Advances in Social Science, Education and Humanities Research, volume 528
Proceedings of the 7th International Conference on Research, Implementation, and Education of Mathematics and Sciences (ICRIEMS 2020)

Assessment Instrument on Measuring Physics Verbal Representation Ability of Senior High School Students

R Adawiyah [1,*] E Istiyono [1]

[1] *Physics Education Graduate School, Universitas Negeri Yogyakarta, Indonesia.*
[*] *Corresponding author. Email: awa23795@gmail.com*

ABSTRACT
The problem questions used in the physics learning assessment process do not only require the ability of students to apply concepts only in solving problems. But more than that, these questions measure how the concept can be applied in various situations, and how students make sense in solving the problems. Representing the same concept and then conveying it in a different form is the meaning of multiple representations. The ability of multi representation is one of the most important abilities to be mastered by students. For this reason, this study aims to develop a question instrument and to determine the characteristics of the verbal representation questions of Physics on Particle Dynamics, Newton's Law of Gravity, and Work and Energy. A total of 39 verbal representation questions were developed and tested on 322 students in Yogyakarta. The selection of the trial subjects was done by using the purposive sampling method. The steps for developing the instrument are, (1) Determination of test objectives, (2) Determination of test materials, (3) Designing test blueprints, (4) Writing test questions based on verbal indicators, (5) Content Validation, (6) Determination test subjects, (7) test implementation, and (8) analyzing test results. A total of 39 verbal representation questions have been developed with the format of two-tier multiple-choice questions. Valid items based on the results of content validation analysis using Aiken's V based on expert judgment and empirically, according to the PCM model. The difficulty index of all items on the test instrument is in a good category between -0.49 to 1.9. Test questions are very suitable for measuring verbal representation skills with an ability range of $-2.3 \leq \theta \leq 1.9$. The results of this test item analysis can be used as a guide in assembling and presenting questions to measure the verbal representation ability of high school students in physics learning.

Keywords: *Development of question instruments, verbal representation skills, physics assessment.*

1. INTRODUCTION

Physics learning is considered to have achieved success if students are able to understand concepts, solve problems and apply them to everyday life [1][2]. Evaluate or reward learning outcomes including efforts to improve learning outcomes [3][4][5][6][7]. The purpose of the learning assessment is used to determine the extent of the achievement of educational objectives or determine the level of learning achievement that has been set [8][9]. The results of the assessment process in the form of information and test results are used to convey information, as discussion material with students, and material accountability to parents to improve students' mastery of concepts and competencies. [9][10][11].

Based on an analysis of the results of the PISA international assessment in 2019 Indonesia received a low average score on the assessment of science competencies [12]. The results of this study can be a reference to the low ability to solve science problems of Indonesian students [12]. Physics National Examination Results from year to year 46,35 (2017), 39,85 (2018), and49, 57 (2019) from these data it can be seen that Physics gets a low average value from year to year [13]. The low average value of Physics in the national exam shows that the learning objectives have not been reached

optimally [13]. One of the factors causes the low results of the two assessments is partly because students are not skilled enough in working on and completing the test questions presented in the assessment process [12][14][15]. PISA questions and National Examination questions contain components of thinking ability ranging from high-level thinking ability (HOTS), representation, problem solving, and reasoning [16][17][18]. The ability of representation is the basic ability of conceptual and procedural understanding used in problem solving [17][19][20][21].

The process of learning physics requires students to use multiple representations [22]. The use of various representations is an important component in the effective learning process in learning physics [23]. Students learn more effectively and meaningfully when they process and obtain information in various ways [24][25][26][27]. Verbal or textual representations are very important in explaining scientific phenomena that can be conveyed in moresimple and easily understood language [28][29][30]. Verbal explanations or statements can be presented in writing or verbally articulated [31][32]. Physics learning really needs verbal ability, because the thought process cannot be done without this ability [33]. If you want to know the representation ability of students in solving problems in the matter of representation, it is necessary to develop an instrument to measure that ability. However, school teachers more often develop assessment instruments to measure and train students' high-level thinking skills, assessment instruments to measure and practice representation skills are rarely developed [27]. Therefore, it is very important to develop instruments and know their characteristics in order to measure and know the abilities of students, especially the most basic representations namely, verbal representations as a form of qualitative problem solving [27][30].

