Jurnal Penelitian dan Evaluasi Pendidikan
Volume 25, No 2, December 2021 (162-169)
Online: http://journal.uny.ac.id/index.php/jpep

Analysis of the difficulty index of item bank according to cognitive aspects during the Covid-19 pandemic

Maizura Fauzie*; Andi Ulfa Tenri Pada; Supriatno
Universitas Syiah Kuala, Aceh, Indonesia
*Corresponding Author. E-mail: maizurafauzie@gmail.com

ARTICLE INFO

Article History
Submitted:
24 July 2021
Revised:
27 November 2021
Accepted:
9 December 2021

Keywords
difficulty index; final semester exam; item bank; Covid-19; cognitive aspect

Scan Me:

ABSTRACT

The Covid-19 pandemic is a major challenge for the education system. The face-to-face learning process shifted to online learning, including the school exams. In Aceh province, the school exams have changed from paper-based and computer-based. This research aims to analyze the difficulty index of an item bank based on cognitive aspects of Bloom's Taxonomy. The study samples included 850 students. The data were the item bank of a final semester exam consisting of 200 multiple-choice items, answer keys, and students' answer sheets. The empirical analysis of the item bank using classical test theory (CTT) found that 141 out of 200 items are valid based on content validity and computing data set using the Aiken's V formula. Item tests have reliability of 0.983. The reliability is calculated using the Kuder-Richardson 21 formula. If the reliability coefficient is r11 ≥ 0.70, then the item is declared reliable. In addition, 62 out of 141 (43.97%) items from the item bank are classified with a moderate difficulty index, and 79 items (56.03%) are categorized with a high difficulty index. The cognitive aspects found in the items are remembering, understanding, applying, and analyzing. Students mostly found items with the cognitive aspects of remembering and understanding are difficult to solve.

How to cite:
Fauzie, M., Pada, A., & Supriatno, S. (2021). Analysis of the difficulty index of item bank according to cognitive aspects during the Covid-19 pandemic. *Jurnal Penelitian dan Evaluasi Pendidikan, 25*(2), 162-169. doi:https://doi.org/10.21831/pep.v25i2.42603

INTRODUCTION

COVID-19 is caused by coronavirus, a group of viruses infecting the respiratory system. The first case of the coronavirus appeared in the province of Wuhan, China. The virus has currently spread to various countries, including Indonesia. It is easily contagious and requires various sectors to immediately take action to prevent wider transmission, including the education sector. Due to its vulnerability even within the school environment, thousands of school closures have been implemented worldwide as COVID-19 currently occurred (Toquero, 2020). With the increasing spread of COVID-19 globally and orders to reduce the corona virus, many schools have also been closed in Indonesia.

Various policy initiatives were launched by the government and educational institutions to continue learning activities and retain the virus. One of the policies is online learning. The face-to-face learning process has shifted into online learning to reduce the spread of COVID-19 (Daniel, 2020; Toquero, 2020). Large-scale national efforts to utilize technology to support distance learning, distance education, and online learning during the COVID-19 pandemic have emerged and developed rapidly (Ali, 2020).

COVID-19 is the biggest challenge of the national education system that has ever happened, including the implementation of school exams. Semester exams in Aceh province were previously paper-based, and it has instantly switched to computer-based at present. The final semester exam utilizes items from item banks prepared by the Forum of the Subject Teachers

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v25i2.42603

or *Musyawarah Guru Mata Pelajaran* (hereafter, MGMP) in each regency/city. Item banks have been calibrated (bin Abd. Razak et al., 2012), and its main purpose is to assemble or construct the tests and provide test suitability both for daily test and end-of-semester assessments (Jiraro, 2014). The item banks prepared in Aceh has similar objectives.

