

Examining the Psychological Factors Impacting Teachers' Constraints in Creating and Administering Final Examinations

Evy Ratna Kartika Waty^{1*}, Yanti Karmila Nengsih², Ciptro Handrianto³, Mega Nurrisalia⁴, Shomedran⁵

^{1,2,4,5} Universitas Sriwijaya, Indonesia

³ Sultan Idris Education University, Malaysia

*Corresponding author email: evyrkwaty@gmail.com

Received: 06-May-2023

Revised: 18-June-2023

Accepted: 02-July-2023

Abstract

Introduction: Examination tests are used by teachers to evaluate and to get valuable insight into their students' challenges and the strategies they find most effective by reflecting on their teaching practices. If the teacher is aware of the student's weaknesses, he or she may quickly strengthen subsequent lessons.

Objectives: The objective of study is to evaluate of teachers' constraints to generate and administer the final examination test in secondary school.

Methods: The study used mixed methods design. These quantitative and qualitative assessments examined how difficult for teachers to design and administer final exam questions by analyzing questionnaires using objective answer sheets and conducting interviews with teachers. This research focused on the Senior High School grade 12 teachers of Religious Education, Citizenship Education, and History in Palembang who created the questions for the Final School Examination.

Results: The findings demonstrated that teachers develop formative and diagnostic tests to identify students' weaknesses. The majority of teachers, when asked how they would overcome limitations in test creation, said they were hesitant to study references about test writing or to attend seminars and training. Teacher meetings, talks among colleagues, and topic-teacher meetings are teachers' most typical means to overcome constraints in developing achievement test items.

Conclusions: Based on the results of questionnaires and these interviews, it is clear that Indonesian teachers need in-service training in assessment.

Keywords: Teachers' constraints, generating test, administering test, final examination test, schools test

1. Introduction

Teachers use assessments to see how their classes stack up against national norms (Rimfeld et al., 2019). This is a challenge that the teacher has created and sent to the class. The collected information may help schools evaluate their children's progress and determine whether they are prepared for the next grade level (Wandasari et al., 2019). However, education specialists and specialist groups (Arikunto, 2005) agree that the culminating examination in schools has significant flaws and should be revised. The current system of testing and evaluating teachers has several things that could be improved.

The exam needed to be constructed more effectively to make informed decisions and take productive actions; quality assessments need solid quantitative qualities like validity and reliability (Rose & Johnson, 2020). The classroom exams utilized up to this point could be better since some teachers are not adept at producing tests (Liu et al., 2020). Furthermore, there is scant evidence that the standardized tests used in schools are reliable (Kokaraki et al., 2019).

Additionally, the range for test and item scores is sometimes different for tests used to evaluate students' abilities (Adom, 2020). A student's performance be inferred from their test results (Tus, 2020). Having to do with studying for and taking one's last exams in school. Sometimes teachers may shift to other schools, and in

such places, they may encounter new teachers who need to learn how to educate or evaluate their students' progress.

Test teams are typically comprised of teachers from all disciplines and are formed at the beginning of each school year (Tanjung et al., 2021). This committee organizes the school exam, and its decisions serve as a guide for the finished product. The next step is the school's preparation of test items, which are then utilized to generate test questions, which are replicated. Sudijono (2008) states that most veteran teachers can create more rigorous examinations. Teachers who regularly administer formative assessments are better prepared to provide valid summative assessments that accurately reflect their students' mastery of course material. The exam's grade also affects the degree to which a test accurately reflects a student's level of knowledge. The teacher's role as the exam's creator is crucial. Teachers should be able to create practical summative assessments after the school year if they can create valid formative assessments throughout the year (Louzi et al., 2022; Saravanan & Vasantha, 2023).

Arguments against the exam include its use of objective test methodologies that do not adequately gauge students' competence (Adom et al., 2020). We also need a modern assessment strategy that goes beyond measuring IQ. According to Sukardi (2008), many teachers favour using objective multiple-choice examinations for their final assessments. Because all they have to do on objective tests is pick the best answer, students will need help demonstrating their true abilities. Students' abilities will decline due to the increased pressure to take examinations at tutoring facilities.

