Development of Instrument to Assess Mathematics Learning Result in 2nd Grade Elementary School

Ilena Dwika Musyafira and Aryani Tri Wrastari
Fakultas Psikologi, Universitas Airlangga, Surabaya, Indonesia
ilena.dwika.musyafira-2019@psikologi.unair.ac.id

Keywords: Elementary School, Instrument, Mathematics

Abstract: This study aims to determine the quality of the 2nd grade elementary school mathematics learning result test instrument from a cognitive perspective. The study was conducted of 33 students at Cita Hati School, Surabaya. This type of research was quantitative descriptive based on reliability, discriminating power, level of difficulty and effectiveness of distractors. Data collection was carried out by online during the Covid-19 pandemic. Development of the instrument starts from curriculum analysis based on Week 3 Lesson Plan owned by SD Cita Hati. Blueprint was prepared which refers to Bloom's taxonomy theory. The validity technique used content validity by two assessment experts. The results showed that instrument is a fairly stable and reliable with α = 0.561. The discriminatory power showed 13 items have very good discriminating power and 3 items have good discriminating power and 11 items have poor power. The level of difficulty in all items were easy to very easy category. The results of the analysis based on effectiveness of distractors indicated most of the questions don't work as a good distractors, thus making the distractors work better than the answer keys.

1 INTRODUCTIONS

Assessment is an activity to collect information from various sources to determine the quality of a learner. The quality of a student can be seen from the learning curriculum, learning programs, school climate, school policies and teacher management in the classroom. Assessment can also be interpreted as an assessment (Poerwanti, 2001). Educational assessment can be carried out using test and non-test instruments to find information in the learning process (Juhairiyah, 2017). Educational assessment is a set of activities carried out to measure the achievement of student learning outcomes. Educational assessment can be carried out by collecting and processing information that can be used as a measure of student success. Some examples of educational measurement tools are; midterm tests, final semester self-assessment tests and school final exams (Permendikbud, 2013). After the assessment is carried out, educators can provide assessments and evaluations of each ability possessed by students (Juhairiyah, 2017).

Measuring is an activity that is carried out to give a value or value to a symptom, including in the learning process. The success of the learning process can be reflected through the numbers obtained from the results of measuring the learning process which are quantitative in nature and will later be compared with certain criteria or benchmarks. Tests are a number of tasks that need to be completed by students in order to measure the level of understanding of a particular teaching (Poerwanti, 2001).

Assessment in class aims to see how far the achievements achieved by students during the learning process takes place and after the learning process takes place. In addition, as an educator, the purpose of carrying out a learning assessment is to continuously monitor the progress of each student and to be able to detect learning difficulties experienced by students. The results of this assessment can also be used as a written report that can be submitted to parents and also the school committee regarding the effectiveness of education that takes place in an educational institution (Poerwanti, 2001). Conducting an assessment at the elementary school level certainly involves three domains namely; cognitive domain, affective domain and psychomotor domain (Juhairiyah, 2017).

Musyafira, I. D., & Wrastari, A. T.
Development of Instrument to Assess Mathematics Learning Result in 2nd Grade Elementary School
In *Proceeding of the International Conference on Current Advancement in Psychology (ICCAP) 2022 - Psychology for Sustainable Recovery in the Life after the Pandemic,* page 272-276

These three domains are used to see how the quality of education has been carried out. Cognitive is related to thinking processes, affective is related to values, behavior and self-awareness and psychomotor includes applications and skills. In this study, the authors developed a test tool as a medium for conducting assessments, especially in educational assessments.

The cognitive domain consists of learning abilities related to thinking processes. The process of thinking in the cognitive domain consists of abilities in developing information processes, building understanding, applying knowledge, being able to solve problems and being able to conduct research. In this cognitive domain, the theory used refers to Bloom's taxonomy which explains the cognitive domain using a hierarchical model (Hoque, 2016). There are six levels of the cognitive domain hierarchy proposed by Bloom namely; knowledge, comprehension, application, analysis, synthesis and evaluation.

