

PAPER • OPEN ACCESS

Physical Reasoning Ability Development of High School Students

To cite this article: Suhariyono and Djemari Mardapi 2019 *J. Phys.: Conf. Ser.* **1233** 012064

View the [article online](#) for updates and enhancements.

You may also like

- [The effectiveness of project-based learning assisted by digital module toward mathematical connection ability](#)
Kusnandar, I Junaedi and A Suyitno
- [Phynox Improved Corrosion Resistance with MPC Initiated from Mixed Monolayers of Phosphonic Acids](#)
Bastien Barthélémy, Frédéric Kanoufi, Catherine Combellas et al.
- [Needs analysis to development of biology module based on problem solving at topics of respiratory and excretory systemf to student of seniorhigh school grade XI](#)
Adela Mulyana and R Sumarmin



ECS Membership = Connection

ECS membership connects you to the electrochemical community:

- Facilitate your research and discovery through ECS meetings which convene scientists from around the world;
- Access professional support through your lifetime career;
- Open up mentorship opportunities across the stages of your career;
- Build relationships that nurture partnership, teamwork—and success!

Join ECS!

Visit electrochem.org/join



Physical Reasoning Ability Development of High School Students

Suhariyono* and Djemari Mardapi
Yogyakarta State University, Indonesia

*Email: Suhariyono87@gmail.com

Abstract. This study has purpose to reveal: (1) the characteristics of physics subject test of high school students of grade X and XI, (2) the equality of vertical test of physics subject of high school students of grade X and XI, (3) the development of the ability of high school students of grade X to grade XI physics subject. The data which was in the form of scores were then processed and synchronized by using a common item non-equivalent groups design by the method of equalizing of mean and sigma based on the Rasch model. The results of this study are: (1) the validity of the physics subjects test of grade X is 0.92 and grade XI is 0.96 (2) the reliability tests is 0.928 for grade X and 0.926 for grade XI; (3) the results of the vertical equalization based on the approach of Rasch model showed the conversion equation for grade X to grade XI was $b^* = b_x - 0.36$; (4) there is a development of the ability of grade X students to grade XI students or the ability of grade XI students is higher than the ability of grade X students.

Keywords: Ability development; Physics subject; Vertical equalization.

1. Introduction

Physics is the science subjects that can develop the ability to think deductive analysis using a variety of natural events and the settlement of the problem, both qualitatively and quantitatively by using math and can develop the knowledge, skills and attitude of confidence. From these definitions, it can be concluded that physics is a science subjects in which to learn the parts of the universe and their interactions, thus requiring the ability to think. Teach thinking skills and combine them with learning materials can help students to become critical thinkers, creative and effective.

The learning process which are currently found less support students to develop the skills of analysis [1]. Because students act more as a receiver of information from teachers and learners in learning activities just listen and record the results of the teacher's explanation. The impact of the learning process such learners lack the analytical skills and problem solving skills. One of the efforts to develop the intellectual potential to the fullest learners is to develop reasoning skills. This reasoning abilities need to be trained in teaching in schools, such as analytical skills.

Analytical skills are essential possessed by students because it can be easier for students looking for a solution to a particular problem in physics. Analysis capabilities can be divided into three, namely the ability to distinguish, organize, and find meaning implied. Analysis capabilities can be improved by training students to get used to work on the problems involving the analysis capability so that students can obtain the expected results.

To improve learning achievement physics subjects teachers need to know how students learn and



what he has learned. By knowing how students learn, teachers will strive to further improve themselves, it is expected of students in the evaluation will receive the expected values. Hence in order to prepare an evaluation tool, teachers can create a higher quality device. Based on this evaluation, the teacher will know how far the knowledge and abilities of students of the subject matter of Physics.

The problem in the study of physics reasoning abilities of high school students (SMA) made in the form of tests. Physical reasoning ability in this case was the physical reasoning ability of students of grade X and grade XI that were located at different levels, but they was assumed to have the same ability. By using the test, someone can reveal the things might not be able to be revealed with other methods. The test is highly desirable existence, so we need standardization of the tests made. To achieve this, it will be made Test of reasoning ability of Physics. The test will be estimated for validity and reliability and conducted equalization.

