CHEMISTRY EDUCATION PRACTICE

FEASIBILITY OF *FOUR-TIER MULTIPLE CHOICE* DIAGNOSTIC TEST FOR IDENTIFICATION CONCEPTIONS ON CHEMICAL EQUILIBRIUM CONCEPT

Reta Nareswari Rachmadhaniar[1], Dian Novita[2]*

[1][2] Program Studi Pendidikan Kimia, Universitas Negeri Surabaya. Jalan Lidah Wetan, Lidah Wetan Kec. Lakarsantri Surabaya, Jawa Timur 60213, Indonesia.
*Coressponding Author. E-mail: diannovita@unesa.ac.id

Abstrak

Tujuan dari penelitian ini adalah untuk memperoleh kelayakan instrumen tes diagnostik *four-tier multiple choice* (4TMC) untuk identifikasi konsepsi pada materi kesetimbangan kimia yang ditinjau dari kriteria validitas, kepraktisan, dan keefektifan. Jenis penelitian ini yaitu penelitian pengembangan ADDIE dengan tahapan *Analyze, Design, Development,* dan *Evaluation.* Validitas diperoleh dari lembar validasi yang terdiri dari validitas konstruk, isi, dan bahasa yang mendapatkan modus = 4, sehingga instrumen *four-tier multiple choice* dikatakan valid. Kepraktisan diperoleh dari angket respon peserta didik yang didukung oleh lembar observasi aktivitas peserta didik. Kepraktisan ditinjau dari angket respon peserta didik sebesar 84% dengan kategori sangat praktis, sedangkan lembar observasi aktivitas peserta didik mendapatkan modus = 1, sehingga instrumen tes diagnostik dikatakan praktis. Keefektifan diperoleh dari uji reliabilitas instrumen tes diagnostik. Nilai *cronbach alpha* yang didapatkan pada uji reliabilitas sebesar 0,798, sehingga instrumen tes diagnostik yang dikembangkan dinyatakan reliabel. Berdasarkan hasil validitas, kepraktisan, dan keefektifan, maka instrumen tes diagnostik *four-tier multiple choice* dinyatakan layak untuk digunakan.

Kata Kunci: kelayakan, *four-tier multiple choice*, konsepsi

Feasibility of Four-Tier Multiple Choice Diagnostic Test for Identification Conception On Chemical Equilibrium Concept

Abstract

The purpose of this study was to obtain the feasibility of a four-tier multiple choice (4TMC) diagnostic test instrument for the identification of conceptions on chemical equilibrium material in terms of validity, practicality, and effectiveness criteria. This type of research is ADDIE development research with the Analyze, Design, Development, and Evaluation stages. Validity is obtained from a validation sheet consisting of construct, content, and language validity which gets mode = 4, so that the four-tier multiple choice instrument is said to be valid. Practicality is obtained from a learner response questionnaire supported by a learner activity observation sheet. Practicality in terms of the student's response questionnaire is 84% with a very practical category, while the student activity observation sheet gets mode = 1, so the diagnostic test instrument is said to be practical. Effectiveness is obtained from the diagnostic test instrument reliability test. The Cronbach alpha value obtained in the reliability test was 0.798, so the diagnostic test instrument developed was declared reliable. Based on the results of validity, practicality, and effectiveness, the four-tier multiple choice diagnostic test instrument is declared suitable for use.

***Keywords**: feasibility, four-tier multiple choice, conception*

_____

INTRODUCTION

According to Permendikbud No. 5 of 2022 regarding the competency standards of graduates at SMA/MA/SMALB/Paket C/other equivalent forms, students must have the competence to demonstrate the ability to analyze complex problems and ideas, conclude the results and convey arguments that support their thoughts based on accurate data. Learners are expected to have good analytical skills and be able to convey ideas related to observed phenomena.

Part of sciences that requires analytical skills in learning, filled with ideas and various interesting applications is chemistry. Conceptual understanding in Chemistry is very important because one concept is interrelated with other concepts. Van Den Berg (1991) explains that conception is an individual interpretation of a scientific concept.

