Development of an instrument measuring the multi representation ability of senior high school students

R Adawiyah[1], E Istiyono[1], I Wilujeng[1] and S Hardiyanti[2]

[1]Physics Education, Universitas Negeri Yogyakarta, Sleman, Indonesia
[2]Physics Education, SMA Negeri 1 Banguntapan, Bantul, Indonesia

Corresponding Author: robiahadawiyah.2018@student.uny.ac.id

Abstract. This study aims to develop an assessment instrument to measure the multi-representation physics skills of senior high school students in grade X on simple harmonic motion material. The method used is instrument development by Djemari Mardapi, (1) Determination of test objectives (2) Determination of material to be tested (3) Arrange the test grid (4) Write item's Question (5) Validate test item, (6) Determination of Test Subjects (7) Test Implementation (8) Analyze Test Question. Based on the result validation of two expert judgement the Aiken's V index value of 10 items are in the range of 0.85-0.91, according to Aiken's V > 0.8 validation included in the very high classification. The level of difficulty in instrumen is good category between -1.53 up to 1.04. and the instrument reliability of 0.84 the assessment instrumen multi-representation skills are classified as very high category.
Keywords: *instrument, physics, multi-representation*

1. Introduction

The purpose of learning physics in senior high school in general is to teach science to students, provide an understanding of concepts, and a number of skills and abilities that are required to enter higher education levels and develop science and technology to foster the ability to think that is useful for solving problems in everyday life [1]. However, physics still gets a bad reputation as a difficult subject for students in school. Teachers and physicists, in general, have obstacles and problems in trying to convey physics more meaningfully [2], [3].

Students learn more effectively and meaningfully when they process information in various ways [4], [5]. In Physics, mastering the concept requires different abilities or ways of understanding or representation or multi representations in each concept being studied. Using multi-representation of students is easier in the process of understanding physics in various ways that can affect cognitive processes in students [5]-[7]. The representations most often used in the physics learning process are mathematical, drawing, graphic, and verbal descriptions [8], [9]. However, the problem now is the limited use of multiple representations in the learning process. Textbooks that have become sources of learning have not contained many strategies and assessments that emphasize the use of multi representations in a comprehensive manner. This becomes a challenge for a teacher to understand the representation strategy to students, so that students are proficient in its use and make relations of these representations comprehensively as an opportunity to understand concepts and communicate them in the systems and processes of a physical concept [10], [11]. Active student involvement in transforming one representation into another is the key to the success of a multi-representation strategy in physics learning [12].

One of the learning activities that require students to be actively involved is assessment. Assessment is carried out to measure and observe student activities and progress and to measure students' learning skills and outcomes [13], [14]. To determine the ability of student representation, it is necessary to develop assessment instruments that can measure these abilities. The use of multi representations in the assessment process helps students in solving physical problems [3].

In the assessment process, the instrument used has not been able to measure the ability of the multi-representation of students. The form of instruments commonly used by teachers in the assessment process is limited to measuring student knowledge and the indicators used are only referred to thinking skills [15],[16]. Teachers in schools more often develop assessment instruments to measure and practice high-level thinking skills of students, penial instruments to measure and practice the ability of representation are still rarely developed. So that the multi-representation ability of students in problem-solving cannot be observed properly [11], [17]-[21]
In line with the problems that have been described above, it is necessary to conduct research to develop instruments that can measure the ability of the multi-representation of students in physics subjects and to get products in the form of instruments that can determine the profile and ability of students in representing a physics concept.

2. Research method
This research aims to develop some product [22]. The resulting product is an instrument to measure students' multi-representation abilities.

