ICIC Express Letters
Volume **15**, Number **9**, September **2021**

ICIC International ©2021 ISSN 1881-803X
pp. **999–1006**

INTELLIGENT COMPUTATIONAL MODEL TO DETERMINE THE ORDER OF THINKING SKILLS OF TEST ITEMS

Hindayati Mustafidah[1,2], Sri Hartati[1,*], Retantyo Wardoyo[1] and Pujiati Suyata[3]

[1]Doctoral Program, Department of Computer Science
Universitas Gadjah Mada
Sekip Utara, Bulaksumur, Yogyakarta 55281, Indonesia
[*]Corresponding author: shartati@ugm.ac.id; rw@ugm.ac.id

[2]Informatics Engineering
Universitas Muhammadiyah Purwokerto
Jl. KH. Ahmad Dahlan Purwokerto, Banyumas, Central Java 53182, Indonesia
h.mustafidah@ump.ac.id

[3]Department of Indonesian Language and Literature Education
Universitas Ahmad Dahlan
Jl. Kolektor Ring Road Selatan, Tamanan Banguntapan Bantul, Yogyakarta 55166, Indonesia
pujiati.suyata@mp.uad.ac.id

Received November 2020; accepted February 2021

Abstract. *Test items as an instrument containing students' thinking abilities must have good quality. Questions that meet the requirements have a balanced order of thinking skills, such as questions on Indonesia's national exam. In analyzing the questions, we have to collect student assignment scores, which require a long time and complicated administration. The intelligent system model, namely text processing and case similarity, can solve this problem. This paper describes how the model works, especially in junior high school math problems in Indonesia. Text processing is used to identify cognitive process dimension (CPD) and knowledge dimensions (KD), while case similarity is used to find solutions according to Revised Bloom's Taxonomy (RBT) category. Therefore, the teacher-made item text can be seen based on the model's performance, whether it is included in the low-order thinking skills (LOTS) or high-order thinking skills (HOTS).*
Keywords: The order of thinking skills, Test item, Intelligent computational, Text processing, Case similarity

1. **Introduction.** Three assessment approaches have to be held to improve the education's quality, namely assessment for learning, assessment of learning, and assessment as learning [1]. One element of developing the primary curriculum model proposed by [2] was evaluating and assessing the learning process implemented. A proper assessment will also get good feedback on the learning conducted, which means good education [3]. Besides that, assessment is also indicated to always go with teaching pedagogy [4]. Assessment can be done by comparing existing data and information with predetermined standards or benchmarks [5].

Test as an assessment instrument is presented in the form of items given to students. Assessment using the written test method is the usual method [6]. Questions must be given according to the content of the lessons learned by students to produce quality questions and can be used to assess different cognitive levels. For this reason, the teaching and learning process and preparation for writing exam questions always refer to the Revised Bloom's Taxonomy (RBT) [7]. The test must have a balanced level of thinking skills, primarily tests in the national examination (UN in Bahasa) as submitted by [8]. This

DOI: 10.24507/icicel.15.09.999

composition had been applied to the UN in 2018 and 2019 [9]. The Ministry of Education and Culture has also increased the portion of questions that require higher thinking skills or known as high-order thinking skills (HOTS) at junior and senior high school levels, by 15-20% [10]. The order of thinking skills in question items is not always measured by students' ability to answer questions. It can still be estimated based on its learning objective [11], as in RBT [12].

The Education Assessment Center [13] reported the average score for the Junior High School Mathematics National Examination in 2019 only reached 47.21, which is still below the median. The low score suspects that schools' exercise items did not have the same quality as the UN questions. Item analysis activities as the assessment are not always carried out because they are very time-consuming and spend a lot of energy because they have to do a trial before the test is given to students. The existence of computer aids has actually refuted these reasons. There have been several computer program application products to help analyze items such as Xcalibre, Iteman, Lertap, CITAS, RASCAL [14], ANATES [15], and QUEST [16]. Nevertheless, teachers are still reluctant to use it due to the lack of teacher mastery of computer operations, as reported by [17,18].

In line with the Digital Revolution-era RI 4.0, artificial intelligence (AI) in various fields is unavoidable [19], included in the teaching and learning [20] and assessment processes [21]. The development of the use of data science and machine learning has penetrated the world of education, such as big data researched by [22], personality classification using k-means [23], and Bayesian networks to trace the state of student knowledge [24].