2. METHOD

The research aims to develop an instrument physics verbal representation. The selection of trial subjects was carried out by purposive sampling method, selecting respondents with certain considerations. The test respondents were students of class XI MIPA who had studied the material to be tried out with the developed question instruments. The research was conducted at SMA N 8 Yogyakarta, SMA N 5 Yogyakarta, and SMA 11 Yogyakarta. Respondents were 322 students. The instrument development stage followed the steps of making the instrument developed [7][27], it can be seen in **Figure 1**. The instrument making stage included: (1) Determination of test objective, (2) Determining of test materials, (3) Designing off test blue-print, (4) Writing test question based on verbal representation indicator, (5) Content Validation, (6) Determination of test subjects, (7) Tesat implementation, and (8) Analyzing test result. The test validity phase includes: (1) content validation of test items and (2) empirical validation test items. The testing phase of the test includes: (1) determining test subject, (2) the trial test, and (3) analyzing the results of the trial [27][34].

2.1 Content validity

The validity of the instrument items was analyzed from the questionnaire data that had been filled in by expert validators. The data obtained is in the form of an assessment of the items with an interval scale of 1 to 4 as well as suggestions and input from expert validators. The data obtained were analyzed using Aiken's V formula (Aiken, 1985). The content validation coefficient obtained was then compared with the content validation criteria by Aiken (1985). The formula proposed by Aiken is as follows:

$$V = \frac{\sum s}{(n(c-1)} \tag{3.1}$$

V = content validity coefficient

s = r-lo, that r= score given by expert and lo= lowest score (1)

c = highest score (4).

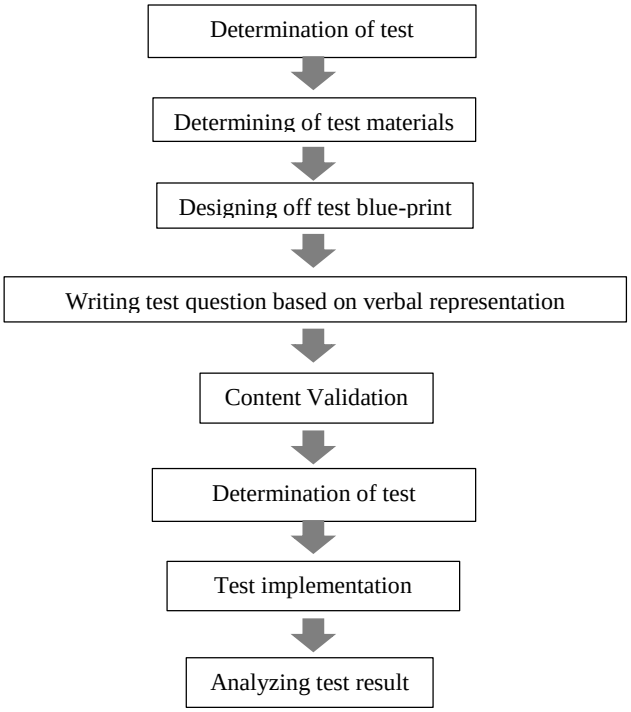


Figure 1. Steps of Verbal Representation Instrumen Development

2.2 Empirical Validation

The empirical validity of the questions is known from the fit of the items with the PCM model or what is called the fit item [1]. 4 categories of polytomus data from the test results of students were analyzed using the Quest program. The results of the empirical validity analysis can be seen in the program output on the value INFIT mean of square and INFIT t.

2.3. Reliability

Reliability is obtained from the output that displays the reliability of estimates of the test instrument set. Reliability estimates according to IRT were calculated based on the reliability of item estimates (item spacing) obtained from the Quest program analysis. then matched with the criteria in Table 1. Following [43]:

Table 1 Value and reliability criteria

Reliability Value (R)	Reliability Criteria
R<0.67	weak
0.67 ≤ R < 0.80	enough
0.80 ≤ R ≤ 0.90	good
0.90 < R ≤ 0.94	very good
R> 0.94	excellent

The suitability of the items with the abilities of students at high, medium, and low levels can be seen from the information function and Standard Error of Measurement (SEM) [2]. The relationship between the information function and SEM can be seen from the Parscale program output.

