



How cell cognitive diagnostic assessment (CDDA) instrument analyzed science pre-service teachers' prior knowledge?

Safira Permata Dewi*, Ermayanti, Lucia Maria Santoso

Biology Education, Faculty of Teacher and Training Education, Universitas Sriwijaya, Indonesia

* Corresponding author: saphire2687@gmail.com

ARTICLE INFO

Article history

Received: 30 November 2020

Revised: 7 March 2021

Accepted: 17 April 2021

Keywords:

Cell cognitive diagnostic

Discrimination index

Difficulties index

Validity

Reliability



ABSTRACT

Increasing the effectiveness of learning is done by exploring the initial understanding of science teacher candidates for the concept of cells to be studied. Exploring the prior knowledge of science teacher candidates about cell concepts can be done using the Cell Cognitive Diagnostic Assessment (CCDA) instrument. This study aims to determine the effectiveness of the CCDA instrument that has been developed. The research sample ($n = 163$) was student science teacher candidates coming from the Department of Chemistry Education, Physics Education, and Biology Education, Faculty Teaching Training and Education, Sriwijaya University, Indonesia. The topics tested include the structure and function of cells, cell membranes and molecular transport, cell reproduction, and cell communication. The results showed that all the items developed were valid, with a high level of reliability (0.86), a very good discrimination index (0.44), and a balance was found between the number of questions classified as difficult and moderate. Although the research results show that this instrument has been valid and reliable, it still needs to be improved on the items so that they can be used in the future.

© 2021 Universitas Negeri Jakarta. This is an open-access article under the CC-BY license (<https://creativecommons.org/licenses/by/4.0>)

Dewi, S. P., Ermayanti, E., & Santoso, L. M. (2021). How CCDA instrument analyzed science pre-service teachers' prior knowledge?. *Biosfer: Jurnal Pendidikan Biologi*, 14(1), 25-35. <https://doi.org/10.21009/biosferjpb.18247>

INTRODUCTION

Assessments are carried out as a form of effort to improve the quality of learning that takes place at every level of education (Harahap, Komala & Ristanto, 2020; Lestari, Ristanto, & Miarsyah, 2019). Assessment is not only carried out at the end of learning to see learning outcomes but is also carried out before learning takes place to analyze the strengths and weaknesses of students related to the concepts being taught (Gurel, et.al., 2015). The implementation of such an assessment provides benefits for teachers and students to facilitate the learning process that will be passed by providing valuable input on various policies taken by the teacher with an orientation to the learning process that will take place (Zhao, 2013).

The results of previous research have illustrated that there is a positive influence between the diagnostic tests conducted by the teacher before the learning process and the learning outcomes achieved by students (Hikmasari, et.al., 2017; Nikmard & Tavassoli, 2019; Esomonu & Eleje, 2020). Of course, this is due to the constraints that may arise from students related to their cognitive abilities that have been reflected in the diagnostic tests previously given (Tan Geok Shim, et.al., 2017). The results obtained from the diagnostic tests that are carried out can influence the learning methods and strategies chosen by the teacher to teach these concepts to students so that they are easy to understand and have an efficient time to complete learning achievement targets (Gurel, et.al., 2015). Approaches that can be chosen to diagnose students' abilities before the learning process is carried out include using the cognitive diagnostic assessment (CDA) approach (Jang, 2008).

This CDA approach provides important information related to which concepts need to be studied more deeply and which concepts can be passed according to student needs (Galvin & Simmie, 2015). This departs from the belief that students have their understanding of various kinds of concepts, both from previous levels of education and from experiences possessed by students (Ekon & Edem, 2018; Pekel, 2019). Of course, it is expected that the understanding possessed by students is the correct understanding of the concept, but it does not rule out misconceptions. Many studies have reported the misconceptions students to have about certain concepts with various tracking methods (Andariana, et.al., 2020; Duda & Adpriyadi, et.al., 2020; Soeharto, et.al., 2020; Mahror & Mahmud, 2020). The ongoing learning process is expected to rectify these misconceptions by choosing the right learning methods and strategies. Besides that, the allocation of time must also be made effective, considering that sometimes many concepts must be learned in a short time.