Several studies related to the classification of items into Bloom's Taxonomy or its revision using computational methods had been conducted. The methods used were support vector machine, Naïve Bayes, and K-nearest neighbor [6], lexical and syntactic extraction [25], text analysis [26], and using keyphrase [27]. The studies were limited to classifying and taking objects of science, social studies, Indonesian and English subjects, unlike math problems, which contain lots of numbers, equations, and pictures.

Based on the description, the proposed model is a novelty of problem-solving to get the order of thinking skills in the item in junior high schools in Indonesia based on its category in RBT. The analysis mechanism using a computational algorithm model was also carried out by [28].

2. **Methodology.** This research is a fundamental research that emphasizes developing an intelligent computation-based model to determine a text item's order of thinking skills. This model's framework was developed based on the idea that the government's questions, namely the UN, are standardized. In contrast, the teacher's questions as feedback for achieving learning objectives must have the same thinking skills' order as the standardized items. In general, this model consists of two main processes, namely, **text processing** and **case similarity** (Figure 1).

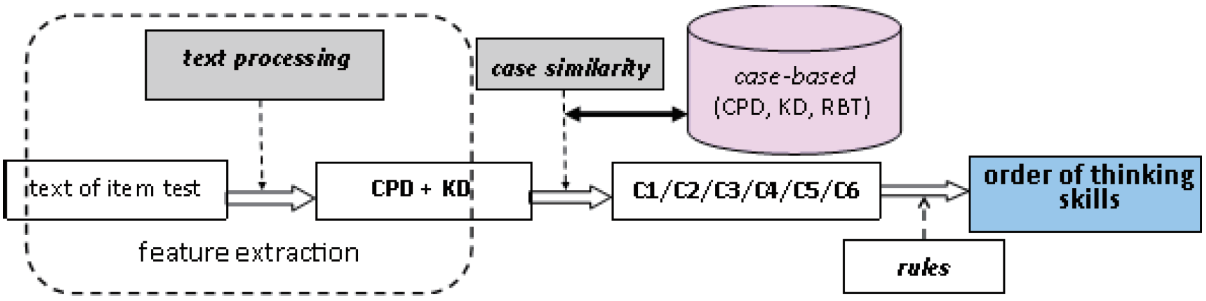

FIGURE 1. Overview and flow of the proposed model to determine the order of thinking skills on the test item process

In the development of the model framework, as shown in Figure 1, we first emphasize how to get cognitive process dimension (CPD) and knowledge dimension (KD), from now on referred to as CPD+KD, through feature extraction using **text processing**. This process aims to identify the item text by the learning objectives framework in RBT. Furthermore, from the CPD+KD obtained as keywords, we carry out a **case similarity** process to get an RBT category, namely C1, C2, C3, C4, C5, and C6, representing remembering, understanding, applying, analyzing, evaluating, and creating [12]. The final step in the model is determining the level of thinking skills using **rules**, whether the text item is at a low or high level of thinking skills.

Data required in this study consists of **case-based** and **test** data. Case-based data are learning objectives of junior high school mathematics questions on the UN in 2019, consisting of 104 item text variations. UN items are used as case-based data because they are standardized questions that have been tested and arranged to follow the signs of a balanced order of thinking skills. This data is acquired from the Education Assessment Center (PUSPENDIK). To find out the RBT category of item text, we do not use a machine-learning algorithm because it does not acquire large data sets [29]. The test data are teacher-made items to be analyzed in their order of thinking skills, referring to the UN questions.

3. **Results and Discussion.** Teachers' text items must have indicators derived from categories and subcategories in learning objectives, classifying the education taxonomy. The revised taxonomy places the "verb" and "noun" components of the original knowledge level into two separate dimensions. This **verb** is a learning objective that describes the cognitive processes dimension (CPD). In contrast, the **noun** describes the knowledge expected to be mastered by students as a knowledge dimension (KD). In other words, CPD describes what students learn to do, while KD describes the knowledge or content that takes place to be discovered. Henceforth, both CPD and KD written in CPD+KD are known as RBT [30]. To get CPD+KD information from text items made by the teacher, we need a text processing method. This method was also used by [31] to classify math problems based on topics and sub-topics. Furthermore, a case similarity process is required between CPD+KD from teacher-made questions with standardized questions to type RBT.

