JERE 8 (2) (2019) 91 - 97

Journal of Educational Research and Evaluation

http://journal.unnes.ac.id/sju/index.php/jere

Three Tier Multiple Choice Test Instrument Design For Analysis Of Student Misconceptions In The Subject Of PPkn

Suci Widiyana ✉, **Ani Rusilowati, AT Sugeng Priyanto**

Universitas Negeri Semarang, Indonesia

Article Info

Article history:
Received 3 January 2019
Approved 02 August 2019
Published 23 August 2019

Keywords:
Test instrument, three tier multiple choice

Abstract

Weaknesses and strengths of mastery of the material students in the learning process can be identified using diagnostic tests. This study aims to determine the appropriateness of the contents of the test instrument three tier multiple choice (3TMC) for students' misconceptions analysis. The method used is Research and Development (R&D) with a 4D model consisting of define, design, development and desemination. The analysis technique used is the content validity, content reliability, different power and difficulty level. Analysis of the content validity of the 3TMC test instrument uses the Aiken formula, the reliability test of the 3TMC test instrument uses Interclass Correlation Coefficient (ICC), while the difference power test and difficulty level use Anates. Content validation data was obtained from 5 validators consisting of 3 expert lecturers and 2 PPKn teachers. The results of expert validation of the 3TMC test instruments obtained 20 valid 3TMC items with 19 items having a value of v of 1.00, 1 item with a value of v of 0.60 in item 14. The results of inter-rater reliability analysis showed an Alpha value of 0.938 with avalue Single Measures of 0.750 ≥ 0.5, so it can be seen that the rater agreed in providing an assessment with the criteria of High Reliability. Based on the results of the study it was found that the developed 3TMC test products are valid with little revision and are reliable for use, so that the 3TMC test instrument can be used for the analysis of students' misconceptions on PPKn subjects.

✉Alamat korespondensi
Kampus Unnes Kelud Raya, Semarang, 50233
E-mail: suciwidiyana.sw@gmail.com

p-ISSN 2252-6420
e-ISSN 2503-1732

INTRODUCTION

Citizenship education is a very important education for students and society in general. Citizenship education has the main components ofcivic knowledge,civic skillsandcivic disposition. The three components of citizenship according to Winarno (2013, p. 26) are closely related to the goal of personal formation of citizens. Thus, evaluation of learning becomes very important to be developed in the subjects of Pancasila Education and Citizenship. The results of the evaluation can be a guide for teachers to focus more attention on students who have not mastered the subject matter as well as to find out the learning difficulties of students.

Learning difficulties for students one of them when students experience misconceptions. Misconception is an understanding of concepts contained in the minds of students that are contrary to scientific concepts, which are influenced by the experience of students (Hammer in Mubarak, 2016, p. 102). Teachers must be sensitive to misconceptions that occur in students, so that teachers are able to design effective learning processes to overcome these misconceptions, thus misconceptions must be identified so that action can be taken to help students replace them with more scientific concepts (Tuysuz, 2009).

One way to identify misconceptions is to use a diagnostic test instrument that is given to students after the learning process is carried out. Diagnostic tests are used to determine which parts of a subject have weaknesses and provide tools to find the causes of these deficiencies and are used to find out the strengths and weaknesses of students in learning (Suwarto, 2013). Rusilowati (2015, p.2) revealed that diagnostic tests are tests that are used to find out the strengths and weaknesses of students when learning something, so the results can be used as a basis for giving follow up.

Diagnostic test three-tier multiple choice is one type of diagnostic test that can be used to identify and measure the misconceptions on the learner. The three-tier test has three levels; the first level which includes a regular multiple choice test, the second level which is a multiple choice question that asks the reason, and the third level which is a scale that asks the level of student confidence for the answers given to the two questions above. Student answers for each item are considered correct when both the choices and the correct reasons are given with a high level of confidence. The three-tier test considered more accurate in capturing students' misconceptions, because they can detect a lack of knowledge percentage by using a level of trust (Gurel, D., K., Eryilmaz, A., & McDermott, L., C, 2015). In Pancasila and Citizenship Education subjects, the use of diagnostic tests can improve the mastery of PPKn subject matter with a score of 84.15 higher than the score of students who do not use diagnostic tests that is 72.35 (Wirasini, Rasyid, & Winarno, 2017, p .464).