Then Bloom's taxonomy was revised and issued the latest version of the cognitive domain hierarchy which can be seen in the image below:

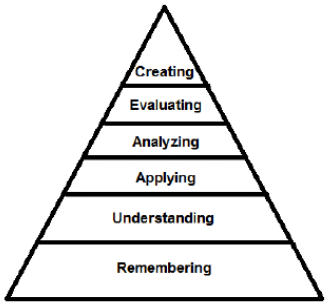

Figure 1: bloom taxonomy arc.

The pyramid shows that the higher the level, the more complex the thinking skills required. An individual cannot reach the highest level without the abilities of a lower level. When an individual rises to a higher level then these skills can be applied to everyday life (Bloom, et al., 1956).

After that Bloom's taxonomy was re-developed from unidimensional to multidimensional. This multidimensional consists of two dimensions, namely, the knowledge dimension and the cognitive dimension. The cognitive dimension refers to Bloom's taxonomy, et al. (1956) which has been revised consists of; remember, understand, apply, analyze, evaluate and create. Meanwhile, the knowledge dimension refers to Anderson & Krathwohl (2001) consisting of; factual knowledge, conceptual knowledge, procedural knowledge and metacognitive knowledge. The following is the use of Bloom's multidimensional taxonomy (see Appendix).

2 METHODS

2.1 Research Subject

The researcher determines the sample criteria that will be involved in this research, including:

- Student of SDK Cita Hati Surabaya
- Grade 2
- Follow the math lessons meeting 3-4

2.2 Data Collection

The online data collection technique used the help of Google forms on 3-4 & 7-8 December 2020. Data collection was carried out 2 times due to the request of the students' parents and the policy of the school that in one data collection, no more than 20 questions were given. So, the researcher divided the questions on two Google forms and at different times but with the same students.

2.3 Process

The first step to do this research were requested to access lesson plan week 3-4 (multiplication 3 &4) in SDK Cita Hati, Surabaya. The next step was submit a research permit, then create blue-print and create items. After the instruments were ready, researcher doing validation and reliability test. The next step was collecting data. The final step was processing data and writing the result.

2.4 Instrument

The following is attached a blueprint consisting of cognitive domains, affective domains and psychomotor domains:

3 RESULT & DISCUSSION

3.1 Validity & Reliability

Researchers use Content Validity (Content Validity) to measure students' ability to work on this test. Where Content Validity is the accuracy / suitability of a measuring instrument as a sample of the item to

be measured. To achieve content validity, a test item or measuring material must represent the overall components to be measured to represent the teaching materials provided and reflect the characteristics of the behavior to be measured (Azwar, 2014).

The instruments were tested by two expert judgements, namely 2 teachers at SDK Cita Hati; Eghita Desiane N, S. Hum & Dita Arum Wulansari, S. Pd. The results showed that the instrument is fairly stable and reliable with α = 0.561

3.2 Descriptive Statistic

The slope of the test taker's score distribution can be seen through the Skew score. Based on the data above, it is known that Skew shows the number -1.360. Negative values indicate that most of the scores are at the top (high scores) of the score distribution. So if you describe the distribution of the test answers, it will be in the form of a curve that leans to the right (negative squint).

Meanwhile, the slope of the score distribution compared to the normal distribution can be seen through the Kurtosis value. Based on the data above, it is known that Kurtosis shows the number 1.516. A positive number indicates that the curve has a sharp peak so that the distribution of the data is uneven. So from the cognitive tests conducted, it is known that the distribution of the participants' answers is not evenly distributed in all sections. It is also known that the maximum score of the cognitive tests carried out is 27 and the minimum score is 20. The average score of the test takers is 25,333 with a median of 26.

Subject categorization refers to the Norm Reference Approach (PAN), where this reference will compare the subject's score with a predetermined group norm. The norm group is obtained from a representative sample for a population. The result showed 10 subjects are included in the high category, 20 subjects in moderate category and 3 subjects in low category.

3.3 Discriminatory Power Analysis

Biser is an index of discriminating power using a biserial correlation coefficient. Distinguishing power is the ability of the items to distinguish the ability of the subject. Good items are items that have a different power > 0.3. This means that 30% of the subjects in the high group were able to answer these items correctly, and 70% of the subjects in the low group were not able to answer this question correctly. A positive score indicates that the test taker who answered the item correctly has a relatively high score on the test. On the other hand, a negative score indicates that the test taker who answered the item correctly obtained a relatively low score on the test.