The case-based development process also uses feature extraction, as explained in Figure 2. In addition to producing learning objectives information, namely CPD+KD, case-based also contains a solution, which is called RBT. Technically, to get CPD+KD was done by developing a dictionary. This dictionary includes a list of terms used both in CPD and KD. Categories in RBT are obtained based on CPD+KD mapped into operational verb tables and knowledge dimensions [12]. As reference knowledge, this case-based accuracy must go through the validation of education experts.

Next, the model looks for CPD+KD in the same case-based as CPD+KD questions made by the teacher. After finding similarities, the model provides RBT category information. The solution was found based on the CPD and KD contained in the teacher's text item according to its category in RBT: C1, C2, C3, C4, C5, or C6. The class in RBT provides information on the order of thinking skills. The classification of low or high order thinking skills can be determined based on the RBT category presented by [7] using a **rule**. This rule does an action; if an item is C4, C5, or C6, it is in the HOTS classification. Otherwise, it is included in the LOTS [32].

RBT employs 25 verbs that create a collegial understanding of student behavior and learning outcomes, with a composition of C1: 2 verbs; C2: 7 verbs; C3: 2 verbs; C4: 3 verbs; C5: 2 verbs; and C6: 3 verbs, while the six verbs are remembering, understanding,

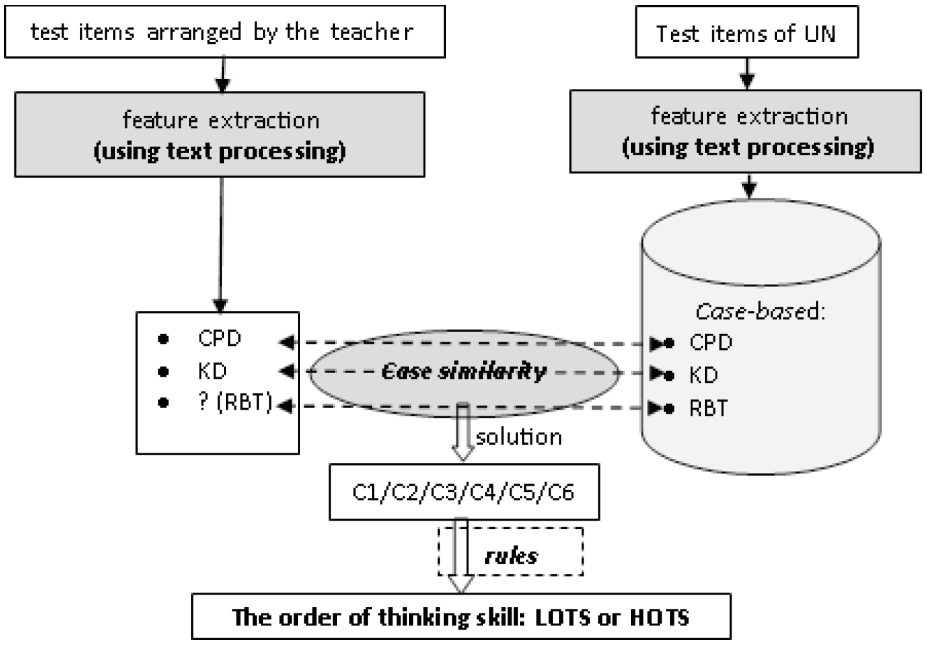

FIGURE 2. Feature extraction of case-based and test case

applying, analyzing, evaluating, and creating, which are the RBT categories. The illustration of the model in analyzing the order of thinking skills in the text item is presented in Figure 3.

As explained in Figure 3, the case of the teacher-made text items is a new case. The model performs initial extraction from the new case to obtain two keywords, verb (CPD) and noun information (KD). KD is the knowledge that students learn, or in other words, it is the subject matter. Examples of KD in math problems are integers (*bilangan bulat*) fractional number (*bilangan pecahan*), exponent (*bilangan berpangkat*), and so on [33]. An example case is shown in Figure 3, given the problem in "Determine the x Value in $x^2 + 2x - 3 = 0$" ("*Tentukan Nilai* x *dalam* $x^2 + 2x - 3 = 0$" in Bahasa). This problem is an assessment format of the learning objectives that are designed. In this case, students learn to find the value of x that satisfies the quadratic equations system. To find answers, students must execute certain techniques or methods, namely systems of quadratic equations. The method is the knowledge used to do something, so this method is included in "procedural knowledge". The combination of the verb "execute" as CPD and the noun "equation" as KD has to map into the taxonomy table [34].