This research was conducted at SMA N 1 Banguntapan, Bantul, Yogyakarta. The subjects of this study were 61 students of class X MIPA 3 and X MIPA 4. The instrument development stage followed the steps of making the instrument developed by Djemari Mardapi [23], it can be seen in figure 1. The instrument making stage included: (1) determining the purpose of the test, (2) determination of competencies to be tested, (3) determination of test material, (4) arrange test specification, (5) writing test questions according to multi-representation indicators. The test analysis phase includes: (1) validation of test items, (2) improvement of items and test assembly. The testing phase of the test includes: (1) determining the subject of the trial, (2) conducting the trial, (3) analyzing the results of the trial. The example for the test item can be seen in figure 2.

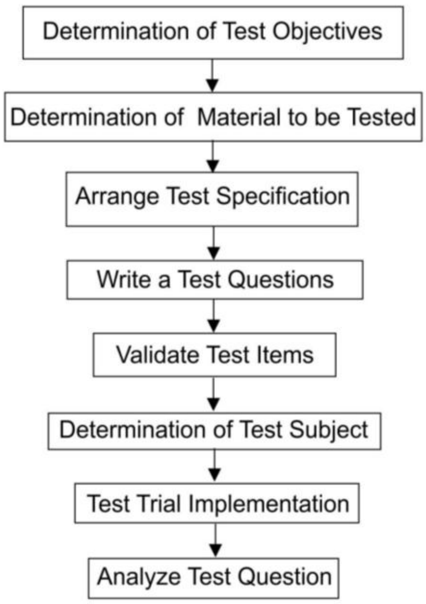
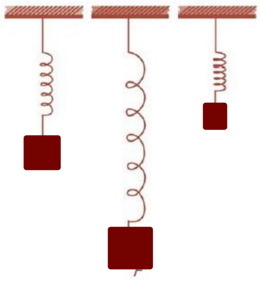


Figure 1. Instrument making steps.

Multiple representation indicator:
Analyze the question of visual form, solve it in verbal form.
Material Indicator:
Effect of mass on changes in spring length
Questions:
Look at the following three spring images!



What causes the change in spring length as shown in the picture, why is the length of each spring different?
All three springs have the same material
a. Acceleration of gravity which causes the length of the spring to increase.
b. The period of the type of object that is hung on a spring.
c. Wire diameter of the spring material.
d. The mass of the weight is suspended on a spring, which results in greater gravity.
e. Spring vibration period.

Figure 2. An item about multiple representation test.

The results of the instrument content validation by experts were analyzed using the Aiken validation index. Aiken's validation aims to measure the extent to which items in the instrument meet the specific purpose of making test items [23]. Analysis of the test items was carried out using the Rasch Model by Quest to determine the profile of students' abilities. Using Quest as an extension of the Rasch model for the 1-PL (one parameter) model, we can use samples that are not too large. The specific sample size for the 1-PL (one parameter) model is the Rasch Model (RM) between 30-300 [24-25]. Item analysis is performed using the QUEST. The central element of the QUEST program follows the Rasch model which can be used on data that are given a score of polytomus [26]. Analysis of the items to be done include the level of difficulty and reliability of the items.

Analysis of the ability of multi-representation of students is done by using SPSS in the form of descriptive statistics, to describe the existence of research data that has been collected without intending to conclude in general [27].

3. Results and Discussion

3.1. Test instrument analysis results

3.1.1. Content validity. The content validity is a validation process that is submitted to two expert judgment to ensure that the items to be used are valid and can measure or reveal a concept [28]. There are 10 items validated by experts. The validation results are then analyzed and used as consideration in revising the product. [24], [25]. The interpretation criteria for validation of the contents of the items are presented in table 1, if the index is less or equal to 0.4, the validation is said to be low or low, 0.4-0.8 is said to be of moderate validation, and if greater than 0.8 is said to be very high or very valid [29].

The validity of the multi-representation test questions was reviewed based on the suitability of the material with the indicators, construction questions, and language. Each item is approved based on three aspects. Based on responses and judgments by 2 expert validators (expert judgment) the Aiken’s V index value of 10 items is in the range of 0.85-0.91, then it is included in the very high classification. The overall index can be seen in table 2.

Table 1. Interpretation criteria for validating item questions.

