

CHAPTER V

CONCLUSION AND SUGGESTIONS

A. Conclusion

This study aimed to examine the psychometric characteristics of the Reading Comprehension section of the English Proficiency Test of Universitas Nahdlatul Ulama Purwokerto (EPTUNU) using Item Response Theory (IRT) with the QUEST program. The analysis focused on four main aspects: test validity, test reliability, item discrimination, and item difficulty, based on response data from 55 test-takers.

The findings indicate that the Reading Comprehension section of EPTUNU demonstrates good construct validity. The overall item fit statistics generated by the QUEST program show that the test fits the Rasch model adequately, as reflected by the Mean Infit MNSQ value within the acceptable range. This result suggests that the test items consistently measure a single underlying construct, namely reading comprehension ability, and that score interpretations derived from the test are psychometrically defensible.

In terms of reliability, the QUEST analysis yielded a high reliability coefficient of 0.91, indicating excellent measurement precision. This finding confirms that the Reading Comprehension section of EPTUNU provides stable and consistent scores and meets internationally accepted standards for educational and language proficiency testing. Therefore, the test is suitable for institutional purposes, such as proficiency benchmarking and graduation requirements.

The item discrimination analysis shows that most items demonstrate ideal or acceptable discrimination, indicating that they are generally effective in differentiating examinees with different levels of reading ability. Although several items were identified as weakly discriminating or overfitting, these results should be interpreted cautiously. Despite the sample size of 55 students being adequate for preliminary Rasch analysis, discrimination estimates in IRT

can still be influenced by response patterns and limited ability variation, and therefore these findings should be viewed as diagnostic rather than conclusive.

Similarly, the item difficulty analysis indicates that the Reading Comprehension test contains a balanced distribution of easy, moderate, and hard items, with the majority of items falling within the ideal difficulty range. Items classified as easy or hard reflect the interaction between item characteristics and the ability level of the sample rather than inherent weaknesses in item construction. Under IRT principles, difficulty estimates become more stable when supported by larger and more diverse samples.

Overall, this study concludes that the Reading Comprehension section of EPTUNU is valid and reliable at the test level. While some item-level indicators, particularly item discrimination, require cautious interpretation, the results provide valuable empirical evidence to support the operational use of the test. Future administrations involving larger and more heterogeneous samples are recommended to further strengthen item calibration and support continuous test development.

B. Suggestions

Based on the findings and conclusions of this study, several suggestions are proposed for the university, the language center administrators and future researchers.

1) For the University

Based on the findings of this study, it is recommended that Universitas Nahdlatul Ulama Purwokerto continue to support systematic evaluation of institutional assessments such as the English Proficiency Test (EPTUNU). Given the pivotal role of the EPTUNU in academic placement, progression, and graduation, the university should encourage periodic psychometric review to ensure that test decisions remain fair, accurate, and defensible. Institutional support in terms of policy, resources, and test development infrastructure will be essential for maintaining the quality and credibility of the EPTUNU over time.

2) For the Language Center Administrators

It is suggested that the Language Center administrators utilize the findings of this study as empirical input for test improvement. Items identified as requiring revision, particularly those with weak discrimination or overfit discrimination values should be reviewed systematically rather than immediately discarded. Administrators are encouraged to revise and pilot these items in future test administrations involving larger and more diverse samples to obtain more stable Item Response Theory (IRT) parameter estimates. Furthermore, the continued use of the QUEST program or similar IRT-based tools is recommended to support evidence-based decision-making in test development and item banking.

3) For Future Researchers

Future researchers are encouraged to extend this study by involving larger sample sizes to improve the accuracy and stability of item parameter estimation. Further research may also analyze other sections of the EPTUNU, such as listening or structure, or apply different IRT models to compare results across analytical approaches. Such studies would contribute to a more comprehensive understanding of the EPTUNU and strengthen its development as a standardized institutional English proficiency assessment.