Several bodies of literature include research analyzing exam questions according to cognitive aspects. These studies mainly deal with cognitive domain categories of entrance exam questions or the relationship between difficulty index (Hingorjo & Jaleel, 2012; Johari et al., 2011; Kibble & Johnson, 2011; Pande et al., 2013; Tan & Othman, 2013; Zainudin et al., 2012). Furthermore, some studies analyzed exam questions based on the cognitive aspect and difficulty index. For example, Veeravagu et al. (2010) found a relationship between the cognitive aspects in Bloom's taxonomy and the learners' performance for the questions of an English reading skills course. They also revealed that the learners had difficulty answering questions requiring higher-level cognitive skills: analysis, synthesis, and evaluation. On the other hand, a study by Kibble and Johnson (2011) aimed to assess how useful two multiple-choice question classification schemes would be in helping us to predict the difficulty of test items. Furthermore, they hypothesized that the cognitive aspect of items could be used to predict item difficulty.

A study conducted by Zainudin et al. (2012) revealed the role of difficulty index as a potential transformation agent for the Management Information Systems course since it has successfully portrayed the level of difficulty for each question and task assigned to students. In another study, Nevid and McClelland (2013) indicated that the learners had difficulty in solving problems of evaluation and explanation at higher-level cognition of Bloom's taxonomy for a psychology course, and these kinds of questions were the most distinctive for high-performing and low-performing learners.

Assessment of learning is one of the essential elements in the process of learning design since it provides feedback to students and teachers and improves the quality of the education system. However, studies analyzing the difficulty of item banks at the school level based on Bloom's taxonomy are still limited. Thus, this current study would contribute to the assessment practices, teachers, test developers, and assessment experts in generating an item bank.

Analyzing the difficulty index means examining the test items to identify low, moderate, and high difficulty items. The difficulty index of the items can be noticed from the students' ability to answer the test items. The item difficulty index is the percentage or proportion of test-takers who correctly answered an item (Hingorjo & Jaleel, 2012; Marie & Edannur, 2015; Mitra et al., 2009; Quaigrain & Arhin, 2017). The higher the percentage of students who correctly answer an item, the easier the item or, the lower the item difficulty, and vice versa (DiBattista & Kurzawa, 2011; Escudero et al., 2000; Marie & Edannur, 2015; Taib & Yusoff, 2014). Thus, it can be concluded that when fewer students can answer the item correctly, the item is hard, or the difficulty level is high.

The difficulty index is represented by p (proportion) and ranged from 0 to 1 (Escudero et al., 2000; Koçdar et al., 2016; Marie & Edannur, 2015). The greater the difficulty index is obtained, the easier the item will be, and thus, it must be revised. An item with $p = 0.00$ means that no student answered the item correctly, and if the $p = 1.00$, it means that all students answered it correctly (Arifin, 2017; DiBattista & Kurzawa, 2011; Johari et al., 2011). Allen and Yen (1979) stated that, in general, the difficulty index of an item should lie within the interval of 0.3 and 0.7. At this interval, the information regarding students' abilities will be maximally obtained (Purnama & Alfarisa, 2020; Salih et al., 2020).

The difficulty index will be analyzed using the difficulty index formula, and it is used to verify the relationship between the difficulty index and the achievement of learning outcomes. The importance of analyzing the difficulty index is to identify the difficulty level of an item, whether it is low, moderate, or high. A good item is an item with neither too low nor too high difficulty. This research is critical because it is a benchmark to determine whether the test

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v25i2.42603

items from the item bank are in accordance with the item-writing guideline and student learning outcomes. In the future, the results of this study can be used to improve the quality of test items to achieve the targets of learning outcomes and provide input to test-makers, teachers, schools, and the government.

RESEARCH METHOD

This study is quantitative descriptive research. The data were taken from an item bank of the final semester examination of the Biology subject. It consisted of 200 multiple-choice items and the answer sheets of Grade 11 students at Islamic senior high schools in Aceh province. The empirical analysis of the item bank using classical test theory (CTT) found that 141 out of 200 items are valid based on content validity and computing data set using the Aiken's V formula. Item tests have reliability of 0.983. The reliability is calculated using the Kuder-Richardson 21 formula. If the reliability coefficient is r11 ≥ 0.70, then the item is declared reliable. However, this article only focuses on discussing the difficulty index of the test items and how they relate to student learning outcomes. Students' responses were analyzed empirically based on a classical test theory to examine the difficulty index.