Creating a high-quality test requires consideration and using several tools (Clifton & Schriener, 2010; Naeem et al., 2012; Osterlind, 1998; Razak et al., 2012). According to Ebel (2007), a good test requires several qualities, including but not limited to the following: relevance, specificity, balance, objectivity, efficiency, equality of opportunity, difficulty level, discrimination index, and dependability. A decent exam based on the issues presented relies heavily on the teacher's expertise and experience. When exams are well designed, they reveal students' actual levels of achievement. For this reason, many teachers place a premium on providing their students with high-calibre instruction in learning and test-taking (Sukardi, 2008; Husin et al., 2023).

Internal and external factors for the teacher significantly impact the quality of the final assessment (Barrett et al., 2015; Grimes et al., 2004; Izci, 2016; Nguyen et al., 2020; Sumak et al., 2017; Sun et al., 2016). According to Anastasi (2007), a teacher's knowledge, emotions, and socioeconomic status all play a role in the quality of the assessment they provide. People believe it is crucial to investigate teachers' histories, such as their highest degree of education, years of experience, frequency of test administration, and amount of training received in the preceding three years. The continued success of the final school examination depends on satisfactory responses to all of these concerns. Internal factors also influence teachers' ability to provide valid assessments. Teacher test quality may be better understood by looking at the teachers' perspectives when designing assessments (Lam, 2019; Pandey et al., 2022).

Teachers may get valuable insight into their students' challenges and the strategies they find most effective by reflecting on their teaching practices. If the teacher is aware of the student's weaknesses, he or she may quickly strengthen subsequent lessons. The teacher is well aware of the material to be covered and the expected level of student comprehension. Therefore, the teacher is aware of the most crucial concepts the students need to learn based on how the course has been structured thus far. Moreover, it is vital to examine the restrictions that teachers experience when it comes to the preparation and administration of final exams. This is because teachers have limited time and resources. Therefore, the goal of this study was to investigate the restrictions that teachers confront when creating and conducting final examinations.

2. Methods

This study used mixed methods quantitative and qualitative approaches. This quantitative review used questionnaires, objective answer sheets, and teacher interviews to determine how hard it is for teachers to make final test items and how much they have to work with what they have (Leavy, 2022). The 60 teachers who participated in this study made questions for the grade 12 Final School Examination in three subjects: religious

education, citizenship education, and history education. These participants were taught in all state high schools in Palembang, Indonesia.

In order to gather data for this study, the researcher created a questionnaire after the experiment was complete. The researcher changed this form to determine what keeps teachers from being limited when taking final tests (Alshurideh et al., 2022; Butler et al., 2022). This form mostly stayed the same for this study because teachers have to work with what they have. Interviews with teachers were used to get information for this study. When the interviewer has planned questions ahead of time that is meant to get only the most essential information, they use an organized interview form. This study used this discussion to ensure teachers' answers are accurate when given directly. This helped teachers get around the problem of being limited when making final tests. Questionnaires and discussion notes with teachers were used to get a detailed study of teachers' problems or limitations when making test items and an idea of what teachers do to get around these problems (Patten, 2016; Vyas, 2022).

3. Results

Teacher Constraints in Test Construction

Some results have been made based on the teacher's answers to the questions and the interviews. Information about the problem of constraints that teachers face when making tests is looked at from three different points of view: the main thing the teacher thinks about when making the test, the problem or constraint that the teacher faces when making the test, and the things the teacher does to get around the constraints.

Teacher's Experience Taking Tests

The experience of teachers in making tests reveals the nature of teachers in making tests, whether formative tests or diagnostic tests, to find out the difficulties experienced by teachers and the experience of making other types of tests such as field tests, attitude or interest tests, high-performance student tests, and Al-Qur'an Reading and Writing tests for Religious Education teachers. The results of the teacher's experience in making tests are shown in Table 1.

Table 1. Teacher's Experience Taking Tests

No.	Questions for teachers	No (k=60)	Per cent (%)
1	Make a formative test in a month at least four to six times	60	100
2	Experience performing diagnostic tests		
	Experienced	19	32
	Not Experienced	41	68
3	If you ever do a diagnostic test		
	One to three times	10	17
	More than four times	9	15
4	Experience doing other tests		
	Orientation test	2	3
	Attitude/interest test	17	28
	Student performance test	16	27
	Al-Quran reading and writing test	20	33
	Never did another test	29	48

Table 1 shows that teachers' test-making experience was good for formative tests (daily tests for each topic). This indicates that for each subject they teach, students have tested their teachers on daily tests in the classroom. Diagnostic tests are used to find out what problems students have with how they are being taught and how they are learning, as well as what causes those problems. If a reason can be found, the teacher can plan and take more

steps to solve the problem. Diagnostic tests usually focus on one topic, and each question is in-depth so that each student's problems with that subject can be found. This test is essential because it can show where students are weak, which gives the teacher a chance to help those students.

