East Asian Journal of Multidisciplinary Research (EAJMR)
Vol. 2, No. 1, 2023 : 95-108

Bloom's Taxonomy Approach to Cognitive Space Using Classic Test Theory and Modern Theory

Helvy Yunida[1]*, Riyan Arthur[2]
Jakarta State University
Corresponding Author: Helvy Yunida hybindjaji@gmail.com

A R T I C L E I N F O

Keywords: Bloom's Taxonomy, Cognitive Space, Classic Test Theory, Modern Theory

Received : 17 November
Revised : 19 December
Accepted: 21 January

A B S T R A C T

Bloom's Taxonomy is an operational verb (KKO) developed by Benyamin S. Bloom, with the aim of developing education which includes cognitive, affective and psychomotor aspects. The aim of this research is Bloom's Taxonomy Approach To Cognitive Space Using Classic Test Theory And Modern Theory. KKO is needed by a teacher, lecturer or facilitator or trainer to make a syllabus and Learning Implementation Plan (RPP). The Classical Theory Test is one of the test models that is still quite useful today which is still used in research to describe an error in measurement that can affect the observed score. This classical test theory is also an attempt to explain errors in a measurement. While the modern theory test is called the system analysis theory or open theory which combines classical and neoclassical theory. Modern organizational theory sees that all elements of an organization are interdependent and inseparable. The research method uses literature review, namely by collecting data techniques looking for materials from journals, books and from the internet (google, google scholar) and others. The results of the research show that data processing uses classical theory tests and modern theories, variations and modifications are obtained that can develop knowledge in the research field.

DOI: https://10.55927/eajmr.v2i1.2331
ISSN-E: 2828-1519
https://journal.formosapublisher.org/index.php/eajmr/index

INTRODUCTION

Bloom's taxonomy is an operational verb (KKO) developed by Benyamin S. Bloom, with the aim of developing education which includes cognitive, affective and psychomotor. KKO is used to evaluate the quality of learning, so that learning can be measured both cognitively, affectively, and psychomotorically. KKO is needed by a teacher, lecturer or facilitator or trainer to make syllabus and lesson plans. There are six levels or levels of taxonomy bloom from the lowest level to the highest level, starting from C1 to C6. The Cognitive Domain Measurement Tool is one of the bloom taxonomic domains. Learning objectives in the cognitive domain in Bloom's taxonomy according to Benyamin S Bloom all activities related to the brain are divided into six levels according to the lowest level to the highest level which is denoted by the letter C (Cognitive). Aspects of the cognitive domain include C1 knowledge, C2 understanding, C3 application, C4 analysis, C5 synthesis and C6 evaluation. The Classical Theory Test is one of the test models that is still quite useful today which is still used in research to describe an error in measurement that can affect the observed score. This classic test theory is also an attempt to explain the error in a measurement. The steps in this study are: making the Cognitive Measuring Tool Construction Instrument Grid, writing scoring items, conducting validity and reliability tests, item count results and true-false item analysis, different test criteria, results of grouping cognitive domains of Bloom's Taxonomy , and Difficulty Level Test. RESEARCH PURPOSES The purpose of this study was to determine the taxonomic bloom approach with a cognitive measurement tool using the classical test theory. Theoritical Review 1. Bloom's Taxonomy defined as the memory of things that have been learned before. This ability is an initial ability including the ability to know as well as convey memories when needed. This includes remembering materials, objects, facts, phenomena, and theories. The learning outcomes of knowledge are low level. 2) Comprehension, defined as the ability to understand material or materials. The process of understanding occurs because of the ability to translate one material into another material. Understanding can also be demonstrated by the ability to estimate trends, the ability to predict the consequences of various causes of a symptom. Learning results from more advanced understanding than simple memory, rote, or lower level knowledge. 3) Application, is the ability to use material that has been learned and understood in concrete or new situations. This ability includes the use of knowledge, rules, formulas, concepts, principles, laws, and theories. Learning outcomes for the ability to apply this level is higher than understanding. 4) Analysis, is the ability to break down material into parts or components that are more structured and easy to understand. Analytical abilities include identifying the parts, analyzing the links between the parts, and identifying or expressing the organization between the parts. The result of learning analysis is a cognitive level that is higher than the ability to understand and apply, because to have the ability to analyze, one must be able to understand the substance as well as its organizational structure. 5) Synthesis, the ability to think which is the opposite of the analytical thinking process, synthesis is a process that combines parts or elements logically so that they are

transformed into a structured pattern or in the form of a new pattern. 6) Assessment or evaluation, is the highest level of thinking in the cognitive domain according to Bloom's Taxonomy. Self-assessment or evaluation is a person's ability to make judgments about a situation, value or idea.

