

THE DEVELOPMENT OF EQUATING APPLICATION FOR COMPUTER BASED TEST IN PHYSICS HOTS CATEGORY

Yetti Supriyati¹, Deni Iriyadi², Ilham Falani³

¹Department of Research and Evaluation in Education, Jakarta State University (Indonesia)

²Universitas Islam Negeri Sultan Maulana Hasanuddin Banten (Indonesia)

³Universitas Indraprasta PGRI Jakarta

y_supriyati@yahoo.com, deni.iriyadi@uinbanten.ac.id, ilhamfalani@gmail.com

Received October 2020

Accepted December 2020

Abstract

This study aims to develop a score equating application for computer-based school exams using parallel test kits with 25% anchor items. The items are arranged according to HOTS (High Order Thinking Skill) category, and use a scientific approach according to the physics lessons characteristics. Therefore, the questions were made using stimulus, options, logical and systematic. This study used Research and Development (R&D) method, which began with an analysis of the current school needs, and takes advantage of technological developments that make it easier for schools to organize. Furthermore, it continues in the literature study to design a school exam model with its devices. The instrument items were analyzed using the RASCH model. In addition, the principle of equating used in application development was the linear equating method.

Keywords – Equating, Application, Computer based test.

To cite this article:

Supriyati, Y., Iriyadi, D., & Falani, I.. (2021). The development of equating application for computer based test in physics hots category. *Journal of Technology and Science Education*, 11(1), 117-128.
<https://doi.org/10.3926/jotse.1135>

1. Introduction

In the industrial revolution 4.0, implementation of examination is proliferating, which provides conveniences including computer-based national exams. In a study, the implementation of computer facilitates data collection, exams administration, data organization, ease of correcting questions, as well as a direct correction system that provides information about the obtained test results (Amin, Ramadhani, Islam, Muhammad & Al, 2018). In fact, Abubakar and Adebayo (2014) showed that the implementation of computer-based exams improved students' experience and abilities in terms of using technology. Other studies also showed that exams using the Computer-Based Test (CBT) concept provided some advantages, such as cost savings, ease of administration, higher accuracy, compatibility of scoring and reporting, flexible exam scheduling, as well as more accessible assessment and reporting of tests. CBT has some advantages such as easiness in accessibility, flexibility, efficiency, and more consistent result than Paper and Pencil Test (Nugroho, Kusumawati & Ambarwati, 2018). Furthermore, it provides precision, innovation,

test creation, more excellent safety, standardization, efficiency, test books and answer sheet elimination, more flexible scheduling, reduces measurement errors, etc. (Hosseini, Abidin & Baghdarnia, 2014). In addition, CBT adherence provides innovations for educational assessment and evaluation.

With computerized exams, students' abilities are quickly and accurately detected without manipulation from any party. The development of exam application has become an innovation to prepare students before taking real exams (Kurniawan, 2015), and facilitates the implementation of CBT (Lumanau, Naga & Arisandi, 2018). Meanwhile, observation results showed that computer-based exam implementation often experience problems, including limited computer availability in schools. Also, the development of exam application can be a solution, because students can directly access and write exam via a personal computer or smartphone. These applications can speed up exams implementation, guarantee questions confidentiality, makes supervision easier, speed up the marking and analysis process without having to be busy with manual corrections, reduce the possibility of malpractice, and the time use (Wijaya, Arifin & Studiawan, 2016). Zeng and Bender (2019) implementation of a CBT system (ExamSoft) at a dental school in the U.S. Guided by the Technology Acceptance Model (TAM) to maximize its potential as an assessment and learning tool.