Many teachers still have not been able to do a test to figure out if a student is having trouble learning. Table 1 shows that many teachers still need to take a diagnostic test. Based on their experience, 41 teachers (68%) have never done a diagnostic test, so they must need the training to do tests to help students who are having trouble.

The teacher's experience with making other tests, like the significant selection test, shows that only two teachers have ever made a major selection test for students to choose a natural science major or a social science major. As for teachers who have taken an attitude or interest test, 17 have done so. On the other hand, 16 teachers who have done it before created the High-Performance Student Selection Test.

It was found that 29 teachers (48%) had never taken a test and did not know how to give introduction tests to students in a natural science major, social science major, attitude or interest tests, or student performance tests. This shows that teachers need the training to make tests that help improve the learning review process. All 20 teachers have taken the Qur'an Reading and Writing Test for Religious Education. Teachers must be trained to make other tests that could improve their knowledge and grading skills more successfully. Table 2 shows the different types of training teachers get. This training includes training on the curriculum and evaluations of the last three years of teaching.

Table 2. Curriculum and Assessment-Related Service Training

No.	Questions for teachers	No (k=60)	Per cent (%)
1	Frequency of training related to curriculum and assessment		
	One To Three Times	22	37
	More Than Four Times	38	63
2	Fields of training attended by teachers		
	Subject	30	50
	Assessment Practice	9	15
	Teaching-Learning Model	14	23
	Teaching Profession	16	27
3	The need for in-service training for teachers is very necessary	60	100
4	Assessment exercises are applied in class		
	Yes	58	97
	No	2	3

Table 2 shows that teachers' training includes curriculum-related training and evaluation. In the last three years, 22 (37%) teachers have been trained one to three times and 38 (63%) teachers have trained more than four times. However, some still need to be able to use the content and testing training they have had in the classroom.

In the field of training related to curriculum and assessment for assessment practice in the proper knowledge of teachers to make acceptable items, there are as many as 30 teachers (50%) who follow the subject area, but only nine teachers (15%) who follow the area of assessment practice. Even though all of the teachers who responded said that training in services related to curriculum and assessment is vital, this is what is needed in the field of training related to the subject area, the field of assessment practice, learning model training, and teaching profession training, so that after teachers get training, they can use it in the classroom.

Constraints in Test Construction

The constraints` teachers are analyzed based on the answers teachers give to questionnaires. These questionnaires ask about the teacher's main concern when making a test, the problems the teacher faces when making the test, and what the teacher does to get around these problems.

Based on the analysis of the questionnaire's results, teachers usually think about all important aspects of test preparation, such as the content of the lessons being tested, the purpose of the test, the syllabus and competence, the item distribution framework, the suitability of the type of test with the purpose, the suitability of the number of items with the test time, the level of difficulty of the item, the discrimination index of the item, and the order of the items. This finding is shown in Table 3.

Table 3. Main Considerations for Teachers Preparing Tests

No	Questions for teachers	Always		Not Sure		Never	
		No	%	No	%	No	%
1	Lesson content is tested	56	93	4	7		
2	The purpose of the test is made	48	80	10	17	2	3
3	Syllabus and Competencies	52	87	7	12	1	1.5
4	Item distribution framework	47	78	11	19	2	3
5	Suitability of test type to purpose	51	84	6	10	3	6
6	Correspondence of the number of items with the test time	51	84	6	10	3	6
7	Item difficulty level	34	57	26	43		
8	Item discrimination index	17	29	41	68	2	3
9	Order of test items	44	73	12	21	4	6
10	Scoring method	44	73	15	26	1	1.5

Based on Table 3, the most important thing for the teacher to do is make a test with questions that other teachers can use as a guide for making good test questions that follow good item-writing methods. When making tests, teachers always think most about the lessons being tested (93%), the test's goal (80%), the test's schedule and the skills that need to be learned (87%), and the item distribution system (78%). More than 70% of teachers think about these things when making tests.