The population of this study was 44926 grade 11 students who took the final exam of Biology subject in Semester I within the scope of the Ministry of Religious Affairs of Aceh. The sampling technique used in this research was cluster random sampling, in which the sample was chosen based on subject availability. Therefore, the research sample involved 850 students.

The difficulty index in a classical test theory is the proportion of who answered correctly (Mardapi, 2015). The formula that is employed to calculate the difficulty index is presented in Formula (1) (Johari et al., 2011; Kumar et al., 2021; Marie & Edannur, 2015; Nitko, 1996). In addition, the criteria of the item difficulty index are described in Table 1.

$$Difficulty\ Index = \frac{\text{the number of students who answer correctly}}{\text{the number of students who took the test}} \quad \ldots\ldots (1)$$

Table 1. Difficulty Index Criteria

Difficulty Index	Category	Modification Results
0.00 – 0.29	High Difficulty	Modify
0.30 – 0.70	Moderate Difficulty	Accept
0.71 – 1.00	Low difficulty	Modify

The analysis of the difficulty index was also carried out by comparing the cognitive domains of the revised Bloom's taxonomy to identify student learning outcomes. The categories are remembering, understanding, applying, analyzing, evaluating, and creating in the revised Bloom's taxonomy (Koçdar et al., 2016).

FINDINGS AND DISCUSSION

Difficulty index measures the difficulty level of an item for students. The difficulty index is calculated for each item by calculating the number of students who answered correctly divided by the number of students taking the test.

The difficulty index that has been calculated is interpreted based on three criteria (Johari et al., 2011; Kumar et al., 2021; Wibawa, 2019). An item with a difficulty index ranging from 0.00 to 0.29 is classified as high difficulty. An item with a difficulty index between 0.30 and 0.70 is considered moderate difficulty, and an item having a difficulty index of 0.71 to 1.00 is categorized as low difficulty (Hingorjo & Jaleel, 2012; Pande et al., 2013).

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v25i2.42603

The findings of the study found that the difficulty index of an item bank of the Semester I final exam in Biology for Grade 11, Islamic Senior High Schools in Aceh province, the academic year of 2019/2020, ranged from moderate to high. In brief, the difficulty index of the item bank is presented in Figure 1.