Each school develops different tests using the same learning indicators, however, the instruments have different quality. The instruments are made of several devices which had anchor items. Based on the results, the number of anchor items was at least 20% of the total (Kolen & Brennan, 2014). This provides an equalization of two or more test sets. Therefore, with the equalization of school examinations, it can provide test equivalence between schools. Several studies have been conducted on this topic, including research by Rahman, Ofianto, & Yefferson (2019) entitled "Development of Higher Order Thinking Skills (HOTS) Test Instruments in Indonesian History Subjects" which resulted in a product in the form of HOTS test instrument used by both teachers and students. Also, research by Zulkifli, Abidin, Razi, Mohammad, Ahmad and Azmi, (2018) entitled "Evaluating the Quality and Reliability of Final Exam Questions for Probability and Statistics Subjects Using the Rasch Model" provided results in the form of questions that have been developed and analyzed with the Rasch measurement model. Liana, Suana, Sesunan and Abdurrahman (2018) research entitled "Development of High-Level Thinking Test Questions for Fluid Materials in Senior High Schools" provided results in the form of questions that have been developed valid with good reliability, and divided into various difficulty levels. Furthermore, research by Widyaningsih and Yusuf (2018), entitled "Analysis of School Physics Laboratory I Module Questions Using the Rasch Model" gave results in the form of questions quality in school laboratory module 1 using the Rasch Model. In this study, the HOTS question was also developed, but with several packages, all of which were parallel, hence they could be horizontally used to equalize scores. Also, the number of questions sets and the presence of anchor items are a differentiator from previous research, and there is also a score equalization application that uses the base of the developed questions. In other words, there is one unit between the questions and the score equalization application. Jodoin, Zenisky, and Hambleton (2006) in the research conducted a psychometric analysis of the CBT instrument that was developed to provide assurance that the instrument was available for use.

Therefore, this study aims to develop an application model of computer-based school exam which has horizontal anchor items in physics subjects. From this study, it is hoped that the computer-based school exams will provide training and experience, and test kits with anchor items will provide security in administering tests. In addition, the items are arranged according to the characteristics of physics subjects and basic competency to be achieved.

2. Method

This study used Research and Development (R&D) method, as well as Borg and Gall's R&D model with qualitative and quantitative approaches. This model can be systematically used to design new products through various stages. R&D aims to develop or validate products used in education, and to find the knowledge that can be practically applied (Gall, Gall & Borg, 2003). This study was conducted in the physics education program of FMIPA, Jakarta State University. Also, it lasted for 12 months, starting from

January 2019 to December 2020. The next steps are the stages carried out in developing the model, which go through the following stages:

2.1. Preliminary Research

Before applying the R&D process, there is a need for special description of the educational product to be developed. After describing the model, various kinds of literature studies were carried out, which were useful in developing the product. This literature study was carried out by studying various theories and previous relevant research.

2.2. Model Development Planning

After the preliminary stage and literature study, the next step is development planning. The essential factors in the planning phase are the estimation of money, human resources and the time needed to develop the product. Also, good planning can help developers avoid waste of work as long as the R&D cycle is established. The plans made are:

2.2.1. Phase I Theoretical Analysis and Preliminary Study

Create an initial product by preparing the necessary materials to support the model by gathering various information and constructing a grid. This is followed by writing items in the HOTS category and making them into 3 test kits, each of which has 25% anchor items. The items prioritize the test construction, which needs to have a stimulus, subject matter and options. For computer-based school exams, a stimulus is made by the characteristics of physics and meets scientific principles, hence the stimulus can be made in the form of videos, animations, charts etc.

2.2.2. Phase II Product Validation

Conduct product validation such as material, media, and assessment in order to review the initial product, and provide input for improvement. When there are deficiencies, an improvement can be made through second revision. However, when there is no revision, then the product is said to be the final.

2.2.3. Phase III Trial

The final product needs to be repeatedly tested and revised in order to obtain a ready-to-use product.

2.3. Validation, Evaluation, and Revision

Validation was carried out on the model to be developed through testing by experts related to the product. Furthermore, their suggestions and input were used for improvements and revisions in the model design. The developed products are in the form of a test assessment model.

2.3.1. Expert Review (Expert Judgment)

The initial field test is a validation of the assessment model, which explains the procedure in developing the model and the experts tasked with observing and providing input. Based on input from these experts, the existing model was revised.

2.3.2. Trial Test

At this stage, product trials were carried out in the field involving high school students. The trial was conducted with each device with as many as 300 students spread across six different schools. Therefore, three valid and reliable test kits were obtained. The input from this field trial will be the basis for improvement and refinement of the final product. After being corrected according to input from the field, the developed HOTS category of computer-based school exam can be considered final and ready to be implemented.

2.3.3. Equalization Tests

The equipment was tested, declared valid, reliable and used for equalization test. Furthermore, the test equalization was carried out to determine the quality equivalence of one exam with the quality of another.