The limited-time allocation and the breadth of material is one of the obstacles faced in the Biology learning process for science teacher candidates at Universitas Sriwijaya, Indonesia. Especially on the topic of cells. The results of observations made on science teacher candidates and science teachers indicate that there are difficulties in understanding the topic of cells. This is due to the abstract, complex concept of cells and the use of terminology that is difficult to remember and understand. Furthermore, the observations indicate that the various mechanisms involved in this concept make the topic of cells even more difficult to understand. One of the efforts that can be made to increase the efficiency of time in the learning process of cell concepts in science teacher candidates can be done by giving Cell Cognitive Diagnostic Assessment (CCDA) questions. It is important to give CCDA to know more about the weaknesses and strengths of the concepts possessed by science teacher candidates.

METHODS

Research Design

This research is quantitative descriptive. This study aims to evaluate the CCDA instrument developed for science teacher candidates. This is done to obtain a complete and clear picture of the instrument being developed. This study shows the validity of the items, the reliability of the

items, the discrimination index, and the difficulty index. The research lasted for four months, from July-October 2020.

Population and Samples

The population in this study were all science teacher candidates at the Faculty of Teacher Training and Education, Sriwijaya University, Indonesia. The research sample consisted of 163 science teacher candidate students from three majors, namely the Department of Chemistry Education, Physical Education, and Biology Education who took General Biology courses in the 2020/2021 academic year.

Instrument

The instrument used in this study was a previously developed CCDA instrument. The question instrument consists of 20 multiple choice questions with five alternative answer choices. The development of questions has gone through the stages of literature analysis, both analysis of supporting books and research results that are relevant and related to the topic under study. The development of the questions used refers to the expected learning objectives in General Biology courses, while the development of answer choices is carried out by considering the results of research related to students' misconceptions about the concept of cells. The topics tested in this study include cell structure and function (9 questions); cell membrane and molecular transport (4 questions); cell reproduction (5 questions), and cell communication (2 questions) (Table 1.). The expert validation (n=1) process has been carried out and it shows that the instrument used has been validated in the valid category (86.42%).

Table 1.

Question distribution based on the topics

No.	Topic	Question Number
1.	Cell structure and function	1, 2, 3, 4, 5, 7, 16, 19, 20
2.	Cell membrane and molecular transport	8, 10, 11, 12,
3.	Cell reproduction	6, 9, 13, 14, 15,
4.	Cell communication	17, 18,

The data collection process was carried out online by considering the current pandemic conditions. Data retrieval was carried out using Google Form media. The sample is given a link to the CCDA instrument used.

Procedure

This research took place through three stages of research, namely: making instruments, collecting data, and analyzing data. The stage of making the instrument is carried out by referring to the Cresswell & Cresswell framework (2017). The data collection stage begins by entering the question instruments that have been made into Google Form. The Google Form is arranged in a quiz format, so that science teacher prospective students immediately get their results after sending their answers. Furthermore, the link to the CCDA question instrument was distributed to the sample to be answered by science teacher prospective students. Testing of this instrument is carried out before the learning process takes place. The third stage is to analyze the data that has been obtained from the response of the research sample.

Data Analysis Techniques

The data that has been obtained from the data collection process is then carried out by calculating the scoring for each item tested. Each correct answer has a value of 1 and the wrong answer is worth 0 so that the total score for each item is obtained. Furthermore, the score is

used to calculate the validity of the items (1), the reliability of the questions (2), the discrimination index (3), and the difficulty index (4).

$$r_{xy} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\{N\sum X^2 - (\sum X)^2\}\{N\sum Y^2 - (\sum Y)^2\}}} \quad (1)$$

$$r_{11} = \left(\frac{k}{k-1}\right) \left(1 - \frac{\sum \sigma_b^2}{\sigma_t^2}\right) \quad (2)$$

$$D = \frac{S_A - S_B}{n} \quad (3)$$

$$p = \frac{R}{T} \quad (4)$$

After calculating the validity, the items are said to be valid if the value of $r_{count} > r_{tabel}$, while to determine the quality of the test reliability, the discrimination index, and the difficulty index are determined by referring to [Tables 2, 3, and 4](#).

