test instruments," in *Journal of Physics: Conference Series*, IOP Publishing, 2019, p. 12050.
- [2] F. V. Vianti, R. Ambarini, and S. N. Ardini, "AN ANALYSIS ON THE VALIDITY TEST QUESTION WITH THE EXAM CONTENT OUTLINE IN FIRST SEMESTER FOR THE XI GRADE STUDENT OF SMA N 3 REMBANG," in *Undergraduate Conference on Applied Linguistics, Linguistics, and Literature*, 2022, pp. 93–103.
- [3] E. Akib, "AN ANALYSIS OF THE VALIDITY OF ENGLISH TEST MADE BY THE TEACHER IN SMP NEGERI 1 BONTOMARANNU," *J. Lang. Test. Assess.*, vol. 2, no. 1, pp. 47–52, 2022.
- [4] J. C. Alderson and A. Beretta, *Evaluating second language education*. Cambridge University Press, 1992.
- [5] R. Fathony, M. A. Bashiri, and B. Ziebart, "Adversarial surrogate losses for ordinal regression," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.
- [6] J. B. Heaton, *Writing English language tests*. 1983.
- [7] A. Harrison, "A language testing handbook," (No Title), 1983.
- [8] I. Farida, "Evaluasi pembelajaran berdasarkan kurikulum nasional." PT. Remaja Rosdakarya, 2017.
- [9] Sugiyono, *Metode Penelitian kuantitatif, kualitatif dan R&D*. Bandung: Alfabeta, 2014.
- [10] R. Bogdan and S. K. Biklen, *Qualitative research for education*. Allyn & Bacon Boston, MA, 1997.
- [11] J. C. Alderson, C. Clapham, and D. Wall, *Language test construction and evaluation*. Cambridge University Press, 1995.
- [12] W. D. Areta, "An Analysis On The Content Validity Of English Summative Test Items At The Even Semester Of The Second Grade Of Junior High School," 2014.
- [13] H. D. Brown and P. Abeywickrama, "Language assessment," *Princ. Classr. Pract. White Plains, NY Pearson Educ.*, 2004.
- [14] D. A. N. Douglas and S. Frazier, "Teaching by Principles: An Interactive Approach to Language Pedagogy .: H. Douglas Brown." Wiley Online Library, 2001.
- [15] H. Douglas, *Principles of language learning and teaching*. ALLYN & BACON, 2021.
- [16] R. Elis Ratnawulan, "Evaluasi pembelajaran dengan pendekatan kurikulum 2013," *Bandung: Pustaka setia*, 2014.
- [17] L. R. Gay, G. E. Mills, and P. W. Airasian, *Educational research: Competencies for analysis and applications*. Pearson, 2012.
- [18] M. S. Haromain, "An analysis on the validity of the summative test in the first year of SMP Al Kamal Jakarta".
- [19] T. P. P. K. Ilmiah, "Pedoman Penulisan Karya Ilmiah," *Bandung UPI*, 2012.
- [20] H. Jeremy, "The practice of English language teaching," 1983.
- [21] M. Mirnawati, "Measuring the English Lecturers' Evaluation Tests Items in English Education Program of Tarbiyah Faculty of State Islamic Institute of Parepare (Analysis of Content Validity of Final Tests Items)." IAIN Parepare, 2020.
- [22] S. M. Shawashreh, "Evaluate the causes of student ' s weakness in writing and speaking skills in English language , case study of Jordanian schools," *Int. J. Educ. Res.*, vol. 7, no. 10, pp. 235–244, 2019.
- [23] S.-W. Su, "The various concepts of curriculum and the factors involved in curricula-making," *J. Lang. Teach. Res.*, vol. 3, no. 1, pp. 153–158, 2012.
- [24] A. Suharsimi, "Prosedur penelitian suatu pendekatan praktik," *Jakarta: Rineka Cipta*, vol. 134, 2006.
- [25] A. Sulaeman, "An Analysis on the Content Validity Of the Summative Test For the First Year Students of SMPN 4 Bekasi," 2008.