2.4. Research Design

No.	Stages	Activities
1	Preliminary studies	<ul style="list-style-type: none"> Needs analysis of Physics school exams The obtained data were quantitatively analyzed
2	Development studies	<ul style="list-style-type: none"> Developing a model school exam for physics lessons Develop computer-based items with a scientific approach Three test items were made with 25% anchor items, and were qualitatively analyzed and validated by PEP experts, physics lecturers and high school / MA teachers. The 3 draft test kits that had 25% anchor items for computer-based school exams were tested on students and analyzed using the RASCH model
3	Implementation	<ul style="list-style-type: none"> Implementation test of computer-based school exams with HOTS category for physics students 3 test sets have valid and reliable 25% anchor items

Table 1. The research design used

2.4.1. Research Chart

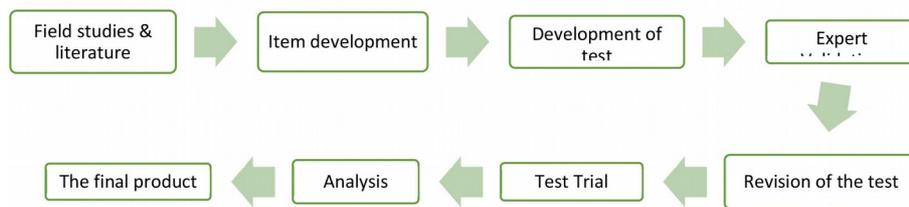


Figure 1. Research Chart

3. Results and Discussion

The result is the development of equivalent 3 test application models, which have horizontal anchor items for computer-based school exams, as well as a score equalization application based on the test sets characteristics. Meanwhile, many experts validated the developed product, and was tested on many students in several schools. Subsequently, the test results were analyzed regarding the items quality and the equivalence level between one test and the other.

3.1. Product Research Application Test Equipment

The developed product has some components, namely:

3.1.1. Homepage



Figure 2. Homepage

The homepage is the first display in any test kit, which can also be referred to as a "cover". On the homepage, there is a "START" button which functions to start the test and takes the user to the next page. Meanwhile, the "PETUNJUK" button functions to provide instructions for using the test kit.

3.1.2. Instructions for User

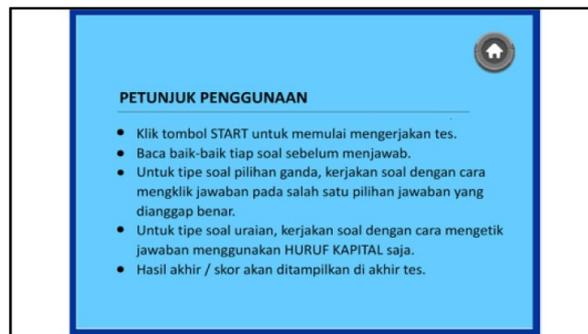


Figure 3. Instructions for User

This section provides information on how to use the exam application.

3.1.3. Student Personal Data Form

The image shows a blue form titled "Form Data Diri Siswa" (Student Personal Data Form). Below the title, it says "Lengkapi nama dan nomor peserta sebelum memulai tes!". There are two input fields: "Nama" with the value "Lelouch" and "No. Peserta" with the value "3217-5-030811-004". At the bottom, there is a blue button labeled "Enter".

Figure 4. Student Personal Data Form

This section asks the students to fill in their bio before starting the test. The requested data in this section are name and number.

3.1.4. Question

In this section, students can randomly answer questions by selecting another number button on the right side. The questions are divided into two:

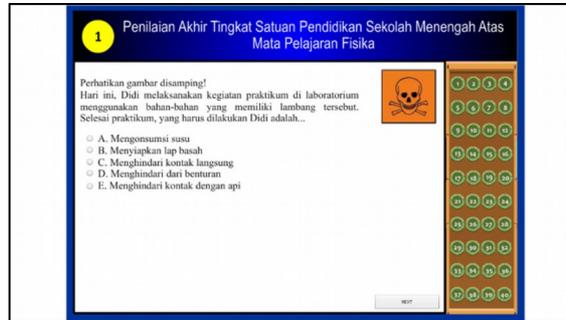


Figure 5. Multiple Choice Questions

In multiple-choice questions, the student answers by choosing the perceived correct option. After selecting the answer, the student clicks the "NEXT" button.