THEORETICAL REVIEW

The learning outcomes of knowledge are low level. 2) Comprehension, defined as the ability to understand material or materials. The process of understanding occurs because of the ability to translate one material into another material. Understanding can also be demonstrated by the ability to estimate trends, the ability to predict the consequences of various causes of a symptom. Learning results from more advanced understanding than simple memory, rote, or lower level knowledge. 3) Application, is the ability to use material that has been learned and understood in concrete or new situations. This ability includes the use of knowledge, rules, formulas, concepts, principles, laws, and theories. Learning outcomes for the ability to apply this level is higher than understanding. 4) Analysis, is the ability to break down material into parts or components that are more structured and easy to understand. Analytical abilities include identifying the parts, analyzing the links between the parts, and identifying or expressing the organization between the parts. The result of learning analysis is a cognitive level that is higher than the ability to understand and apply, because to have the ability to analyze, one must be able to understand the substance as well as its organizational structure. 5) Synthesis, the ability to think which is the opposite of the analytical thinking process, synthesis is a process that combines parts or elements logically so that they are transformed into a structured pattern or in the form of a new pattern. 6) Assessment or evaluation, is the highest level of thinking in the cognitive domain according to Bloom's Taxonomy. Self-assessment or evaluation is a person's ability to make judgments about a situation, value or idea.

METHODOLOGY

This research method is descriptive by using a literature review. What is meant by descriptive according to Vinayak Bairagi (2019) is: "generally used in business analysis or social problems. This type of research does not have any control over the parameters or variables. It just tries to represent or analyze the previous and or current facts. Whereas what is meant by a literature review according to Vinayak Bairagi (2019) is: "a comprehensive study of technical and authorized content related to research keywords. Where the author collects materials in the form of reputable national and international journals and from books whose data is relevant to the title.

RESULTS

In this study, we used CTT and IRT to examine the replicability of the psychometric properties of the SWLS in a sample of South African teachers. The results obtained support the findings in the literature regarding the psychometric properties of the SWLS as examined through CTT . First, the

mean life satisfaction score for the current sample of teachers in South Africa was found to be significantly lower than the values reported in other countries. Thirdly, both the CTT and the IRT confirmed that the SWLS demonstrated sufficient validity. With regard to construct validity, all items were found to highly correlate with the total scale. the scalability coefficient of the individual items (Hi) indicated that all items contributed to the measurement of life satisfaction.. (Tyrone B. Pretorius 1, 2022) This study examined the reliability and validity of the Swedish version of the self-administered 36-item version of the WHODAS 2.0 by using classical test theory. results demonstrate that the translated form, as applied to a psychiatric outpatient population who assessed their daily functioning, has satisfactory psychometric properties. The strong correlation between the SDS and WHODAS ratings supports its validity as a self-report instrument in all stages of disease. (Robin Midage, 2021)

the authors conclude that the three measures can be used interchangeably. The (Julio César Acosta-Prado 1, 2021) We examined the reliability based on Classical Test Theory calculating Cronbach's alpha for each test, including other indicators such as Difficulty indexes, Discrimination indexes, Means and Standard Deviations. About the reliability of test scores, a widely accepted intervals for describing internal consistency using Cronbach's alpha is as follows:≥0.9 excellent;≥0.8 and<0.9 good;≥0.7 and<0.8 acceptable;≥0.6 and<0.7 questionable;≥0.5 and <0.6 poor;>0.5 unacceptable. Before implementing the changes and the corrections two of them were already acceptable, (Xabier Larrucea1, 2021) Information we gained about the ESSP Social Isolation scale from the CFA and IRT analyses. With the use of a combination of CFA and IRT. The Social Isolation scale had good overall fit according to both CFA and IRT analyses. The CFA approach provides additional information about local fit based on how well the model reproduced the analysis correlation matrix, they can help researchers improve factor models. The IRT model also indicated that individual items had good fit, according to S-χ2-RMSEA values. Findings from both the CFA and IRT analyzes also suggest that the five social isolation items are adequately differentiated among levels of the latent construct or θ. In the IRT framework, (Bowen, 2021).