Item challenge level (43%) and item discrimination index (68%) were considered when making the test. The survey results showed that 40–70% of the people who filled it out were unsure if these things were considered when the test was made.

Table 4. Teacher Constraints on Preparing Tests

No	Questions for teachers	Always		Not Sure		Never	
		No	%	No	%	No	%
1	Determine the purpose of the test	30	50	14	23	16	26
2	Determine the type of test that fits the purpose	18	30	19	32	23	38
3	Determine the number of items and test time	16	26	16	26	28	48
4	Create an item distribution framework	20	33	14	23	26	44
5	Compose test items	27	45	12	20	21	35
6	Determine the validity and reliability of the test	9	15	13	21	38	64
7	Assessing the effectiveness of distractions	10	17	14	23	36	60
8	Choose the appropriate words on the test	14	23	21	34	25	43
9	Assess test quality	13	23	17	27	30	50

10	Analyze self-made tests	11	19	15	24	34	57
11	Make a report of test results	32	54	18	29	10	17

Based on Table 4, it is known that many teachers experience problems or constraints. If teachers do not consider it essential to answer never, then it does not become a rule in arranging or constructing tests, especially in the following matters:

- determining the validity and reliability of tests (64%)
- evaluating the effectiveness of distractions (60%)
- analyzing self-made tests (57%)
- evaluating the quality of the test (50%)
- determining the number of items with the test time (48%)
- determining the item distribution framework (44%)
- choosing the appropriate words on the test (43%)
- determining the type of test with the purpose (38%)
- arranging test items (35%)
- determining the purpose of the test (26%)

To overcome constraints in test construction, teachers talk with fellow teachers, talk with experts, read test-writing reference books, and attend various seminars and training. The results of this study were obtained from the teacher's questionnaire in Table 5.

Table 5. Teachers' Efforts to Overcome Constraints

No	Questions for teachers	Always		Not Sure		Never	
		No	%	No	%	No	%
1	Discussions with fellow teachers	43	72	17	28		
2	Discussion with experts	13	22	26	44	21	34
3	Read the test writing reference	24	40	32	53	4	7
4	Discussion during a teacher meeting	49	82	11	18		
5	Meeting subject teachers	31	52	28	47	1	1.5
6	Attend seminars and training	19	32	34	57	7	11

Based on Table 5, the most common way teachers overcome constraints in building achievement test items is through discussions during teacher meetings (82%), discussions among fellow teachers (72%), and subject-teacher meetings (52%).

However, despite teachers' constraints, there is still less discussion with experts (34%) who say they have never discussed with experts. About half of the teachers (53%) reported needing clarification after reading references about test writing and attending seminars and training (57%) to overcome constraints in test construction.

Although the findings from the questionnaire show that teachers generally do not have significant constraints in preparing for the test, the findings from interviews with teachers are the opposite. Interviews show that teachers need to understand reasonable construction procedures. This is clearly illustrated in the excerpt from the following passage:

"I... when making the questions for the final school exam lacked the time and lacked ability or understanding, most of the test questions was taken from the book I bought at the bookstore; this way is

faster and easier. Some teachers create question items from the question bank or change them to adapt from the guidebook (Respondent 18: Religious Education Teacher).

Most teachers copy and paste questions from market books for their students' final exams because it is easier and faster. Some teachers utilize item banks or make changes to content from reference books. As a result, the teacher may only sometimes be held accountable for typos on tests since they may assume such errors are usual while designing exam questions. The error was caused by a misunderstanding between the teacher and the administrator who typed the exam (Handrianto & Salleh, 2019; Xasanova, 2020). Furthermore, teachers, there are several inconsistencies in the creation of exam questions, implying that we need to pay attention to the topical relevance of the information.

"I... stated that it was a normal thing, and it was just a mistake in typing in the marks, so it wasn't entirely the teacher's fault." This is due to the lack of cooperation between teachers and administrative staff who type the test (Respondent 18: Religious Education Teacher).

Teachers, in reality, there are many inconsistencies in the creation of test questions, which indicates that we need to focus on whether or not the content is relevant to each topic. Teachers often only choose resources that may easily be turned into exam questions and are appropriate. They know that many teachers show inconsistency when designing exam questions, which implies they often overlook the need to cover each subject thoroughly (Heller & Finley, 1992). Typically, teachers only cover appropriate material that can easily be turned into test questions. Teachers also consistently rely on the teacher's handbook, despite needing a firm grasp on the following topics: estimating the reliability of tests, designing test questions, and utilizing personal experience.