There are three categories of conception, namely understanding the concept, not understanding the concept, and misconceptions. Learners' conceptions that are based on scientific concepts are called conceptual understanding; conceptions that are not based on scientific concepts are called not understanding concepts. Learners' incomprehension of a scientific concept results in learner errors in interpreting a concept. Learners' interpretations that are not by scientific concepts can lead to misconceptions. Misconception is defined as an understanding that is not by the understanding accepted by scientific experts (Winarni & Syahrial, 2016).

Based on the results of pre-research conducted on March 8, 2023, at SMAN 16 Surabaya, it shows that most students consider chemical equilibrium material to be difficult to understand because many factors affect the shift in the direction of chemical equilibrium. In the initial diagnostic test conducted, students were identified as having misconceptions about each factor affecting the shift in equilibrium, where these misconceptions occurred in macroscopic, submicroscopic and symbolic representations. Misconceptions in the concentration factor amounted to 21.38%, in the pressure and volume factor amounted to 11.72%, and in the temperature factor amounted to 15.86%.

Based on an interview with one of the chemistry teachers at SMAN 16 Surabaya, it is known that so far the teacher has never conducted an assessment with a diagnostic test instrument. The teacher explained that students often experience misconceptions related to factors that affect the shift of equilibrium. It is known that students can understand pressure and volume factors when viewed from symbolic representations, but there are often misunderstandings in submicroscopic representations that can lead to misconceptions. Learners are also often reversed in understanding the concept of pressure factor and volume factor because the two concepts are opposite to each other.

To identify students' conceptions of this material, a proper instrument is needed to help teachers find out whether students have misconceptions or not. The right instrument is a diagnostic test, in agreement with Treagust (2010) who revealed that a good method to identify students' misconceptions in learning is a diagnostic test. Diagnostic tests are tests used to determine the strengths and weaknesses of students when learning a concept, so that the results can be used as a basis for teachers to conduct follow-up.

The development of diagnostic test instruments has been done from time to time, such as two-tier, three-tier, and four-tier multiple choice. Each instrument has advantages and disadvantages, so it continues to undergo development. In the two-tier multiple choice diagnostic test, the categories of misconceptions and do not understand the concept are difficult to distinguish when students answer incorrectly. In the three-tier mutiple choice diagnostic test, confidence in answering the first and second levels is answered simultaneously (Pujayanto et al., 2018). To overcome the shortcomings of previously developed diagnostic tests, namely two-tier and three-tier multiple choice, a four-tier multiple choice diagnostic test was developed.

Four-tier multiple choice (4TMC) is a four-level multiple choice test that can be used to diagnose misconceptions. Ismail (2015) revealed that the *four-tier test* is the development of a *three-tier test* equipped with a *confidence rating* on the reason for the answer so that the level of confidence in the answer and the reason for the answer is more accurate. In the *four-tier multiple* choice (4TMC), the first tier contains questions that have various answer options, the second tier contains the level of confidence of students in answering questions in the first tier, the third tier contains the choice of reasons for answering the first tier, and the fourth tier is the level of confidence of students in choosing reasons in the third tier (Agustin et al., 2022). According to Fariyani (2015), the four-level diagnostic test has

several advantages over other diagnostic tests, namely: (1) it can diagnose misconceptions more deeply, (2) it can find out more about the understanding of concepts that students have because the level of confidence in answers and reasons is differentiated, (3) it can identify parts of the material that need further emphasis, (4) develop better lesson plans as an effort to reduce misconceptions that occur.

Based on the description of the importance of identifying students' conceptions to determine the level of understanding of chemical equilibrium material, especially on factors that affect the equilibrium shift, the researchers are interested in developing a *four-tier multiple choice* diagnostic test instrument for identification of conceptions on the material of factors that affect the equilibrium shift.

