At-Ta'dib. Vol. 19. No. 1, Juni 2024
ISSN: 0216-9142
DOI: http://dx.doi.org/10.21111/attadib.v19i1.11970
Available online at:
https://ejournal.unida.gontor.ac.id/index.php/tadib/index
e-ISSN: 2503-3514

Assessing the Effectiveness of CAT and MCAT in Measuring Students' Understanding in Technology Education

Muslem Daud*
Universitas Serambi Mekkah, Banda Aceh, Indonesia
muslem.daud@serambimekkah.ac.id

Bor-Chen Kuo
National Taichung University of Education, Taiwan
kbc@mail.ntcu.edu.tw

Jamaluddin
Universitas Islam Negeri Ar-Raniry, Banda Aceh, Indonesia
jamaluddin@ar-raniry.ac.id

Muhammad Duskri
Universitas Islam Negeri Ar-Raniry, Banda Aceh, Indonesia
m.duskri@ar-raniry.ac.id

Cut Morina Zubainur
Universitas Syiah Kuala, Banda Aceh, Indonesia
cutmorinazubainur@usk.ac.id

Jalaluddi
Universitas Serambi Mekkah, Banda Aceh, Indonesia
Email: jalaluddin@serambimekkah.ac.id

Received February 15 2024, Accepted June 08 2024

Abstract

This research aims to do two things. First, it aims to make a computer adaption test (CAT) for Islamic subjects for the santri in ABC Pesantren. Second, it aims to see the results of the CAT test for Islamic subjects in the production of Learning Medical Records. The records is for diagnosing students' abilities. This research consists of two parts using the research

and development (R&D) approach: 1). We collected the data by observing, documenting, and interviewing. We analyzed the data using the concept of reliability among the raters/respondents. We used the generalizability coefficient in Genova 3.1. We analyzed the results using numbers. They describe data from RMP technology education. The research shows that: (1) the CAT can be made based on user needs. It's Internet-based, user-friendly, interactive, highly secure, and easily accessible. (2) The CAT can recognize three different users: school management, teachers, and students. (3) The expert evaluation's quantitative data has high consistency: 0.76. It comes from single, small group, and large group outcomes. (4) The CAT can accurately estimate students' abilities and record learning records. The software uses the triangle decision tree method. It can pick test items and measure ability well. This is clear from the correlation (r) between ability and the software's CAT scores.

Keywords: *CAT, MCAT, assesment, effective*

INTRODUCTION

Education in the pesantren environment, especially in the field of technology, is a subject at the primary and secondary levels that is developed from the main (fundamental) teachings contained in Islam. Therefore, it is an inseparable part of the Islamic teachings. It is given in accordance with the guidance that people must learn in order to realize people who are devoted to Allah SWT and have a noble character, and it aims to produce useful people. Pesantren education, especially in a modern environment, emphasizes how students can master Islamic studies while being able to practice them in everyday life in the midst of society in this modern era. Thus, education in the pesantren environment emphasizes not only the cognitive aspect, but more importantly the affective and psychomotor aspects.[1]

CAT can be applied to all school subjects. With CAT in pesantren subjects, the results of learning medical record (RMP) can

[1] *Corresponding Author Ulfatun Wahidatun Nisa, "ISLAMIZATION OF KNOWLEDGE AND ITS CHALLENGE," dalam *PROCEEDING OF INTERNATIONAL CONFERENCE ON EDUCATION, SOCIETY AND HUMANITY* (INTERNATIONAL CONFERENCE ON EDUCATION, SOCIETY AND HUMANITY, Probolinggo: UNUJA, 2023), 2.

be seen as the output of CAT, which is very useful for teachers to diagnose students' abilities during the implementation of teaching and learning process.[2]

As seen from various fields, the development of technology today is getting faster and faster. In aspects of human life where manual data processing is gradually abandoned and replaced by computer systems, the computer field is developing very rapidly. The industrial era 4.0 is developing very rapidly. It includes connectivity devices or networks in data acquisition and processing, automatic network devices, Internet of Things (IoT), big data, cloud computing, artificial intelligence, machine learning, and cybersecurity systems.