The model mapped the two keywords to the developing dictionary. This mapping aims to get CPD_KD labels. Next, the model performs a case similarity process to get the RBT category as the solution from the teacher's items to UN questions. This similarity process is calculated using Equation (1) [35]. Simultaneously, the case solution, RBT, is obtained based on the similarity degree's maximum value, as in Equation (2) [35].

$$SM(\hat{e}, e_i) = \frac{\text{common}}{\text{common} + \text{different}} \tag{1}$$

$$SM(\hat{e}, e_i) = \max\left\{SM\left(\hat{e}, e_j\right), e_j \in Cl_i\right\}; \quad i = 1, 2, 3, \ldots, k \tag{2}$$

where $\hat{e}$ represents a query case or new case; e_i represents the CPD of $\hat{e}$ which is similar to the number of j CPDs in case-based (e_j); "common" represents the number of features whose value is the same between $\hat{e}$ and e_i; "different" represents the number of features whose value is different between $\hat{e}$ and e_i; Cl is a category class that has similarities to CPD in the case $\hat{e}$. As stated in the method, there are 104 data cases in case-based, with each case consisting of 3 classes, namely CPD (Cl_1), KD (Cl_2), and RBT (Cl_3). For the given example, a new case $\hat{e}$ = (determine, equation, ?) or (*menentukan*, *persamaan*,