METHOD

The research method used in this research is the *Research and Development* (R&D) research method. This research model is the ADDIE *(Analyze, Design, Development, Implementation, Evaluation)* development model without going through the implementation stage, because testing in this study was not on a large scale. At the development stage, a limited trial was conducted on January 8-12, 2024 in class XII IPA 1 SMAN 16 Surabaya with 30 students who had received chemical equilibrium material. Singarimbun and Effendi (1995) explain that the minimum number of questionnaire trials is at least 30 respondents. The procedure of this research can be seen in Figure 1.
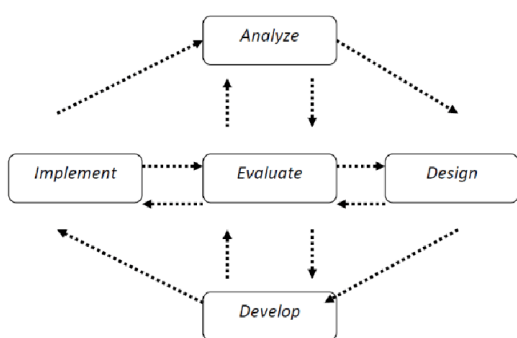
(Tegeh et al., 2015)

Figure 1. Research Procedure

The ADDIE development model steps in this study are limited only to the *development* stage, while the evaluation stage is carried out at each stage in the ADDIE development model.

There are several instruments in this study, namely the review sheet instrument, validation sheet, learner response questionnaire, learner activity observation sheet, and *four-tier multiple choice* diagnostic test instrument. The review sheet is used to collect suggestions and input as a refinement of the initial draft of the *four-tier multiple choice* diagnostic test instrument developed. The validation sheet was used to assess the validity of the *four-tier multiple choice* diagnostic test instrument based on aspects of content, construct, and language validity. Validation was carried out by three validators, namely two chemistry lecturers and one chemistry teacher. The students' response questionnaire was filled in by students after the diagnostic test. The observation sheet was assessed by 5 observers by making observations during the diagnostic test trial. Data from *four-tier multiple choice* diagnostic tests were used to identify the category of students' conceptions and calculate the reliability of the diagnostic test instruments developed.

The data analysis technique in this study, namely the data from the diagnostic test instrument review sheet, was analyzed descriptively and qualitatively related to suggestions or input from the reviewer regarding the diagnostic test instrument developed. Data from the validation sheet obtained from three validators were used to determine the validity of the *four-tier multiple choice* (4TMC) diagnostic test instrument developed. The validity results were analyzed by assessing the presence or absence of each indicator using a Likert rating scale in the following table:

Table 1. Likert Scale Validation Sheet

Skala	Indikator
1	Very Less
2	Less
3	Good
4	Very Good

(Sugiyono, 2013)

The validation result data is in the form of ordinal data which can be analyzed by determining the mode on each aspect. If the aspect assessed by the validator has a mode score ≥ 3, then the aspect is declared valid. If the aspect assessed by the validator has a mode score < 3, then the aspect is declared valid. If there are aspects that do not meet the valid requirements, then improvements and validation must be made again until they reach the specified criteria (Lutfi, 2021).

Data from students' questionnaires are used to measure the practicality of the *four-tier multiple choice* diagnostic test instrument. The score obtained will determine the percentage of practicality of the *four-tier multiple choice* diagnostic test instrument with the following formula:

Practicality percentage$=\frac{\sum score\ obtained}{\sum maximum\ score} \times 100\%$

The calculation of this score is based on the calculation of the Guttman scale as follows:

Table 2. Guttman Scale of Response Questionnaire

Answer	Positive Answer Score	Negative Answer Score
Yes	1	0
No	0	1

The percentage results were used to determine the practicality of the *four-tier multiple choice* diagnostic test instrument developed using the following categories:

Table 3. Practicality Category

No.	Persentase	Kategori
1.	0-20%	Very Bad
2.	21-40%	Bad
3.	41-60%	Fair
4.	61-80%	Practical
5.	81-100%	Very Practical

(Sugiyono, 2013)

The data from the observation of students' activities were used as supporting data for the students' response questionnaire in measuring the practicality of the *four-tier multiple choice* diagnostic test instrument developed. The observation results were analyzed based on the calculation of the Guttman scale as in Table 2. If the aspect assessed by the observer has a mode score = 1, then the aspect is declared practical. If the aspect assessed by the observer has a mode score = 0, then the aspect is declared impractical (Lutfi, 2021).

