ANALYSIS OF COGNITIVE ASSESSMENT INSTRUMENTS IN GENERAL BIOLOGY LEARNING MEDIA

Trio Ageng Prayitno [1*], Nuril Hidayati [2]

[1,2]IKIP Budi Utomo, Jl. Simpang Arjuno 14B Malang
*Corresponding author: trioageng@gmail.com

ARTICLE INFO:

ABSTRACT

The existence of cognitive assessments in learning media that does not require much consideration for their feasibility as a measuring tool is one of the problems that arise because it can cause misinformation obtained by the teacher. The purpose of this study was to determine the construct validity and content of the cognitive assessment instrument and the validity of the questions, the reliability of the questions, the distinguishing power of the questions, and the level of difficulty of the questions on the cognitive assessment instruments to be used in general biology learning media. The research method used is descriptive quantitative through a survey involving one expert and 58 students of the Biology Education study program of IKIP Budi Utomo. The research limitation is a cognitive assessment of microbiology material in general biology courses. The type of instrument analyzed is a test. The results obtained are that the cognitive assessment instrument has a valid level of validity from the aspect of construct and content, the validity of the questions in the valid category, the reliability of the questions is moderate, the discriminatory power of the questions is good, and the level of difficulty of the questions is moderate. The conclusion is that the cognitive assessment instrument is appropriate to be used as an appropriate measuring tool for microbiology material in general biology learning media.

INTRODUCTION

Learning media is explicitly developed following the learning objectives to be achieved. Learning media must contain appropriate and accurate material and be equipped with appropriate assessments (Prayitno & Hidayati, 2017; Hidayati et al, 2019; Hidayati & Irmawati, 2019; Prayitno & Hidayati, 2020). Assessment is one of the crucial things to measure the learning objectives that have been determined. A poor assessment will not provide helpful information in the learning process, and it can be said that the results obtained are invalid (Pratama, 2019). Assessment assessments can be in the form of questions in tests and assignments (Arafah, 2018).

Some of the learning media developed only focus on material content and have not paid attention to the importance of preparing the correct assessment. Existing research related to media development, including Asyhari & Silvia (2016), Sa'diyah et al (2016), Tamimiya et al (2017), has not shown the results of the assessment of the test instrument used. Not many studies have tested and analyzed specific assessment assessments on the learning media that will be developed.

One of the existing ones is the analysis of instruments for the completeness of learning media conducted by Hidayati & Irmawati (2020), but this research is only limited to the material of the cardiovascular system. Other research only discusses instrument analysis without any follow-up on the unity between the assessment and the learning media used (Taherdoost, 2018; Azizah et al, 2018; Walid et al, 2019). The importance of analyzing the instrument for the assessment is to produce the right measuring tool and become an integral part of the learning media used. Meanwhile, this study will focus on analyzing cognitive assessment instruments on microbiology material in general biology learning media. This cognitive assessment instrument will be analyzed in depth by analyzing construct and content validity, question validity, question reliability, the discriminatory power of questions, and the level of difficulty of the questions. This study aimed to determine the construct and content validity of the instrument and the reliability, the differentiating power, and the level of difficulty of the questions. The instrument will be used in general biology learning media to produce an appropriate cognitive assessment.

METHOD

The research method used is descriptive quantitative research with survey methods to experts and students of the Biology Education study program at IKIP Budi Utomo. The research sample consisted of 1 expert competent in educational evaluation to check the construct and content validity of the instrument by filling out a validation questionnaire and 58 students who were taken using a random sampling technique to work on the questions. The instrument to be analyzed is a cognitive assessment instrument. The type of instrument analyzed was in the form of a test with 25 multiple choice questions and five answer options constructed based on learning achievement and cognitive indicators. Cognitive assessment instruments were analyzed based on validity, reliability, discriminating power, and difficulty level. The data obtained are in the form of validity assessment data by experts and the results of student answers to 25 questions in multiple-choice.