Validity	Interpretation of item validity
V < 0.8	Very High
0.4-0.8	High
V ≤ 0.40	Low

Table 2. Item reliability classification.

Reliability	Classification
0.80-1.00	Very high
0.60-0.80	High
0.40-0.60	Medium
0.20-0.40	Low
0.00-0.20	Very low

3.1.2. Level of Difficulty. The level of difficulty of the items is an opportunity to correct a problem at a certain ability level expressed in the form of an index [30]. In the Rasch model, an item is fit if it has an index value between -2 to 2 (with a chance <0.05) [26-27]. The difficulty level of an item is determined by the proportion of the number of test-takers who correctly answers an item to all test takers. From the results of the interpretation of 10 items, the level of difficulty is included in both categories. This is evidenced by all items fit with the Rasch model because it has a range of values of -2 to 2 which can be seen in table 3 and graphs of the difficulty level of the items can be seen in figure 3.

Table 3. The result of Aiken validity and item difficulty level.

Item	Aiken V	Category	Result of Difficulty Level
1	0.91	Very high	-0.02
2	0.85	Very high	1.04
3	0.90	Very high	-0.78
4	0.89	Very high	-0.89
5	0.90	Very high	0.06
6	0.89	Very high	0.31
7	0.85	Very high	-1.53
8	0.88	Very high	-1.25
9	0.91	Very high	-0.58
10	0.92	Very high	0.64

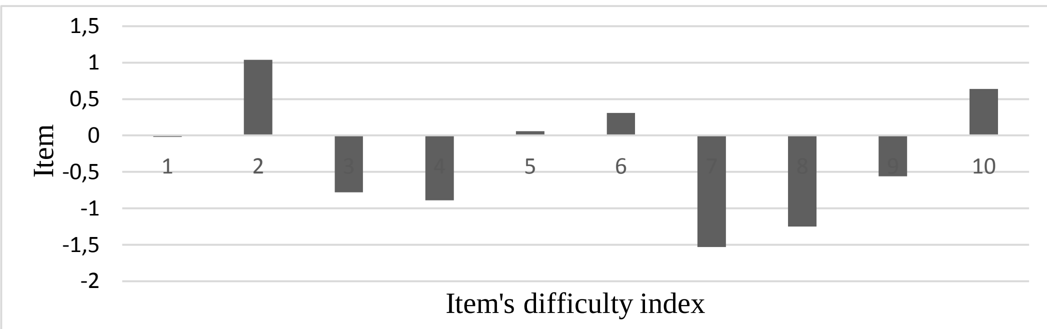


Figure 3. Item’s difficulty level.

3.1.3. Goodness fit. Goodness fit testing is carried out on each item as a whole using rules developed by Adam and Khoo [28]. Goodness of fit is known from the mean value of INFIT Mean of Square (MNSQ Mean INFIT) along with its standard deviation. Testing the determination of fit items if the MNSQ INFIT value is between 0.77 to 1.30 [28].

Estimated Value of Item Mean INFIT MNSQ 1.00 with a standard deviation of 0.18 according to the item fit standard and according to the Rasch model. The higher value more better and gives consistent results. This result is also determined by the characteristics of the sample. The lower means also more samples for trials items that do not provide the expected information [32].

Table 4. Goodness fit and reliability analysis results.

No	Information	Estimasi Item	Estimasi Testi
1	Mean IINFIT MNSQ and standart deviation	1.0 $\pm$ 0.18	1.0 $\pm$ 0.26
2	Outfit MNSQ and standart deviation	0.95 $\pm$ 0.25	0.95 $\pm$ 1
3	Reliability	0.84	

3.1.4. Reliability. Reliability allows us to understand some of the results that can be trusted [32]. There are 61 test data, 10 items with a probability level of 0.5 in accordance with the maximum likelihood principle obtained reliability values [31], [33]. Based on the results of the analysis, the reliability of the multi-representation instrument set was 0.84. In accordance with table 2, the reliability value is classified as a high category. The reliability results can be seen in table 4.