After the diagnostic test trials were carried out, the student's answers were analyzed to determine the reliability of the test instruments developed. The following is the interpretation of students' answers when working on the 4TMC diagnostic test.

Table 4. Combination of Student Answers

Answer	Answer Confidence Level	Reason	Reason Confidence Level	Criteria
Correct	High	Correct	High	Understand the Concept
Correct	Low	Correct	Low	
Correct	High	Correct	Low	
Correct	Low	Correct	High	
Correct	Low	Wrong	Low	Doesn't Understand the Concept
Wrong	Low	Correct	Low	
Wrong	Low	Wrong	Low	
Correct	High	Wrong	Low	
Wrong	Low	Correct	High	
Correct	Low	Wrong	High	
Correct	High	Wrong	High	
Wrong	High	Correct	Low	
Wrong	High	Correct	High	Misconceptions
Wrong	High	Wrong	Low	
Wrong	Low	Wrong	High	
Wrong	High	Wrong	High	

(Fariyani et al., 2015)

Test reliability is carried out by the *Internal Consistency* method, which uses one instrument that is tested only once. If students answer correctly on tier one and three and are sure on tier 2 two and four (conceptual understanding category), then a score of 1 is given, while other than the combination of these answers, namely the category of not understanding the concept and misconceptions will be given a score of 0. Testing to measure how consistent (reliable) an instrument developed can use SPSS *software*. SPSS test results are analyzed with the condition

that if the *Cronbach alpha* value is > 0.6, the instrument is said to have good reliability (Sujarweni, 2014). According to Ristianti & Fathurrochman (2020), the level of instrument reliability based on the *Cronbach Alpha* value can be seen in the following table:

Table 5. Instrument Reliability Level

Cronbach Alpha	Reliability Level
0,00-0,20	Less Reliable
0,201-0,40	Moderately Reliable
0,401-0,60	Reliable Enough
0,601-0,80	Reliable
0,801-1,00	Very Reliable

RESULTS AND DISCUSSION

Analyze Stage

At this stage, a literature review was conducted to study the theories related to the product to be developed, namely the *four-tier multiple choice* diagnostic test. In addition, a field study through pre-research was conducted to collect data and information. It is known that students of SMAN 16 Surabaya have never carried out a diagnostic test for chemical equilibrium material. When the initial diagnostic test was conducted, it was found that most students were identified as having misconceptions about the material factors that affect the shift of equilibrium.

After the analysis stage is carried out through literature review and field studies through pre-research and interviews with teachers, an evaluation stage is then carried out to assess whether the analysis carried out by with the needs or needs improvement. The use of diagnostic test instruments in schools has never been done and it is known that some students still experience misconceptions and do not understand the concept of chemical equilibrium material. Understanding the wrong concept or not according to the theory is fatal, so an instrument is needed that can identify students' conceptions of the material, namely a diagnostic test instrument.

Design Stage

Based on the analysis that has been done, the next stage is the *design* stage, namely designing products to identify students' conceptions. The product designed is a diagnostic test instrument because diagnostic tests are considered the right instrument to identify misconceptions. This was stated by Treagust (2010) who explained that diagnostic tests are a good method for identifying students'

misconceptions about a material. The diagnostic test in this study is a *four-tier multiple choice* type, where this test consists of four tiers in it. The following is the stage of designing a *four-tier multiple choice* diagnostic test instrument.

- Determine the material used in the preparation of diagnostic test questions.
- Make a concept map that shows the sequence in understanding the concept of factors that affect the shift of equilibrium. The concept sequence is used as a reference in making the sequence of questions on the diagnostic test.
- Determine the learning outcomes of the material tested in the diagnostic test and the question indicators as a measure of the success of the learning outcomes in the material tested in the diagnostic test.
- Designing a diagnostic test question grid consisting of title, education unit, subject, class/semester, subject matter (submaterial), question number, question indicator, question items, and answers.
- Designing diagnostic test items with a *four-tier multiple choice format* consisting of four tiers of questions with macroscopic, submicroscopic, and symbolic representations.