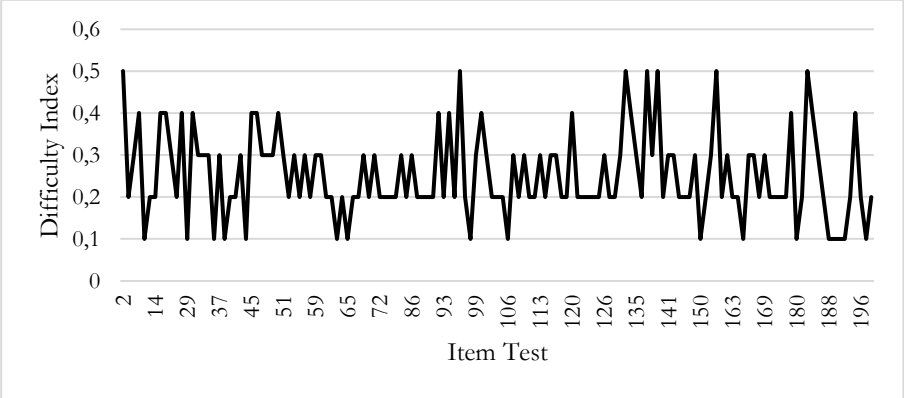

Figure 1. The Difficulty Index of the Item Bank

Figure 1 shows the difficulty index of each item, ranging from 0.1 to 0.5. It indicates that the item bank consisted of items with high and moderate difficulty. For example, in item 2 in Figure 1, the number of students who answered was 200, and 106 answered correctly. Then, the calculation to determine the difficulty index of item 2, calculation using Formula (1) is used, so the difficulty index of Item 2 = 106/200=0.5. This means that item 2 is moderate. The moderate difficulty index ranges from 0.30 to 0.70, so item 2 is classified as accepted. Based on the difficulty index theory by Anon (Johari et al., 2011), the higher the difficulty index is, the easier the item will be. The smaller difficulty index indicates that the item is more difficult.

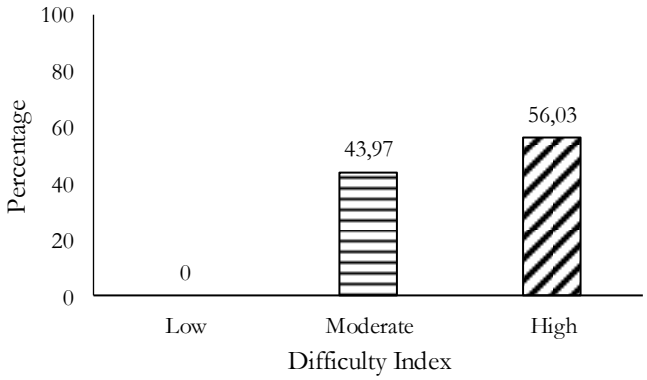

Figure 2. Percentage of Difficulty Index

Figure 2 shows that the difficulty of 43.97% of the items is moderate and 56.03% is high, and no item with low difficulty. A good item is an item that is not too easy or not too hard. Allen and Yen (1979) stated that, in general, the preferable difficulty index ranges from 0.30 to 0.70. At this interval, information on a student's ability will be optimally gained. The items (62 items) with moderate difficulty can be included in the item bank, but the items (79 items) with high difficulty should be omitted.

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v25i2.42603

The difficulty index relates to a balanced proportion of multiple-choice items with low, moderate, and high difficulty. Sudjana (2017) stated that the difficulty index of an item bank should be 3-4-3, meaning that 30% low difficulty items, 40% moderate difficulty items, and 30% high difficulty items, or it can follow a 3-5-2 pattern. Meanwhile, Rao et al. (2016) mentioned that the proportion of difficulty index of an item bank could be 30:70, 30% high, and 70% moderate difficulty items.

The item bank in this study consists of 0% low, 43.97% moderate, and 56.03% high difficulty items (Figure 2). This finding is not aligned with the aforementioned theory since the difficulty index is not proportionally distributed. The higher proportion of high difficulty items in the item bank results from many students answering the items incorrectly.