Tests are a number of questions that have a right or wrong answer [2]. Allen & Yen in [3] states that the test is a device for obtaining a sample of an individual's behavior. Tests can also be interpreted as the number of questions that need answers or responses from the test participants. The test aims to measure the level of a person's ability to reveal certain aspects of the test participants.

In 1950 Dr. Georg Rasch a mathematician from Denmark dealing with analysis of exam elementary school students at different grade exam used to use the same question that is not based on the subject matter in accordance kelasnya. analisis raw score took him on a new finding that opportunities for students to answer the same items as when the student's ability compared dxengan about the difficulty level, so it can be written

Equalization of the test consists of two kinds, namely horizontal and vertical equalization [4]. Horizontal equalization performed on scores from the test instrument which has the same item difficulty by groups of participants at the same level. Vertical equalization is done to reveal the development of the ability of students, although students are in a different class levels, but these instruments measure the same trait.

Vertical equalization is applied at different grade levels at study subjects in high school physics will help teachers obtain information related to the development of the ability of physics students after learning process [5]. Information student skill development can be used as reference material in the teacher determine the next steps and can be used as an evaluation for teachers to improve the quality of learning in the classroom. Therefore, this paper presents physical reasoning ability development of high school students.

The rest of this paper is organized as follow: Section 2 presents literature review. Section 3 describes the proposed research method. Section 4 presents the obtained results and following by discussion. Finally Section 5 concludes this work.

2. Related Works

Tests are a number of questions that have a right or wrong answer [2]. The test is a device for obtaining a sample of an individual's behavior [3]. Tests can also be interpreted as the number of questions that need answers or responses from the test participants. The test aims to measure the level of a person's ability to reveal certain aspects of the test participants.

Equalization is a statistical process that is used to adjust the scores on such tests instrument RPA-instrument so that the scores of tests that can be primarily used to exchange [6]. Equalization can also be interpreted as an empirical procedure to determine the relationship between the raw scores of the two test instruments so that the relationship was able to express instrument test scores which one can be enacted in other instruments [7]. Scores on the assessment of measurement results of education can be compared statistically [8]. This method is called by connecting two tests (linking). The term refers to a hubngan linking between the scores of two tests. Often these two associated tests measure the same construct, but for specific purposes, associate two different test construction. Equalization is not intended to adjust for differences in the assessment and should be applied only to tests designed to the same specifications [9]. Equalization is the process of adjusting statistical score to make scores of different forms of tests [10].

Interest equalization is to produce a score that can be interchangeable. A measure can be interchanged with one another size if the size is obtained from the same construct (eg, length), and the same size. Equalization or can be called concordance would happen if measuring the same construct across different subpopulations. Prediction score the equalizing just hope when the two scores measure the same thing without error with the same reliability as well.

There are four basic principles of equalization [4]. The first is the principle of equality (equity), on the condition of the test participants in each group with the same capabilities, condition score on a test frequency distribution after transformed Y equals the frequency distribution of scores on test X. Secondly, the principle of invariance of the population, which mean that equivalency relationships in transformation do not concern on the group of population used. The third is the principle of symmetry, equalization can be carried on back and forth, regardless of where the test labeled X and Y. The fourth principal is unidimensional, synchronized test devices which measure the same ability.

There are two types equalization, types of vertical and horizontal [4]. In Vertical, two or more test score equivalent is a test that measures the different levels, there is a higher or more lace than others. In the horizontal equalization, two or more test score equivalent is the test that measures the level/ grade which is in the same position. A basic element in the vertical equalization is the use of a set of anchor tests given to two or more groups of examinees (test participants) [11].

The development of the equalization procedure is intended to transfer the scores from test instruments with different degree of difficulty to get to a single scale. The implementation of the equalizing score of two or more in need of the same items, the items shall be the same, whether the charge, the contents of the test, the words or the construction of a sentence. Assessment strategies needed to guide the anchor point decide grains used as an anchor point [12]. Items along part of an item from each of the test instruments.

3. Proposed Method

This study was an exploratory quantitative research since it explores the students' ability in physics reasoning ability tests measurement of high school students (SMA). It also covers development since this study develop the test instruments for measuring the reasoning ability of Physics. The data obtained from tests were then pass through the equalization process by using mean and sigma method on the theory of item response. The accuracy of the method was checked by using Root Mean Square Different (RMSD). The information about the development of reasoning ability of Physics can be obtained from the equalizing score and scale of students in the both classes involved in this study. In some parts, qualitative approach was also carried out to interpret the data obtained from the test.