Figure 6. Problem Brief Description

In the short description questions, the student answers by typing the correct answer in the box provided. After typing, the student clicks the "NEXT" button to save.

3.1.5. Test Score

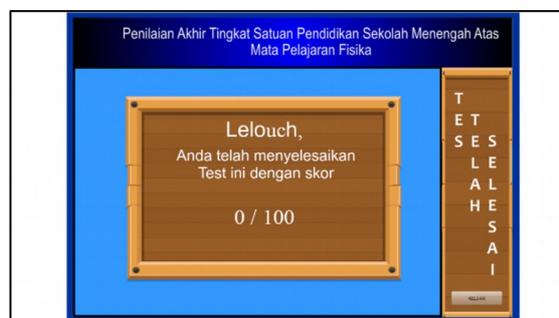


Figure 7. Test Score

In the end, the student will be notified about the results/scores.

3.2. Research Product Equating Application

Based on accuracy level of the equating method, a Computer-Based Test Value Conversion Application was compiled in the HOTS High School Physics Category using the linear method. The following is a display of the application:



Figure 8. Display equating application

The picture above shows how the application was created. Meanwhile, several steps need to be carried out in using the App as follows:

1. Input the number of students whose grades are to be converted.
2. Input the code for the test set done by the student.
3. Click "create table", the number of students inputted will appear.
4. Click "convert", the application will automatically convert the student's score for the test device code which has been inputted to the equivalent value on the test kit.
5. Also, a statistical description for value group for each test set is displayed, including mean, standard deviation, variance, maximum value, and minimum value.
6. Users can also download the conversion result, which is the output of the above application in excel by clicking "Download".

3.3. Theoretical Validation

The developed product then undergo theoretical validation stage for quantitative assessment, and the results determine the test kits quality. This assessment was carried out using a questionnaire containing 11 statement items, and the rating scale used was a scale of 1 - 5 ranging from "very bad" to "very good".

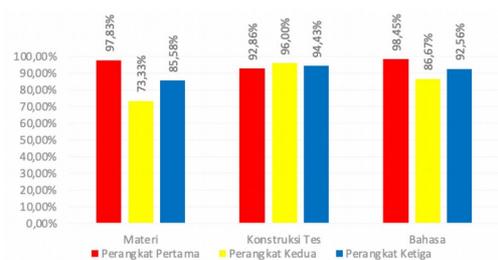


Figure 9. Expert Theoretical Validation Graph

The theoretical validation for the three test sets showed excellent results with a mean score of 96.38% for the first instrument, 85.33% for the second, and 90.86% for the third. Furthermore, the assessed aspects in theoretical validation include material, question construction, and language aspects. The details are as follows: (1) the material was 97.83% (very good) for the first instrument, 73.33% (good) for the second, and 85.58% (very good) for the third. (2) The construction of questions was 92.86% (very good) for the first, 96.00% (very good) for the second, and 94.43% (very good) for the third. (3) Language of 98.45% (very good) for the first instrument, 86.67% (very good) for the second, and 92.56% (very good) for the third.

The next theoretical validation was by educational practitioners. The assessment was carried out using a questionnaire containing 11 statement items and the rating scale used was 1 - 5 ranging from "very bad" to "very good".

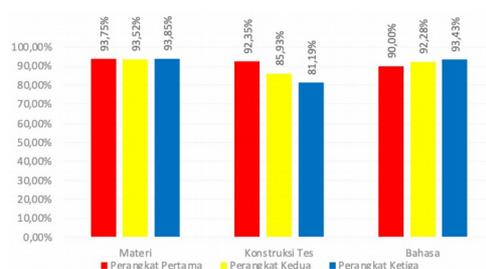


Figure 10. Graph of Educational Practitioners Theoretical Validation

The theoretical validation by educational practitioners for the three test sets showed excellent results with a mean score of 92.03% for the first instrument test, 90.58% for the second, and 89.49% for the third. Aspects that are assessed in theoretical validation by educational practitioners include material, question construction, and language aspects. The details are as follows: (1) The material was 97.75% (very good) for the first instrument test, 93.52% (good) for the second, and 93.85% (very good) for the third. (2) The construction of questions was 92.35% (very good) for the first instrument test, 85.93% (very good) for the second, and 81.19% (very good) for the third. (3) Language of 90.00% (very good) for the first, 92.28% (very good) for the second, and 93.43% (very good) for the third.