"The teachers... lack understanding of the procedure for calculating test reliability; in making test questions, we often use personal experience and always use the teacher's handbook (guidebook). This is a guideline that personal experience and the teacher's handbook (guidebook) are very helpful for teachers in making good test questions" (Respondent 29: Citizenship Education Teacher).

When developing test items, teachers typically refer to the teacher's handbook and have no idea how to calculate the test's dependability based on their experience. This highlights the value of a teacher's handbook and personal experience in developing high-quality evaluation measures (Looney, 2011). Teachers also seldom perform the following to analyze the effectiveness of their assessments:

"We teachers... have never calculated and analyzed the quality of the constructed test (validity, reliability, discrimination index, level of difficulty, and effectiveness of distractions). Because a lot of time is spent managing administrative matters and a lot of time is used to teach in other places" (Respondent 41: History Education Teacher).

Following that, teachers rarely conduct statistical analyses on test quality (such as validity, reliability, power, difficulty, and distractor effectiveness) because time spent teaching elsewhere detracts from time spent on administrative responsibilities. Findings from interviews with teachers show that teachers lack an understanding of good construction procedures. Interviews with teachers also show that in building tests, teachers lack time and expertise; most test items are taken from books on the market. In making test items, the teachers are not consistent, ignore the depth of the content of each topic, and only choose content that is easy to construct.

The findings of the interview with the teacher showed that the teacher never calculated and analyzed the quality of the test he constructed, including validity, reliability, item discrimination index, item difficulty level, and the effectiveness of distractions in administrative matters. Excerpts from interviews with teachers give an impression of the need for in-service training related to assessment for teachers in the Indonesian city of Palembang.

4. Discussion

The most significant issues with school administration are the dearth of physical facilities and infrastructure, the absence of a collection of educational technology resources, and the ineffective management of school resource centres (Halina, 2001; Handrianto et al., 2019). Additionally, educational technology resources are lacking. She

claims that teachers lack motivation, are underpaid, and are not recognized for their work in creating engaging instructional aids. These factors may reduce teachers' awareness of and interest in utilizing technological media (Sudarsana et al., 2019). Not only is it essential for a teacher to understand the challenges or limitations that arise when devising or designing tests for his students' learning, but he or she must also be capable of overcoming these challenges and limitations (Hang & Van, 2020).

The findings of this study provide additional information that can be used to learn more about the capabilities of teachers. According to Kusmiyati (2005), the quality of this test is either attributable to the teacher's specific abilities or to the fact that these abilities are still relatively low. Because it is challenging to establish measurement instruments at the level of synthesis and evaluation (Manullang, 2006), the relationship between these three concepts is an intriguing topic that needs to be investigated (Hamidah, 2004). It must be examined to comprehend the purpose of evaluation completely.

According to the survey results used by the researcher, it comes out that teachers also encounter limitations when creating tests, but they have never done so by creating a distracting answer option (60%) or determining if the test form is appropriate for its purpose (38%). Handmade tests are rarely graded by teachers (57%) due to the amount of time required to create a test distribution plan (44%), select words that can be used in writing (43%), and anticipate the number of questions that can be asked in the allotted time (48%). Constructing a prototype distribution system (44%) and choosing the appropriate words when composing (43%). A significant problem for many teachers is that they do not entirely comprehend the purpose of the exam and do not know how to set it up most effectively. Overall, the queries and methods for assessing knowledge are subpar, as are the assessments created by teachers. This is because teachers can pose inquiries using this method. The results of this study's queries indicate that half of the Palembang Senior High School teachers find it difficult to determine whether the test they created is accurate. The study results indicate that teachers only consider the topic validity of a test when creating a measurement instrument (Ismail et al., 2021).

This is a problem with limits teachers are less aware of, so 64% of teachers reported that they had never assessed the accuracy of the tests they created. In addition, 48% of teachers said they had never calculated the number of tasks they could complete within the available time. In addition, teachers reported that it was difficult to plan prior to the distribution of exams. Therefore, the teacher's system for test distribution does not contribute significantly to the test's validity, and it is difficult to create.