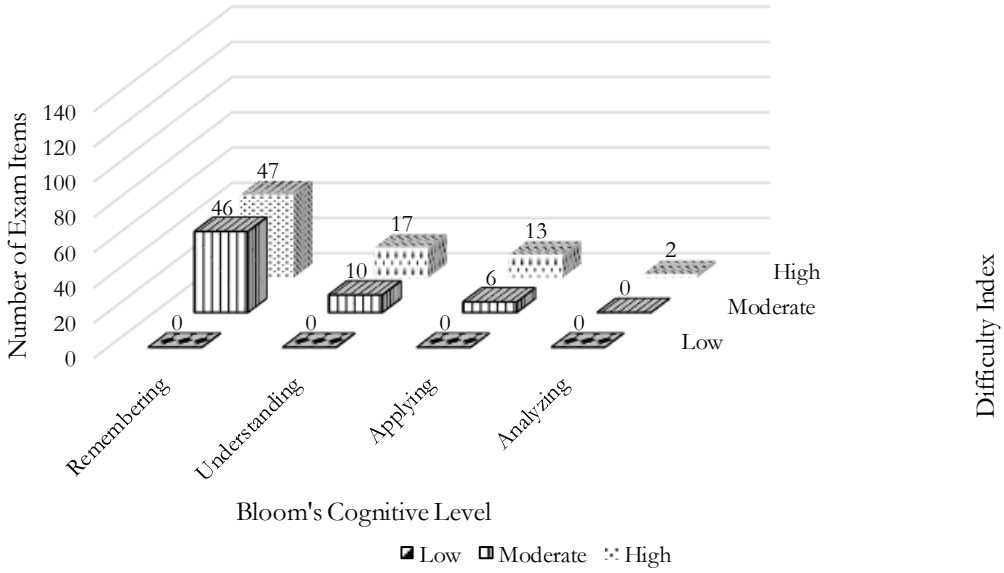

Figure 3. Distribution Difficulty Index Based on Their Bloom's Taxonomy Score

The analysis results showed that the high difficulty items dominate the item bank (DI < 0.3). Those items should be deleted. These 79 items are too difficult for students. Nevertheless, the item analysis based on cognitive aspects of Bloom's taxonomy revealed that, on average, the items fall into Lower Order Thinking (LOT) aspects: remembering (47 items), understanding (17 items), and applying (13 items), as indicated in Figure 3. The items that students are difficult to answer dominantly lie in remembering and understanding aspects. The other two items are at the analyzing aspect.

The analysis results reveal that low-level cognitive questions are not always easy questions and vice versa. The possible factor contributing to the incorrect students' answers is that students have not yet completely understood the lessons, or the lessons might be new for them (or not yet taught at schools). Besides, students have not mastered the concepts, and they might answer the test by guessing. In this case, teachers should be aware of the strengths and weaknesses of each indicator being assessed. Therefore, teachers should find ways to address the problem and more appropriate teaching methods.

The Higher Order Thinking (HOT) items require cognitive aspects of analyzing, evaluating, and creating. Out of 141 valid items based on content validation, only two HOT items (analyzing). The recommended portion is 30% items for remembering and understanding aspects, 40% items for applying and analyzing aspects, and 30% items for evaluating and creating aspects (Septiana, 2016).

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v25i2.42603

Uneven distribution of Bloom's taxonomy cognitive aspects in the item bank is likely caused by insufficient knowledge of the team who made the test about the criteria of good items and adopting items from textbooks or online sources without considering the proportion of items based on the cognitive aspects. The test criteria at the senior high school level should include HOT items.

CONCLUSION

Based on empirical analysis with a classical test approach, the study shows that out of 200 items, 141 items are valid based on the content validation. Among those valid items, 62 (43.97%) are high, and 79 (56.03%) are moderate difficulty items. Cognitive aspects embedded in the item bank vary from remembering, understanding, applying to analyzing. There are 47 items that fall into the remembering aspect, 17 are included in the understanding aspect, 13 are categorized as the applying aspect, and two items are in analyzing aspect. Most of the items that students were challenging to answer are classified into remembering and understanding aspects. The analysis gives us the information that low-level cognitive questions are not always easy questions and vice versa.

REFERENCES

Ali, W. (2020). Online and remote learning in higher education institutes: A necessity in light of COVID-19 pandemic. *Higher Education Studies*, *10*(3), 16–25. https://doi.org/10.5539/hes.v10n3p16

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Brooks/Cole. https://books.google.co.id/books?id=cgElAQAAIAAJ&hl=id&source=gbs_navlinks_s

Arifin, Z. (2017). Kriteria instrumen dalam suatu penelitian. *Jurnal THEOREMS (The Original Research of Mathematics)*, *2*(1), 28–36. https://jurnal.unma.ac.id/index.php/th/article/view/571

bin Abd. Razak, N., bin Khairani, A. Z., & Thien, L. M. (2012). Examining quality of Mathemtics test items using Rasch model: Preminarily analysis. *Procedia - Social and Behavioral Sciences*, *69*, 2205–2214. https://doi.org/10.1016/j.sbspro.2012.12.187