Developing computing technology also has significant implications for developing education. One of them is the convenience of testing. One solution for conducting digital assessments is the presence of a computer-based testing system, often known as computer-based testing (CBT).

In the CBT type of testing, each participant works on a certain number of items with a fixed number of items. According to Santoso, administering tests that give the same number of items to each test-taker using CBT is considered less efficient, especially for low and high ability test-takers.[3] The reason for this is that some items may not be able to provide information for discrimination between test takers within a particular ability range. High ability test takers are given some easy items that they have a low chance of answering incorrectly. Thus, such items provide no information about their ability. Conversely, low ability test takers are given some difficult items on which they have a small chance of answering correctly. As a result, incorrect answers provide little information about their

[2] M. Phankokkruad dan K. Woraratpanya, "An Automated Decision System for Computer Adaptive Testing Using Genetic Algorithms," dalam *2008 Ninth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing* (2008 Ninth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, Phuket, Thailand: IEEE, 2008), 48, https://doi.org/10.1109/SNPD.2008.118.

[3] F. M. Lord, *Applications of Item Response Theory To Practical Testing Problems*, 0 ed. (Routledge, 2012), 34, https://doi.org/10.4324/9780203056615.

ability. Giving test-takers the same number of items also results in ineffective tests, because tests that give the same items ignore the diversity of test-takers' abilities, when in fact test-takers' abilities are diverse.

Since the 1980s, a computerized testing system based on adaptive test design has been in development in the United States, popularly called computerized adaptive testing (CAT). CAT is in contrast to CBT, where the computer is only used for the transfer of tests from paper forms to computer monitors. In CAT, the computer is set up and used to display items that match the individual ability of the test taker; because the test is adapted to the ability of the test taker, there are no questions that are harder or easier. For test takers with high ability, only high difficulty items are given, while easy items do not need to be given, and vice versa for test takers with low ability, only low difficulty items are given, while difficult items do not need to be given. This makes the CAT more efficient, and it also makes the CAT more reliable, because the measurement error will be smaller, since each individual will only be given items that are appropriate for his or her ability.[4]

A test is a set of questions with correct or incorrect answers. A test is also defined as a set of questions to be answered, or a set of statements to be answered, designed to measure someone's level of ability or to reveal certain aspects of the person tested (the testee). Test results are information about the characteristics of a person or group of people. Tests are a way of estimating the level of human ability indirectly, through a person's response to a series of stimuli or questions. Therefore, reliable and valid tests are needed to obtain accurate information.

As technology developed, so did the use of computers in testing. Initially, computers were used only to automate common measurement activities. Tests that were originally on paper were moved into the computer. This use of computers is called

[4] Neville Bennett, Walter R. Borg, dan Meredith D. Gall, "Educational Research: An Introduction," *British Journal of Educational Studies* 32, no. 3 (Oktober 1984): 32, https://doi.org/10.2307/3121583.

computerized testing or computerized-based testing (CBT) and is the first generation of computer use for testing. Bunderson, Inouye, & Olsen in Santoso stated that "several advantages of CBT, namely, increasing standardization, increasing test security, improving test display capabilities, minimizing measurement error, and speeding scoring and interpretation.[5]

Computerized Adaptive Testing (CAT) is the second generation of the use of computers for testing in Santoso (2010). Adaptive means that the items given are adjusted to the ability level of each test taker or tailored testing. In CAT, based on Item Response Theory (IRT), items are not simply entered into the computer, but the computer is set to select and present items according to test takers' estimated ability levels. As a result, individual test takers with high ability levels are presented with items that are more difficult than individuals with low ability levels. Conversely, individual test takers with low ability levels will receive easier items than individual test takers with high ability levels. Therefore, the CAT is more efficient in estimating the ability of test subjects with a smaller number of items than the P&P test and the CBT without reducing the measurement accuracy.[6]

The current modern test theory is item response theory (IRT), which makes several basic assumptions. The most important assumption is that the chance that someone will answer one item correctly is not determined by the chance that they will answer another item correctly, which is known as the independent assumption. This modern theory attempts to develop an analysis that provides an estimate of a person's ability without being influenced by the measuring instrument used. Similarly, the item statistics should be independent of the characteristics of the person being measured.