Rasch analysis also suggested that each factor was unidimensional: the items under the same factor of the MISS-LG were grouped together to demonstrate the same constructs. The concurrent validity of the MISS-LG was also supported by the significant correlations of the three MISS-LG factor scores with HHRS-Homosexuality and AAQ scores. The CFA results indicated that most items had an acceptable factor loading. (Chi Zhang1, 2021)

this is the first study to evaluate the psychometric properties of the GAD-7 among medical university students combining CTT and IRT. We observed a higher prevalence of general anxiety disorders (65.72%) than previous reports in China. The overall α coefficient of the GAD-7 was 0.93 and the Guttmann's coefficient was 0.89. The summed GAD-7 score was statistically significant correlated with scores of the PHQ-9 ($r = 0.78$, $P < 0.001$), PSS-10 ($r = 0.71$, $P < 0.001$), AIS-8 ($r = 0.67$, $P < 0.001$), and SWLS-5 ($r = -0.38$, $P < 0.001$). A

CFA with a weighted least square estimation was used to test the one-factor structure of the GAD-7. The modification index between item 3 and item 4 was 83.52, and the model was modified by establishing the residual covariation correlation between the two items. The adaptability of the modified model was then significantly improved ($\chi 2/df = 3.48$, CFI = 0.97, NFI = 0.96, RMSEA = 0.05) and the factor loading of each item in the CFA model was >0.6. This index CFA mated that the unidimensional structure showed excellent suitability to the data. (Cheng-Fang Yen 1, 2021) (Chi Zhang1, 2021)

this is the first study to evaluate the psychometric properties of the GAD-7 among medical university students combining CTT and IRT. We observed a higher prevalence of general anxiety disorders (65.72%) than previous reports in China. The overall α coefficient of the GAD-7 was 0.93 and the Guttmann's coefficient was 0.89. The summed GAD-7 score was statistically significant correlated with scores of the PHQ-9 ($r = 0.78$, $P < 0.001$), PSS-10 ($r = 0.71$, $P < 0.001$), AIS-8 ($r = 0.67$, $P < 0.001$), and SWLS-5 ($r = -0.38$, $P < 0.001$). A CFA with a weighted least square estimation was used to test the one-factor structure of the GAD-7. The modification index between item 3 and item 4 was 83.52, and the model was modified by establishing the residual covariation correlation between the two items. The adaptability of the modified model was then significantly improved ($\chi 2/df = 3.48$, CFI = 0.97, NFI = 0.96, RMSEA = 0.05) and the factor loading of each item in the CFA model was >0.6. This index CFA mated that the unidimensional structure showed excellent suitability to the data. (Cheng-Fang Yen 1, 2021)

Results from both the EFA and MSA suggested that the factor structure of the AVUPS did not align with the conceptual model. Interestingly, the results from the EFA and MSA generated different factor structures. The EFA resulted in a four-factor structure with 26 items while the MSA was unidimensional and included 20 items. However, using the results from the EFA to inform MSA, resulted in the validation of three Mokken subscales (IU, PB, and well-being) and one whole scale (AVUPS). The 20-item AVUPS (i.e., whole scale) is a strong Mokken scale, which met the criteria for a MHM and displayed excellent internal consistency and excellent 2 weeks test–retest reliability. (Justin Mason1*, 2021) Our goal in developing UCD-11 was to offer a straightforward, descriptive measure that can be used by teams as part of their own reporting processes or alternatively by researchers who may apply it to written reports of design and development processes. UCD-11 is intended as a complement to—not a replacement for—detailed descriptions of the design and development processes of personal health tools and is intended to be applied at the end of a project. As stated earlier, it is a descriptive, not normative, measure. Although Higgins and colleagues [61] offered evidence that higher UCD-11 scores are associated with positive implementation outcomes of a personal health tool, we do not have evidence that higher scores necessarily indicate higher-quality design and development processes. (Holly O Witteman1, 2021) The level of difficulty of the even-even UAS test questions made by the mathematics teacher for grade VIII at SMP Negeri 6 Baubau for the 2020/2021 academic year can be seen by the level of difficulty of the items showing the number of questions that

fall into the difficult category of 13 items (65%) and the medium category of 7 items (35%)). With an average difficulty level of questions (mean P) of 0.283, this shows that the even-even UAS questions made by the mathematics teacher in class VIII at SMPN 6 Baubau have a difficulty level in the difficult category. For the even-even UAS test questions made by the mathematics teacher for class VIII at SMP Negeri 6 Baubau, the ratio of difficulty: medium is 65%: 35%. (Rasmuin 1, 2022)