Table 2.

Interpretation reliability index

Nilai	Category
0,80 < r11 ≤ 1,00	Very high
0,60 < r11 ≤ 0,80	High
0,40 < r11 ≤ 0,60	Moderate
0,20 < r11 ≤ 0,40	Low
0,00 < r11 ≤ 0,20	Very low

(Rahmah, et.al., 2020)

Table 3.

Interpretation discrimination index

Discrimination Index	Item evaluation
0,40 and above	Very good item accepted
0,30 - 0,39	Reasonably good, but subject to improvement
0,20 - 0,29	Marginal item, usually need and subject to improvement
Below 0,19	Poor items to be rejected or improved by revision

(Puthiaparampil, et.al., 2017)

Tabel 4.

Interpretation difficulty index

Difficulty index	Interpretation
0 - 0,3	Difficult
0,31 - 0,79	Moderate
0,8 - 1	Easy

(Rahmah, et.al., 2020)

RESULTS AND DISCUSSION

Making a cognitive diagnostic assessment instrument is one way that can be done to find out students' prior knowledge before the learning process is carried out. Many research results that have been conducted show that extracting students' prior knowledge has an impact on increasing the effectiveness of the learning process (van Riensen, et, al., 2019; Baek, et, al., 2015; Hailikari, et, al., 2008). The development of the cognitive diagnostic assessment instrument is carried out by considering the needs to be achieved from the learning process and also the study of the content of the concepts being studied (Cresswell & Cresswell, 2017). The CCDA

instrument development process has been carried out and produced 20 multiple choice questions with five alternative answer choices. The evaluation of instrument development begins with expert validation and the small group testing that has been done previously. Furthermore, the instrument was tested on a larger scale, namely the field group stage to see the effectiveness of the instrument in terms of item validity, reliability, discrimination index, and difficulty index.

The CCDA instrument developed is divided into four topics, namely cell structure and function; cell membrane and molecular transport; cell reproduction; and cell communication. Testing the effectiveness of this instrument was carried out through the distribution of the instrument online due to the current Covid-19 pandemic. The medium used to distribute the instrument is Google Form. The test results on 163 science teacher candidate students are presented in Table 5.

Table 5.
Result of CCDA Analysis Item Question

Item No.	Item Validity			Reliability test	Index Discrimination		Index Difficulty		
	r_{xy}	r_{count}	r_{tab}		Inter	Value	Inter	Value	Inter
1	0,16	2,12		Valid		0,09	Poor	0,83	Easy
2	0,26	3,42		Valid		0,48	Very Good	0,71	Moderate
3	0,59	9,29		Valid		0,52	Very Good	0,28	Difficult
4	0,56	8,56		Valid		0,48	Very Good	0,23	Difficult
5	0,51	7,61		Valid		0,70	Very Good	0,35	Moderate
6	0,65	10,95		Valid		0,57	Very Good	0,21	Difficult
7	0,59	9,16		Valid		0,48	Very Good	0,25	Difficult
8	0,33	4,43		Valid		0,27	Marginal	0,43	Moderate
9	0,45	6,45		Valid		0,45	Very Good	0,31	Moderate
10	0,55	8,38	1,65	Valid	0,86	0,45	Very Good	0,26	Difficult
11	0,62	10,08		Valid		0,48	Very Good	0,21	Difficult
12	0,35	4,72		Valid		0,41	Very Good	0,51	Moderate
13	0,46	6,58		Valid		0,43	Very Good	0,33	Moderate
14	0,63	10,23		Valid		0,48	Very Good	0,24	Difficult
15	0,55	8,39		Valid		0,36	Good	0,21	Difficult
16	0,38	5,13		Valid		0,34	Good	0,39	Moderate
17	0,59	9,18		Valid		0,57	Very Good	0,28	Difficult
18	0,55	8,41		Valid		0,48	Very Good	0,26	Difficult
19	0,68	11,93		Valid		0,41	Very Good	0,15	Difficult
20	0,34	4,58		Valid		0,41	Very Good	0,63	Moderate
Mean						0,44	very good	0,35	Moderate