3. RESULT AND DISCUSSION

3.1 Test Development Result

This verbal representation problem was developed, total of 39 verbal representation questions with the format of two-tier multiple-choice questions. Distribution of the indicators items is presented in **Table 2.**

Table 2 The distribution of items test

Aspect	Verbal Indicator	Newton's laws of motion Indicator	Newton's laws of gravitation	Work and Energy
Verbal Representasion	Finding concepts verbally through the presentation of physical phenomena, data, and information presented.	Identify the application of the principle of Newton's 1st law (law of inertia) in everyday life.	States Newton's Law about gravity.	Formulate the relationship between force, energy, work and power in a simple form.
	Explain the phenomenon of concepts, data, and information in the form of words (verbal).	Identify the application of Newton's 2nd law in everyday life.	Analyzing the force of gravity associated with the force between two massy particles and their application.	Formulating the concept of power in the form of equations and their relation to work and energy
	Identifying a concept from the text (verbal) or picture explains it in verbal form.			Shows work links with kinetic energy.
	Interpret the meaning of a concept in written form (verbal).	Identify the application of Newton's third law in everyday life.	Analyze the strength of the gravitational field and the acceleration of gravity.	Shows work links with potential energy.
	Resolving visual form problems into verbal forms.			Implements the law of conservation of mechanical energy in daily life.

3.2 Instrument Validity

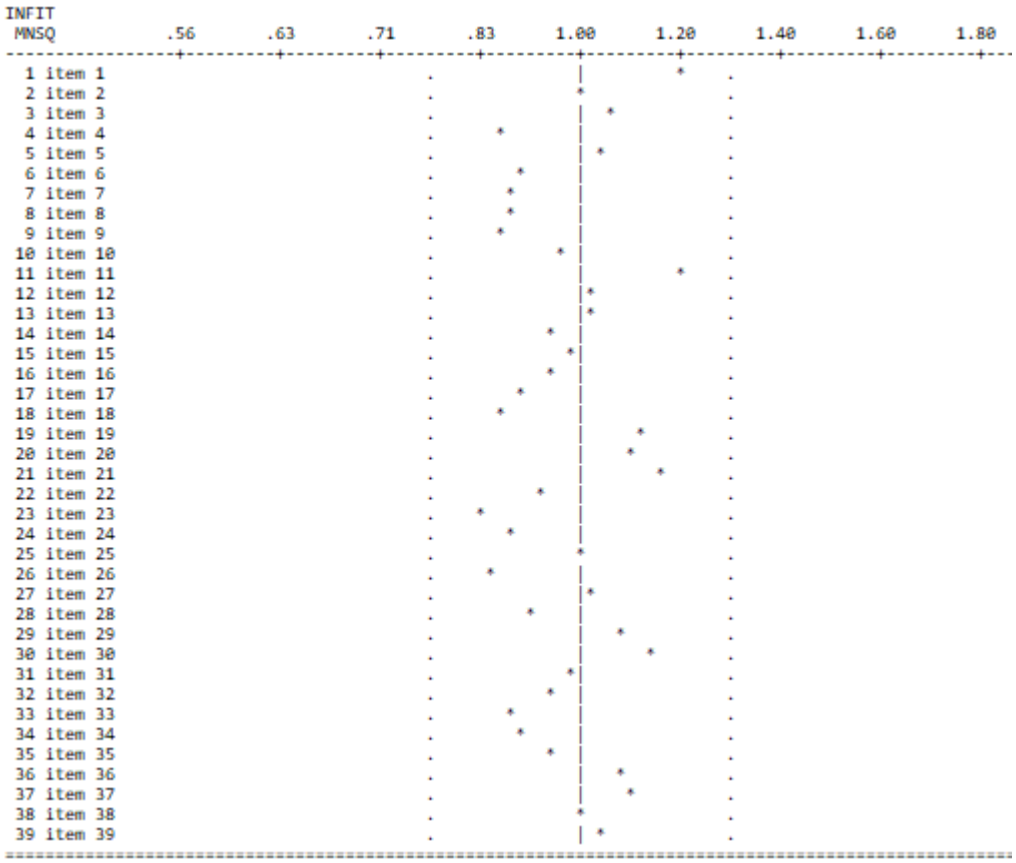


Figure 2. Infit MNSQ of items

Table 3. Item Estimation Results and Verbal Test Estimates

Item Parameter	Item Estimate	Testi Estimate
Mean Value and INFIT MNSQ standard deviation	0.98 ± 0.10	1.00 ± 0.33
Average value and standard deviation of INFIT t	-0.06 ± 1.10	0.07 ± 0.97
Reliability	0.81	