At the final stage of *design*, an evaluation was carried out by the reviewer, namely the supervisor to get suggestions for improving the initial draft of the diagnostic test instrument.

Development Stage

At this stage, the product will be realized based on the design that has been made previously, namely the four-tier multiple choice diagnostic test instrument to identify students' conceptions. After making improvements according to the reviewer's suggestions and input, the diagnostic test instrument was then validated by three validators before a limited trial was conducted. The validated diagnostic test instrument was then tested on students.

Validity

The validation stage was carried out by three validators, namely two chemistry lecturers and one chemistry teacher. Three aspects are assessed during the validation process of diagnostic test instruments, namely content, construct, and language validity. Content validity is a validity criterion that assesses the extent to which parts of the instrument represent components of the overall content of the object to be measured. Construct validity is a validity

criterion that shows the extent to which the instrument assesses a certain theoretical ability or construct to be measured. As for language validity where every question, answer choice, and reason choice on the diagnostic test must use sentences that are easy to understand so that there is no error in interpreting the question.

The following are the results of the validity of the *four-tier multiple choice* diagnostic test instrument:

Table 6. Validity Assessment Results

| Question Item | Mode | | |
	Content Validity	Construct Validity	Language Validity
Q1	4	4	4
Q2	4	4	4
Q3	4	4	4
Q4	4	4	4
Q5	4	4	4
Q6	4	4	4
Q7	4	4	4
Q8	4	4	4
Q9	4	4	4
Q10	4	4	4
Q11	4	4	4
Q12	4	4	4
Q13	4	4	4
Q14	4	4	4
Q15	4	4	4

Based on Table 5, it is known that all items get mode 4 with very good categories for aspects of content, construct, and language validity. Based on the data analysis of the validation results, it can be concluded that the diagnostic test instrument in terms of content, construct, and language validity aspects is declared valid because it has a mode score ≥ 3 (Lutfi, 2021).

Practicality

The practicality of the developed *four-tier multiple choice* diagnostic test instrument is obtained from the analysis of the learner response questionnaire supported by the learner activity observation sheet. The students' response questionnaire was given in the form of a *google forms link* which was filled in by 30 students after the *four-tier multiple choice* diagnostic test. 10 statements that must be filled in the learner response questionnaire. The results of the analysis of the learner response questionnaire are presented in the following table.

Table 7. Response Questionnaire Analysis Results

Statement	Skor	Skor Max	%
1	29	30	96,7%
2	30	30	100%
3	28	30	93,3%
4	25	30	83,3%
5	25	30	83,3%
6	27	30	90%
7	21	30	70%
8	30	30	100%
9	26	30	86,7%
10	11	30	36,7%
% Average			84%

Based on Table 6 above, the overall percentage of practicality is 84%, so the *four-tier multiple choice* diagnostic test instrument developed can be said to be very practical. A diagnostic test instrument is said to be practical if the percentage of practicality is $\geq 61\%$.

The practicality of the diagnostic test instrument developed is supported by the student activity observation sheet filled out by 5 observers. All observers observed the activities of students when working on the *four-tier multiple choice* diagnostic test. From the assessment of the five observers, the mode of each observed aspect is then determined. On the observation sheet there are 6 activities for each factor that affects the shift in equilibrium. The following are the results of the analysis of the student activity observation sheet:

Table 8. Observation Sheet Analysis Results

No.	Answer the Question	Mode
1.	Concentration Factor	1
2.	Pressure And Volume Factors	1
3.	Temperature Factor	1

Based on Table 7 above, it is known that in working on diagnostic test questions for each factor that affects the equilibrium shift, all aspects observed have a mode of 1. The observation results show that students work on all *four-tier multiple choice* diagnostic test questions by the instructions provided and take place in an orderly manner. This supports the results of the students' response questionnaire related to the diagnostic test instrument developed. Based on the analysis of the student's response questionnaire supported by the observation sheet, it can be concluded that the developed *four-tier multiple choice* diagnostic test instrument is declared very practical.