The origin of IRT is the combination of a version of the phi-gamma law with item factor analysis. This was later called Latent Trait Theory and is now commonly known as Item Response Theory

[5] Lord, *Applications of Item Response Theory To Practical Testing Problems*, 32.

[6] Bor-Chen Kuo, Muslem Daud, dan Chih-Wei Yang, "Multidimensional Computerized Adaptive Testing for Indonesia Junior High School Biology," *EURASIA Journal of Mathematics, Science and Technology Education* 11, no. 5 (25 Agustus 2015): 31, https://doi.org/10.12973/eurasia.2015.1384a.

.[7] According to Hambleton, Swaminathan, & Rogers (1991), item response theory rests on two propositions. (A) that the performance of an individual on an item is predicted by an underlying set of factors referred to as latent factors, and (B) that the relation between the performance of an individual on an item and the underlying set of factors corresponds to some monotonic increasing functional graph referred to as the item response curve. This item characteristic curve illustrates that the probability of answering an item correctly increases as the ability level of the test taker increases.[8]walla

The difficulty of an item is defined as the proportion or percentage of test takers who answer a particular test item correctly. The parameter b refers to the point on the ability scale at which a respondent has a 50% chance of answering the item correctly. The larger b, the more difficult the item. When items are assumed to have varying parameters of discriminability (expressed by a), the 2P model is more appropriate. The assumption used to obtain good question quality, especially in terms of question difficulty, is that there is balance in addition to meeting validity and reliability. The intended balance is the existence of questions that contain easy, medium, and difficult questions in proportion. The level of difficulty of the question is seen from the ability or capacity of the students to answer it, not from the point of view of the teacher as a question maker. An important issue in analyzing the difficulty of questions is to determine the proportion and criteria for questions that are easy, moderate, and difficult.

The discriminatory power (D) of a question is the ability of a question to distinguish between test takers who are good (high performance) and test takers who are less good (low performance). Suryabrata in Mansyur states that "the main purpose of looking for discrimination power is to determine whether the item has the ability

[7] Herbert Fliege dkk., "Evaluation of a Computer-adaptive Test for the Assessment of Depression (D-CAT) in Clinical Application," *International Journal of Methods in Psychiatric Research* 18, no. 1 (Maret 2009): 76, https://doi.org/10.1002/mpr.274.

[8] Virgil Zeigler-Hill dan Marion T. Wallace, "Self-Esteem Instability and Psychological Adjustment," *Self and Identity* 11, no. 3 (Juli 2012): 47, https://doi.org/10.1080/15298868.2011.567763.

to distinguish groups on the aspect being measured, according to the differences that exist in these groups".

Discrimination analysis (item discrimination) examines items with the goal of knowing the question's ability to discriminate between able and less able respondents. that item banks for CAT purposes should contain items that have an even distribution of discrimination power ranging from 0.4 to 2.0.[9]

Computer Assisted Test (CAT) is a comparative study of countries that have used Computer Assisted Test (CAT) for example the Civil Service Commission in the Philippines. Computer Assisted Test (CAT) was developed with several modifications to fit the norms, situations and conditions of the existing employment system in its application in Indonesia. CAT testing aims at expediting the examination and review of test results, providing uniform test results throughout the country, presenting uniform results, and eliminating the possibility of collusion, corruption, and favoritism in the selection process of future officials. One of the most important advantages of CAT is the effectiveness and efficiency in the implementation of the test, which is reflected in the fact that the results of the test are available immediately upon the solution of the problem. Computer Assisted Test (CAT) application, as the passing grade will appear at the end of the question.

Computer Assisted Test (CAT) is a method of administering tests that is assisted by a computer. Computer Assisted Test (CAT) is used as a testing medium and its forms vary, ranging from the simplest is the computer displays test questions that replace paper answer sheets to being used to build student skills with active questions and answers.