The data analysis technique was carried out with the help of SPSS to determine the value of the validity and reliability of the cognitive assessment instrument and ANATEST to determine the distinguishing power and level of difficulty of the cognitive assessment instrument. Giving the value of construct and content validity by experts can be done by referring to a rating scale of 1 (invalid), 2 (less valid), 3 (quite valid), 4 (valid), and 5 (very valid). The reference in giving a valid category on the question if the value is above the r table is declared valid, whereas if the value is below the r table, it is declared less valid. The reference for giving the question reliability criteria is the value of $0.000 – 0.400$ (low), $0.401 – 0.700$ (medium), and $0.701 – 1,000$ (high). The reference for giving the criteria for discriminating power is the value of differentiating power 0.199 (very low), the value of $0.200 – 0.299$ (low), the value of $0.300 – 0.399$ (medium), and the value of the power of difference 0.400 (high). The reference for giving the criteria for the level of difficulty is $0.000 – 0.250$ (difficult), of $0.251 – 0.750$ (medium), and a value of $0.751 – 1,000$ (easy).

RESULTS AND DISCUSSION

The data obtained from the results of this study include two parts, namely the data from the analysis of construct and content validity by experts and data from the analysis of validity, reliability, distinguishing power, and the level of difficulty of students' answers to the questions being worked on. Table 1 summarizes the results of the analysis of construct and content validity by experts on cognitive assessment instruments developed on microbiology material in general biology courses.

Table 1. Summary of Construct and Content Validity Analysis Results by Experts

No	Aspect	Description	Score
1	Relevance and Representation	Conceptual definition of the instrument	5
		Operational definition on instrument	4
		Rating scale on questions	5
		Instrument Function	5
		Instructions for respondents	4
		Representation of the number of items	5
		Answer format	5
		Scoring	5
		The population sample used	5
		Time needed	4
		Average: 4,7 (Criteria Valid)	
2	Grammar Accuracy	The use of sentences on the instrument	5
		Average: 5 (Criteria Valid)	
3	The suitability of the questions with the Learning Objectives	Students are able to describe the definition of microbiology and the object of microbiology discussion (questions number 1, 2, 3, 4)	5
		Students are able to distinguish the concepts of abiogenesis and biogenesis as theories of the origin of living things in the history of microbiology (questions number 5, 6)	5
		Students are able to analyze Koch's postulates as the basis for microbiological experiments in the laboratory (questions 7, 8, 9,10)	5
		Students are able to distinguish the characteristics of viruses, archaebacteria, and eubacteria (questions numbers 11,12,13,14, 15, 16, 17, 18, 19)	4
		Students are able to distinguish the characteristics of fungi, protozoa, and algae (20, 21, 22, 23, 24, 25)	4
		Average: 4,8 (criteria Valid)	

From the analysis results above, it is known that the construct and content validity of the cognitive assessment instruments used in microbiology material in general biology courses meet the valid aspects. Aspects of relevance and representation show valid results, which means that the developed instrument has met the eligibility criteria and can be used as a good measuring tool. A good instrument must meet a valid element of construct validity (Hidayati & Irmawati, 2020). The section on the use of language in the instrument also fulfills the valid aspect. One of the requirements of construct validity is related to the use of language, sentence structure, vocabulary, and clarity that students can understand (Hidayat, 2015). Instruments in the form of questions tested have been following the learning objectives determined on microbiology material in general biology courses with valid criteria results. The results of this study are in line with the results of research by Hidayati & Irmawati (2020) and Kusumawati & Hadi (2018) that the

evaluation questions given to students must be under the learning objectives or learning outcomes that have been set in each course. Under the purpose of testing construct validity, the above mechanism is to evaluate the validity empirically so that it can become the right instrument (Strauss & Smith, 2009; Firdaos, 2017). The instrument must meet the element of construct validity to be precise in carrying out a measurement, starting from the process of identifying the construct, defining and developing it according to theory, and determining the type of instrument to be used (Flake et al, 2017; Moafian et al, 2019).

Second data is the analysis of the validity, reliability, differentiating power, and difficulty level of the questions. The data were obtained from 58 students. Then the data were processed using SPSS and ANATEST to obtain the analysis results in Table 2.

Table 2. Results of Question Validity, Question Reliability, Differential Power of Questions, and Difficulty Level of Instruments