Effectivity

The effectiveness of the *four-tier multiple choice* diagnostic test instrument developed is obtained from the instrument reliability test. Students work on the diagnostic test instrument developed and the results of the answers are then processed or tested for reliability. The reliability test was carried out using SPSS to determine the level of accuracy of the diagnostic test questions developed. The following are the results of the *four-tier multiple choice* diagnostic test instrument reliability test:

Table 9. Reliability Test Results

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of items
,798	,775	15

Item-Total Statistic

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Soal1	2,93	8,961	,116	,810
Soal2	2,90	7,610	,641	,765
Soal3	3,10	8,645	,411	,787
Soal4	3,17	9,247	,176	,799
Soal5	3,17	9,385	,051	,803
Soal6	2,80	8,303	,322	,795
Soal7	3,07	8,478	,436	,785
Soal8	2,97	7,895	,577	,772
Soal9	2,97	7,964	,546	,775
Soal10	2,97	7,620	,704	,761
Soal11	3,10	9,541	-,084	,814
Soal12	2,90	7,472	,701	,759
Soal13	2,93	7,926	,532	,776
Soal14	3,00	8,552	,319	,793
Soal15	2,83	8,006	,443	,784

Based on Table 8 above, it can be seen that the diagnostic test instrument developed has a *Cronbach alpha* value of 0.798. This shows that the *four-tier multiple choice* diagnostic test instrument developed is said to be reliable because of the value of *Cronbach alpha* > 0.6 (Sujarweni, 2014). Based on the reliability test results obtained, it can be concluded that the diagnostic test instrument developed is said to be effective because it has good reliability. This is relevant to research conducted by Izzah and Madlazim (2019), the diagnostic test instrument developed to detect misconceptions with a *four-tier* format has a *Cronbach alpha* value of 0.843, so the instrument can be said to be very reliable.

CONCLUSIONS

Based on the analysis and discussion that has been done, it can be concluded that the development of *four-tier multiple choice* diagnostic test instruments for identification of conceptions on chemical equilibrium material is feasible to use. The feasibility of diagnostic test instruments is reviewed from the aspects of validity, practicality, and effectiveness. Based on the validation sheet in terms of content, construct, and language validity aspects, the mode score of 4 is obtained, so that the instrument developed is declared valid. Based on the results of the student response questionnaire, the overall level of practicality of the diagnostic test instrument developed was 84%. The practicality of the diagnostic test instrument is also supported by the student activity observation sheet and gets mode = 1, so the diagnostic test instrument is declared very practical. Based on the results of the reliability test, the *Cronbach alpha* value is 0.798, so the diagnostic test instrument is declared reliable and effective to use.

REFERENCE

Agustin, Utami., Endang, Sulasiningsih., Sri Nurhayati., & Nanik Wijayanti. (2022). Pengembangan Instrumen Tes Diagnostik Four Tier Multiple Choice Untuk Identifikasi Miskonsepsi Siswa Pada Materi Kesetimbangan Kimia. *Journal of Chemistry In Education*, 11(1).

Bernal-Ballen, A., & Ladino-Ospina, Y. (2019). Assessment: A suggested strategy for learning chemical equilibrium. *Education Sciences*, 9(3).

Fariyani, Qitshi., Ani, Rusilowati., & Sugianto. (2015). Pengembangan Four-Tier Diagnostic Test Untuk Mengungkap Miskonsepsi Fisika Siswa SMA Kelas X. *Journal of Innovative Science Education*, 4(2).

Fitri, N.S., Samsuddin, A., & Nugraha, M.G. (2017). Identifikasi Miskonsepsi dan Penyebab Miskonsepsi Siswa Menggunakan Four-Tier Diagnostic Test pada SubMateri Fluida Dinamik: Azas Kontinuitas. *Jurnal Fisika Universitas Pendidikan Indonesia*, 3(2), 175-180.