On the other hand, it is common knowledge that when creating test questions, many teachers do not consider how much of each subject's material is addressed. These results are the consequence of speaking with numerous teachers. Most of the time, teachers select appropriate topics that can be swiftly converted into test questions. Thus, a study claiming that teachers find it difficult to generate test questions that require reasoning and evaluation is credible (Wilson & Narasuman, 2020; Zainil et al., 2023). This concerns the final exam Palembang High School teachers created, as there are not enough.

The findings of this study indicate that the study's analysis of test item answer choice distractions was inadequate. This is because the teacher has not properly set up the distractions, and the distractions for students who cannot answer test questions are also distinct (Lang, 2020; Nengsih et al., 2022). In other words, 36% of teachers who have difficulty coming up with distractions are unsure, and 60% of them never examine how well creating distractions aids students in answering test questions. By utilizing reference books and their teaching experience, teachers can circumvent the problem of insufficient time when creating tests. The researchers conducted the test's validity, reliability, discrimination index, degree of difficulty, and how well the distractions functioned by interviewing the teachers who administered it. The researchers discovered that many teachers had never estimated and graded the test's quality. In sum, according to the teachers, much time is spent on routine duties, and much time is spent teaching in other locations.

5. Conclusion

The results of this study give us more information that can be used to understand better what teachers can do. According to the questionnaire results, half of the teachers at the school found it hard to tell if the information was true. The grade of the test made by the teacher depends a lot on both things inside the teacher and things

outside of the teacher. People also think that the teacher's experience with making and giving final tests is important for study. This study adds to the existing amount of information that can be used to learn more about what teachers can do. This study shows how teachers create tests, whether they are formative or diagnostic, to find out where students are having trouble. It also shows how teachers have made other kinds of tests, like field tests, interest or attitude tests, tests for high-performing students, and Al-Qur'an reading and writing tests for Religious Education teachers.

References

1. Adom, D., Mensah, J. A., & Dake, D. A. (2020). Test, Measurement, and Evaluation: Understanding and Use of the Concepts in Education. *International Journal of Evaluation and Research in Education*, 9(1), 109-119.
2. Alshurideh, M. T., Alzoubi, H. M., El Khatib, M., Ghazal, T. M., Al-Dmour, N. A., Sattar, O., & Kukunuru, S. (2022). An Experimental Evaluation on Resource Attribute, Internal Risks and Regime Structure of R&D Association-Including Exploration of Moderating Effect of Association Management Capability, Psychological Impact on Staff. *Journal for ReAttach Therapy and Developmental Diversities*, 5(2s), 201-215.
3. Anastasi, A. (2007). *Psychological Testing*. New York: Mc. Millan Pub.Co Inc.
4. Arikunto, S. (2005). *Penilaian Program Pendidikan*. Jakarta: Bina Aksara.