Number of Question	Validity (r table 0,2586)	Distinguish (%)	Difficulty level	Reliability
1	0,308 (valid)	37,50 (good)	moderate	0.670 (fair)
2	0,169 (less valid)	12,50 (less)	moderate	
3	0,437 (valid)	25,00 (fair)	very difficult	
4	0,483 (valid)	56,25 (very good)	moderate	
5	0,450 (valid)	43,75 (very good)	moderate	
6	0,413 (valid)	43,75 (very good)	moderate	
7	0,488 (valid)	43,75 (very good)	difficult	
8	0,301 (valid)	31,25 (good)	moderate	
9	-0,037 (less valid)	6,25 (less)	moderate	
10	0,192 (less valid)	12,50 (less)	difficult	
11	0,280 (valid)	18,75 (less)	moderate	
12	0,372 (valid)	43,75 (very good)	difficult	
13	0,360 (valid)	43,75 (very good)	difficult	
14	0,382 (valid)	43,75(very good)	moderate	
15	0,490 (valid)	37,50 (good)	difficult	
16	0,389 (valid)	25,50 (less)	difficult	
17	0,217 (less valid)	25,50 (less)	moderate	
18	0,305 (valid)	37,50 (baik)	difficult	
19	0,280 (valid)	25,50 (less)	sangat sukar	
20	0,350 (valid)	56,25 (very good)	difficult	
21	0,163 (less valid)	0,00 (less)	difficult	
22	0,315 (valid)	31,25 (good)	moderate	
23	0,222 (less valid)	37,25 (good)	moderate	
24	0,229 (less valid)	31,25 (good)	moderate	
25	0,256 (less valid)	25,00 (less)	moderate	

The cognitive assessment instrument used was in the form of multiple-choice questions with five answer choices. Questions in the form of multiple-choice have a high level of consistency. This is one of the reasons why the instrument used is in the form of multiple-choice (Zhongshannvga, 2007). Multiple choice questions only have one correct answer and are equipped with many alternative distractors (Kumar et al, 2016).

The analysis of the validity of the questions obtained that as many as 68% of the instruments showed valid criteria, while 32% showed less valid criteria. Valid criteria are obtained if the questions on the instrument being tested are at numbers above the r table, namely 0.2586 (df N-2, from a total of 58) respondents, while if the results of the calculation of the questions are below the r table, the conclusion is less valid. There are eight questions (2, 9, 10, 17, 21, 23, 24, and 25) with less valid criteria which will later be improved in writing systematics, sentence structure, and clarity of questions, as well as ensuring that respondents answer seriously. and the time used is sufficient (Arifin, 2017; Dewi & Sudaryanto, 2020). Good validity results indicate that the questions made meet good quality to measure specific aspects (Mokshein et al, 2019).

The importance of testing the cognitive assessment instrument will be used to see the level of validity (Kereh et al, 2015). The instrument must meet the validity criteria both from construct validity and content validity. The results of this study are in line with the results of research by Walid et al (2019) that construct, and content validity were used to obtain the right questions as a measuring instrument. Instruments with valid content validity can be used as independent measuring tools in a learning process (Van Lankveld et al, 2017).

Based on the results of the SPSS analysis in Table 2, it is known that the reliability of the instruments analyzed is 0.670 with moderate criteria. The reliability value will be better if it is close to 1 and vice versa (Nuryani, 2019). The reliability value shows that the questions in the instrument can give relatively the same results if used many times (Pratama, 2019). The reliability value can be increased by using questions with a high level of validity and consistency of questions (Puspitasari et al, 2019).

The differentiating power of the questions in the cognitive assessment instrument showed 32% very good, 28% good, 4% moderate, and 36% poor. Unprecedented power is used to determine the difference between students with high abilities and groups of students with fewer abilities (Pratama, 2019). The difference in power is in the less significant category between the high and low

ability groups, both being able to answer correctly or unable to answer correctly (Arafah, 2018). The instrument must have very good to moderate discrimination criteria in order to be able to distinguish between the upper and lower groups (Mujianto, 2017; Kusumawati & Hadi, 2018).

The level of difficulty in the questions on the instrument is 56% moderate, 36% difficult, and 8% difficult. Questions on the instrument must have a moderate difficulty level so that students do not despair in taking the test (Mujianto, 2017). The difficulty level is the ability to correctly answer a question at a certain ability level (Arafah, 2018). The difference in the level of difficulty can be caused by the placement of the order of questions (Debeer & Janssen, 2013). Cognitive assessment instruments that have been tested and analyzed for validity, reliability, discriminating power and difficulty level can be used as an appropriate measuring tool in a learning process (Lia et al, 2020).

CONCLUSION

Experts' construct and content validity on cognitive assessment instruments meet the valid aspects. The validity of the questions on the cognitive assessment instrument has met the valid criteria with improvements to the eight questions that have less valid criteria. The value of the reliability of the questions on the cognitive assessment instrument is in the medium category. Most of the questions on cognitive assessment instruments are in the very good and good categories. Most of the cognitive assessment instrument questions were in the medium category. Overall, the cognitive assessment instrument on microbiology material in general biology learning media deserves to be used as an appropriate measuring tool in learning.

