

Efforts to Improve the English Test Reliability in SMP Negeri 1 Manisrenggo Klaten

Efendi Hidayatullah

Magister Pendidikan Bahasa Inggris, Universitas Sarjanawiyata Tamansiswa

Yogyakarta

Email koresponden : effendihidayatullah@gmail.com

ARTICLE INFO

Riwayat Artikel

Artikel masuk : 2024-04-28

Artikel direview : 2024-06-11

Artikel diperbaiki: 2024-06-28

Artikel diterima : 2024-06-29

Kata Kunci

Implementasi

Kebebasan

Usaha

Pendaftaran

Pupuk Organik

ABSTRAK

This study aims to attempt to improve the reliability of English language tests. The qualitative research focused on the results of the English reliability test in the ninth grade of SMP Negeri 1 Manisrenggo Klaten. The study's results were analyzed and discussed in depth to find implications and contributions to efforts to improve the reliability of English tests, especially in grade nine. This study found reliability issues such as Student-Related Reliability, Rater Reliability, Test Administration Reliability, and Test Reliability Efforts to improve English test reliability at SMP N 1 Manisrenggo School, Klaten, can be enhanced by focusing on test content reliability.

keyword: Reliability, English language, test

Penelitian ini bertujuan untuk Upaya meningkatkan reliability tes bahasa Inggris . Metode penelitian yang digunakan adalah kualitatif, dengan fokus pada hasil test bahasa Inggris pada kelas Sembilan SMP Negeri 1 Manisrenggo Klaten. Hasil penelitian dianalisis dan dibahas secara mendalam untuk menemukan implikasi dan kontribusi terhadap Upaya meningkatkan keandalan tes bahasa Inggris terutama pada kelas sembilan. Pada penelitian ini ditemukan masalah reability seperti Reliabilitas Terkait Siswa, Reliabilitas Rater, Reliabilitas Administrasi Tes, dan Reliabilitas Tes Upaya untuk meningkatkan reliabilitas tes bahasa Inggris di Sekolah SMP N 1 Manisrenggo, Klaten, dapat ditingkatkan dengan memfokuskan pada reliabilitas isi tes.

Kata kunci: Reliability, Bahasa Inggris, tes



INTRODUCTION

English proficiency tests are essential to know candidates' listening, reading, writing, and speaking skills. They determine how effectively a candidate can communicate with classmates and teachers in English. Good tests are important indicators of academic progress; tests help educators determine student grade level placement, design specialized instruction, set learning goals, and monitor progress (Deegan, 2014). Sound tests have some characteristics, such as reliability. Cypress (2017) has defined reliability as the "worthiness with which a measuring device measures something; the degree to which a test or other instrument of evaluation measures consistently whatever it does in fact measure." According to (Moskal & Leydens 2019) defined, "the term reliability means the consistency with which a set of test scores measure whatever they do measure."

Efforts to improve the reliability of English tests in schools are crucial for ensuring accurate assessments of students' language proficiency (Alharbi, 2015). This paper will focus on improving the reliability of English tests in School SMP N1 Manisrenggo, Klaten. The reliability of a test is determined by its consistency in producing stable results, and factors such as the skills being tested and the student's language proficiency can affect it. To improve the reliability of English tests, clear instructions for each assignment, well-crafted questions, and seeking feedback from students and colleagues can be beneficial. This paper will analyze the factors affecting the reliability of English tests in schools and provide recommendations for improving the reliability of English tests in School SMP N1 Manisrenggo, Klaten.

Testing refers to measuring the result of students' learning in the teaching-learning process. Consequently, the teachers should be able to arrange and analyze a good test (Boone, 2016). Therefore, teachers' accuracy and carefulness may significantly impact the increase in the quality of teaching, particularly in judging students' abilities (Ertürk, 2015). This information benefits both students in their learning and teachers in their teaching. It can be feedback for the teachers, who are responsible for meeting the instructional objectives, while for the students, it illustrates their performance. A good test should fulfill specific criteria; according to Rusilowati et al., a good test has four criteria: validity, reliability, level of difficulty, and discrimination power. Concerning the requirements of a good test above, the writer was focused on the data. The valid test must be reliable. However, the mean scores of a reliable test are not required to be very close to the test criteria to qualify as valid. A valid test has a low dispersion of scores around or near the requirements. A heavily skewed test may be reliable, but it is invalid. The research questions proposed to align with the primary objective of what efforts were conducted to improve the test reliability. This research improves the reliability of the tests written at SMP Negeri 1 Manisrenggo Klaten.

RESEARCH METHODS

This research was conducted using a quantitative approach. Reliability tests typically use a qualitative research design (Noble & Smith, 2015). Qualitative research aims to produce facts that can be relied upon and trusted. The research data collection technique uses an interview with one of the English teachers and writing an English test at SMP Negeri 1 Manisrenggo Klaten. The participant in this research on English test reliability is a 9th-grade student; one of the primary purposes of the research is to analyze the reliability of the tests that students have written and take the necessary precautions if the tests are not found to be reliable, and efforts to improve the English test reliability in students.

RESULT AND DISCUSSIONS

Table 1. The frequency distribution of the English test

NO	Components	First test	Second test
1	Number of respondents	18	22
2	Number of test scores	1152	1848
3	Mean	72	84

The given data presents the results of two tests conducted on respondents. The first test had 18 respondents and 1152 test scores, with a mean score of 72. The second test had 22 respondents and 1848 test scores, with a mean score of 84. The data does not provide any information about the nature or purpose of the tests or the characteristics of the respondents.

The results of research using the interview method on English teachers at SMP N1 Manirenggo. Based on the data collection results by interview in grade 9 at SMP N1 Manis renggo, researchers found question reliability, such as Student-Related Reliability, Rater Reliability, Test Administration Reliability, and Test Reliability.