The results showed that all the CCDA items were valid. This is indicated by the magnitude of the rcount value which is greater than the rtab as the reference value. This means that both the expert judgment and the calculation of each item developed have been valid to diagnose students' cognitive abilities in the concept of cells. One of the factors that determine the validity of an item is the readability of the item being developed (Maizeli, et, al., 2020; Roy, et, al., 2020; Oliffe, et, al., 2019; Taherdoost, 2016)). Before entering the field group stage, the questions have gone through the small group evaluation stage and have gone through the readability test. The readability test results lead to various kinds of revisions that are needed so that the sample does not feel confused and ambiguous about the word chosen in the question.

The validity of each developed CCDA item is also supported by the relatively high-reliability value (0.86). Of course, this also shows that in general the quality of the CCDA instruments developed is good. The high-reliability value indicates that this instrument will show relatively the same results even though it has been tested many times on various samples. Broadly speaking, the development of CCDA questions carried out has shown good results in terms of their validity and reliability. This finding is corroborated by other findings related to the discrimination index and the item difficulty index. Only one of the 20 questions developed required revision because this item had a poor discrimination index and an easy level of difficulty.

Several research results have shown a relationship between the incidence of discrimination and the difficulty level of the questions developed (Mehta & Mokhasi, 2014; Kheyami, et al., 2017; Velou & Ahila, 2020; Toksoz & Ertunc, 2017; Pande, et al., 2013; Sim & Rasiah, 2006; Dhakne-Palwe, et.al., 2015). Both the discrimination index and the problem difficulty index are tools used to get an idea of whether the developed questions have been constructed properly or not (Quaigrain & Arhin, 2017). The discrimination index and the difficulty index of this item also affect the validity of the CCDA items developed (Mahjabeen, et.al., 2018; Kheyami, et al., 2017; Dhakne-Palwe, et.al., 2015).

The calculation of the discrimination index is carried out to get an idea of whether the items developed have been able to distinguish between the upper and lower groups seen from the results of each question. The results of this study indicate that as many as 90% (18 questions) are in the good and very good categories which indicate that these questions can be used for CDA testing in the future, as recommended by Licona-Chavez, et.al. (2020). Meanwhile, 5% of the questions were included in the marginal category and 5% were included in the poor category, which still needed to be modified to the developed items. When viewed from the average discrimination index of all developed questions it has reached a value of 0.44 which indicates that these items have a very good ability to distinguish between the upper and lower groups.

Some questions that still need to be revised due to a very low discrimination index can be done by changing the choice of answers that are not chosen at all by students or have an efficiency level of less than 5% and changing questions that have a low level of difficulty into more difficult questions (Kehoe, 1995). Making alternative answers to trick the correct answer in multiple-choice questions is not an easy thing to do. The more answer choices that must be made to deceive the correct answer, the more difficult it is and the less likely it is to be chosen (Kheyami, et al., 2017). Referring to the results of the validity and reliability of the test, questions that are included in the poor category must be modified only. The dominant item discrimination index is classified as marginal, indicating that the questions have not been able to distinguish students from the high group and students who are in the low group (Kheyami, et al., 2017; Toksoz & Ertunc, 2017). The low ability of the questions to distinguish between the number of questions developed can be due to the subjectivity factor of the students who are the research sample and can be caused by the questions developed which are classified as difficult and easy questions so that students with high abilities find it difficult to answer the questions given or Even students who have the low ability can still answer the question or even because of the subjectivity of the examiner for some evaluation cases with description assessment techniques (Taib & Yusoff, 2014).

The results of this study also indicate that the developed CCDA instrument has a balance between the difficult (50%) and the difficult (45%) questions. The problem with the lowest difficulty index is shown in question number 19 and the question with the highest difficulty index is found in question number 1. The interesting thing about this finding is that the two questions have the same topic of study, namely cell structure, and function. This research indicates that science teacher candidates have a good understanding of cell size in general, but

science teacher candidates still do not have sufficient knowledge of the detailed organelles in cells even though this material has been studied at previous levels of education. This finding is certainly a very valuable finding for the effectiveness of the learning process that will take place. It can be noted that the learning that will take place places more emphasis on the detailed structure than the general structure of the cell.