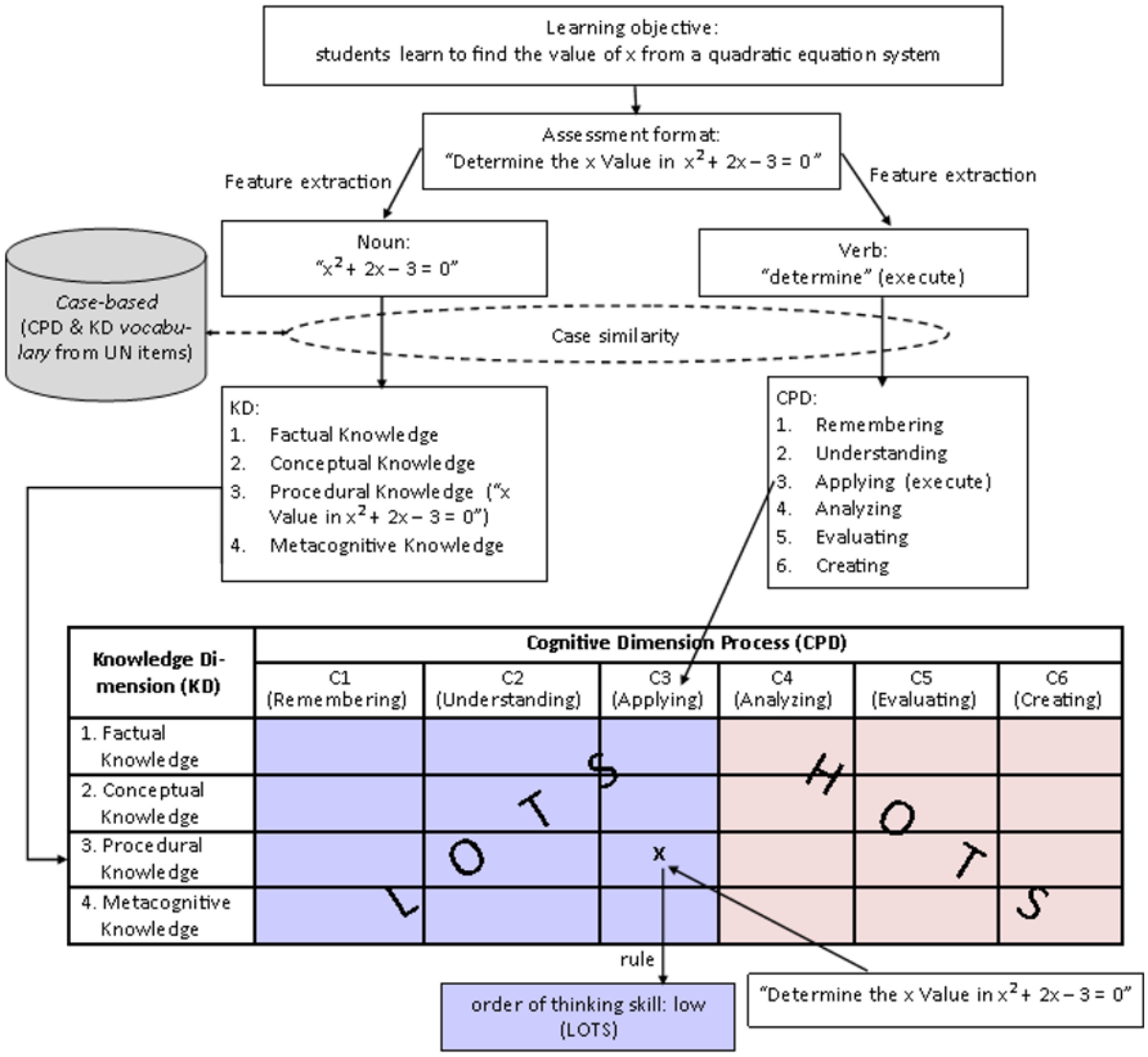

FIGURE 3. Illustration of the process of thinking skills order analysis in the test item text (example of a junior high school math problem)

?) in Bahasa, after following the equation, 12 similar cases are retrieved from class Cl_1: $e_1, e_2, \ldots, e_{12}$ as in Table 1.

Based on the twelve cases in Table 1 that are similar, there is one case with the most significant degree of similarity, namely e_4 = (determine, equation, C3). The solution is found, which is included in the RBT C3 (applying). Thus, the example item test belongs to items with a low order of thinking skills or LOTS. Finally, the implementation model's framework in analyzing the order of thinking skills with item text input is illustrated in Figure 4.

Figure 4 explains the model's whole framework from the beginning of the input process, namely the teacher's item, to analyze the order of thinking skills. The model performs feature extraction of new cases, namely the teacher-made item text to get the CPD and KD labels. Furthermore, the model processes the similarity of these two label entities' cases with the entities in a case-based (UN questions) to get a solution, the RBT category. Using the rule obtained information on the order of thinking skills in the item's text, namely LOTS.

4. **Conclusion.** The text processing and case similarity as a part of the intelligent system can be considered a model for analyzing items test to determine the teacher's questions'

TABLE 1. Case-based data and the similarity degree of a new case ($\hat{e}$) to case-based (e_i)

Case_Id (i)	Cases in case-based (e_i): (Cl_1, Cl_2, Cl_3)	Label	Degree of similarity SM ($\hat{e}, e_i$)
1	(determine, fractional number, C2)	C2_CK	$1/2 = 0.5$
2	(determine, exponent, C2)	C2_CK	$1/2 = 0.5$
3	(determine, radical number, C3)	C3_PK	$1/2 = 0.5$
4	**(determine, equation, C3)**	**C3_PK**	**1**
5	(determine, relations or functions, C2)	C2_CK	$1/2 = 0.5$
6	(determine, lines and angles, C2)	C2_CK	$1/2 = 0.5$
7	(determine, flat shapes, C2)	C2_CK	$1/2 = 0.5$
8	(determine, centralizing data, C2)	C2_CK	$1/2 = 0.5$
9	(determine, ratio, C2)	C2_CK	$1/2 = 0.5$
10	(determine, sequence and series, C2)	C2_CK	$1/2 = 0.5$
11	(determine, algebraic form, C2)	C2_CK	$1/2 = 0.5$
12	(determine, set, C2)	C2_CK	$1/2 = 0.5$
⋮	⋮	⋮	⋮
104	(conclude, centralizing data, C4)	C4_CK	0

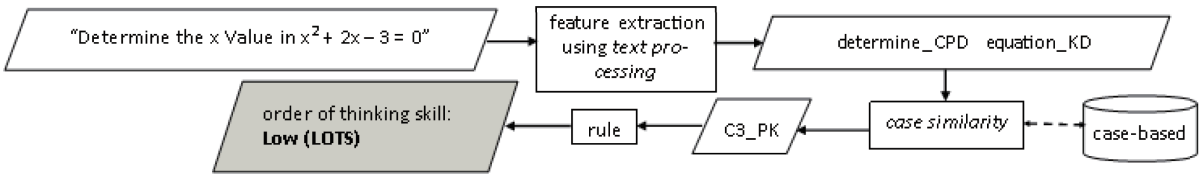

FIGURE 4. The model implementation framework analyzes the text of the item to get information on thinking skills

order of thinking skills. Based on the classification of learning objectives in the Revised Bloom's Taxonomy (RBT), it can be seen that the order of thinking skills contained in a test item is lower-order (LOTS) or higher-order (HOTS). Although the examples illustrated were the case of junior high school mathematics item text, to analyze the level of skill of elementary school and high school math items can use this model. The reason is that the text model and structure of math questions for elementary and high schools have similarities with junior high school mathematics questions. Because this model can only identify items in the form of text, future work develops a model that can identify images and numbers, considering that there are also pictures and numbers and even non-linear equations in math problems.