3.4. Rasch Model Analysis

The validated products were then revised with suggestions from experts. Subsequently, the revised product was tried out on many grade 12 students, and the results were analyzed using the Rasch model analysis. In the study, the products were 3 test kits with Anchor Items Design. These kits then went through the trial phase, and the results were analyzed using the Rasch model analysis.

The trials were carried out in six different high schools with a total of 300 participants (50 participants per school) for each test set. The obtained data were subsequently analyzed using the WinSteps software. The results of the Rasch model analysis can be seen in the following table.

Rasch analysis	Results
Unidimensionality	The first analysis was carried out on the results of the unidimensionality value, which aims to determine whether the developed instrument can be effectively used for necessary measurements. The analysis results showed that the raw variance value was 77.4% for the first test set, 84.3% for the second, and 83.9% for the third. This showed that the developed instrument is accepted with special criteria and can make necessary measurements.
Item Reliability	Item reliability provides information about the consistency of the questions contained in the instrument. The analysis results showed that the item reliability value is 0.92 for the first test set, 0.92 for the second, and 0.93 for the third. This showed that the developed questions have an excellent reliability level.

Rasch analysis	Results
Fit Items	The fit item provides information about the quality of the developed items. Meanwhile, the analysis results showed that each developed question meets at least one of the criteria for the item to be declared valid. Therefore, the questions are declared valid and can function normally in making measurements.
Measure items	Item measure provides information about the difficulty level of each developed item. Also, the difficulty level of the item is divided into four, namely very easy, easy, difficult, and very difficult (Sumintono & Widhiarso, 2015).
Differential Item Functioning (DIF)	DIF provides information about bias of the developed items. In this study, two demographic data were used, namely gender (male and female) and school origin (Six schools were initialized to SMA A, SMA B, SMA C, SMA D, SMA E, and SMA F). <i>Gender</i> The analysis results for gender demographics showed that each developed item has a DIF value of more than 0.05. This means that the developed questions are not biased towards sex differences. <i>School Origin</i> The analysis results for demographics of school origin also showed that each developed item has a DIF value of more than 0.05. This means that the developed questions are not biased towards differences in school origins.
Person Reliability	Person reliability provides information about the consistency of the answers given by the students. The analysis results showed that the person reliability value was 0.46 for the first instrument test, 0.47 for the second, and 0.44 for the third.
Person Measure	The person measure provides information about the student's ability. This ability is divided into three categories, namely high, medium, and low.
Person Fit	Person fit provides information about students who have inappropriate response patterns. The unsuitable response pattern is the mismatch of the student's ability to the given answers. The analysis results showed that each person fulfils at least one of the criteria for the person to have an appropriate response pattern. Therefore, each student has a response that matches their ability.
Scalogram	The Scalogram provides more information about students who have inappropriate response patterns. This can be in the form of students who are lucky in answering questions (lucky), careless in answering questions (careless), or cheating. The Scalogram results showed that none of the students had an unsuitable response pattern. Therefore, in this case, there are no lucky, careless, or cheating ones.

Table 2. Rasch Analysis

3.5. Equalization Tests

The test equipment was analyzed using Rasch Model, and re-analyzed to determine the equality level between one test set and the other. The results of the test kit equivalence used as a reference in making the score equalization application are as follows.

3.5.1. First and Second Test Form

The first analysis was to determine the equivalence between the first and second test set. The equations for the test set are as follow.

$$Y = 1.034x - 3.985y = 1.034x - 3.985$$

Where: y is the value obtained for the first test set, and x is the value obtained for the second. For example, when a student scores 80 on the second test set, then using the above equation, the score will be 78.735 on the first test set. Although there are differences in values, slight differences are considered equal.

3.5.2. Second and Third Test Form

The second analysis was to determine the equivalence between the second test set and the third. The similarities for both the second and third are as follows.

$$Y = 0.988x + 0.928y = 0.988x + 0.928$$

Where: y is the value obtained for the second test set, and x for the third. For example, when a student scores 80 on the third test set, then using the above equation, the score will be 79.968 on the second. Although there are differences in values, but they are negligible and are considered equal.