Table 6.

Example of question

Question Number 1	Question Number 19
The correct statement regarding cell size is ...	The main functions of the mitochondrial are
a.* Cells are microscopic in size so they must be viewed under a microscope (82,82%)	a. Perform intracellular digestion (6,13%)
b. The cell is the smallest structural unit and can be seen with the naked eye (3,68%)	b. Produces enzymes to form ATP (51,53%)
c. The small, visible cells that make up all organism (12,88%)	c.* Compartments convert glucose into a simple structure (14,72%)
d. Cells are medium in size so that they can accommodate a wide variety of organelles (0%)	d. Produce various kinds of molecules needed in metabolic processes (24,54%)
e. Cells are large and continue to grow in the organism (0,61%)	e. Play a role in inheritance (3,07%)

Of course, the evaluation that is carried out when developing the questions cannot be done once, but it takes several evaluations to produce a good instrument that can be used in the long term. The choice of answer choices as a distraction from the correct answer is one of the difficult tasks so that the quality of the instrument can be ascertained both from the discrimination index and the difficulty index of the questions (Quaigrain & Arhin, 2017). The problem difficulty index is determined by the student's achievement in solving the questions being tested. Of course, easy questions do not require a lot of effort to solve these problems, on the contrary, difficult questions require more effort to solve these problems (Rahmah, et.al., 2020). The complexity of the relationship between various important aspects that must be reviewed when question development is carried out shows that item analysis is important to do to improve the quality of evaluations carried out, both in the form of diagnostic tests and formative tests to see learning outcomes.

CONCLUSION

The results of this study indicate that the instrument for the CCDA questions developed is valid and reliable for diagnosing students' prior knowledge before the learning process takes place. Some analyzes that are also important in the process of developing a question instrument include the calculation of the discrimination index and the difficulty index. Broadly speaking, the CCDA instrument developed has a good discrimination index which indicates that this question has been able to distinguish students from upper and lower groups. In addition, this CCDA instrument also has a problem difficulty index which is dominated by difficult and medium questions. This instrument can be used for future student diagnostic testing although some revisions must be made to improve the quality of the CCDA instrument.

ACKNOWLEDGMENT

Researchers are grateful to Universitas Sriwijaya, Indonesia for funding this research through the 2020 Science, Technology, and Arts Research Grant.