Based on the results of the analysis of the level of difficulty, differentiating power and effectiveness of the distractor, it was carried out using the Anates V4 program. Furthermore, the analysis based on Bloom's taxonomy is done manually. The following is the result of an analysis of multiple choice items for class VII SMP Negeri 16 Bintan. For example, in question number 31 out of 62 test takers, only 9 test takers could answer correctly and 53 students answered incorrectly. Judging from the proportion of the answers to these questions, they fall into the category of very difficult questions because the difficulty level of these questions is 0.14 or 14.52%. This question is in the very difficult category because it is at a difficulty level index between 0.00-0.19 (Nurgiyantoro, 2017: 219) because of the total number of students only 9 people were able to answer correctly. Therefore, questions in this very difficult category should be discarded or not used again. This shows that this item is quite good at differentiating the ability of students in the upper group and the lower group in line with the opinion of Nurgiyantoro (2017: 222) that questions that are at an index of 0.20-1.00 are questions that have good discriminating power and are appropriate to use. . Discussion of question number 1 results in the percentage of distractor A being 6.45% which has a category of functioning distractors based on the category given Arikunto (2012: 234) the percentage of distractors is more than 5%. The same thing happened to distractor C of 16.1% and distractor D of 9.68%. So, question number 1 has an effective distractor function that is effective or works well. (Guiliani Marescosthi Angelina1, 2022)

Classical test theory introduces three concepts: test scores, correct scores, and error scores. Models of various shapes have been formulated based on this theory. For example, we often refer to the classical test model, which is a simple linear model in which the postulates relate the apparent test score (X) to the sum of two unobservable variables, the pure score (T), and the error score (E), that is, X = T + E. There are seven kinds of assumptions that exist in this classic test theory. Allen & Yen (1979:57) outlines the assumptions of classical theory. (Setyawarno, 2017) The results of the analysis of daily science test items using Anates software on the effectiveness aspect of the distractor are interpreted into 5 categories, namely: very good, good, fair, not good and very bad. In Anates software itself the number of students who answered each option is directly listed on the each answer option, but the names and groups of students who answered each answer option are unknown. (Joelan Maulidina Fiska1, 2021)

The research results show that each item has a CVR value greater than 0.78, so that the item has good content validity. Thus, based on the content validity analysis using the Aiken index and the Lawsche CVR formula, it shows

that each item of the test instrument for the ability to solve mathematics problems can reach the measurement stage of the test instrument. the test reliability index using the internal consistency approach was the Alpha-Cronbach method. The assumptions underlying the classical test theory using this method are that there is no correlation between the actual score and the error score. (Muhammad Rais Ridwan1, 2021)

The statistics produced by CTT-based classics include the level of difficulty of the items, the differentiability of the items, the distribution of answers, the reliability of the test, and the standard error of measurement. In addition to item and test statistics,Classic also provides test participant score analysis results in an easy-to-understand graphical form. Several conveniences are offered by Classica in reporting the results of item analysis and tests. Classica, which is a GUI-based program, makes it easy for users to perform data analysis without syntax or code. The development of Classics is expected to initiate the development of other programs, particularly the psychometric analysis program with a modern test theory approach. Qualitative analysis through a review of the items made by expert judgment, quantitative analysis of the items to determine the characteristics of the items quantitatively. Quantitative analysis of try out questions with a classical test theory approach with the QUEST program to see the level of difficulty of the items. The difficulty level of items in classical test theory can be interpreted as the percentage of test takers who answered certain test items correctly. Good item difficulty index between 0.30 and 0.70 (Dian Normalitasari Purnama1, 2020)

From the results of the analysis obtained using the classical test theory approach consisting of the level of difficulty, discriminating power, and the effectiveness of deceiving three tier test questions (three-level test) level I (content tier) final semester exam (UAS) in Basic Physics 1 for students UIN Alauddin Makassar shows that there are 11 items that have a good category, there are 9 items in the bad category. In the three-tier test questions at level II (reason tier) the research results obtained show that there are still many giving wrong reasons, especially on questions that are in the cognitive domains C1, C2, C3, and C4. Based on the research results obtained, it is known that there are still many who give less confidence, especially on questions that are in the C3 and C4 cognitive domains. (Anas Irwan, 2022)

The steps taken are checking assumptions, estimating the characteristics of the items and the abilities of the participants, and checking the suitability of the model. The percentage of items that can be described by the 1PL, 2PL, and 3PL IRT models is 37.14%; 74.29%; and 82.86%. Based on these results, the appropriate model to describe the characteristics of the items is the logistic parameter response theory model. Good questions are questions number 2, 4, 7, 12, 19, 23, 32, and 33. Good questions are numbers 5, 6, 8, 9, 10, 11, 13, 16, 17, 18, 21, 22, 24, 25, 26, 27, 28, 29, 30, 34 and 35. While the items that cannot be described in the IRT 3PL model are numbers 1, 3, 14, 15, 20, and 31. Keywords: Classical Test Theory, Item Response Theory (Anggreyani, 2020)

Using the KBK-BA test has an internal reliability of 0.892. The difficulty level ranges from 0.244 to 0.732 and the different power ranges from 0.269 to