3.2. The results of multi-representation

Analysis of the ability of multi representation of students is done by using SPSS in the form of descriptive statistics. Descriptive statistics are methods related to data collection and presentation. Estimated data are variable types, statistical summaries (mean, median, mode, standard deviation, and so on), distribution, and pictorial representation (graph), without any probabilistic formulas [23], [34] [35]. The results of the multi-representation capability analysis are described as follows.

3.2.1. Class MIPA 3. Based on table 5 shows the number of respondents (N) or students in class X MIPA 3 there are 30, of the 30 respondents the smallest value (minimum) is 13 and the largest value (maximum) is 80 maximum and minimum scores. The average value of 30 respondents (mean) was 50.267 with a standard deviation of 19.511. Table 6 shows the results of Skewness and Kurtosis which is a measure of data to see whether the data is normally distributed or not. Data is said to be normally distributed if the value of Skewness and Kurtosis is close to zero [36]. Table 6 shows the Skewness and Kurtosis values respectively -0.022 and -0.903. This value is close to zero, so the data is normally distributed. The graph of the distribution of grades for class X MIPA 3 is presented in Figure 4. Table 7 shows the ability of students from the results of the test scores obtained. Table 7 also shows the frequency and percentage of students who are in the low to high categories.

Table 5. Descriptive statistics for MIPA 3

Lebel	N	Minimum	Maximum	Mean	Std.Deviation
Score of Class MIPA 3	30	13.00	80.00	50.2667	19.51115
Valid N (listwise)	30				

Table 6. Skewness kurtosis descriptive statistics.

Lebel	N	Skewness		Kurtosis	
	Statistic	Statistic	Std. Error	Statistic	Std. Error
Nilai MIPA 3 valid n (listwise)	30	-0.022	0.427	-0.903	0.883

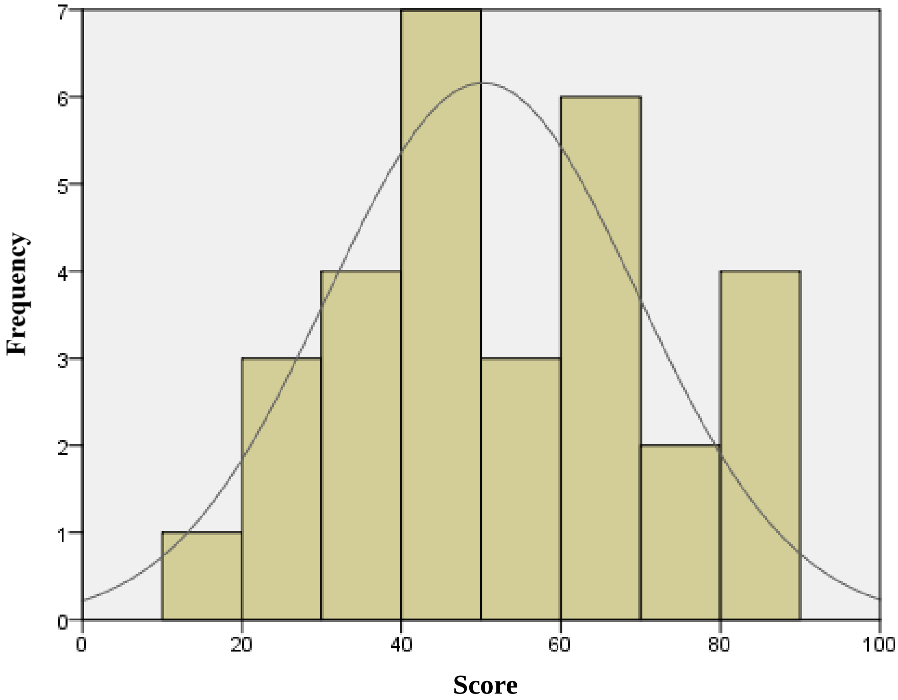


Figure 4. Graphic distribution of Class MIPA 3.