REFERENCES

- Andariana, A., Zubaidah, S., Mahanal, S., & Suarsini, E. (2020). Identification of biology students' misconceptions in human anatomy and physiology course through three-tier diagnostic test. *Journal for the Education of Gifted Young Scientists*, 8(3), 1071–1085. <https://doi.org/10.17478/JEGYS.752438>
- Baek, Y., Xu, Y., Han, S., & Cho, J. (2015). Exploring Effects of Intrinsic Motivation and Prior Knowledge on Student Achievements in Game-based Learning. *The Smart Computing Review, October*, 368–377. <https://doi.org/10.6029/smartcr.2015.10.001>
- Cresswell, J. W., & Cresswell, J. D. (2018). *Research design: Qualitative, quantitative and mixed methods approaches* (5th Edition). Thousand Oaks, CA: Sage.
- Dhakne-Palwe, S., Gujarathi, A., & Almale, B. (2015). Item Analysis of MCQs and Correlation between Difficulty Index, Discrimination Index and Distractor Efficiency in a Formative Examination in Community Medicine. *Journal of Research in Medical Education & Ethics*, 5(3), 254. <https://doi.org/10.5958/2231-6728.2015.00052.9>
- Duda, H. J., & Adpriyadi, A. (2020). Students' Misconception in Concept of Biology Cel. *Anatolian Journal of Education*, 5(1), 47–52. <https://doi.org/10.29333/aje.2020.515a>
- Ekon, E. E., & Edem, N. B. (2018). Conceptual Change Pedagogy and Its Effects On Students' Cognitive Achievement and Interest in Biology. 9(2), 3407–3413.
- Esomonu, N. P.-M., & Eleje, L. I. (2020). Effect of Diagnostic Testing on Students' Achievement in Secondary School Quantitative Economics. *World Journal of Education*, 10(3), 178. <https://doi.org/10.5430/wje.v10n3p178>
- Galvin, E., & Mooney Simmie, G. (2015). Identification of Misconceptions in the Teaching of Biology: A Pedagogical Cycle of Recognition, Reduction and Removal. *Higher Education of Social Science*, 8(2), 1–8. <https://doi.org/10.3968/6519>
- Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. *Eurasia Journal of Mathematics, Science and Technology Education*, 11(5), 989–1008. <https://doi.org/10.12973/eurasia.2015.1369a>
- Hailikari, T., Katajavuori, N., & Lindblom-Ylanne, S. (2008). The relevance of prior knowledge in learning and instructional design. *American Journal of Pharmaceutical Education*, 72(5). <https://www.ajpe.org/lookup/doi/10.5688/aj7205113>
- Harahap, L. J., Komala, R., & Ristanto, R. H. (2020). Assessing critical thinking skills and mastery concepts: The case of ecosystem. *Edusains*, 12(2), 223-232. <https://doi.org/10.15408/es.v12i2.16544>
- Hikmasari, P; Kartono; Mariani, S. (2017). Analyze of Diagnostic Assessment and Remedial Teaching Result of Mathematics Problem Solving Achievement by Problem Based Learning Model. *Unnes Journal of Mathematics Education*, 6(2), 215–222. <https://doi.org/10.15294/ujme.v6i2.15576>

- Jang, E. E. (2008). A framework for cognitive diagnostic assessment. *Towards Adaptive CALL: Natural Language Processing for Diagnostic Language Assessment*, (January 2008), 117–131.
- Kehoe, J. (1995). *Basic Item Analysis for Multiple-Choice Tests*. Basic Item Analysis for Multiple-Choice Tests. Eric Development Team.
- Kheyami, D., Jaradat, A., Al-Shibani, T., & Ali, F. A. (2018). Item analysis of multiple choice questions at the department of paediatrics, Arabian gulf university, Manama, Bahrain. *Sultan Qaboos University Medical Journal*, 18(1), e68–e74. <https://doi.org/10.18295/squmj.2018.18.01.011>
- Lestari, P., Ristanto, R. H., & Miarsyah, M. (2019). Metacognitive and conceptual understanding of pteridophytes: Development and validity testing of an integrated assessment tool. *Indonesian Journal of Biology Education*, 2(1), 15-24. <http://dx.doi.org/10.31002/ijobe.v2i1.1225>
- Licon-Chávez, A. L., Montiel Boehringer, P. K., & Velázquez-Liaño, L. R. (2020). Quality assessment of a multiple-choice test through psychometric properties. *MedEdPublish*, 9(1), 1–12. <https://doi.org/10.15694/mep.2020.000091.1>
- Mahjabeen, W., Alam, S., Hassan, U., Zafar, T., Butt, R., Konain, S., & Rizvi, M. (2018). Difficulty index, discrimination index and distractor efficiency in multiple choice questions. *Ann. Pak. Inst. Med. Sci., January*, 310–315.
- Mahrer, N., & Mahmud, S. N. D. (2020). Secondary school students' cognitive structures and misconceptions in respiration topic. *Humanities and Social Sciences Reviews*, 8(3), 1272–1284. <https://doi.org/10.18510/HSSR.2020.83130>
- Maizeli, A., Nerita, S., & Afza, A. (2020). An analysis of cognitive assessment readability toward biology learning outcome and process evaluation course. *Journal of Physics: Conference Series*, 1521(4), 0–5. <https://doi.org/10.1088/1742-6596/1521/4/042014>
- Mehta, G., & Mokhasi, V. (2014). Item analysis of multiple choice questions-an assessment of the assessment tool. *International Journal of Health Science and Research*, 4(7), 197–202.
- Nikmard, F., & Tavassoli, K. (2020). The Effect of Diagnostic Assessment on EFL Learners' Performance on Selective and Productive Reading Tasks. *Journal of Modern Research in English Language Studies*, 7(1), 79–104. <https://doi.org/10.30479/jmrels>.
- Oliffe, M., Thompson, E., Johnston, J., Freeman, D., Bagga, H., & Wong, P. K. K. (2019). Assessing the readability and patient comprehension of rheumatology medicine information sheets: A cross-sectional Health Literacy Study. *BMJ Open*, 9(2), 1–10. <https://doi.org/10.1136/bmjopen-2018-024582>
- 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>
- Pekel, F. O. (2019). Misconceptions and “learning doctors.” Problems of Education in the 21st Century, 77(1), 5–7. <https://doi.org/10.33225/PEC/19.77.05>