0.66. The test can distinguish the low group from the high group, and the distractor functions well. From the analysis with item response theory, the most rational logistic parameter model is the 2PL model. There is a positive linear correlation between the discriminating power of rpbis and parameter a 2PL, and a negative linear correlation between the degree of difficulty p and parameter b 2PL (Ramalis, 2015)

When person parameters were assumed to be unknown and items parameters to be either known or not, the power achieved using IRT or CTT were similar and always lower than the expected power using the well-known sample size formula for normally distributed endpoints. The number of items had a substantial impact on power for both methods. Without any missing data, IRT and CTT seem to provide comparable power. The classical sample size formula for CTT seems to be adequate under some conditions but is not appropriate for IRT. In IRT, it seems important to take account of the number of items to obtain an accurate formula. (Véronique Sébille 1, 2010). The results of the construct validity test showed that of the 40 items used, there were 24 items (60%) valid questions and 16 items (40%) invalid items. The results of the reliability analysis used Ms. Excel 2010 obtained rcount = 0.6045. Based on the rtable reference for determining the reliability of the data with the conditions rcount > rtable, so that the question is quite reliable. Difficulty level Based on the difficulty level test using Ms. Excel 2010 shows that out of 40 multiple choice questions, 1 item (2.%) belongs to the easy category, 17 items (42.5%) fall into the medium category and 22 items fall into the difficult category (55 %), the results of the analysis of discriminatory power using Ms. Excel 2010 shows that 1 item (2.5%) has very good discriminating power, 5 items (12.5%) have good discriminating power, 8 b.

Based on the results of the analysis obtained from the scalogram, information was obtained that there were 4 respondents who indicated a person fit order. This means that out of 50 respondents there were 46 who answered items consistently (no one answered carelessly). Overall, the results of the analysis of the items and respondents indicated that there were no items that were missed to be answered by the respondents or all respondents filled out all the items in full. The item reliability value of 0.86 indicates that the quality of the items in this instrument is high. However, out of 50 respondents, 4 were inconsistent and out of 25 items, 2 items were identified as inappropriate with the model. So that the other 23 items have accuracy with the model and are indeed quality items. Furthermore, the respondent's reliability value was 0.80 indicating that the consistency of the respondents' answers was high. In other words, respondents answered all items seriously (not carelessly). (1Amalia Rahmayani, 2022)

The Rasch Model with the help of Winstep software obtained results where the two methods obtained reliability results using the SPSS 22.0 program with Cronbach's Alpha (KR-20) the problem was said to be reliable if rcount> 0.05 (Table 2). This means that the reliability analysis of the questions with SPSS 22.0 obtained 0.828 is reliable with a high category. Meanwhile, test the reliability of the questions with the Rasch model with the help of Winstep

software obtained 0.69, which means the items are reliable in the sufficient category. One of the factors that distinguishes the reliability results between the analysis using SPSS 22.0 and the Rasch model with the help of Winstep software is that from the results of the validation test using SPSS 22.0 there are 20 valid items with 3 more items than the analysis using the Rasch model with a total of 17 valid items. So, if the reliability test is carried out using SPSS 22.0 the level of reliability obtained is also greater, namely 0.828 (high category) compared to the analysis using the Rasch model, which obtains 0.69 (sufficient category). Table 2. Differences in the reliability test with SPSS 22.0 and the Rasch model SPSS 22.0 Rasch model Cronbach's alpha Category Item reliability Category 0.828 High 0.69 Enough The description shows that students' understanding of concepts is one of the goals of the learning process where these goals can be achieved using Bloom's Taxonomy. The following research implication is understanding (Diona Amelia, 2015)

a concept is an ability that must be mastered by students. Because a concept is related to other concepts. With Bloom's taxonomy it can be seen the level of understanding of concepts in students. In this study, the level of understanding of students' concepts was obtained at a moderate level with 54%. The percentage of students with low conceptual understanding ability is 18%. Students with high conceptual understanding of 28%. Students with a high level of understanding can fulfill all indicators of conceptual understanding based on Bloom's Taxonomy well. Students who have moderate conceptual understanding abilities can fulfill several indicators of conceptual understanding based on Bloom's Taxonomy. Students with low understanding only meet the indicators of remembering and understanding. (Budi Murtiyasa1, 2022). Based on the details of the analysis results in table 1 and table 2, it can be seen that the questions presented in the daily tests on the subject of the set already contain a higher cognitive level, namely the cognitive level of understanding (C2) and the cognitive level of application (C3). The average percentage of cognitive level in class VII-A is as follows: (a) the cognitive level of knowledge (C1) is 92.5%; (b) cognitive understanding level (C2) of 61.2%; and (c) application cognitive level (C3) of 71.2%. The mistakes that many students made in questions that contained the concept of intersection and combination, and students were also less able to understand the language posed in the questions presented. The results of the analysis show that the total percentage of questions in C1 and C2 is 34.20%. When compared with the proportion of questions from the Basic Competency Achievement analysis, this percentage is close to the proportion of questions that should be. The total percentage of questions at level C3 and C4 is 65.81%. This percentage is greater than the proportion of questions that support the achievement of Basic Competence, which is 40%. The percentage of questions at level C3 is 61.94%, meaning that more than half of the questions accumulate at level C3. Questions at the cognitive level C5 and C6 were not found. In fact, one of the Basic Competences is "Making a mathematical model of problems related to one-variable linear equations and inequalities". Operational word make is at level