The V'Aiken index value of 39 items is in the range of 0.85-0.91, then it is included in the Very Hight classification [35]. The analysis of the result test shows that the empirical validity of test for each item and for the whole test can be proven. According to Adam and Khoo the empirical validity of the problem is determined based on the *goodness of fit* of the *partial credit model* (PCM) [36]. The item or testi passes and is declared fit with the MNSQ INFIT range limit from 0.77 to 1.30 [37]. Based on the provisions of Adam & Khoo and Hambelton, all test items fit the PCM model [36][37].

3.3 Difficulty Index

The difficulties index of the item is between -0.68 to 0.57. Thus, based on the level of difficulties (-2.0<b<2.0). The 39 items are categorized good items from the instruments [37][38]. **Figure 3.** shows the level of difficulties in indicators items test.

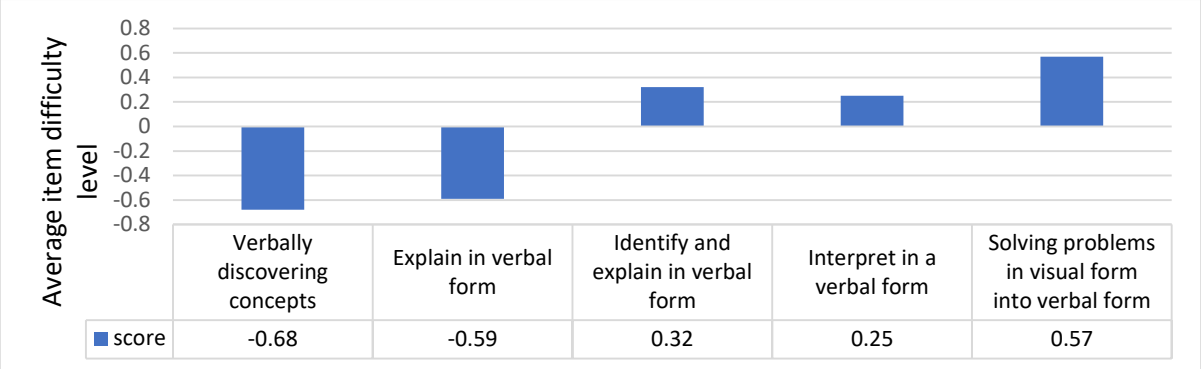
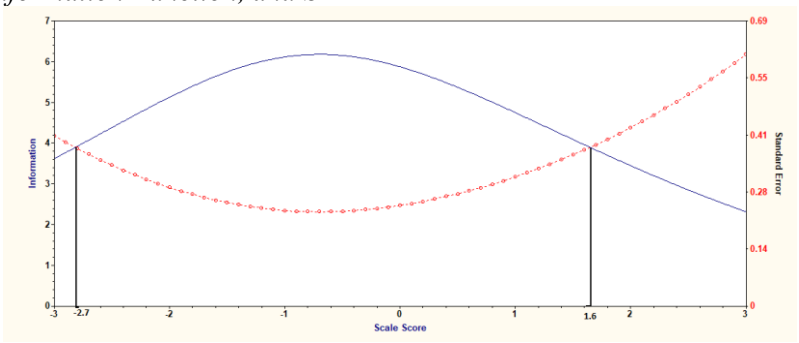


Figure 3. Graphic Result of Difficulty Indexs

3.4 Reliability, Information Function, and SEM



Fugure 4. Information Function and SEM

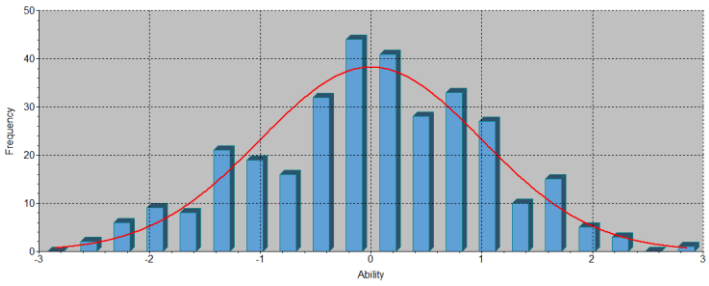


Figure 5. Histogram capability distribution

The results of the analysis show the reliability value of the test instrument is 0.81, the reliability results meet the requirements of a good instrument [43]. Information and SEM relationship curves obtained are inversely proportional. The higher the SEM, the lower the information function will be, the lower the SEM, the higher the information function will be obtained [37]. The intersection of the curve in **Figure 4** shows the suitability of the test developed for participants with abilities of around $-2.7 \leq \theta \leq 1.6$. While the distribution index histogram is difficult to see in **Figure 5**.