Table 7. Frequency score of MIPA 3.

Test Score	Frequency	Category	Percentage (%)
13	1	Very low	3.3
20	2	Very low	6.7
26	1	Very low	3.3
33	4	Very low	13.3
40	4	Very low	13.3
46	3	Very low	10.0
53	3	Very low	10.0
60	4	Very low	13.3
66	2	Low	6.7
73	2	Medium	6.7
80	4	High	13.3

3.2.2. Class MIPA 4. Based on table 8 shows the number of respondents (N) or students in class X MIPA 4 there are 31, of the 31 respondents the smallest value (minimum) is 46 and the largest value (maximum) is 80 maximum and minimum scores. The average value of 31 respondents (Mean) was 66.58 with a standard deviation of 9.05. Table 9 shows the results of skewness and kurtosis which is a measure of data to see whether the data is normally distributed or not. Data is said to be normally distributed if it has a skewness and kurtosis value close to zero [36]. Table 9 shows the skewness and kurtosis values respectively -0.482 and -0.628. This value is close to zero, so the data is normally distributed. The graph of the distribution of grades for class X MIPA 4 is shown in Figure 5. Table 10 shows the ability of students from the test scores obtained. Table 10 also shows the frequency and percentage of students who are in the low to high categories.

Table 8. Descriptive statistics for MIPA 4

Label	N	Minimum	Maximum	Mean	Std.Deviation
Score of class MIPA 4	31	46.00	80.00	66.5806	9.05087
Valid N (listwise)	31				

Table 9. Skewness kurtosis descriptive statistics

Lebel	N	Skewness		Kurtosis	
	Statistic	Statistic	Std. Error	Statistic	Std. Error
Score of class MIPA 4	31	-0.482	0.421	-0.628	0.821
valid n (listwise)	31				

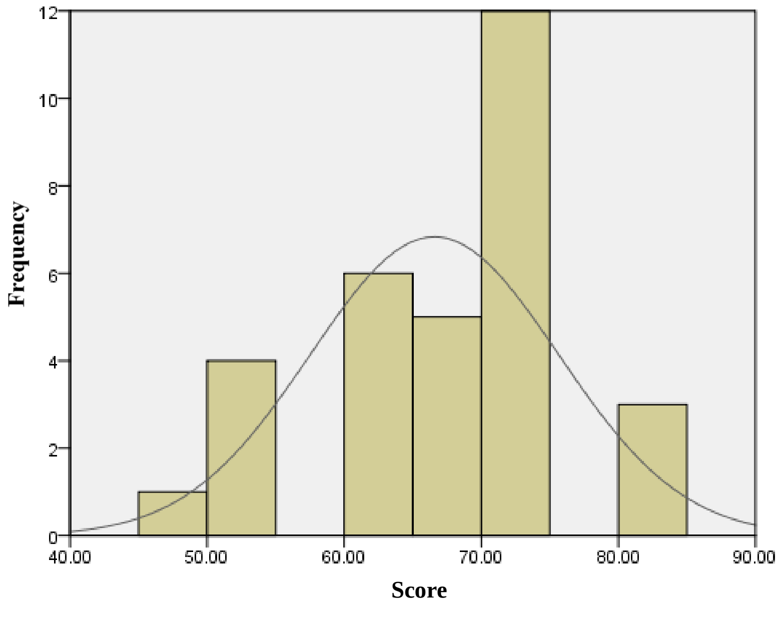


Figure 5. Graphic distribution of class MIPA 4.

Table 10. Frequency score of MIPA 4.

Test Score	Frequency	Category	Persentase (%)
46	1	Very low	3.3
53	4	Very low	6.7
60	6	Very low	3.3
66	5	Low	13.3
73	12	Medium	13.3
80	3	High	10.0

4. Discussion

The validity of the contents of the test has been carried out by 2 experts (expert judgment) and is proven by the results of the Aiken's V analysis of 10 items in the range of 0.85-0.91, then it is included in the very high classification. The overall Aiken's V index can be seen in table 3.