C6. So, the one-variable linear equations and inequalities chapter should have test questions (Ina Magdalena 1, 2020)

competence at the level of questions C6. Mastery of the cognitive domain of students, including student behavior shown through intellectual aspects, such as knowledge and thinking skills. The knowledge and skills of students can be seen from the development of the theories possessed by students, as well as the thinking memory of students who can store the new things they receive. For example, new students learn about the definition of drama, theater, and stage layouts. In general, students with strong cognitive domains can memorize and understand the definitions they just know. In addition, the ability of students to remember the theory they just got is very strong. (Ruwaida, 2019).

CONCLUSIONS AND RECOMMENDATIONS

There are six domains of bloom's taxonomy used in this study with a cognitive approach. Even though all of these domains are not always used, all of them will be adjusted to the learning objectives. Statistical data processing of classical theoretical tests combined with modern theoretical tests, according to research results, produces better and more varied output. These modifications were made to further develop statistical data processing related to research methods and types of processing and applications used. Based on the results of these studies, it was found that there are many ways to process data that can be used in testing the results of tests of classical theories and modern theories, so that the knowledge of statistical data processing that is related to methods and application usage can be further developed, both in terms of validity, reliability, normality, homogeneity and linearity, as well as to test research with other purposes, so as to produce data that can be accounted for by the processing results numerically and categorically.

FURTHER STUDY

Therefore, questions in this very difficult category should be discarded or not used again. This shows that this item is quite good at differentiating the ability of students in the upper group and the lower group in line. Furthermore, the respondent's reliability value was 0.80 indicating that the consistency of the respondents' answers was high. In other words, respondents answered all items seriously (not carelessly). Overall, the results of the analysis of the items and respondents indicated that there were no items that were missed to be answered by the respondents or all respondents filled out all the items in full.

REFERENCES

Amalia Rahmayani, 2. 3. (2022, Agustus). Analisis Kualitas Butir Soal Ulangan Harian Matematika Di Kelas IV MI Al-Islamiyah Menggunakan Rasch Model. JURNAL PERSEDA, Volume V, Nomor 3, DESEMBER 2022 : 170-177. Dipetik Desember 19, 2022

Budi Murtiyasa1, N. K. (2022, September). Analisis Kemampuan Pemahaman Konsep Pada Materi Bilangan Berdasarkan Taksonomi Bloom. AKSIOMA: Jurnal Program Studi Pendidikan Matematika(Received 16

Diona Amelia, S. A. (2015, Maret). Analisis Hasil Belajar Matematika Siswa Pada Pokok Bahasan Himpunan Berdasarkan Ranah Kognitif Taksonomi Bloom Kelas VII-A di SMPN 14 Jember.

Nita Fitria1*, N. F. (2022, Desember). Rasch Model Analysis of the Social Prejudice Scale. Jurnal Konseling Gusjigang(Vol. 8, No. 2, Desember 2022), Hal. 42-51.

Rosid Bahar 1, E. I. (2021, November). Analisis Karakteristik Soal Ujian Sekolah Hasil Musyawarah Guru Matematika Di Tasikmalaya. Aksioma: Jurnal Program Studi Pendidikan Matematik(2660-2674 Received 29 October 2021; Received in revised form 16 November 2021; Accepted 28 December 2021), Volume 10, No. 4, 202. doi:

Véronique Sébille 1, 2. ,.-B. (2010). Methodological issues regarding power of classical test theory (CTT) and item response theory (IRT)-based approaches for the comparison of patient-reported outcomes in two groups of patients - a simulation study. BMC Medical Research Methodology 2010

Anas Irwan, A. P. (2022). Analisis Kualitas Soal Three Tier Test Menggunakan Teori Respon Butir Klasik.JurnalPendidikanFisika(Vol.10No.2).