- Puthiaparampil, T., Rahman, M. M., & Lim, I. F. (2017). From Item Analysis to Assessment Analysis: Introducing New Formulae. *MedEdPublish*, 6(1), 1–12. <https://doi.org/10.15694/mep.2017.000007>
- 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). <https://doi.org/10.1080/2331186X.2017.1301013>
- Rahmah, N., Yusrizal, & Syukri, M. (2020). Analysis of multiple-choice question (MCQ) of physics final examination in senior high school. *Journal of Physics: Conference Series*, 1460(1). <https://doi.org/10.1088/1742-6596/1460/1/012143>
- Roy, D., State, W., Zhang, Z., Ma, M., Arnaoudova, V., Panichella, A., Panichella, S., Gonzalez, D., Mirakhorli, M., & Mirakhorli, M. (2020). DeepTC-Enhancer: Improving the Readability of Automatically Generated Tests. *ACM Reference Format*, 1. <https://doi.org/10.1145/3324884.3416622>
- Sim, S. M., & Rasiah, R. I. (2006). Relationship between item difficulty and discrimination indices in true/false-type multiple choice questions of a para-clinical multidisciplinary paper. *Annals of the Academy of Medicine Singapore*, 35(2), 67–71.
- Soeharto, Csapó, B., Sarimanah, E., Dewi, F. I., & Sabri, T. (2019). A review of students' common misconceptions in science and their diagnostic assessment tools. *Jurnal Pendidikan IPA Indonesia*, 8(2), 247–266. <https://doi.org/10.15294/jpii.v8i2.18649>
- Sun, Y., & Suzuki, M. (2013). Diagnostic Assessment for Improving Teaching Practice. *International Journal of Information and Education Technology*, 3(6), 607–610. <https://doi.org/10.7763/ijiet.2013.v3.345>
- Taherdoost, H. (2018). Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research. *SSRN Electronic Journal*, September. <https://doi.org/10.2139/ssrn.3205040>
- Tan Geok Shim, G., Shakawi, A. M. H. A., & Azizan, F. L. (2017). Relationship between Students' Diagnostic Assessment and Achievement in a Pre-University Mathematics Course. *Journal of Education and Learning*, 6(4), 364. <https://doi.org/10.5539/jel.v6n4p364>
- 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>
- Toksöz, S., & Ertunç, A. (2017). Item Analysis of a Multiple-Choice Exam. *Advances in Language and Literary Studies*, 8(6), 141. <http://dx.doi.org/10.7575/aiac.all.v.8n.6p.141>
- Van Riesen, S. A. N., Gijlers, H., Anjewierden, A. A., & de Jong, T. (2019). The influence of prior knowledge on the effectiveness of guided experiment design. *Interactive Learning Environments*, 0(0), 1–17. <https://doi.org/10.1080/10494820.2019.1631193>

- Velou, M. S., & Ahila, E. (2020). Refine the multiple-choice questions tool with item analysis. *International Archives of Integrated Medicine*, 7(8), 80–85. Retrieved from https://www.iaimjournal.com/wp-content/uploads/2020/08/iaim_2020_0708_13.pdf
- Zhao, Z. (2013). An Overview of Studies on Diagnostic Testing and its Implications for the Development of Diagnostic Speaking Test. *International Journal of English Linguistics*, 3(1), 41–45. <https://doi.org/10.5539/ijel.v3n1p41>