Purworejo is a district in Central Java Province of Indonesia. There are eleven state high school (SMA) and 14 private high school (SMA) in Purworejo. The population in this study were all high school (SMA) students either in the state or private high school (SMA) in Purworejo based on the response test of reasoning ability of physics subject. The selection of school and grade taken to the sample in this study was done by using proportional stratified random sampling. Based on the population size obtained, then taken proportionally from each stratum to be sampled. The size of the sample taken were 352 students of grade X and 352 students grade XI.

The testing of the validity in this study was done by using content validity test. The validity of the contents on this test instrument was emphasized, therefore, it was done in two ways namely a qualitative study of the test instrument and process them through a formula [13]. Qualitative research which was used based on judgment of experts with the criteria of examination based on aspects of materials, construction, and language. The qualitative research was conducted by three experts, consisting of two high school physics lecturers who have experience and are competent in their field and one expert lecturers of measurement and assessment.

The testing of reliability of instrument in this study was estimated by using the formula of Kuder Richardson (KR 20). Kuder-Richardson formula was used since each component of test was item that given a score of dichotomous, which consists of 0 and 1 [14].

4. Results and Discussion

This section presents the results obtained and following by discussion.

Table 1. Expert Validation of Instrument Grade X

No	R ₁	R ₂	R ₃	S ₁	S ₂	S ₃	Σs	V	No.	R ₁	R ₂	R ₃	S	S ₂	S ₃	Σs	V
1	4	3	3	3	2	2	7	0.78	21	4	3	4	3	2	3	8	0.89
2	4	3	4	3	2	3	8	0.89	22	4	3	3	3	2	2	7	0.78
3	4	4	3	3	3	2	8	0.89	23	4	4	4	3	3	3	9	1.00
4	4	4	4	3	3	3	9	1.00	24	4	3	4	3	2	3	8	0.89
5	4	3	3	3	2	2	7	0.78	25	4	3	4	3	2	3	8	0.89
6	4	3	3	3	2	2	7	0.78	26	4	4	4	3	3	3	9	1.00
7	4	3	3	3	2	2	7	0.78	27	4	4	4	3	3	3	9	1.00
8	4	4	4	3	3	3	9	1.00	28	4	3	4	3	2	3	8	0.89
9	4	4	3	3	3	2	8	0.89	29	4	3	4	3	2	3	8	0.89
10	4	4	3	3	3	2	8	0.89	30	4	4	4	3	3	3	9	1.00
11	4	4	4	3	3	3	9	1.00	31	4	4	4	3	3	3	9	1.00
12	4	3	4	3	2	3	8	0.89	32	4	4	4	3	3	3	9	1.00
13	4	3	4	3	2	3	8	0.89	33	4	4	4	3	3	3	9	1.00
14	4	3	4	3	2	3	8	0.89	34	4	4	4	3	3	3	9	1.00
15	4	3	4	3	2	3	8	0.89	35	4	3	4	3	2	3	8	0.89
16	4	3	3	3	2	2	7	0.78	36	4	4	4	3	3	3	9	1.00
17	4	3	3	3	2	2	7	0.78	37	4	4	4	3	3	3	9	1.00
18	4	4	4	3	3	3	9	1.00	38	4	3	4	3	2	3	8	0.89
19	4	4	4	3	3	3	9	1.00	39	4	4	4	3	3	3	9	1.00
20	4	4	4	3	3	3	9	1.00	40	4	3	4	3	2	3	8	0.89
mean										4	3.5	3.75	3	2.5	2.75	8.25	0.92

The validity of test of reasoning ability of Physics of high school students of grade X was analyzed with a model of expert judgment by three experts (see Table 1). Two experts in the field of Physics and one expert in the field of measurement. The classification of validity can be seen in the following Table 2.