Acknowledgment. Thanks to Ministry of Education and Culture of Republic Indonesia (Kemdikbud-ristek) who has funded this research and Educational Assessment Center (PUSPENDIK) which supports this research as a data provider. The authors also gratefully acknowledge the reviewers' helpful comments and suggestions, which have improved the presentation.

REFERENCES

[1] Emtha, *Three Assessment Approaches in the 2013 Curriculum*, https://emtha1110.blogspot.com/2017/05/tiga-pendekatan-penilaian-dalam.html, Accessed on April 16, 2019.

[2] R. W. Tyler, *Basic Principles of Curriculum and Instruction*, University of Chicago Press, Chicago, 2013.

[3] C. R. Gareis and L. W. Grant, *Teacher-Made Assessments: How to Connect Curriculum, Instruction, and Student Learning*, 2nd Edition, Taylor & Francis, New York, 2015.
[4] N. H. Tien, D. T. Anh, M. Van Luong, N. M. Ngoc and N. Thanh, Formative assessment in the teacher education in Vietnam, *J. Hunan Univ. Nat. Sci.*, vol.47, no.8, 2020.
[5] BOSTES, *Assessment for, as and of Learning*, Board of Studies NSW, New South, 2012.
[6] D. A. Abduljabbar and N. Omar, Exam questions classification based on Bloom's taxonomy cognitive level using classifiers combination, *J. Theor. Appl. Inf. Technol.*, vol.78, no.3, pp.447-455, 2015.
[7] L. W. Anderson and D. R. Krathwohl, *Platform Framework for Learning, Teaching and Assessment (Revised Bloom's Educational Taxonomy)*, Pustaka Pelajar, Yogyakarta, 2010.
[8] Nizam, *Number of UNAS Questions and Level of Difficulty not Changed*, https://www.jpnn.com/news/jumlah-soal-unas-dan-tingkat-kesulitan-tak-diubah, Accessed on April 15, 2019.
[9] T. Suprayitno, *The Difficulty Level of 2019 UN Question Has not Changed, This Is the Composition of the Question*, https://www.kemdikbud.go.id/main/blog/2019/03/tingkat-kesulitan-soal-un-2019-tidakberubah-ini-komposisi-soalnya, Accessed on April 09, 2019.
[10] H. Muhammad, *Ministry of Education and Culture Adds Difficulty Level of UN Questions Next Year*, https://www.cnnindonesia.com/nasional/20180423160243-20-292892/kemendikbud-tambah-tingkat-kesulitan-soal-un-tahun-depan, Accessed on April 15, 2019.
[11] Sukardi, *Evaluasi Pendidikan: Prinsip dan Operasionalnya (Educational Evaluation: Principles and Operations)*, Bumi Aksara, Jakarta, 2008.
[12] L. W. Anderson and D. R. Krathwohl, *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Education Objectives*, Addison Wesley Longman, Inc., United States, 2001.
[13] PUSPENDIK, *National Examination Result Report*, https://hasilun.puspendik.kemdikbud.go.id/#2019!smp!daya_serap!99&99&999!T&03&T&T&1&!1!&, Accessed on May 28, 2020.
[14] Assessment Systems Corporation, *The World's Best Psychometric Software, in Your Hands*, https://assess.com/psychometric-software/, Accessed on September 15, 2020.
[15] K. To and Y. Wibisono, *Anates Online for Objective Tests*, ver 0.5, C00200400291-338, 2005.
[16] R. J. Adams and S.-T. Khoo, *Quest: The Interactive Test Analysis System*, Education Resources Information Center, ED362553, 1993.
[17] B. Subali and P. Suyata, Standardization of school based assessment, *J. Penelit. dan Eval. Pendidik.*, vol.17, no.1, pp.1-18, 2013.
[18] A. Muhson, B. Lestari, Supriyanto and K. Baroroh, Anbuso software as a practical and applicative item analysis tool, *Proc. of the National Economic Education Seminar FE UNY Educator Professionalism in the Dynamics of Educator*, pp.1-16, 2015.
[19] M. E. Auer, D. Guralnick and I. Simonics, *Teaching and Learning in a Digital World: Proc. of the 20th International Conference on Interactive Collaborative Learning – Volume 2*, Springer International Publishing, Switzerland, 2018.
[20] M. E. Auer and R. B. Kalyan, *Cyber-Physical Systems and Digital Twins: Proc. of the 16th International Conference on Remote Engineering and Virtual Instrumentation*, Springer International Publishing, Switzerland, 2020.
[21] M. Auer and T. Tsiatsos, *The Challenges of the Digital Transformation in Education: Proc. of the 21st International Conference on Interactive Collaborative Learning (ICL2018) – Volume 1*, Springer International Publishing, Switzerland, 2020.
[22] S. Khan and S. Alqahtani, Big data application and its impact on education, *Int. J. Emerg. Technol. Learn.*, vol.15, no.17, pp.36-46, DOI: 10.3991/ijet.v15i17.14459, 2020.
[23] A. Talasbek, A. Serek, M. Zhaparov, S. Moo-Yoo, Y. K. Kim and G. H. Jeong, Personality classification experiment by applying k-means clustering, *Int. J. Emerg. Technol. Learn.*, vol.15, no.16, pp.162-177, 2020.
[24] Q. Zhang, D. Yang, P. Fang, N. Liu and L. Zhang, Develop academic question recommender based on Bayesian network for personalizing student's practice, *Int. J. Emerg. Technol. Learn.*, vol.15, no.18, pp.4-19, 2020.
[25] S. F. Kusuma, D. Siahaan and U. L. Yuhana, Automatic Indonesia's questions classification based on Bloom's taxonomy using natural language processing a preliminary study, *Proc. of International Conference on Information Technology Systems and Innovation (ICITSI2015)*, pp.1-6, 2015.
[26] S. F. Kusuma, R. Z. Alhamri, D. O. Siahaan, C. Fatichah and M. F. Naufal, Indonesian question generation based on Bloom's taxonomy using text analysis, *Proc. of International Seminar on Intelligent Technology and Its Application (ISITIA2018)*, pp.269-274, 2018.
[27] B. D. Wijanarko, Y. Heryadi, H. Toba and W. Budiharto, Automated question generating method based on derived keyphrase structures from Bloom's taxonomy, *ICIC Express Letters*, vol.14, no.11, pp.1059-1067, 2020.