Good items are items that have a moderate level of difficulty so that they are not too difficult and not too easy for students [37]. The difficulty level b for good items varies between -2.00 to 2.00. Items with a difficulty level of -2.00 mean that the item is very easy, while a value of 2.00 means that the problem is very difficult [26-27]. Based on the results of the analysis of the level of difficulty items that have been developed belong to both categories because it is in the range of -2 to 2. This is evidenced by all items fit with the Rasch model because it has a range of values of -2 to 2 which can be seen in table 3 and graphs item's difficulty level can be seen in figure 3.

The result of goodness fit, the item has an INFIT value of MNSQ 1.00 with a standard deviation of 0.18 which is located between the item acceptance limits or fit according to the Rasch model (between 0.77 to 1.30) means that all items are 10 items fit and can be accepted. The reliability of the multi-representation test is quite high at 0.84 which confirms that the measurement results with this instrument are reliable. Tests that have a reliability coefficient of at least 0.80 testing results with these tests can be used to make decisions about individuals [38],[39].

The role of multi representation in the learning process helps in the process of solving problems (problem-solving) better [22],[40],[41]. Several studies in physics learning show that multi-representation significantly impacts student learning outcomes [13], [22], [42]. The results obtained from 61 students there are 7 students have high multi-representation ability, 14 students have medium multi-representation ability, and 40 students have low multi-representation ability. Thus, this instrument can be used to measure and explore information about the ability of multi-representation physics in high school students.

5. Conclusion

Based on the results of the study it can be concluded: (1) the multi-representation instrument was developed in the form of multiple choice in the representation of verbal, images, and graphics for simple harmonic motion material. (2) the multi-representation instrument has fulfilled the content validity of expert judgment and has obtained empirical validity. (3) The level of difficultyin in instrumen is good category between -1.53 up to 1.04. and the instrument reliability of 0.84 the assessment instrumen multi-representation skills are classified as very high category. The results obtained from 61 students there are 7 students have high multi-representation ability, 14 students have medium multi-representation ability, and 40 students have low multi-representation ability. Thus, this instrument can be used to measure and explore information about the ability of multi-representation physics in high school students.

References

[1] Damiano and Agustinus Cristhian 2017 *Pengetahuan guru tentang tujuan pembelajaran fisika dan pengaruhnya terhadap proses pembelajaran (Studi Kasus Mengenai 2 Guru Fisika)* (Yogyakarta: Universitas Sanata Dharma) pp 1
[2] Faragher S 2014 *Understanding assessment in primary education* (Los Anggles: Sage) p 5
[3] Euler M 2004 *Proceeding of the Internasional School of Physics Enrico Fermi (Verenna)* (Amsterdam: IOS Press) https://doi.org/10.3254/978-1-61499-012-3-175
[4] Treagust Duit and Fischer 2017 *Multiple representations in physics education* (Berlin: Springer) pp 41
[5] Airey J and Linder C 2009 *Jurnal of Research in Science Theaching* **46** 27-49 https://doi.org/10.1002/tea.20265
[6] Abdurrahman, Liliasari and Waldrip 2011 *Cakrawala Pendidikan* **30** 30-45