ACKNOWLEDGMENT

REFERENCES

Arafah, K. (2018). Test Instrument Development for Science Process Skills in Mechanic Wave Topic for Grade XI SMA Student. *In: Proceeding Book Volume 1 1ˢᵗ International Conference On Educational Assessment And Policy (ICEAP 2018)*. http://eprints.unm.ac.id/16667/

Arifin, Z. (2017). Kriteria instrumen dalam suatu penelitian. *Jurnal Theorems*, 2(1), 301743.

Asyhari, A., & Silvia, H. (2016). Pengembangan media pembelajaran berupa buletin dalam bentuk buku saku untuk pembelajran IPA terpadu. *Jurnal Ilmiah Pendidikan Fisika Al-Biruni [Journal of Physics Education Al-Biruni]*, 5(1), 1-13. https://doi.org/10.24042/jpifalbiruni.v5i1.100

Azizah, Z. F., Kusumaningtyas, A. A., Anugraheni, A. D., & Sari, D. P. (2018). Validasi preliminary product Fung-Cube pada pembelajaran fungi untuk siswa SMA. *Jurnal Bioedukatika*, 6(1), 15-21. https://doi.org/10.26555/bioedukatika.v6i1.7364

Debeer, D., & Janssen, R. (2013). Modeling item-position effects within an IRT framework. *Journal of Educational Measurement*, 50(2), 164-185. https://doi.org/https://doi.org/10.1111/jedm.12009

Dewi, S. K., & Sudaryanto, A. (2020). Validitas dan Reliabilitas Kuisioner Pengetahuan, Sikap dan Perilaku Pencegahan Demam Berdarah. *In Prosiding Seminar Nasional Keperawatan Universitas Muhammadiyah Surakarta* (pp. 73-79). https://publikasiilmiah.ums.ac.id/bitstream/handle/11617/11916/Call For Paper NEW-78-84.pdf?sequence=1

Firdaos, R. (2017). Developing and testing the construct validity instrument of Tadzkiyatun Nafs. *Addin*, 11(2), 433-462. https://doi.org/10.21043/addin.v11i2.2497

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*, 8(4), 370-378.https://doi.org/10.1177/1948550617693063

Hidayat, P. (2015). Pengembangan Instrumen Baku Penilaian Kualitas Lembar Kerja Siswa Tematik Sub Sains Sekolah Dasar Kelas Tinggi. *Al-Bidayah: jurnal pendidikan dasar Islam*, 7(2).

Hidayati, N., & Irmawati, F. (2019). Developing digital multimedia of human anatomy and physiology material based on STEM education. *JPBI (Jurnal Pendidikan Biologi Indonesia)*, 5(3), 497-510. https://doi.org/10.22219/jpbi.v5i3.8584

Hidayati, N., & Irmawati, F. (2020). A Survey on the SETS-Based Human Anatomy and Physiology Course: Analysis of Instruments for Assessing Critical Thinking Skills Using Multimedia. *In Proceeding Biology Education Conference: Biology, Science, Enviromental, and Learning* (Vol. 17, No. 1, pp. 112-119). https://jurnal.uns.ac.id/prosbi/article/view/54048

Hidayati, N., Pangestuti, A. A., & Prayitno, T. A. (2019). Edmodo mobile: Developing e-module biology cell for online learning community. *Biosfer: Jurnal Pendidikan Biologi*, 12(1), 94-108. https://doi.org/10.21009/biosferjpb.v12n1.94-108

Kereh, C. T., Tjiang, P. C., & Sabandar, J. (2015). Validitas dan Reliabilitas Instrumen Tes Matematika Dasar yang Berkaitan dengan Pendahuluan Fisika Inti. Jurnal Pendidikan, 2(01). https://core.ac.uk/download/pdf/267822994.pdf

Kumar, H., Kumar, S., Dalabh, M., & Ahmad, J. (2016). Measurement and Evaluation in Education.