Based on the author's preliminary study at the Pagar Alam College of Technology on New Student Admissions conducted at the campus of the Pagar Alam College of Technology, which is currently conducted with a computer-assisted test. The results

[9] David J. Weiss, "Computerized Adaptive Testing for Effective and Efficient Measurement in Counseling and Education," *Measurement and Evaluation in Counseling and Development* 37, no. 2 (Juli 2004): 11, https://doi.org/10.1080/07481756.2004.11909751.

showed that from the percentage of prospective students of Pagar Alam Technology High School who registered per year seen from the Academic Section data that in 2012 the number of registrants was 30 people, 2013 as many as 180, in 2014 as many as 212, in 2015 as many as 212, in 2015 as many as 212, and in 2014 as many as 3,000.

Based on this data, the increase in the number of registrants makes it necessary to improve the speed, accuracy, effectiveness and efficiency of testing services. This is evidenced by the implementation of the freshman entrance exams in 2015 and 2016, which used computerized testing, allowing for more effective and efficient administration of the exams, as well as the ability to know the outcome of the exams within several days, compared to the implementation of the exams in 2012, 2013, and 2014, which still used the worksheet.

The computer-based test is administered through a computer where users can see, hear, interact, and control the media display. Without these characteristics, existing products are polymedia or mixed media. Polymedia is media that is not controlled by a computer system, but is a mixture of several media and tools, such as television, radio, textbooks, etc. Multimedia products must have relationships that allow the user to move from one interface to another, and at any given time be supported by separate structures and spaces. Without this property, the product is more like a book. The product must have the Navigation Tools property, where navigation is an icon, button, hot spot, or active device. Active means that the icon or button attempts to connect the user to the multimedia product. The media is interactive, where the user can interact with and control the media. Without these characteristics, the product is more of a non-interactive medium, such as television (TV) [10]Original tools, layouts, graphics, content, videos and animations were not designed attractively in the development of Computer Assisted Test (CAT) applications. Further

[10] Wim J. Van Der Linden dan Peter J. Pashley, "Item Selection and Ability Estimation in Adaptive Testing," dalam *Computerized Adaptive Testing: Theory and Practice*, ed. oleh Wim J. Van Der Linden dan Gees A.W. Glas (Dordrecht: Springer Netherlands, 2000), 99, https://doi.org/10.1007/0-306-47531-6_1.

developments will develop a more dynamic Computer Assisted Test (CAT) in terms of more attractive and convenient tools, layouts, graphics, content, videos and animations.

METHOD

This type of research uses research and development (R&D). In the development of CAT software for Islamic Religious Education (PAI) Class VII at the junior high school education level, there are two stages of implementation to be carried out, namely the first stage is the product development stage and the second stage is the product implementation stage. The development procedure in this study was carried out within a certain period of time. The conducted research phase was selected as essential and has to pass through a product design, i.e. (1) needs analysis and identification, (2) system design, (3) product manufacture (coding), (4) product feasibility test, (5) product testing, (6) implementation, and (7) product revision.

The purpose of the test results is to find out the performance capabilities of the program so that the results can be recommended for standardization and be used to support improvements in testing/evaluation/licensing and quality of testing instruments. The school where the CAT was administered was selected using a purposive sampling technique. That is, a school that has a computer lab and is equipped with a LAN (Local Area Network). The implementation site was ABC Pesantren, which has a computer laboratory equipped with a LAN network.

The purpose of the test design was the determination of the capability and feasibility of the CAT software. The testing process was conducted using a set of computer units and was conducted in the computer laboratory. In the implementation, each student is prompted by the computer for some information including: login name, email, first name, last name, user level, and password. The student then creates an account with the user name and password. After the student logs in, the computer displays questions taken from the question bank in CAT. In this study, questions on Islamic religious education topics are used. The results of the student's

answers are given a value of 1 if the answer is correct and a value of 0 if the answer is incorrect, the ability results (θ) are then converted into a score with a range of 0 - 100. The score conversion calculation is shown in Table 1 below.