https://doi.org/10.21831/cp.v1i1.4189
[7] Linder C 2013 *European Journal of Science and Mathematics Education* **1** 43-49 https://doi.org/10.1002/tea.20265
[8] Hill M, Sharman M, O'Byrne J and Airey J 2014 *International Journal of Innovation in Science an mathematics Education* **22** 22-24
[9] Kim K, Clarianay, Roy B and Kim Y 2018 *Educational Tecnology Research & Development* **67** 105-22 https://doi.org/10.1007/s11423-018-9626-6
[10] Susac A, Bubic, Andreja, Martinjak, Petra and Planinic M 2017 *Phys. Rev. Phys. Educ Res* **13** https://doi.org/10.1103/PhysRevPhysEducRes.13.020125
[11] Meltzer D E 2005 *American journal of Physics* **73** 463 https://doi.org/10.1119/1.1862636
[12] Ainswort S 2006 *Learning and Instruction* **16** 183-198 https://doi.org/10.1016/j.learninstruc.2006.03.001
[13] Ainswort S 1999 *Computers & Education* **33** 131-52
[14] Chin C 2007 *Proc. the 1st International Seminar on Science Education (Bandung)* (Bandung: Graduate School Indonesia University of Education)
[15] Nieminen P, Savinainen A and Viiri J 2010 *Phys. Rev. PER.* **6** http://dx.doi.org/10.1103/PhysRevSTPER.6.020109
[16] Widiati I 2015 *Jurnal Pengajaran MIPA* **20** http://dx.doi.org/10.18269/jpmipa.v20i2.571
[17] Prain V, Tytler R and Peterson S 2009 *International Journal of Science Education* **31** https://doi.org/10.1080/09500690701824249
[18] Irwandani 2014 *Jurnal Ilmiah Pendidikan Fisika: Al- Biruni* **3** https://doi.org/10.24042/jpifalbiruni.v3i1.64
[19] Chang H Y 2018 *Sci. Edu.* **102** 1129-49. http://dx.doi.org/10.1002/sce.21457
[20] Waldrip B, Prain V and Carolan J 2010 *Res. Science Education* **40** 65-80 https://doi.org/10.1007/s11165-009-9157-6
[21] Lodico M G, Dean T S and Katherine H V 2010 *Methods in educational research: from theory to practice* (New Jersey: Wiley)
[22] Istiyono E and Suyoso 2019 *Proc. International Seminar on Science Education (Yogyakarta)* vol 1233 (Bristol: IOP Publishing) p 1-9 https://doi.org/10.1088/1742-6596/1233/1/012038
[23] Istiyono E, Djemari M and Suparno 2014 *Jurnal Penelitian dan Evaluasi Pendidikan* **18** 1-12 https://doi.org/10.21831/pep.v18i1.2120
[24] Keeves J P and Masters G N 1999 *Advences in messurement in educational research and assasement* (Amsterdam: Elsevier Science) pp 12-3
[25] Bond T G and Fox C M 2007 *Applying the rasch model: fundamental meassurement in the human science* (New Jersey: Laurance Erlbaum Associates) pp 43
[26] Adam R J and Kho S T 1996 *Acer quest version 2.2 cemberwell* (Victoria: The Australian Councilfor Educational Research) pp 86
[27] Aiken L R 1985 *Educational and physicological Measurement* **45** 131-42 https://doi.org/10.1177/0013164485451012
[28] Hambelton R K, Swaminathan H and Rogers H J 1991 *Foundamental of item responses theory* (Newbury Park: Sage Publication)
[29] Adam R J and Khoo S T 1996 *Acer Quest. The interactive test analysis sytem* (Victoria: The Australia Council for Educational Research) pp 1-30
[30] Wright B D 1982 *Rating scale analisis* (Chicago: Mesa Press)
[31] De C J L and Gire E 2005 *Proc. Physics Education Research Confereance (California)* (New York: AIP Publishing) https://doi.org/10.1063/1.2177019
[32] Kohl P B and Finkelstein 2006 *Phys. Educ. Res.* **2** https://doi.org/10.1103/PhysRevSTPER.2.010106
[33] Neiminen P, Savinainen A and Viiri J 2010 *Physics Review Special Topics. Phys. Educ. Res* **6** 1-12 https://doi.org/10.1103/PhysRevSTPER.6.020109
[34] Savinainen A, Nieminen P, Mäkynen A and Viiri J 2013 Phys. Edu. **48** 372.

https://doi.org/10.1088/0031-9120/48/3/372