5. Barrett, P., Davies, F., Zhang, Y., & Barrett, L. (2015). The impact of classroom design on students' learning: Final results of a holistic, multi-level analysis. *Building and Environment*, 89(1), 118–133.
6. Butler, C., Wilson, P., Abrahamson, V., Mikelyte, R., Gage, H., Williams, P., ... & Barclay, S. (2022). Methodology. In *Optimum models of hospice at home services for end-of-life care in England: a realist-informed mixed-methods evaluation*. National Institute for Health and Care Research.
7. Clifton, S. L., & Schriener, C. L. (2010). Assessing the quality of multiple-choice test items. *Nurse teacher*, 35(1), 12–16.
8. Ebel, R. L. (2007). *Essentials of Education Measurement*. New Jersey: Prentice Hall Inc.
9. Grimes, P. W., Millea, M. J., & Woodruff, T. W. (2004). Grades—Who's to blame? Student evaluation of teaching and locus of control. *The Journal of Economic Education*, 35(2), 129–147.
10. Halimah, R (2001), *Faktor-faktor Kekangan Penggunaan Teknologi Pendidikan Dalam Pengajaran dan Pembelajaran di Kalangan Guru-guru Sekolah Menengah Daerah Bentong*, Tesis (Teknikal) Kolej UTT Hussein Onn.
11. Hamidah .(2004).*Metode Penelitian Kualitatif*. Malang: Universitas Muhammadiyah.
12. Handrianto, C., Jusoh, A. J., Nengsih, Y. K., Alfurqan, A., Muslim, M., & Tannoubi, A. (2021). Effective pedagogy in primary education: A review of current literatures. *Abjadia: International Journal of Education*, 6(2), 134-143. <https://doi.org/10.18860/abj.v6i2.12978>
13. Handrianto, C., & Salleh, S. M. (2019). The environmental factors that affect students from outside java island to choose yogyakarta's bimbel. *International Journal of Environmental and Ecology Research*, 1(1), 27-32. Retrieved from: <http://www.environmentaljournal.in/article/view/5/1-1-14>
14. Hang, L. T., & Van, V. H. (2020). Building Strong Teaching and Learning Strategies through Teaching Innovations and Learners' Creativity: A Study of Vietnam Universities. *International Journal of Education and Practice*, 8(3), 498-510.
15. Heller, P. M., & Finley, F. N. (1992). Variable uses of alternative conceptions: A case study in current electricity. *Journal of Research in Science Teaching*, 29(3), 259–275.
16. Husin, A., Maharani, S. D., Raharjo, M., Yosef, Y., Sumarni, S., & Handrianto, C. (2023). Prospects for implementation of green campus in education and research pillars at edupark fkip Unsri become edutourism. *International Journal of Professional Business Review*, 8(4), e01597. <https://doi.org/10.26668/businessreview/2023.v8i4.1597>
17. Ismail, I., Nursalam, N., Angriani, A. D., & Kusumayanti, A. (2021). Development of measurement tool for understanding, application, and reasoning mathematics of madrasah ibtdaiyah students. *Kreano, Jurnal Matematika Kreatif-Inovatif*, 12(1), 26-38.
18. Izci, K. (2016). Internal and External Factors Affecting Teachers' Adoption of Formative Assessment to Support Learning. *Online Submission*, 10(8), 2774–2781.
19. Kokaraki, N., Hopfe, C. J., Robinson, E., & Nikolaidou, E. (2019). Testing the reliability of deterministic multi-criteria decision-making methods using building performance simulation. *Renewable and Sustainable Energy Reviews*, 112, 991-1007.
20. Kusmiyati. (2005). *Karakteristik Butir Tes Ujian Akhir IPA SLTA Buatan Guru*. Tesis Magister, tidak diterbitkan. Universitas Negeri Yogyakarta: Yogyakarta.