Table 1. Conversion of Ability (θ) into score (y)

Type	Learning mastery		
	Unfinished	Finished	Finished Very Well
Ability (θ)	-4,00 ≤ θ 1,00	-1,00 ≤ θ 1,00	1,00 ≤ θ ≤ 4,00
Normal Reference Valuation (x)	0,0 ≤ x M-0,5s	M-0,5s ≤ x M+1,5s	M+1,5s ≤ x ≤100
Interval Score 0 – 100	0,0 ≤ y 37,0	37,0 ≤ y 62,5	62,5 ≤ y ≤ 100

Description:
N = Value
M = Mean
s = Standard deviation
y = 12.5 θ + 50

This research was conducted in: (a) ABC Pesantren Computer Lab, for CAT software development process; (b) PAI Teacher User Trial Computer Lab and ABC Pesantren Student User Trial Computer Lab as place for CAT software product implementation/research in real situations.

The subjects used in the research for the pilot test were: (a) 8 PAI teachers for the process of information extraction and identification of system requirements needed in the CAT software and also to obtain information on the selection of test materials, (b) IT experts to determine the feasibility of the CAT software, and (c) 20 pesantren to test the functionality and performance of the CAT software program in creating a learning medical file (RMP) for diagnosis of pupil ability.

RESULT & DISCUSSION

Table 2. CAT Validation

No	Respondent	Aspect	Mean	$p\hat{}_{2p}$*	Note
1	Expert	Usage Performance	4,0	0,73	High
		Display Performance	3,9	0,48	Very High
		Relevance of Test Materials	4,3	0,70	High
		Expediency	4,0	0,75	High
2	Small group	Usage Performances	3,7	0,78	High
		Display Performance	3,9	0,7	High
		Relevance of Test Materials	3,9	0,70	High
		Expediency	4,0	0,82	High
3	Large group	Usage Performance	4,5	0,87	High
		Display Performance	4,2	0,74	High
		Relevance of Test Materials	3,8	0,85	High
		Expediency	4,0	0,63	Very High

Based on the table above, it shows that the respondents had a fairly high and high consistency in validating the CAT software. The grand mean average value of all aspects is: 4.02, while the grand mean value of CAT software can be used for research.

Table 3. Student trial results

Students	Theta	Score	Test Score	Notes
1	-0,5	43,75	70	Well done
2	0,1	51,25	95	Well done
3	-0,2	47,5	73	Well done
4	-1,5	31,25	60	Well done
5	-0,7	41,25	68	Well done
6	-0,9	38,75	70	Well done
7	1,5	68,75	97	Excellent completion
8	-0,2	47,5	78	Well done

9	0	50	90	Well done
10	-0,8	40	63	Well done
11	-0,6	42,5	83	Well done
12	-0,7	41,25	88	Well done
13	-0,6	42,5	88	Well done
14	0	50	95	Well done
15	1,4	67,5	100	Excellent completion
16	0,1	51,25	90	Well done
17	0,2	52,5	80	Well done
18	-0,9	38,75	80	Well done
19	-0,5	43,75	95	Well done
20	-0,1	48,75	78	Well done

The ability of the CAT PAI program to predict student ability using the test items was determined through beta testing. The students' names were represented by numerical attributes to protect against unwanted things. The results of the program's work, tested on a total of 20 students in the trial, are shown in Table 3 above. The test results are as follows: (1) on average, the subjects received 15 PAI items, the least number of subjects received 13 items, and the most number of subjects received 18 items; (2) the lowest ability of the tenth subject is 1.5, while the highest ability of the forty-sixth subject is 1.5; the average ability of the subjects is 1; and (3) 18 subjects passed the test, and 2 subjects passed the test.

It can be seen that the inference system based on the CAT program has been successful in selecting test items that are appropriate and in accordance with the ability of the participants based on the results of the data analysis of the 20 students. Participants of medium ability are given items of medium difficulty. Participants with low ability are given items of low difficulty. On the other hand, the number of items each participant receives is different (not equal). Participants with high ability or participants with low ability receive many fewer items than participants with medium ability. This shows that the PAI CAT built using the decision triangle tree also worked well with the participants' abilities (adaptive).