Kusumawati, M., & Hadi, S. (2018). An analysis of multiple choice questions (MCQs): Item and test statistics from mathematics assessments in senior high school. *REiD (Research and Evaluation in Education)*, 4(1), 70-78. https://doi.org/10.21831/reid.v4i1.20202

Lia, R. M., Rusilowati, A., & Isnaeni, W. (2020). NGSS-Oriented chemistry test instruments: validity and reliability analysis with the rasch model. *REiD (Research and Evaluation in Education)*, 6(1), 41-50. https://doi.org/10.21831/reid.v6i1.30112

Moafian, F., Ostovar, S., Griffiths, M. D., & Hashemi, M. (2019). The construct validity and reliability of the 'characteristics of successful efl teachers questionnaire (Coseflt-q)'revisited. Porta Linguarum Revista Interuniversitaria de Didáctica de las Lenguas Extranjeras, (31), 53-73. https://dialnet.unirioja.es/descarga/articulo/7015456.pdf

Mokshein, S. E., Ishak, H., & Ahmad, H. (2019). The use of rasch measurement model in English testing. Jurnal Cakrawala Pendidikan, 38(1), 16-32. https://doi.org/10.21831/cp.v38i1.22750

Mujianto, S. (2017). Analisis daya beda soal. taraf kesukaran, butir tes, validitas butir tes, interpretasi hasil tes valliditas ramalan dalam evaluasi pendidikan. *Jurnal Manajemen Dan Pendidikan Islam*, 2(2), 2.

Nuryani, N. (2019). Validitas dan Reliabilitas Kuesioner Pengetahuan, Sikap dan Perilaku Gizi Seimbang Pada Remaja. Ghidza: Jurnal Gizi dan Kesehatan, 3(2), 37-46. https://doi.org/10.22487/ghidza.v3i2.19

Pratama, D. (2019). Analysis Of Clasical Test Theory (CTT) Approch on Academic Ability Test Instrument. *JISAE: Journal of Indonesian Student Assessment and Evaluation*, 5(2), 43-54. https://doi.org/10.21009/jisae.052.05

Prayitno, T. A., & Hidayati, N. (2017). Pengembangan multimedia interaktif bermuatan materi mikrobiologi berbasis edmodo android. Bioilmi: Jurnal Pendidikan, 3(2), 86-93. https://doi.org/https://doi.org/10.19109/bioilmi.v3i2.1399

Prayitno, T. A., & Hidayati, N. (2020). Multimedia development based on science technology engineering and mathematics in microbiology learning. *JPBIO (Jurnal Pendidikan Biologi)*, 5(2), 234-247. https://doi.org/10.31932/jpbio.v5i2.879

Puspitasari, E. D., Susilo, M. J., & Febrianti, N. (2019). Developing psychomotor evaluation instrument of biochemistry practicum for university students of biology education. *REID (Research and Evaluation in Education)*, 5(1), 1-9. https://doi.org/10.21831/reid.v5i1.22126

Sa'diyah, W., Suarsini, E., & Ibrohim, I. (2016). Pengembangan modul bioteknologi lingkungan berbasis penelitian matakuliah bioteknologi untuk mahasiswa S1 Universitas Negeri Malang. *Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan*, 1(9), 1781-1786.

Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual review of clinical psychology*, 5, 1-25. https://doi.org/10.1146/annurev.clinpsy.032408.153639

Taherdoost, H. (2016). Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research. How to test the validation of a questionnaire/survey in a research (August 10, 2016). https://doi.org/10.2139/ssrn.3205040

Tamimiya, K. T., Ghani, A. A., & Putra, P. D. A. (2017). Pengembangan modul pembelajaran IPA berbasis SETS untuk meningkatkan collaborative problem solving skills siswa SMP pada pokok bahasan cahaya. Jurnal Pembelajaran Fisika, 5(4), 392-398. https://jurnal.unej.ac.id/index.php/JPF/article/view/4344

Van Lankveld, W., Jones, A., Brunnekreef, J. J., Seeger, J. P., & Bart Staal, J. (2017). Assessing physical therapist students' self-efficacy: measurement properties of the physiotherapist self-efficacy (PSE) questionnaire. BMC medical education, 17(1), 1-9. https://doi.org/10.1186/s12909-017-1094-x

Walid, A., Sajidan, S., Ramli, M., & Kusumah, R. G. T. (2019). Construction of the assessment concept to measure students' high order thinking skills. *Journal for the Education of Gifted Young Scientists*, 7(2), 237-251. https://doi.org/10.17478/jegys.528180

Zhongshannvga, L. L. (2007). Designing and Revising a Multiple Choice Vocabulary Test.