3.5 Discussion

The validity of the contents was submitted to two experts, to ensure each item was appropriate and valid was used [38][39]. A total of 10 items were validated by experts and then analyzed, then used as a reference to revise the problems [1][2]. The validity of this item in terms of several aspects, namely aspects of language, aspects of conformity with material indicators and indicators of verbal representation, fiber construction aspects of test items. Based on responses and judgments by 2 expert judges (expert judgment) the V'Aiken index value of 39 items is in the range of 0.85 to 0.91, then it is included in the classification of Very Hight or very valid [35].

The test run was conducted on 322 high school grade X students in Yogyakarta. Analysis of the test results shows that the empirical validity test for each item and for the whole test can be proven. According to Adam and

Khoo, the empirical validity of the problem is determined based on the *goodness of fit* of the *partial credit model* (PCM) model [36]. The suitability of the respondents (testi) to the model was carried out to find out the relationship between the answers and the respondents with the level of ability to the level of difficulty of the items. Goodness of fit is known based on the output of the Quest program by looking at the average value of INFIT Mean of Square (INFIT MNSQ) along with its standard deviation (INFIT Mean of INFIT t). Estimation results obtained can be seen in **Table 3**. Verbal estimation shows the estimated value of items is 0.98 with standard deviation 0.10 and estimated testies are 1.00 with standard deviation 0.33, the average value of INFIT t items is -0.06 with standard deviation 1.10 and testi 0.07 with standard deviation 0.97. In addition, in both of them there were no respondents who got a score of 0 (zero) or a perfect score, so that no respondents were discarded. In the Quest program it is determined that an item or testi / case / person passes the goodness of fit test if the average value of the MNSQ INFIT is around 0.77 to 1.30 [35][36]. Then the average value of INFIT t is around ± 2 (rounding from ± 1.96) because it uses an error or alpha of 5% [38][39]. Therefore, all test items match the PCM model based on the specified Quest program criteria. The suitability of each test item with the model can also be seen from the Quest INFIT MNSQ output, where if the item has an INFIT MNSQ value in the interval 0.77-1.30 it is said to be fit [36]. This interval is described as a dotted line that forms two boundary lines and the item is said to fit if it is between these two lines [40][41]. The distribution of MNSQ INFIT values can be seen in **Figure 2**.

The results of the analysis show the reliability value of the test instrument is 0.81, the reliability results meet the requirements of a good instrument. The range of values for reliability is 0-1. The closer to 1, the measurement shows consistent results, and vice versa [39][42][43]. Information and SEM relationship curves obtained are inversely proportional. The higher the SEM, the lower the information function will be obtained, and vice versa [36]. The intersection of the curve at **Figure 4** shows the suitability of the test developed for participants with abilities of around $-2.3 \leq \theta \leq 1.9$. The difficulty index distribution histogram can be seen in **Figure 5**. The difficulty index distribution is seen following the normal distribution.

4. CONCLUSION

Based on analysis of the data, it can be concluded that:

1. These 39 item verbal representation problems are developed with the format of two-tier multiple-choice questions, and indicators of Newton's laws of motion material, Newton's laws of gravitation, Work and Energy.
2. The characteristics of test are as follows:
3. Test items is valid based on contens validity using Aiken's V by expert judgment and empirically, fit witht PCM model.
4. The difficulty index of all the items in the instrument test is in good category between -0.68 to 0.57.
5. The items test is reliable is base on the reliability of estimate, information function and SEM. The results of the analysis show that the reliability value of the test instrument is 0.81, the reliability results meet the requirements of a good instrument. The items test is very appropriate to measure verbal representation ability whose range of ability is from $-2.3 \leq \theta \leq 1.9$.

REFERENCES

[1] Istiyono, E. (2018). Pengembangan Instrumen Penilaian dan Analisis Hasil Belajar Fisika dengan Teori Tes Klasik dan Modern. UNY Press.