The precision and accuracy of the CAT model inference system in selecting test items in accordance with students' ability can be seen from the results of the beta test (students), with the results that students with high pure test scoring (PTS) at school had high ability, students with medium PTS at school had medium ability, and students with small PTS at school had small ability. The correlation value (r) between the results of the ability estimation on the CAT and the NUM PAI at school is quite high, namely r = 0.69.

CONCLUSION

Based on the results of research and discussion, it can be concluded that Computer Adaptive Test (CAT) developed for Islamic Religious Teaching (PAI) subjects in Pesantren is web based and the development uses PHP programming language and the database system uses My SQL. The results of the learning medical record consist of components: question number, ID, question number, true/false statement, time, difficulty level, differential power, guessing rate, competency standards (SK), basic competencies (KD), indicators, and theta. The accuracy and precision of the CAT model inference system in selecting test items according to students' abilities is realized from the results of the beta test (students) with the result that students who have high pure test scores (NUM) of PAI in school also have high abilities (θ), The correlation value (r) between the results of the ability estimation (θ) on the CAT and the pure test scores (NUM) of PAI in students' schools was quite high, namely r = 0.69.

REFERENCES

Amirzan. (2021). Implementing Discipline In The Covid-19 Period In Improving Student Achievement, . *Jurnal Serambi Ilmu, Vol 22, No 2*.

Ashabul Kahfi. (2020). Tantangan Dan Harapan Pembelajaran Jarak Jauh Di Masa Pandemi Covid 19, . *Dirasah, Vol. 03 No.2*.

Assidiqi, M. H. , & Sumarni, W. (2020). Pemanfaatan Platform Digital dalam Pembelajaran Daring di masa Pandemi Covid-19. . *Prosiding Seminar Nasional Pascasarjana UNNES*.

Balasheva, I. V., & Pretrova, N. F. (2016). Psychological stability of a personality and capability of tolerant interaction as divers manifestations of tolerance, International . *Journal of Environmental and Science Education, Vol. 11 (10)*.

Basar, A. M. (2021). Problematika Pembelajaran Jarak Jauh Pada Masa Pandemi Covid-19. *Edunesia : Jurnal Ilmiah Pendidikan*, *2*(1), 208–218. https://doi.org/10.51276/edu.v2i1.112

Bennett, Neville, Walter R. Borg, dan Meredith D. Gall. "Educational Research: An Introduction." *British Journal of Educational Studies* 32, no. 3 (Oktober 1984): 274. https://doi.org/10.2307/3121583.

Daud, M. , Mok M.C.M, & Kuo, B. C. ,. (2015). The Construction of an online curriculum-based multidimensional computerized adaptive testing, I. *International Multidisciplinary Conference (Pp.322-328)*.

Daud, M. , Mok M.C.M., & Kuo, B. C. ,. (2015). The Construction of an online curriculum-based multidimensional computerized adaptive testing, In International Multidisciplinary Conference (pp.322-328). . *Banda Aceh. Universitas Serambi Mekkah*.

Daud, M. , Mok, M. C. M., Yusrizal, Y. , & Kuo, B. C. ,. (2018). *Computerised Adaptive Testing: Application and Implications for Teaching and Learning in Indonesia, in Routledge International Handbook of Schools and Schooling in Asia, .* Routledge Inc.

Daud, M. , Mutia, R., Imran, Israwati, Yusrizal, & Sariakin. (2020). An Evaluation of Indonesian Students' Self-Confidence in Learning Mathematics, . *Proceeding ICMR* .

Fitria, S., & Surya, D. (2021). The Impact of Covid-19 Pandemic on Students' Sedentary Behavior in Indonesia. *JURNAL SERAMBI ILMU*, *22*(1), 88–97. https://doi.org/10.32672/si.v22i1.2732

Fliege, Herbert, Janine Becker, Otto B. Walter, Matthias Rose, Jakob B. Bjorner, dan Burghard F. Klapp. "Evaluation of a Computer-adaptive Test for the Assessment of Depression (D-CAT) in Clinical Application." *International Journal of Methods in Psychiatric Research* 18, no. 1 (Maret 2009): 23–36. https://doi.org/10.1002/mpr.274.