Anggreyani, A. (2020). Penerapan Teori Uji Klasik dan Teori Respon Butir dalam Mengevaluasi Butir Soal (Studi Kasus : Soal Ujian Akhir Semester Tingkat Persiapan Bersama Institut Pertanian Bogor Mata Kuliah Fisika Tahun Ajaran 2008/2009). Jurnal IPB, C=Scientific Repository.

Bowen, G. J. (2021, Juli 22). Item Response Theory and Confirmatory Factor Analysis: Complementary Approaches for Scale Development. Journal of Evidence-Based Social Work.

Cheng-Fang Yen 1, 2. ,.-T.-T.-Y. (2021, Desember 18). Measure of Internalized Sexual Stigma for Lesbians and Gay Men (MISS-LG) in Taiwan: Psychometric Evidence from Rasch and Confirmatory Factor Analysis. International journal environment and Publik health.

Chi Zhang1, 2. T. (2021, Mei 19). Reliability, Validity, and Measurement Reliability Reliability, Validity, and Measurement Invariance of the General Anxiety Disorder Scale Among Chinese Medical University Students. ORIGINAL RESEARCH.

Dian Normalitasari Purnama1, F. A. (2020). Karakteristik Butir Soal Try Out Teori Kejuruan Akuntansi Smk Berdasarkan Teori Tes Klasik Dan Teori Respons Butir. Jurnal Pendidikan Akuntansi Indonesia, Vol. 18, No. 1, Tahun 2020, Hal. 36 - 46. Dipetik Oktober 6, 2020.

Giani Giani, Z. Z. (2015). Analisis Tingkat Kognitif Soal-Soal Buku Teks Matematika Kelas Vii Berdasarkan Taksonomi Bloom. JPM (Jurnal Penddikan Matematika), Vol 9, No 2 (2015).

Guiliani Marescosthi Angelina1, L. E. (2022). Analisis Butir Soal Menggunakan Teori Pengukuran Klasik Dan Taksonomi Bloom Pada Ujian Akhir Semester Mata Pelajaran Bahasa Indonesia Kelas Vii Sekolah Menegah Pertama Negeri 16 Bintan Tahun Pelajaran 2020/2021. (VOL: 3, NO: 1, TAHUN: 2022 1).

Hayat, B. (2021). Klasika: Program Analisis Item dan Tes dengan Pendekatan Klasik. JP3I (Jurnal Pengukuran Psikologi dan Pendidikan Indonesia), 10(1), 2021, 1-11.

Holly O Witteman1, e. a. (2021). An 11-Item Measure of User- and Human-Centered Design for Personal Health Tools (UCD-11): Development and Validation. JOURNAL OF MEDICAL INTERNET RESEARCH W.

Julio César Acosta-Prado 1, 2. ,.-M. (2021, November 12). Psychometric Properties of Heavy Work Investment Measures: A Systematic Review. Sustainability.

Joelan Maulidina Fiska1, Y. H. (2021, Juli). Analisis Butir Soal Ulangan Harian Ipa Menggunakan. Jurnal Natural Science Educational Research. Dipetik Oktober 6, 2022

Justin Mason1*, S. C. (2021, Januari 25). Construct Validity and Test–Retest Reliability of the Automated Vehicle User Perception Survey. ORIGINAL RESEARCH article.

Iska Amalia1, S. M. (2022). Analisis Distribusi Taksonomi Bloom Soal Penilaian Akhir Semester Matematika di SMP Negeri 1 Wonopringgo 2020/2021. Jurnal Pendidikan Matematika(Vol. 2 No. 02 (2022)). Dipetik Oktober 15, 2022

Robin Midhage, L. H. (2021, Mei). Psychometric evaluation of the Swedish self rated 36-item version of WHODAS 2.0 for use in psychiatric populations – using classical test. Nordic Journal of Psychiatry(09 May 2021. 09 May 2021.).

Rosid Bahar 1, E. I. (2021, November). Analisis Karakteristik Soal Ujian Sekolah Hasil Musyawarah Guru Matematika Di Tasikmalaya. AKSIOMA: Jurnal Program Studi Pendidikan Matematik(2660-2674 Received 29 October 2021; Received in revised form 16 November 2021; Accepted 28 December 2021), Volume 10, No. 4, 202. doi:

Muhammad Rais Ridwan1, 2. E. (2021). Test Items Analysis of Mathematical Problem Solving Ability using a Classical Test Theory Approach. Jurnal Pendidikan MIPA(22 (1), 2021), hal 98-111.

Ramalis, T. R. (20015, Desember). Karakteristik Pengembangan Tes Keterampilan Berpiki Kritis Bumi dan Antariksa Untuk Calon Guru. Jurnal Penelitian dan pengembangan jurusan fisika(Volume 1 Nomor 2, Desember 2015).