3.5.3. Third and First Test Form

The third analysis was to determine the equivalence between the third test set and the first. The equations for the third test set and the first are as follows.

$$Y = 0.939x - 5.994y = 0.939x - 5.994$$

Where: y is the value obtained for the third test set, and x is for the first. For example, when a student scores 80 on the first test set, then using the above equation, the score will be 81,114 on the third. Although there are differences in values, they are negligible and are considered equal.

3.6. Discussion

This study produced three equivalent test device application models with horizontal anchor items for computer-based school exams. Subsequently, many experts validated the developed product, then tested on many students that spread across several schools. The results were then analyzed for items quality and equivalence level between one test and another. Based on the validation results, the three test sets showed that the developed sets are theoretically feasible and can be implemented in school exams.

The test instrument was analyzed using the Rasch Model which was then re-analyzed to determine the equality level between one test and the other. The first analysis is knowing the equivalence between the first and second test set, while the second is knowing the equivalence between the second test set and the third. Furthermore, the third analysis is knowing the equivalence between the third and the first test set. Although there are differences in value between the sets, but they are negligible and can be considered to be equal.

The analysis results showed that the raw variance value was 77.4% for the first, 84.3% for the second, and 83.9% for the third test set. This showed that the developed instrument is accepted with special criteria, and can make necessary measurements. Also, the results showed the item reliability value of 0.92 for the first test set, 0.92 for the second, and 0.93 for the third. This showed that the developed questions have very good reliability level. The analysis results showed that every developed question meets at least one criteria for the item to be declared valid. Therefore, the developed questions are all valid and can function normally in making measurements. Also, the difficulty level of the items are divided into four, namely very easy, easy, difficult, and very difficult. Other analysis related to items that are potentially biased for gender demographics and school origin also showed that each developed item has a DIF value higher than 0.05. This means that the developed questions are not biased towards sex differences.

Apart from the items analysis, there was also analysis on the person test participants. The analysis results showed that the person reliability value is 0.46 for the first test set, 0.47 for the second, and 0.44 for the third. Also, the student's ability was divided into three categories, namely high, medium, and low. The analysis results showed that each person fulfils at least one of the criteria to have the appropriate response pattern. Therefore, each student has a response pattern that matches their ability. The scalogram results confirmed that none of the students had inappropriate response pattern. In this case, there were no students who are lucky, careless, or cheating.

The test used in this study was focused on measuring Higher Order Thinking Skill in Physics. Meanwhile, the development of similar applications can be carried out on subjects that have the same characteristics as physics such as chemistry and mathematics. However, test development will experience challenges for test instruments in the social field. Also, the social field allows student's answers to be more flexible. Therefore, a specific and complex strategy is needed to formulate the items and make measurements in order to obtain equivalent test.

In addition, the limitation of this study is that the equating application development only used a linear method. There is a need for further research in order to compare the accuracy of equalizing scores between the linear equalization method and others. This aims to increase the accuracy of the equations carried out, therefore the developed application can be more accurate.

4. Conclusion

Based on the results and discussion, it can be concluded that the developed items on each test set referred to the curriculum used in Indonesia, and has been declared feasible by expert. The items developed on each test set use a stimulus that helps the student's logical reasoning. Furthermore, the stimulus used was in the form of images, animations, videos, tables, or graphics. Also, the developed three test instruments can make necessary measurements, and the items it contained are valid and reliable. Three of these items are divided into two types of question items, namely anchor and non-anchor. All the item types are valid and reliable with the same level of equality. Even though the developed test kits are equivalent and are valid with reliable terms, the research can still be expanded by re-developing the anchor item questions in vertical form.

Acknowledgements

We sincerely thank Universitas Negeri Jakarta for funding this research.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received financial support for the research, authorship, and/or publication of this article from Universitas Negeri Jakarta.