Kuo, Bor-Chen, Muslem Daud, dan Chih-Wei Yang. "Multidimensional Computerized Adaptive Testing for Indonesia Junior High School Biology." *EURASIA Journal of Mathematics, Science and Technology Education* 11, no. 5 (25 Agustus 2015). https://doi.org/10.12973/eurasia.2015.1384a.

Lord, F. M. *Applications of Item Response Theory To Practical Testing Problems*. 0 ed. Routledge, 2012. https://doi.org/10.4324/9780203056615.

Mok, M. M. C., & Zhu, J. (2014). Cluster analysis of attitudes towards feedback and their mathematics achievement: a study of Hong Kong primary students. . *Asia Pacific Journal of Educational Development, 3(2), 1-13*.

Phankokkruad, M., dan K. Woraratpanya. "An Automated Decision System for Computer Adaptive Testing Using Genetic Algorithms." Dalam *2008 Ninth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing*, 655–60. Phuket, Thailand: IEEE, 2008. https://doi.org/10.1109/SNPD.2008.118.

Prandika, D. D., & Amrullah, M. (2021). The Effect of the Covid 19 Pandemic on the Online Learning Process. *Proceedings of The ICECRS*, *10*, 63–71. https://doi.org/10.21070/icecrs20211058

Rafiq, R., & Shah, P. A. ,. (2015). *Teachers' role and stress relationship: A Review, The International Journal of Human Resource Management and Industrial Research, Vol 2, (1), 8-17*.

Riskiana Rahma, & Nursiwi Nugraheni. (2020). Strategi Pembelajaran Guru Sekolah Dasar Di Era Pandemi Covid-19,. *Prosiding Webinar Nasional IAHN-TP Palangka Raya 2020*.

Sarah, S. (2022). High Order Thinking Skills (Hots) Of Ibtidaiyah Madrasah Teacher Education Program (Pgmi) Study Program Students During The Covid-19 Pandemic, . *Jurnal Serambi Ilmu, Vol 23, No 1 (2022)*.

Thu Nguyet, P. T., & Daud, M. (2021). Computer Adaptive Test Development To Assess Students' Psychology. *JURNAL SERAMBI ILMU*, *22*(1), 139–149. https://doi.org/10.32672/si.v22i1.2760

Ulfatun Wahidatun Nisa. "ISLAMIZATION OF KNOWLEDGE AND ITS CHALLENGE." Dalam *PROCEEDING OF INTERNATIONAL CONFERENCE ON EDUCATION, SOCIETY AND HUMANITY*. Probolinggo: UNUJA, 2023.

Van Der Linden, Wim J., dan Peter J. Pashley. "Item Selection and Ability Estimation in Adaptive Testing." Dalam *Computerized Adaptive Testing: Theory and Practice*, disunting oleh Wim J. Van Der Linden dan Gees A.W. Glas, 1–25. Dordrecht: Springer Netherlands, 2000. https://doi.org/10.1007/0-306-47531-6_1.

Walter, O. B., Becker, J., Bjorner, J. B., Fliege, H., Klapp, B. F., & Rose, M. (2007). Development and evaluation of a computer adaptive test for 'Anxiety' (Anxiety-CAT). *Quality of Life Research*, *16*(S1), 143–155. https://doi.org/10.1007/s11136-007-9191-7

Weiss, David J. "Computerized Adaptive Testing for Effective and Efficient Measurement in Counseling and Education." *Measurement and Evaluation in Counseling and Development* 37, no. 2 (Juli 2004): 70–84. https://doi.org/10.1080/07481756.2004.11909751.

Zeigler-Hill, Virgil, dan Marion T. Wallace. "Self-Esteem Instability and Psychological Adjustment." *Self and Identity* 11, no. 3 (Juli 2012): 317–42. https://doi.org/10.1080/15298868.2011.567763.