[28] G. Chen and J. Kang, Universal formulations and computation algorithm for three-phase space vector PWM ripple current analysis, *International Journal of Innovative Computing, Information and Control*, vol.14, no.2, pp.587-602, 2018.
[29] D. Rothman, *Artificial Intelligence by Example: Acquire Advanced AI, Machine Learning, and Deep Learning Design Skills*, 2nd Edition, PACKT Publishing, Birmingham, Mumbai, 2020.
[30] O. F. Tutkun, D. Guzel, M. Koroğlu and H. Ilhan, Bloom's revised taxonomy and critics on it, *TOJCE Online J. Couns. Educ.*, vol.1, no.3, pp.23-30, 2012.
[31] J. Sadik, A. E. Wijaya and A. Wibowo, Classification of mathematics question using natural language processing and text mining, *ICIC Express Letters*, vol.14, no.5, pp.423-430, 2020.
[32] Y.-J. Lee, M. Kim, Q. Jin, H.-G. Yoon and K. Matsubara, *East-Asian Primary Science Curricula: An Overview Using Revised Bloom's Taxonomy*, Springer, Singapore, 2017.
[33] BSNP, *Junior High School/Madrasah Tsanawiyah National Examination Grid for the 2018/2019 Academic Year*, Jakarta, 2018.
[34] D. J. Flinders and P. B. Uhrmacher, *Curriculum and Teaching Dialogue*, 13th Edition, Information Age Publishing, Inc., Charlotte, North Carolina, 2011.
[35] S. K. Phal, S. C. K. Siu, S. K. Pal and S. C. K. Shiu, *Foundations of Soft Case-Based Reasoning*, John Wiley & Sons, Inc., Hoboken, New Jersey, 2004.