Based on these results it is known that there are 13 questions that have different power in the very good category, question number 5,6,7,9,12,17,18,19,20,21,25,26,27. In addition, there are 3 questions that have different power in the good category, questions number 3,13 and 15. These questions are acceptable because they can differentiate the abilities of the subject. However, there are also 11 questions that must be discarded because they have differentiating power in the bad category or negative value, question number 1,2,4,8,10,11,14,16,22,23,24. These questions cannot be used to differentiate subject abilities.

3.4 The Level of Difficulty

Proportion Correct is the proportion of students who answered correctly on the test items. Proportion correctly describes the level of difficulty of a question. The level of difficulty of the question is the opportunity to answer a question correctly at a certain level of ability which is usually expressed in the form of an index. This index of difficulty level is generally expressed in the form of a proportion whose magnitude ranges from 0.00 to 1.00 (Aiken (1994: 66). The greater the index of difficulty level obtained from the calculation results, the easier the problem is.

Based on these results it is known that there are no questions with moderate difficulty levels. The questions that are made are in the easy to very easy category. Questions number 6, 14, 16, 25 and 26 are suggested to be revised because they are not able to measure students' actual abilities.

3.5 Effectiveness of Distractors

Based on these results it is known that the choice of answers from most of the questions cannot function as a distractor properly. Because the answer choices were not chosen at all by the subject. In addition, in question number 16 it is necessary to re-check the answer key. Because subjects with high or low scores choose the wrong answer choice (distractor). So that makes the distractor work better than the answer key.

4 CONCLUSION

The results showed that the instrument is fairly stable and reliable with α = 0.561. The

discriminatory power showed 13 items have very good discriminating power and 3 items have good discriminating power and 11 items have poor power. The level of difficulty in all items were easy to very easy category. The results of the analysis based on effectiveness of distractors indicated most of the questions don't work as a good distractors, thus making the distractors work better than the answer keys.

Due to the Covid-19 pandemic, aspects of online or offline learning are needed that can be used as evaluation. There will be different cultures among virtual and face-to-face learning cultures. Therefore, researchers need to develop research designs that are in accordance with current conditions.

REFERENCE

Anderson, L., & Krathwohl, D. (2001). A taxonomy for learning, teaching and assesing: A revision of Bloom's Taxonomy of educational objectives. New York: Longman.

Azwar, S. (2014). Tes Prestasi. Yogyakarta: Pustaka Belajar

Bloom, B., Engelhart , M., Furst, E., Hill, W., & Krathwohl , D. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain. New York: David McKay.

Hoque, E. (2016). Three Domains of Learning: Cognitive, Affective and PSychomotor . The Journal of EFL Education and Research (JEFLER), 45-52.

Juhairiyah. (2017). Asesmen Konten Isi Bidang Studi. Jurnal Pedagogik, 62-80.

Permendikbud. (2013). Peraturan Mentri Pendidikan dan Kebudayaan No. 66 Tentang Standar Penilaian . Jakarta : Kemendikbud.

Poerwanti, E. (2001). Evaluasi pembelajaran, Modul Akta mengajar. UMM Press.

APPENDIX

Knowledge Dimension	Cognitive Dimension					
	Remember	Understand	Apply	Analyze	Evaluate	Create
Factual Knowledge						
Conceptual Knowledge						
Procedural Knowledge						
Metacognitive Knowledge						

Blue Print

	Remember	Understand	Apply	Analyze	Evaluate	Create
Factual						Sc
Concept		G1	G4		Sb	
Procedural	Sa		G2	G3		
Metacognitive				Sd		

Information:
G1: Students are able to do multiplication 3 and 4
G2: Students are able to do skip counting 3 and 4
G3: Students are able to group the same values and can do the division
G4: Students are able to apply their knowledge of multiplication 2,5 and 10
Sa: Students are consistently able to analyze groups and members of the multiplication column
Sb: Students are able to use skip counting to find answers to multiplication 3 and 4
Sc: Students are able to make multiplication numbers in random order according to instructions
SD: Students are able to use multiplication sets