This study aims to developinstrument design three-tier multiple choice a valid and reliableto identify the profile of students' misconceptions in PPKn class X SMK subjects. So, it can help PPKn teachers in revealing the profile of understanding of students at the vocational level.

METHODS

Study was conducted at SMK Negeri Moenadi and SMK Negeri 1 Bawen. This type of research is development research. The development model used is a 4-D model with a modification of the Tessmer formative research type. The development model used is a 4-D model consists of four stages: define, design, develop, and dissemination (Thiagarajan, Semmel, & Semmel, 1974, p.5). In the study of the use of the development model used is a 4-D model carried out to thestage develop. The stages of the research consisted of define in the form of needs analysis and literature study, thestage design in the form of developing a prototype of the diagnostic test instrument and thestage developing in the form of validation, product assessment and small-scale and large-scale product testing. The product produced in this study is atest instrument three tier multiple choice to identify the profile of students'

misconceptions on national integration material in the frame of Bhineka Tunggal Ika.

The research subjects in this study were grade X students of SMK N H. Moenadi and SMK N 1 Bawen Agribusiness and Horticulture Study Program 2018/2019. The research object is a test instrument for misconception analysis with a three tier multiple choice test (3TMC) model. The sampling technique is random.

Data collection methods used in the form of tests, questionnaires, and observations. Questionnaires were distributed to students, consisting of assessment questionnaires and response questionnaires. The assessment questionnaire was distributed to small-scale trial subjects and the response questionnaire was distributed to large-scale trial subjects. Data collection instruments included a 3TMC model misconception analysis product, an expert validation sheet, and a student response questionnaire. The data analysiswas conducted by validity test expert technique, reliability test, different power, difficulty level and students' response analysis to 3TMC test instrument products.

RESULTS AND DISCUSSION

Contents Validity of the 3TMC Diagnostic Test Instrument The 3TMC

Diagnostic test is a test with the first level being an ordinary multiple choice test, the second level is a question containing the reasons for the reasons for the first level, and the third level is a scale regarding the level of trust or confidence of students in answering questions and reasons on the two levels above. Researchers compile for a level of confidence consisting of 6 scales that are very very unsure, very unsure, not sure, sure, very sure, and very very sure.

The instrument was developed based on the Core Competencies and Basic Competencies that have been set, which are then used to compile the syllabus of subjects. Syllabus is the main reference for the preparation of diagnostic test instruments developed, the syllabus covers the identity of subjects, core competencies, basic competencies, learning materials, indicators of competency achievement, time allocation, learning activities and learning resources. The syllabus is then used as a reference in compiling the instrument lattice. The material used as a study in this study is national integration material within the frame of Bhineka Tunggal Ika. Each question was compiled by following Bloom's cognitive domain in the revised cognitive domain of Anderson and Krathwohl (2017, p.99-133).

Validation of the 3TMC test instrument consists of 3 aspects of assessment namely material, constructive, and language aspects. Validity test conducted by 5 experts consisting of three academics and two practitioners, can be seen in Table 1.

Table 1. Expert Data on Test Validity Content

Expert	Competency Expert	Educational	Description
Expert 1	Academics and Practitioners	S3	Lecture
Expert 2	Academics and Practitioners	S3	Lecture
Expert 3	Academics and Practitioners	S2	Lecture
Expert 4	Practitioners	S2	Teacher
Expert 5	Practitioners	S2	Teacher

The experts gave responses from the instrument of content validity to the diagnostic test instrument of three tier multiple choice PPKn class X SMK subjects. the results of the analysis of expert response data can be seen in table 2.