Table 2. Expert Validation of Instrument of Grade XI

No	R ₁	R ₂	R ₃	S ₁	S ₂	S ₃	Σs	V	No	R ₁	R ₂	R ₃	S ₁	S ₂	S ₃	Σs	V
1	4	4	3	3	3	2	8	0.89	21	4	4	4	3	3	3	9	1.00
2	4	4	3	3	3	2	8	0.89	22	4	4	4	3	3	3	9	1.00
3	4	4	3	3	3	2	8	0.89	23	4	4	3	3	3	2	8	0.89
4	4	4	3	3	3	2	8	0.89	24	4	4	3	3	3	2	8	0.89
5	4	4	4	3	3	3	9	1.00	25	4	4	4	3	3	3	9	1.00
6	4	4	4	3	3	3	9	1.00	26	4	4	4	3	3	3	9	1.00
7	4	4	4	3	3	3	9	1.00	27	4	4	3	3	3	2	8	0.89
8	4	3	3	3	2	2	7	0.78	28	4	4	3	3	3	2	8	0.89
9	4	4	4	3	3	3	9	1.00	29	4	4	3	3	3	2	8	0.89
10	4	4	4	3	3	3	9	1.00	30	4	4	4	3	3	3	9	1.00
11	4	3	4	3	2	3	8	0.89	31	4	4	4	3	3	3	9	1.00
12	4	3	4	3	2	3	8	0.89	32	4	4	3	3	3	2	8	0.89
13	4	4	4	3	3	3	9	1.00	33	4	4	4	3	3	3	9	1.00
14	4	4	4	3	3	3	9	1.00	34	4	4	4	3	3	3	9	1.00
15	4	4	4	3	3	3	9	1.00	35	4	4	4	3	3	3	9	1.00
16	4	3	4	3	2	3	8	0.89	36	4	4	4	3	3	3	9	1.00
17	4	4	4	3	3	3	9	1.00	37	4	4	4	3	3	3	9	1.00
18	4	4	4	3	3	3	9	1.00	38	4	4	4	3	3	3	9	1.00
19	4	4	4	3	3	3	9	1.00	39	4	4	4	3	3	3	9	1.00
20	4	4	4	3	3	3	9	1.00	40	4	4	4	3	3	3	9	1.00
mean										4.	3.	3.7	3.	2.9	2,7	8.63	0.96

As can be seen in the Table 1 above that we obtained the average index of aiken'V test of Physics reasoning ability of high school students of grade X is 0.92 which considered in the very high category. Meanwhile, in Table 2, for Grade XI we obtained the average index of aiken'V test is 0.96 which considered in the very high category.

The realibility test of physical ability of grade X is 0.926 and grade XI is 0.95 and this result is classified into the high reliability. In accordance with the Table. 3 Criteria for the level of reliability based on the value of Alpha [8].

Table 3. The Classification of Reliability

coefficient	level Reliability
0 to 0.20	Very low
0.21 to 0.40	Low
0.41 to 0.70	moderate
0.71 to 0.90	High
0.91 to 1.00	Very high

Table 4. The Results of the Reliability of the program Iteman

Score	Alpha	SEM	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	S-B Random	S-B First-Last	S-B Odd-Even
Scored items	0.926	2.695	0.878	0.557	0.936	0.935	0.715	0.967

Score	Alpha	SEM	Split-Half (Random)	Split-Half (First-Last)	Split-Half (Odd-Even)	S-B Random	S-B First-Last	S-B Odd-Even
Scored items	0.928	2.442	0.834	0.634	0.861	0.909	0.776	0.926

The level of difficulty on the test instrument of grade X obtained moderate mean as well as the difficulty level of grade XI.

Table 5. The Level of difficulty of instrument of Grade X

Item No	The difficulty level (Logit)	Category	Item No.	The difficulty level (Logit)	Category
1	-0.96	moderate	21	0.77	moderate
2	-0.61	moderate	22	-0.01	moderate
3	-0.35	moderate	23	0.33	moderate
4	-0.87	moderate	24	-0.10	moderate
5	-0.44	moderate	25	0.16	moderate
6	-0.27	moderate	26	0.16	moderate
7	0.24	moderate	27	0.07	moderate
8	0.16	moderate	28	0.42	moderate
9	0.16	moderate	29	0.77	moderate
10	0.07	moderate	30	-0.01	moderate
11	0:50	moderate	31 *	0.24	moderate
12	0.59	moderate	32 *	-0.18	moderate
13	0.77	moderate	33 *	-0.27	moderate
14	0.68	moderate	34 *	-0.52	moderate
15	1:06	Hard	35 *	-0.61	moderate
16	0.87	moderate	36 *	-0.70	moderate
17	0.68	moderate	37 *	-0.52	moderate