Rasmuin 1, S. L. (2022, Mei 1). Tingkat Kesulitan Soal Buatan Guru Bidang Studi Matematika Menurut Teori Tes Klasik pada Tingkat SMP di Kota Baubau. nal Akademik Pendidikan Matematika, Volume 8, Nomor 1, Mei 2(Volume 8, Nomor 1, Mei 2022), Halaman 33-40. DipetikMei1,2022

Ruwaida, H. (2019). Proses Kognitif dalam Taksonomi Bloom Revisi : Analisis Kemampuan Mencipta (C6) Pada Pembelajaran Fikih Di MI Miftahul Anwar Desa Banua Lawas. :Jurnal Ilmiah Pendidikan Madrasah Ibtidaiya(Vol. 4, No. 1, Juli-Desember 2019). Dipetik Desember 19, 2022

rated 36-item version of WHODAS 2.0 for use in psychiatric populations – using classical test. Nordic Journal of Psychiatry(09 May 2021. 09 May 2021.).

Vinayak Bairagi, M. V. (2019). Research Methodology, A Practical and Scientific Approach. New York, Francis: CRC Press Taylor & Francis Group Boca London New York.

Setyawarno, D. (2017). Penggunaan AplikasiSoftware Iteman (Item and Test Analysis) untuk Analisis Butir Soal Pilihan Ganda Berdasarkan Teori Tes Klasik. Jurnal Ilmu Fisika dan Pembelajarannya(Vol 1 No 1 (2017)). doi:DOI:

Tyrone B. Pretorius1, A. P. (2022, Juli 19). Assessing the cognitive component of subjective well being: Revisiting the satisfaction with life scale with classical test theory and item response theory. African Journal of Psychological Assessment.

Xabier Larrucea1, J. ·. (2021, April 29). A Pragmatic Approach for Evaluating and Accrediting Digital Competence of Digital Profles: A Case Study

of Entrepreneurs and Remote Workers. Technology, Knowledge and Learning (2022) 27:843–878.

Creswell, J. W. (2017). Research Design, Pendekatan Metode Kualitatif, Kuantitatif, dan Campuran. (A. F. Pancasari, Penerj.) Yogyakarta, Indonesia: Edisi 4, cetakan ke 11 Pustaka Pelajar. Dipetik Oktober 14, 2022

Dali, S. N. (2022). Teori Sekor Pada Pengukuran Mental (Edisi Ketiga ed.). Jakarta, DKI Jakarata, Indonesia: PT Nagarani Citrayasa. Dipetik Oktober 15, 2022

Dian NF, M. (2021, Novemver). Taksonomi Bloom: Model Dalam Merumuskan Tujuan Pembelajaran. (D. B. Sari, Penyunt.) Dipetik Oktober 15, 2022, dari Pusdiklat Perpustakaan Nasional Republik Indonesia:

Faiz Zulkifli, R. Z. (2022, Maret). A Revised Bloom's Taxonomy: An Essential Approach to Constructing Assessment Questions for the Probability and Statistics Course. International Academic Research in Business ang social Science(Vol. 12, No. 3, 2022), 588 –602.

Guadalupe Elizabeth Morales-Martinez and Janneth Trejo-Quintana, D. J.-C.-H. (2021, Februari). Chronometric Constructive Cognitive Learning Evaluation Model: Measuring the Construction of the Human Cognition Schema of Psychology Students. International Journal of Learning, Teaching and Educational Research(Vol. 20, No. 2, February 2021), pp. 1-21.

Guba, E. G. (1989). Fourth Generation Evaluation. The Publishers o f Professional Social Science Newbury Park ■ London • New Delhi. Dipetik Oktober 15, 2022

Priatno, D. (2018). SPSS Panduan Mudah Olah Data bagi Mahasiswa & Umum. Yogyakarta, Istimewa Yogyakarta, Indonesia: Andi. Dipetik Oktober 15, 2022

Riska Amalia1, S. M. (2022). Analisis Distribusi Taksonomi Bloom Soal Penilaian Akhir Semester Matematika di SMP Negeri 1 Wonopringgo 2020/2021. Jurnal Pendidikan Matematika(Vol. 2 No. 02 (2022)). Dipetik Oktober 15, 2022,

Sulistio, D. (2022, Juni). Kemampuan Mahasiswa Menganalisis Tes (Soal) Bahasa Indonesia Sesuai Aspek Ranah Kognitif Model Anderson dan Krathwohl. Jurnal i Pendidikan Bahasa Indonesia dan bahasa Asing(Vol.5,No.1,2022), Page: 1-21.