Daniel, S. J. (2020). Education and the COVID-19 pandemic. *PROSPECTS*, *49*(1–2), 91–96. https://doi.org/10.1007/s11125-020-09464-3

DiBattista, D., & Kurzawa, L. (2011). Examination of the quality of multiple-choice items on classroom tests. *The Canadian Journal for the Scholarship of Teaching and Learning*, *2*(2), 4. https://doi.org/10.5206/cjsotl-rcacea.2011.2.4

Escudero, E. B., Reyna, N. L., & Morales, M. R. (2000). The level of difficulty and discrimination power of the Basic Knowledge and Skills Examination (EXHCOBA). *Revista Electrónica de Investigación Educativa*, *2*(1), 11–29. https://redie.uabc.mx/redie/article/view/15

Hingorjo, M. R., & Jaleel, F. (2012). Analysis of one-best MCQs: The difficulty index, discrimination index and distractor efficiency. *JPMA. The Journal of the Pakistan Medical Association*, *62*(2), 142–147. https://jpma.org.pk/article-details/3255

Jiraro, P. (2014). A development of measurement and evaluation standards and item bank approach model for teachers in Thai secondary schools. *Procedia - Social and Behavioral Sciences*, *116*, 547–556. https://doi.org/10.1016/j.sbspro.2014.01.256

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v25i2.42603

Johari, J., Sahari, J., Wahab, D. A., Abdullah, S., Abdullah, S., Omar, M. Z., & Muhamad, N. (2011). Difficulty index of examinations and their relation to the achievement of programme outcomes. *Procedia - Social and Behavioral Sciences*, *18*, 71–80. https://doi.org/10.1016/j.sbspro.2011.05.011

Kibble, J. D., & Johnson, T. (2011). Are faculty predictions or item taxonomies useful for estimating the outcome of multiple-choice examinations? *Advances in Physiology Education*, *35*(4), 396–401. https://doi.org/10.1152/advan.00062.2011

Koçdar, S., Karadağ, N., & Şahin, M. D. (2016). Analysis of the difficulty and discrimination indices of multiple-choice questions according to cognitive levels in an open and distance learning context. *Turkish Online Journal of Educational Technology*, *15*(4), 16–24. http://www.tojet.net/articles/v15i4/1542.pdf

Kumar, D., Jaipurkar, R., Shekhar, A., Sikri, G., & Srinivas, V. (2021). Item analysis of multiple choice questions: A quality assurance test for an assessment tool. *Medical Journal Armed Forces India*, *77*, S85–S89. https://doi.org/10.1016/j.mjafi.2020.11.007

Mardapi, D. (2015). *Pengukuran, penilaian, dan evaluasi pendidikan*. Nuha Medika. https://opac.perpusnas.go.id/DetailOpac.aspx?id=1162325

Marie, S. M. J. A., & Edannur, S. (2015). Relevance of item analysis in standardizing an achievement test in teaching of Physical Science in B.Ed syllabus. *I-Manager's Journal of Educational Technology*, *12*(3), 30–36. https://doi.org/10.26634/jet.12.3.3743

Mitra, N. K., Nagaraja, H. S., Ponnudurai, G., & Judson, J. P. (2009). The levels of difficulty and discrimination indices in type a multiple choice questions of pre-clinical semester 1 multidisciplinary summative tests. *International E-Journal of Science, Medicine & Education*, *3*(1), 2–7. http://mymedr.afpm.org.my/publications/44080

Nevid, J. S., & McClelland, N. (2013). Using action verbs as learning outcomes: Applying Bloom's taxonomy in measuring instructional objectives in Introductory Psychology. *Journal of Education and Training Studies*, *1*(2), 19–24. https://doi.org/10.11114/jets.v1i2.94