18	-0.01	moderate	38 *	-0.61	moderate
19	1.25	Hard	39 *	-1.14	Easy
20	-0.61	moderate	40 *	-1.14	Easy

Table 6. The Level of difficulty of instrument of Grade XI

Item No	The difficulty level (Logit)	Category	Item No.	The difficulty level (Logit)	Category
1 *	-1.36	Easy	21	0.19	moderate
2 *	-1.72	Easy	22	-0.95	moderate
3 *	-1.50	Easy	23	0.19	moderate
4 *	-1.43	Easy	24	0.65	moderate
5 *	-0.87	moderate	25	-0.26	moderate
6 *	-0.18	moderate	26	0.00	moderate
7 *	-0.43	moderate	27	1.49	Hard
8 *	1.10	Hard	28	0.30	moderate
9 *	-0.58	moderate	29	0.19	moderate
10 *	1.74	Hard	30	0.30	moderate
11	0.00	moderate	31	0.19	moderate
12	0.53	moderate	32	-0.43	moderate
13	-1.50	Easy	33	1.28	Hard
14	1.28	Hard	34	1.74	Hard
15	0.65	moderate	35	-0.26	moderate
16	0.19	Easy	36	-1.72	Easy
17	1.49	Hard	37	-0.26	moderate
18	0.00	moderate	38	0.00	moderate
19	0.09	moderate	39	-0.26	moderate
20	0.30	moderate	40	-0.18	moderate

Vertical equalization was done by equalizing the level of difficulty on the parameter items together or anchor of test instrument of grade X into the test instrument of grade XI. This equivalency test was conducted on the value of the item difficulty level shared items that have been analyzed back to exclude items that do not fit the model. Vertical equalization method used was the average of sigma method. This method concern on the mean and standard deviation for each test instrument. The following are summaries of equivalency of test instrument of grade X and grade XI seen based on the level of difficulty of items (b) anchor in the Table 7.

Table 7. The Average level of difficulty of Grade X and Grade XI

No grain	The difficulty level X (Anchor Test class X)	No grain	The difficulty level Y (Anchor Test XI)
31	-0.72	1	-0.88
32	-0.53	2	-1.49
33	-0.87	3	-1.03
34	-0.56	4	-1.26
35	-0.28	5	-0.36
36	0.17	6	-0.85
37	0.45	7	0.03
38	1.01	8	0.76
39	-0.01	9	-0.29
40	-0.05	10	0.35
b1	-0.14	b2	-0.50
S1	0.58	S2	0.73

Based on the above table, obtained the average level of difficulty of the test instrument of grade XI for the anchor item of (b2) is -0.50 logit whereas, the average level of difficulty of the test instrument of grade X for the anchor item of (b1) the logit -0.14. The scores of the level difficulty of anchor being equivalent obtained from the mean scores of level of difficulties at the anchor items of grade X and XI for each item. Next, the equalization of equation was calculated by using the mean and sigma method by calculating the constants α and β . In a logistic model 1PL (Rasch), a constant value of $\alpha = 1$, so that the constant β translation into:

$$\begin{aligned}\beta &= b_2 - \alpha (b_1) \\ &= -0.50 - (1 \times -0.14) \\ &= -0.36\end{aligned}$$

The average level of difficulty on the test instrument of grade X is -0.33 logit (485 based on scale of WITs) and the average level of difficulty 0.00 logit (500 based on scale of WITs) on the test instruments of grade XI shown Figure 1.