21. Lam, R. (2019). Teacher assessment literacy: Surveying knowledge, conceptions and practices of classroom-based writing assessment in Hong Kong. *System*, 8(1), 78–89.
22. Lang, J. M. (2020). *Distracted: Why students can't focus and what you can do about it*. Hachette UK.
23. Leavy, P. (2022). *Research design: Quantitative, qualitative, mixed methods, arts-based, and community-based participatory research approaches*. Guilford Publications.
24. Liu, Z. Y., Lomovtseva, N., & Korobeynikova, E. (2020). Online learning platforms: Reconstructing modern higher education. *International Journal of Emerging Technologies in Learning (iJET)*, 15(13), 4-21.
25. Looney, J. (2011). Developing High-Quality Teachers: teacher evaluation for improvement. *European Journal of Education*, 46(4), 440–455.
26. Louzi, N., Alzoubi, H. M., Alshurideh, M. T., El Khatib, M., Ghazal, T. M., & Kukunuru, S. (2022). Psychological & Prototypical Model of Execution Management evaluation for the framework Development. *Journal for ReAttach Therapy and Developmental Diversities*, 5(2s), 216-223.
27. Manullang. (2006). *Pengetahuan Evaluasi Guru SMA di Medan Indonesia*. Makalah Lokakarya Sistim Jaminan Mutu Proses Pembelajaran.Solo: UNS.
28. Naeem, N., van der Vleuten, C., & Alfari, E. A. (2012). Faculty development on item writing substantially improves item quality. *Advances in health sciences education*, 17, 369-376.
29. Nengsih, Y. K., Handrianto, C., Nurrisalia, M., Waty, E. R. K., & Shomedran, S. (2022). Media and resources development of android based interactive digital textbook in nonformal education. *Journal of Nonformal Education*, 8(2), 185-191. <https://doi.org/10.15294/jne.v8i2.34914>
30. Nguyen, D., Harris, A., & Ng, D. (2020). A review of the empirical research on teacher leadership (2003–2017) Evidence, patterns and implications. *Journal of educational administration*, 58(1), 60–80.
31. Osterlind, S. J. (1998). *What is constructing test items?* (pp. 1-16). Springer Netherlands.
32. Pandey, D. K., Prasad, M. S., Krishna, K. V. S. M., Saxena, S., & Jain, V. (2022). An Assessment of Positive and Negative Aspects of Organisational Politics: A Systematic Literature Review on Psychological Wellbeing of Employees. *Journal for ReAttach Therapy and Developmental Diversities*, 5(2s), 80-85.
33. Patten, M. L. (2016). *Questionnaire research: A practical guide*. Routledge.
34. Razak, N. B. A., bin Khairani, A. Z., & Thien, L. M. (2012). Examining Quality of Mathematics Test Items Using Rasch Model: Preliminary Analysis. *Procedia-Social and Behavioral Sciences*, 69, 2205-2214.
35. Rimfeld, K., Malanchini, M., Hannigan, L. J., Dale, P. S., Allen, R., Hart, S. A., & Plomin, R. (2019). Teacher assessments during compulsory education are as reliable, stable and heritable as standardized test scores. *Journal of Child Psychology and Psychiatry*, 60(12), 1278-1288.
36. Rose, J., & Johnson, C. W. (2020). Contextualizing reliability and validity in qualitative research: Toward more rigorous and trustworthy qualitative social science in leisure research. *Journal of leisure research*, 51(4), 432–451.
37. Saravanan, B., & Vasanth, S. (2023). Assessment of Adaptability through Psychological Well-Being during Covid Crisis for Job Retention. *Journal for ReAttach Therapy and Developmental Diversities*, 6(3s), 619-629.
38. Sudarsana, I. K., Nakayanti, A. R., Sapta, A., Satria, E., Saddhono, K., Daengs, G. A., ... & Mursalin, M. (2019, November). Technology application in education and learning process. In *Journal of Physics: Conference Series* (Vol. 1363, No. 1, p. 012061). IOP Publishing.
39. Sudijono, A. (2009). *Pengantar Evaluasi Pendidikan*. Jakarta: Rajawali Press.
40. Sukardi. (2008). *Evaluasi Pendidikan: Prinsip dan Operasionalnya*. Jakarta: Bina Aksara.
41. Šumak, B., Pušnik, M., Heričko, M., & Šorgo, A. (2017). Differences between prospective, existing, and former users of interactive whiteboards on external factors affecting their adoption, usage and abandonment. *Computers in Human Behavior*, 72, 733-756.
42. Sun, H., Steinkrauss, R., Tendeiro, J., & De Bot, K. (2016). Individual differences in very young children's English acquisition in China: Internal and external factors. *Bilingualism: Language and Cognition*, 19(3), 550-566.
43. Tanjung, E. F., Harfiani, R., & Sampedro Hartanto, H. (2021). Formation Of Soul Leadership Model In Indonesian Middle Schools. *Educational Sciences: Theory and Practice*, 21(1), 84-97.
44. Tus, J. (2020). Academic stress, academic motivation, and its relationship on the academic performance of the senior high school students. *Asian Journal of Multidisciplinary Studies*, 8(11), 29–37.
45. Vyas, A. (2022). Developmental Assessment of American Jewish Belonging Behaviour with Special Reference to Autism Spectrum Disorder-A Case Study. *Journal for ReAttach Therapy and Developmental Diversities*, 5(1s), 87-94.

46. Wandasari, Y., Kristiawan, M., & Arafat, Y. (2019). Policy evaluation of school's literacy movement on improving the discipline of state high school students. *International Journal of Scientific & Technology Research*, 8(4), 190-198.
47. Wilson, D. M., & Narasuman, S. (2020). Investigating Teachers' Implementation and Strategies on Higher Order Thinking Skills in School Based Assessment Instruments. *Asian Journal of University Education*, 16(1), 70-84.
48. Xasanova, D. (2020). The role of test making in learning process. *The Light of Islam*, 2(3), 166–174.
49. Zainil, M., Kenedi, A. K., Rahmatina, R., Indrawati, T., & Handrianto, C. (2023). The influence of a STEM-based digital classroom learning model and high-order thinking skills on the 21st century skills of elementary school students in Indonesia. *Journal of Education and e-Learning Research*, 10(1), 29-35. <https://doi.org/10.20448/jeelr.v10i1.4336>