Ilham, I.P.P., & Novita, D. (2020). Validitas Dan Reliabilitas Instrumen Tes Diagnostik Four-Tier Multiple Choice (4TMC) pada Konsep Laju Reaksi. *Unesa Journal of Chemical Education*, 9(2), 222-227.

Izzah, Nurul & Madlazim. (2019). Development Four-Tier Diagnostic Test Instruments On

Newton's Law Of Gravitation Concept. *Jurnal Inovasi Pendidikan Fisika*, 8(2), 552-555.

Lutfi, Achmad. (2021. *Research and Development (R&D): Implikasi dalam Pendidikan Kimia*. Jurusan Kimia FMIPA Universitas Negeri Surabaya.

Nofrion. (2016). *Komunikasi Pendidikan: Penerapan Teori dan Konsep Komunikasi dalam Pembelajaran*. Jakarta: Kencana.

Nurulwati, A. Rahmadani. (2019). Perbandingan Hasil Diagnostik Miskonsepsi Menggunakan Three Tier Dan Four Tier Diagnostic Test Pada Materi Gerak Lurus. *Jurnal Pendidikan Sains Indonesia*, 7(2), 101-110.

Perwitasari, A.D. (2015). Pengembangan Tes Diagnostik Berbasis Web pada Materi Termodinamika untuk Mengidentifikasi Tingkat Pemahaman Konsep Siswa. *Jurnal Pendidikan Fisika,* 2(1).

Pujayanto., Budiharti, R., Radiyono, Y., Nuraini, N.R.A., Putri, H.V., Saputro, D.E., & Adhitama, E. (2018.) Pengembangan Tes Diagnostik Miskonsepsi 4 Tahap materi Kinematika. Cakrawala Pendidikan: *Jurnal Ilmiah Pendidikan*, 37(2), 238-249.

Riduwan. 2015. *Dasar-Dasar Statistika*. Bandung: Alfabeta.

Ristianti, Dina.,& Irwan, Fathurrochman. (2020). *Penilaian Konseling Kelompok*. Yogyakarta: Deepublish.

Rusilowati, A. (2015). Pengembangan Tes Diagnostik Sebagai Alat Evaluasi Kesulitan Belajar Fisika. *Prosiding Seminar Nasional Fisika Dan Pendidikan Fisika*, 6(2), 2-7.

Rusman. (2017). *Belajar dan Pembelajaran Berorientasi Standar Proses Pendidikan*. Jakarta: Kencana.

Savira, Intan., et al. (2019). Desain Instrumen Tes Three Tiers Multiple Choice Untuk Analisis Miskonsepsi Siswa Terkait Larutan Penyangga. *Jurnal Inovasi Pendidikan Kimia*, 13(1), 2277-2286.

Singarimbun, M., & Efendi. (1995). *Metode Penelitian Survey*. Jakarta: PT. Pustaka LP3ES.

Sujarweni, Wiratna. 2014. *Metodologi Penelitian*. Yogyakarta: Pustaka Baru Press.

Tegeh, I.M., Jampel, I.N. and Pudjawan, K., 2015. Pengembangan buku ajar model penelitian pengembangan dengan model ADDIE. In *Seminar Nasional Riset Inovatif,* 4(208), 208-216.

Treagust, D., & Tsui, Chi-Yan. 2010. Evaluating Secondary Student's Scientific Reasoning in Genetics Using a Two-Tier diagnostic Instrument. *International Jurnal of Science Education*, 32(8),1073-1098.

Utari, Juli I & Ermawati, Frida U. (2018). Pengembangan Instrumen Tes Diagnostik Miskonsepsi Berformat Four-Tier untuk Materi Suhu, Kalor dan Perpindahannya. *Jurnal Inovasi Pendidikan Fisika*, 7(3), 434-499.

Van Den Berg, Euwe. (1991). *Miskonsepsi Fisika dan Remediasi*. Salatiga: Universitas Kristen Satya Wacana.

Winarni, Sri., & Syahrial. (2016). Miskonsepsi Kimia yang Disebabkan Pernyataan Nonproposisi. *Jurnal Pendidikan Sains*, 4(4), 122–12