Table 2. Analysis Results of the Validity of Content

Item	Aiken V Index	Conclucions	Item	Aiken V Index	Conclucions
1	1,00	Valid	11	1,00	Valid
2	1,00	Valid	12	1,00	Valid
3	1,00	Valid	13	1,00	Valid
4	1,00	Valid	14	0,60	Valid
5	1,00	Valid	15	1,00	Valid
6	1,00	Valid	16	1,00	Valid
7	1,00	Valid	17	1,00	Valid
8	1,00	Valid	18	1,00	Valid
9	1,00	Valid	19	1,00	Valid
10	1,00	Valid	20	1,00	Valid

Based on the analysis of the data in Table 2, it can be obtained information on expert attachment. The validity of the contents of the 3TMC diagnostic test instrument in PPKn Vocational Class X subjects was obtained that 19 items were declared valid with a V of 1.00. 1 item is declared valid with a value of 0.60 in item 14, with moderate criteria. Item 14 continues to be used with improvements based on advice from experts. The validity of the content is related to the rational analysis of the domain to be measured to determine the representation of the instrument with the ability to be measured (Retnawati, 2016, p.17). Furthermore, Budiman & Jailani's research (2014) shows that the instrument is feasible to use if the coefficient of Aiken's V is more than 0.3.

Reliability of The 3TMC Diagnostic Test Instrument

Reliability is the level of questionin assessing what you want to assess. A test can be said to be reliable if it always gives the same results when tested on the same group at different times or opportunities (Arifin, 2009). The reliability test of the 3TMC diagnostic test instrument in PPKn subjects uses the SPSS 16 application. Based on the analysis, the results show that the Alpha reliability coefficient is 0.938 and Single Measures shows a value of 0.750. Alpha reliability ≥ 0.7, it can be said that expert judgment is reliable with high criteria. ICC value ≥ 0.5, it can be said that the rater agreed in providing an assessment with high reliable criteria.

The 3TMC diagnostic test instrument in the revised PPKn class X vocational subjects based on expert advice, was then tested on a large scale subject of 214 students. Students carry out the test for 90 minutes and work on 18 items. The 3TMC diagnostic test instrument in PPKn subjects in class X Vocational High School has a fairly high reliability coefficient of 0.75. The instruments compiled have a level of reliability and consistency of measurements (Azwar, 2016, p.111). Instruments that have a reliability coefficient close to number 1 then the instrument has a high level of reliability and consistency of judgment.

Difference of The 3TMC Diagnostic Tests Instrument

Analysis ofanalysis of different power is performed twice, namely on a small scale and large scale. On a small scale consisting of 30 students and on a large scale consisting of 214 students. The results of different power analyzes in small scale trials can be seen in Table 3.

Table 3. Results of Small Scale Difference Power Analysis

Item	Differential Power	Information	Item	Differential Power	Information
1	0,62	Questions received	11	0,12	Questions discarded
2	0,62	Questions received	12	0,75	Questions received
3	0,50	Questions received	13	0,50	Questions received
4	0,50	Questions received	14	0,37	Questions accepted need

					to be corrected
5	0,00	Questions discarded	15	0,37	Questions accepted need to be corrected
6	0,25	Questions accepted need to be corrected	16	0,62	Questions received
7	0,62	Questions received	17	0,25	Questions corrected
8	0,62	Questions received	18	0,50	Questions received
9	0,25	Soal diterima perlu diperbaiki	19	0,75	Questions received
10	0,5	Questions received	20	0,62	Questions received

The results of the analysis of the different power of thediagnostic test instrument three tier multiple choice in PPKn class X SMK subjects showed that from 20 items there were 13 items received, 4 items were accepted with revision, 1 item was corrected and 2 items were discarded. Researchers make improvements in accordance with the results of the analysis of different power and discard items that are not able to distinguish the ability of different power and discard items that are not able to distinguish the ability of students. The researcher discards item number 5 and 11. The researcher discards 2 item that is unable to distinguish ability because each achievement indicator is still represented. The instrument will be revised before large-scale trials are carried out. The results of different power analyzes in large scale trials can be seen in Table 4.