Nitko, A. J. (1996). *Educational assessment of students*. Merrill Publishing Company. https://books.google.co.id/books/about/Educational_Assessment_of_Students.html?id=CDFLAAAAYAAJ&redir_esc=y

Pande, S. S., Pande, S. R., Parate, V. R., Nikam, A. P., & Agrekar, S. H. (2013). Correlation between difficulty & discrimination indices of MCQs in formative exam in Physiology. *South-East Asian Journal of Medical Education*, *7*(1), 45–50. https://doi.org/10.4038/seajme.v7i1.149

Purnama, D. N., & Alfarisa, F. (2020). Karakteristik butir soal try out Teori Kejuruan Akuntansi SMK berdasarkan teori Tes Klasik dan teori Respons Butir. *Jurnal Pendidikan Akuntansi Indonesia*, *18*(1), 36–46. https://doi.org/10.21831/jpai.v18i1.31457

Quaigrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, *4*(1), 1301013. https://doi.org/10.1080/2331186X.2017.1301013

Rao, C., Kishan Prasad, H., Sajitha, K., Permi, H., & Shetty, J. (2016). Item analysis of multiple choice questions: Assessing an assessment tool in medical students. *International Journal of Educational and Psychological Researches*, *2*(4), 201–204. https://doi.org/10.4103/2395-2296.189670

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v25i2.42603

Salih, K. M. A., Jibo, A., Ishaq, M., Khan, S., Mohammed, O., AL-Shahrani, A., & Abbas, M. (2020). Psychometric analysis of multiple-choice questions in an innovative curriculum in Kingdom of Saudi Arabia. *Journal of Family Medicine and Primary Care*, *9*(7), 3663. https://doi.org/10.4103/jfmpc.jfmpc_358_20

Septiana, N. (2016). Analisis butir soal Ulangan Akhir Semester Biologi tahun pelajaran 2015/2016 kelas X dan XI pada MAN Sampit. *Edu Sains: Jurnal Pendidikan Sains Dan Matematika*, *4*(2), 115–121. https://e-journal.iain-palangkaraya.ac.id/index.php/edusains/article/view/514

Sudjana, N. (2017). *Penilaian hasil proses belajar mengajar*. Remaja Rosdakarya. https://rosda.co.id/beranda/438-penilaian-hasil-proses-belajar-mengajar.html

Taib, F., & Yusoff, M. S. B. (2014). Difficulty index, discrimination index, sensitivity and specificity of long case and multiple choice questions to predict medical students' examination performance. *Journal of Taibah University Medical Sciences*, *9*(2), 110–114. https://doi.org/10.1016/j.jtumed.2013.12.002

Tan, Y. T., & Othman, A. R. (2013). The relationship between complexity (taxonomy) and difficulty. *AIP Conference Proceedings 1522*, 596–603. https://doi.org/10.1063/1.4801179

Toquero, C. M. (2020). Challenges and opportunities for higher education amid the COVID-19 pandemic: The Philippine context. *Pedagogical Research*, *5*(4), em0063. https://doi.org/10.29333/pr/7947

Veeravagu, J., Muthusamy, C., Marimuthu, R., & Michael, A. S. (2010). Using Bloom's taxonomy to gauge students' reading comprehension performance. *Canadian Social Science*, *6*(3), 205–212. http://www.cscanada.net/index.php/css/article/view/j.css.1923669720100603.023

Wibawa, E. A. (2019). Karakteristik butir soal tes ujian akhir semester Hukum Bisnis. *Jurnal Pendidikan Akuntansi Indonesia*, *17*(1), 86–96. https://doi.org/10.21831/jpai.v17i1.26339

Zainudin, S., Ahmad, K., Ali, N. M., & Zainal, N. F. A. (2012). Determining course outcomes achievement through examination difficulty index measurement. *Procedia - Social and Behavioral Sciences*, *59*, 270–276. https://doi.org/10.1016/j.sbspro.2012.09.275

ISSN (print) 2685-7111 | ISSN (online) 2338-6061