References

- Abubakar, A.S., & Adebayo, F.O. (2014). Using Computer-Based Test Methods for the Conduct of Examination in Nigeria: Prospects, Challenges and Strategies. *Mediterranean Journal of Social Sciences*, 5(2), 47-56. <https://doi.org/10.5901/mjss.2014.v5n2p47>
- Amin, M., Ramadhani, B., Islam, U., Muhammad, K., & Al, A., (2018). *Implementation of Online Exam Application System for Junior High Schools (SMP)*. Implementation Of Online Exam Application System.
- Gall, M.D., Gall, J.P., & Borg, W.R. (2003). *Educational Research: An Introduction*. Boston: Pearson Education, Inc.
- Hosseini, M., Abidin, M.J.Z., & Baghdarnia, M. (2014). Comparability of Test Results of Computer-Based Tests (CBT) and Paper and Pencil Tests (PPT) among English Language Learners in Iran. *Procedia - Social and Behavioral Sciences*, 98(1), 659-667. <https://doi.org/10.1016/j.sbspro.2014.03.465>
- Jodoin, M.G., Zenisky, A., & Hambleton, R.K. (2006). Comparison of the psychometric properties of several computer-based test designs for credentialing exams with multiple purposes. *Applied Measurement in Education*, 19(3), 203-220. https://doi.org/10.1207/s15324818ame1903_3

- Kolen, M.J., & Brennan, R.L. (2014). *Test Equating, Scaling, and Linking: Methods and Practices* (3rd ed). New York: Springer Verlag New York Inc. <https://doi.org/10.1007/978-1-4939-0317-7>
- Kurniawan, H. (2015). Design of Application for Trial Test (Try Out) for New Student Admissions Selection (SPMB) and Web-Based National Examination (UN). *In the National Seminar on Informatics*, 402-407.
- Liana, N., Suana, W., Sesunan, F., & Abdurrahman, A. (2018). Pengembangan Soal Tes Berpikir Tingkat Tinggi Materi Fluida Untuk SMA. *Journal of Komodo Science Education*, 1(1), 66-78
- Lumanau, J.C., Naga, D.S., & Arisandi, D. (2018). Development of Website-Based Exam Simulation Application Programs at Candra Naya High School. *Computer Science and Information Systems*, 6(2), 104-109.
- Nugroho, R.A., Kusumawati, N. S., & Ambarwati, O.C. (2018, February). Students perception on the use of computer based test. *IOP Conference Series: Materials Science and Engineering*, 306(1), 012103. <https://doi.org/10.1088/1757-899X/306/1/012103>
- Rahman, A., Ofianto, O., & Yetferson, R. (2019). Pengembangan Instrumen Tes Higher Order Thinking Skills (HOTS) Pada Mata Pelajaran Sejarah Indonesia. *PAKAR Pendidikan*, 17(1), 47-59.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan rasch pada assessment pendidikan*. Jakarta: Trim Komunikata.
- Widyaningsih, S.W., & Yusuf, I. (2018). Analisis Soal Modul Laboratorium Fisika Sekolah I Menggunakan Rasch Model. *Gravity: Jurnal Ilmiah Penelitian dan Pembelajaran Fisika*, 4(1). <https://doi.org/10.30870/gravity.v4i1.3116>
- Wijaya, A.Y., Arifin, Z.A., & Studiawan. H. (2016). Development and Management of Online Exam Applications for Madrasahs as a Solution to Various Problems That Arise When Evaluating Students' Teaching and Learning Activities. *Journal of Community Service LPPM UNTAG*, 2(1).
- Zheng, M., & Bender, D. (2019). Evaluating outcomes of computer-based classroom testing: Student acceptance and impact on learning and exam performance. *Medical teacher*, 41(1), 75-82. <https://doi.org/10.1080/0142159X.2018.1441984>
- Zulkifli, F., Abidin, R.Z., Razi, N.F.M., Mohammad, N.H., Ahmad, R., & Azmi, A.Z. (2018). Evaluating quality and reliability of final exam questions for probability and statistics courses using Rasch model. *International Journal of Engineering & Technology*, 7(4.33), 32-36.

Published by OmniaScience (www.omniascience.com)

Journal of Technology and Science Education, 2021 (www.jotse.org)



Article's contents are provided on an Attribution-Non Commercial 4.0 Creative commons International License.

Readers are allowed to copy, distribute and communicate article's contents, provided the author's and JOTSE journal's names are included. It must not be used for commercial purposes. To see the complete licence contents, please visit <https://creativecommons.org/licenses/by-nc/4.0/>.