[2] Mardapi, Djemari. (2017). Pengukuran Penilaian dan Evaluasi Pendidikan. Yogyakarta: Parama Publishing

[3] Bower, J. A., & Justice, C. (n.d.). Developing Research & Thinking Skills through Active Learning and Assessment in Undergraduate Criminology Core. 1–25

[4] Suryan, I. (2016). Pengembangan Instrumen Penilaian Sikap Ilmiah Pada Pembelajaran Dengan Model Latihan Penelitian Di Sekolah Dasar. 217–227.

[5] Raharjo, S. B. (2018). Kontribusi Delapan Standar Nasional Pendidikan terhadap Pencapaian Prestasi Belajar. Jurnal Pendidikan Dan Kebudayaan, 20(4), 470. https://doi.org/10.24832/jpnk.v20i4.160

[6] Sugiyanto, Kartowagiran, Jailani, 2019. Pengembangan Model Evaluasi Proses Pembelajaran Matematika di SMP Berdasarkan Kurikulum 2013. Jurnal penelitian dan Evalusi Indonesia

[7] Istiyono, E., & Istiyono, E. (2017). The Analysis of Senior High School Students ' Physics HOTS in Bantul District Measured Using PhysReMChoTHOTS. 070008. https://doi.org/10.1063/1.4995184

[8] Arikunto, Suharsimi. (2013). Dasar-Dasar Evaluasi Pendidikan. Jakarta: Bumi Aksara.

[9] Basuki, Ismet & Haryanto. (2017). *Assasement Pembelajaran*. Bandung:Rosda

[10] Sadtyadi, H., & Kartowagiran, B. (2019). Pengembangan Instrumen Penilaian Kinerja Guru Sekolah Dasar Berbasis Tugas Pokok

Dan Fungsi. Jurnal Penelitian Dan Evaluasi Pendidikan, 18(2), 290–304. https://doi.org/10.21831/pep.v18i2.2867

[11] Raharjo, S. B. (2018). Kontribusi Delapan Standar Nasional Pendidikan terhadap Pencapaian Prestasi Belajar. Jurnal Pendidikan Dan Kebudayaan, 20(4), 470. https://doi.org/10.24832/jpnk.v20i4.160

[12] Puspendik Balitbang Kemendikbud. 2015. Result from PISA.

[13] Puspendik Balitbang Kemendikbud. 2018.Ujian Nasional

[14] Hamsarudin. 2016. Meningkatkan Kemampuan Representasi Matematis dan Self- Confidence Melalui Pembelajaran Concrete-Representational-Abstract (CRA) pada Siswa Sekolah Menengah Pertama. Universitas Pendidikan Indonesia

[15] Wardani & Rumiati. 2011. Instrumen penilaian hasil belajar matematika SMP: Belajar dari PISA dan TIMSS

[16] Docktor, J. L., & Mestre, J. P. (2014). Synthesis of discipline-based education research in physics. 020119, 1–58. https://doi.org/10.1103/PhysRevSTPER.10.020119

[17] Astuti, Lusia Triatmi, dkk. 2015. Kemampuan Representasi Matematis Siswa Melalui Pembelajaran Berbasis Masalah di Sekolah Menengah Atas.

[18] Nasrul, M. (2016). Representasi Matematis Siswa SMA dalam Memecahkan Masalah Persamaan Kuadrat Ditinjau dari Perbedaan Gender. 7(2), 145–152.

[19] Suratman, D. (2010). Pemahaman Konseptual dan Pengetahuan Prosedural Materi Pertidaksamaan Linier Satu Variabel Siswa Kelas VII SMP. Pontianak: Universitas Tanjungpura. Jurnal Pendidikan Matematika.

[20] Panjaitan, R. L. (2012). Pemahaman Konseptual Matematis yang Dikaji Menurut Tingkat Kemampuan Siswa pada Materi Pertidaksamaan Linear Satu Variabel di Kelas VII SMP Negeri 9 Pontianak (skripsi). Pontianak: Untan.

[21] Alhadi. S, Sugiatno, Suratman. D. (2013). Pemahaman Konseptual Siswa Dikaji dari Representasi Matematis dalam Materi Fungsi Kuadrat di SMA. Pontianak: Jurnal pendidikan matematika

[22] Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representation. *Learning and instruction,* 16,(3), 183-198.