1. Student-relate reliability

Known as one of the common problems of the current age, stress negatively affects student-relate reliability. Stress refers to the emotional and spiritual tension caused by some problems (Irene & Wisesa, 2020). Stress is a phenomenon that affects student's behavior, professional performance, and communication with colleagues (Clipa & Boghean, 2015). At SMP N1 Manisrenggo, the student-psychological problem is a lack of confidence to learn English, and problems at home are parents' lack of support and supervision in learning, so this becomes a problem point in student-relate reliability. Student-related dependability is established when students from various linguistic backgrounds consistently get similar scores on reading comprehension tests. This implies that the exam reliably assesses English reading skills across different student demographics.

2. Rater Reliability

Rater Reliability is the level of test measurement error that rarely occurs (Livingston et al., 2018). The test given is based on the material discussed, and the assessment rubric must be clear to each teacher because it has different assessment criteria. Assessment criteria are the specific standards or guidelines used to evaluate the performance or quality of a task, product, or outcome, and these criteria help ensure consistency and fairness in assessment by providing clear expectations for what constitutes success Sometimes (Lund & Kirk, 2019). The questions given are not by the lattice given; in other words, the summative test given has not been taught by the teacher. Teachers have a standardized rubric used by all teachers to produce the same assessment for assessment instruments. Therefore, each student cannot be given a separate assessment rubric.

3. Test Administration Reliability

According to Shohamy (2020), test administration reliability affects students' ability to test; calmness and comfort must be prioritized for student tests, such as building work, vehicle noise, and adequate media such as clear sound, straightforward pictorial questions, and clear or easy-to-read questions content. The reliability of the test's administration is demonstrated when students perform similarly whether taking the test in person or online. A high degree of test administration reliability is indicated by consistency in scores across these different modes of administration.

4. Test reliability

On the reliability test, students also experience problems such as requiring more time for some subjects such as math, physics, chemistry, and English (Amalia, 2020). Amalia also define that test reliability in a grammar assessment covering tenses, sentence structure, and punctuation is demonstrated when students achieve similar scores on multiple occasions under similar conditions. Excellent test reliability is demonstrated when the student scores an average of 72 on an English test one week and scores about 84 on the same test the following week.

CLOSING

1. Conclusion

In an English language proficiency test, such as the end of the chapter or the semester test, inter-rater reliability is demonstrated when test takers from different language backgrounds consistently score at similar levels on sections that assess skills such as reading, writing, listening, and speaking. This consistency suggests that the test accurately measures English proficiency across diverse student populations. Rater reliability is demonstrated when different raters consistently assign similar scores to responses of equal quality in an English language proficiency test in which human raters score responses. Test administration dependability is established when students receive similar scores regardless of which testing facility they use. Consistency in results across multiple testing sites suggests that external factors such as test setting and administration techniques have no substantial effect on test performance. In conclusion, each reliability element in the context of an English proficiency exam assures that the assessment appropriately assesses language abilities consistently across various students, raters, testing settings, and testing dates.

2. Suggestions

Based on the findings of the data collection on the subject, future researchers can conduct further validity tests to ensure that the exam effectively assesses the desired components of English language competence, such as reading, writing, listening, and speaking skills. This includes determining how accurately the test predicts real-world language performance and consequences. Conduct longitudinal research to investigate the consistency of English proficiency results over time and how they relate to language acquisition and development. This could include following test takers' progress over several months or years to determine how their competence develops.

REFERENCES

- Alharbi, H. A. (2015). Improving Students' English Speaking Proficiency in Saudi Public Schools. *International Journal of Instruction*, 8(1), 105–116.
- Amalia, K. (2020). *AN ANALYSIS ON THE STUDENTS' ERROR IN USING IRREGULAR VERB OF THE NINTH GRADE OF MTS AL-HIKMAH BANDAR LAMPUNG AT THE SECOND SEMESTER IN THE ACADEMIC YEAR OF 2019/2020*. UIN Raden Intan Lampung.
- Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE—Life Sciences Education*, 15(4), rm4.
- Clipa, O., & Boghean, A. (2015). Stress factors and solutions for the phenomenon of burnout of preschool teachers. *Procedia-Social and Behavioral Sciences*, 180, 907–915.
- Cypress, B. S. (2017). Rigor or reliability and validity in qualitative research: Perspectives, strategies, reconceptualization, and recommendations. *Dimensions of Critical Care Nursing*, 36(4), 253–263.
- Deegan, G. M. (2014). *Distributed instructional leadership and student achievement at an urban school*. Grand Canyon University.

- Ertürk, N. O. (2015). Testing your tests: Reliability issues of academic English exams. *International Journal of Psychology and Educational Studies*, 2(2), 47–52.
- Irene, I., & Wisesa, A. (2020). Analyzing How Spiritual Intelligence Affecting Stress Coping in Stress Management. *Jurnal Ilmu Sosial Politik Dan Humaniora*, 3(2), 1–12.
- Livingston, S. A., Carlson, J., & Bridgeman, B. (2018). Test reliability-basic concepts. *Research Memorandum No. RM-18-01*. Princeton, NJ: Educational Testing Service, 8.
- Lund, J. L., & Kirk, M. F. (2019). *Performance-based assessment for middle and high school physical education*. Human Kinetics Publishers.
- Moskal, B. M., & Leydens, J. A. (2019). Scoring rubric development: Validity and reliability. *Practical Assessment, Research, and Evaluation*, 7(1), 10.
- Noble, H., & Smith, J. (2015). Issues of validity and reliability in qualitative research. *Evidence-Based Nursing*.
- Shohamy, E. (2020). *The power of tests: A critical perspective on the uses of language tests*. Routledge.