Table 4. Results of Large Scale Difference Power Analysis

Item	Differential Power	Information	Item	Differential Power	Information
1	0,48	Questions Received	10	0,47	Questions Received
2	0,48	Questions Received	11	0,57	Questions Received
3	0,35	Questions received need to be corrected	12	0,75	Questions Received
4	0,35	Questions received need to be corrected	13	0,47	Questions Received
5	0,31	Questions received need to be corrected	14	0,60	Questions Received
6	0,41	Questions Received	15	0,64	Questions Received

7	0,50	Questions Received	16	0,64	Questions Received
8	0,50	Questions Received	17	0,43	Questions Received
9	0,43	Questions Received	18	0,50	Questions Received

The results of the analysis of the large scale difference power test of the 3TMC diagnostic test instrument in the PPKn subjects in class X SMK showed that out of 18 items there were 15 items accepted, and 3 items were accepted with revision.

Characteristics of 3TMC diagnostic

Test items good test questions must be valid and reliable, in addition to that test questions must have a good level of difficulty and differentiation. Difficulty level and distinguishing features are characteristic test items, includingdiagnostic test questions three tier multiple choice.

Difficulty level analysis is carried out twice, namely on a small scale and large scale. On a small scale consisting of 30 students and on a large scale consisting of 214 students. The results of different power analyzes in small scale trials can be seen in Table 5.

Table 5. Results of Small Scale Difficulty Analysis Results

Item	Difficulty Level	Information	Item	Difficulty Level	Information
1	0,40	Medium	11	0,43	Medium
2	0,37	Medium	12	0,50	Medium
3	0,80	Easy	13	0,20	Difficult
4	0,33	Medium	14	0,40	Medium
5	0,23	Difficult	15	0,27	Difficult
6	0,33	Medium	16	0,70	Medium
7	0,37	Medium	17	0,43	Medium
8	0,40	Medium	18	0,53	Medium
9	0,30	Difficult	19	0,53	Medium
10	0,33	Medium	20	0,53	Medium

The analysis shows that of the 20 items there are 1 item that has difficulty in the Easy category, namely number 3. 15 items have difficulty in the Medium category, namely numbers 1, 2, 4, 6, 7, 8,10, 11, 12 , 14, 16, 17, 18, 19, and 20. 4 items have difficulty levels in the Difficult category namely numbers 5, 9, 13, and 15.

The researcher analyzes the level of difficulty from the results of large-scale trials. The analysis showed that 18 items had varying degrees of difficulty in the easy, medium, and difficult categories. The results of the analysis of the difficulty level of large-scale trials can be seen in Table 6.

Table 6. Results of Analysis of Large-Scale Difficulties

Item	Difficulty Level	Information	Item	Difficulty Level	Information
1	0,38	Medium	10	0,59	Medium
2	0,49	Medium	11	0,30	Difficult
3	0,70	Easy	12	0,51	Medium
4	0,48	Medium	13	0,38	Medium
5	0,45	Medium	14	0,70	Medium
6	0,52	Medium	15	0,43	Medium
7	0,45	Medium	16	0,46	Medium
8	0,51	Medium	17	0,70	Medium
9	0,49	Medium	18	0,60	Medium

The analysis shows that of the 18 items there are 1 item that has difficulty in the Easy category, namely number 3. 16 items have difficulty in the Medium category, namely numbers 1, 2, 4, 5, 6, 7, 8, 9, 10 , 12, 13, 14, 15, 16, 17, and 18. 1 item has a difficulty level in the ifficult category namely number 11.

Characteristics of the 3TMC diagnostic test instrument in PPKn subjects in class X SMK large-scale trial trials are divided into two analyzes namely analysis of different power and difficulty levels.

Based on Table 4, 18 items have varying degrees of difficulty in the difficult, medium, and easy categories. The level of difficulty of the items has a purpose to determine the possibility of items that are biased and can assemble tests that have the accuracy of the question data (Rusilowati, 2017, p.36).