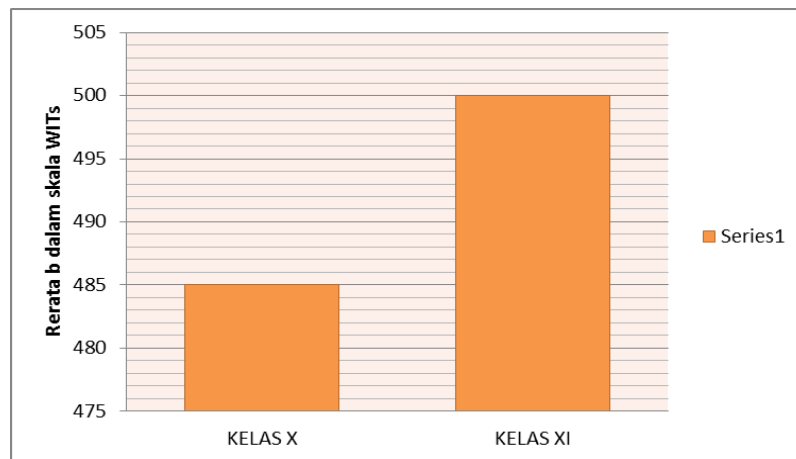


Figure 1. Comparison of Levels of Class X and XI Difficulties

5. Conclusion

Based on calculations, it was found that the average scores of the ability of students grade X in the prior comparable scale is -1.06. Whereas the average ability of students grade XI based on the scale is -1.10. Furthermore, to compare the ability of grade X and grade XI conducted equivalency scale of ability. The scale of abilities of grade X can be compared to the ability of students of grade XI who has been synchronized by using the equation that has been calculated above, it is as $\theta^* = \theta_1 - 0.36$. Based on the equation $\theta^* = \theta_1 - 0.36$, means that if the value of $\theta_1 = 0$. then the value of $\theta^* = -0.36$. This indicates if the ability of grade X is 0 then it was equivalent to the ability of students of grade XI -0.36. Based on these equations, it can be concluded that the ability of grade X become lower than grade XI if it had been converted into grade XI. The scale of abilities of grade X that has been converted to grade XI can then be compared. The comparison of the ability scale was conducted to determine whether there is a development on the students' abilities.

Acknowledgement

This research is fully supported by Affiliation Research Grant.

References

- [1] Fatmawati, H. (2014). Analisis Berpikir Kritis Siswa dalam Pemecahan Masalah Matematika Berdasarkan Polya pada Pokok Bahasan Persamaan Kuadrat. *Jurnal Elektronik Pembelajaran Matematika*, 2(9), 911–922.
- [2] Mardapi, D. (2008). *Teknik penyusunan instrumen tes dan non tes*. Yogyakarta: Nuha medika.
- [3] Allen, M. ., & Yen, W. . (1979). *Introduction to measurement Theory*. Monterey: Brook/Cole Publishing Company.
- [4] Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory Principles and Applications*. Boston: Kluwer Nijhoff Publishing.
- [5] Croker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehard and Winston Inc.
- [6] Kolen, M. J., & Robert L Brennan. (2004). *Springer Series in Statistics*.
- [7] Peterson, N. S., & Kolen, M. J. (1989). *Scaling, Norming, and equating dalam R.L Linn (Eds) Educational Measurement*. New York: Macmillan Publishing Company.
- [8] Retnawati, H. (2014). Teori Respons Butir dan Penerapannya: untuk Peneliti, Praktisi Pengukuran dan Pengujian, Mahasiswa Pascasarjana, 15.
- [9] Muraki, E., Hombo, C. M., & Lee, Y. W. (2000). Equating and linking of performance assessments. *Applied Psychological Measurement*, 24(4), 325–337. <https://doi.org/10.1177/01466210022031787>
- [10] Kim, J., Madison, W., Hanson, B. A., & McGraw-hill, C. T. B. (2002). Test Equating Under the Multiple-Choice Model, 26(3), 255–270.
- [11] Baker, F. B. (1993). Equating Tests Under The Nominal Response Model. *Applied Psychological Measurement*, 17(3), 239–251. <https://doi.org/10.1177/014662169301700305>
- [12] Kopf, J. (2013). Anchor methods for DIF detection : A comparison of the iterative forward , backward , constant and all-other anchor class Anchor methods for DIF detection : A comparison of the iterative forward , backward , constant and all-other anchor class, (141).
- [13] Lewis R. Aiken. (1981). from the SAGE Social Science Collections . All Rights Reserved . *Personality and Social Psychology Bulletin*, 07(04), 565–571. <https://doi.org/0803973233>
- [14] Ebel, R. L., & D A Frisbie. (1986). *Essentials of educational measurement*. Englewood,NJ: Prentice-Hall Inc.