[23] Hill, M., Sharma, M., O'Byrne, J., & Airey, J. (2014). Developing and evaluating a survey for representational fluency in science. International Journal of innovation in science and mathematics education, 22 (5), 22-24.

[24] Airey, J & Linder, C 2009 A diciplinary Discourse Perspective on University Science Learning: Achieving Fluency in a Critical Constellation of Modes. *Jurnal of Research in Science Theaching*, **46** (1), 27-49.

[25] Abdurrahman, Liliasari, & Waldrip. 2011. Implementasi Pembelajaran berbasis Multi Representasi untuk Peningkatan Penguasaan Konsep Fisika Kuantum. Cakrawala Pendidikan. 30, (1), 30-45.

[26] Linder, C 2013 Diciplinary Discourse, representation, and apprepresentation in the theaching and learning of science. European Journal of Science and Mathematics Education, 1 (2), 43-49)

[27] Adawiyah, R., Istiyono E., Wilujeng, I., Hardiyanti, S. 2019 Development of an Instrument measuring the multi representation ability of senior high school students. J. Phys. Conf. 1440 012028. https://iopscience.iop.org/article/10.1088/1742-6596/1440/1/012028/meta

[28] Ott, N., Brünken, R., Vogel, M., & Malone, S. (2018). Multiple simbolic representations: The combination of formula and text supports problem solving in the mathematical field of propositional logic. Learning and Instruction, 58, 88-105.

[29] Forbes, M., & Choi, Y. (2017). Verb physics: Relative physical knowledge of actions and objects. arXiv preprint arXiv:1706.03799

[30] Namdar, B., & Shen, J. (2016). Intersection of argumentation and the use of multiple representations in the context of socioscientific issues. *International Journal of Science Education*, *38*(7), 1100-1132.

[31] Rangkuti, A. (2014, Januari). Representasi Matematis. Forum Pedagogik, VI(1).

[32] Treagust, David F., Duit, Reinders., Fischer, Hans E. (2017). Multiple Representations In Physics Education. Springer: Volume. 10.

[33] Widiati, Indah.2015. Mengembangkan Kemampuan Representasi Matematis Siswa Sekolah Menengah Pertama Melalui Pembelajaran Kontekstual. Jurnal Pengajaran MIPA.

[34] Istiyono, Edi 2017 The analysis of senior high school students' physics HOTS in Bantul Dsitrict measured using PhysReMChoTHOTS AIP Conference Proceedings doi: http://dx.doi.org/10.1063/1.4995184

[35] Aiken, L. R. 1985. Three Coeficients for Analyzin the Reliability and validity of Raitings. *Educational and physicological Measurement*. p 66

[36] Adam, R. J & Khoo S. T 1996 Acer Quest. Yhe Interactive Test Analysis Sytem, (Victoria: The Australia Council for Educational Research) p 1-30

[37] Hambelton, R. K., & Swaminathan, H. (1985). Item Respon Theory. Boston: Kluwer Nijjhoff

Publisher.

[38] Keeves, J.P. & Masters, G. N. 1999. *Advences in Messurement in Educational Research and Assasement*. (Amsterdam: Pergamon, An Imprint of Elsevier Science) p 12-13

[39] Bond, T.G. & Fox, Ch.M. 2007. *Applaying The Rasch Model: Fundamental Meassurement in the human Science* (Mahwah, New Jersey: Laurance Erlbaum Associates. Publisher) p 43

[40] Widoyoko, Eko Putro 2012 Teknik Penyusunan Instrumen Penelitian, (Yogyakarta: Pustaka Pelajar) p 167

[41] Edi Istiyono and Suyoso 2019 The Developing and Calibration of PhysEDiTHOTS Based on IRT and IQF for Students HOTS Diagnostic *J.Phys.:Conf. Ser.* 1233 012038

[42] Subali, B. & Pujiati, S 2012 *Pengembangan Item Tes Konvergen Dan Divergen Dan Penyelididkan Validitasnya Secara Empiris*, (Yogyakarta: Diandara Pustaka Indonesia) p 11

[43] Sumintono, B. & Widhiarso, W. (2013). Aplikasi Model Rasch untuk Penelitian Ilmu-Ilmu Sosial. Jakarta: TrimKom.