The distinguishing features of thediagnostic test instruments three tier multiple choice developed were mostly in the accepted and accepted categories with improvements. Problems with good differentiation can distinguish smart students and students who are not good enough. This is consistent with what was expressed by Arikunto (2009) that the power of differentiation is used to see the extent to which the ability of the item is able to distinguish between high-capacity students and low-ability students. A good test question is one that can distinguish students who really master the material or not, if the test questions cannot distinguish students who are smart and those who are not smart then the test objectives will not be achieved.

CONCLUSIONS

The 3TMC instrument developed for analyzing students' misconceptions in XK Vocational High School subjects was said to be feasible based on the results of validation obtained 20 items of 3TMC valid questions with 19 items having a v value of 1.00, 1 item with a value of v of 0.60 in item 14. The results of the analysis reliability between rater showed Alpha value of 0.938 with Single Measures value of 0.750 ≥ 0.5, so it can be seen that rater agreed in providing an assessment with the criteria of High Reliability. The results of the different power analysis showed that 15 items were accepted and 3 items were accepted with revision. Difficulty level analysis results showed 18 items had varying degrees of difficulty in the difficult, medium, and easy categories.

ACKNOWLEDGMENTS

Researcher would like to thank SMK N H. Moenadi and SMK N 1 Bawen for granting permission to conduct research, and for not forgetting the researchers' gratitude to Mr Rodhi and Mrs. Isnanti for providing support for the implementation of this research.

REFERENCES

Anderson, L. W., & Krathwohl, D. R. 2017. Kerangka Landasan Untuk Pembelajaran, Pengajaran, dan Aasesmen (Revisi Taksonomi Pendidikan Bloom). Yogyakarta: Pustaka Pelajar.

Arifin, Z. 2009. Evaluasi Pembelajaran. Bandung: Rosda.

Arikunto, S. 2009. Dasar-Dasar Evaluasi Pendidikan.Yogjakarta: Bumi Aksara

Azwar S. (2016). Reliabilitas dan Validitas. yogyakarta: pustaka pelajar.

Budiman, A. & Jailani. 2014. Pengembangan Instrumen Asesmen Higher Order Thinking Skill (HOTS) Pada Mata Pelajaran Matematika SMP Kelas VIII Semester 1.

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 & Technology Education. 11(5), 989-1008.

Mubarak, S., Endang, S., & Edy, C. 2016. Pengembangan Tes Diagnostik Three Tier Multiple Choice Untuk Mengidentifikasi Miskonsepsi Peserta Didik Kelas XI. Journal of Innovative Science Education. 5 (2). 101 – 110.

Retnawati, H. 2016. Analisis Kuantitatif Instrumen Penelitian (Panduan Peneliti, Mahasiswa, Dan Psikometrian). Yogyakarta: Parama Publishing.

Rusilowti, A. 2015. Pengembangan Tes Diagnostik Sebagai Alat Evaluasi Kesulitan Belajar Fisika. Prosiding Seminar Nasional Fisika dan Pendidikan Fisika (SNFPF). 6 (1). 2302 – 7827.

Rusilowati, A. 2017. Pengembangan Instrumen Penilaian. Semarang: Unnes Press.

Suwarto. 2013. Pengembangan Tes Diagnostik dalam Pembelajaran. Yogyakarta: Pustaka Pelajar.

Thiagarajan, S., Semmel, D., & Semmel, M. I. 1974. Instructional Development for Training Teachers of Exceptional Children. Indiana University.

Tuysuz, C. 2009. Development Of Two-Tier Diagnostic Instrumen And Assess Student Misunderstanding In Chemistry. Scientific Research and Essay. 4(6). 626 – 631.

Winarno. 2013. Pembelajaran Pendidikan Kewarganegaraan: Isi, Strategi, dan Penilaian. Jakarta: Bumi Aksara.

Wirasini, A. D., Rasyid, M. A., & Winarni. 2017. Pengaruh Penilaian Diagnostik Terhadap Penguasaan Kompetensi Mensintesiskan Kewenangan Lembaga Lembaga Negara Menurut UUD Negara Republik Indonesia Tahun 1945. PKn Progresif. 12(1), 453-465.