JP3I (Jurnal Pengukuran Psikologi dan Pendidikan Indonesia)
DOI: http://dx.doi.org/10.15408/jp3i.v13i1.36745
http://journal.uinjkt.ac.id/index.php/jp3i

Modelling of Colorado Learning Attitude Science Survey in Indonesian Version: A Study with Applying Item Response Theory

Mutmainna[1,4], Mutahharah Hasyim[2], Syamsuriwal[3], Bayu Setiaji[4], Farida Agus Setiawati[4]

[1] Universitas Sulawesi Barat
[2] Universitas Negeri Makassar
[3] Universitas Tadulako
[4] Universitas Negeri Yogyakarta

mutmainna.2022@student.uny.ac.id

Abstract

Colorado Learning Attitudes about Science Survey (CLASS) is an instrument designed to explore students' perceptions of physics and assess how closely their beliefs correspond with those of professional physicists. Before the development of CLASS, several similar instruments were developed in the field of Physics Education, such as the Maryland Physics Expectation (MPEX), Views About Science Survey (VASS), and Epistemological Beliefs Assessment for Physical Science (EBAPS). Adams et al. developed CLASS in 2006 by evaluating these three instruments. Since then, CLASS has been extensively studied for its use in research, especially in Physics Education. It has also been applied in other fields and translated into several languages. As a form of community strengthening, this article attempts to report the research findings related to using the CLASS instrument that has been translated into Indonesian. A total of 292 undergraduate students from four universities were sampled in this study. The respondents in this study were students who had enrolled in the Fundamental of Physics course. The data obtained were analysed using the Grade Response Model (GRM) method after comparison with several other methods for polytomous scale with Item Response Theory (IRT) like Partial Credit Model (PCM), Rating Scale Model (RSM), and Generalized Partial Credit Model (GPCM). The research results show the model considered most suitable is GRM The research results show that among the four models of approach and based on the criteria used, the model considered most suitable is GRM. The research also shows that the number of items declared consistent with the model does not cover all CLASS items but rather some items. This finding indicates that further exploration of the CLASS instrument items is needed, especially in the Indonesian version. The findings of this study also add to the wealth of knowledge related to the quality assessment of the CLASS instrument through the modern test theory approach (IRT). Thus, the CLASS instrument can be considered a standard instrument used globally across various populations.

Keywords: CLASS Indonesian version, physics education, IRT, polytomous response

Abstrak

Colorado Learning Attitudes about Science Survey (CLASS) adalah instrumen yang dirancang untuk mengeksplorasi persepsi siswa terhadap fisika dan menilai seberapa dekat keyakinan mereka dengan keyakinan fisikawan profesional. Sebelum berkembangnya CLASS, beberapa instrumen serupa dikembangkan dalam bidang Pendidikan Fisika, seperti Maryland Physics Expectation (MPEX), Views About Science Survey (VASS), dan Epistemological Beliefs Assessment for Physical Science (EBAPS). Adams *dkk. mengembangkan CLASS pada tahun 2006 dengan mengevaluasi ketiga instrumen ini. Sejak saat itu, CLASS telah dipelajari secara ekstensif untuk digunakan dalam penelitian, khususnya dalam Pendidikan Fisika. Ini juga telah diterapkan di bidang lain dan diterjemahkan ke dalam beberapa bahasa. Sebagai bentuk penguatan komunitas, artikel ini mencoba melaporkan temuan penelitian terkait penggunaan instrumen CLASS yang telah diterjemahkan ke dalam bahasa Indonesia. Sebanyak 292 mahasiswa S1 dari empat universitas dijadikan sampel dalam penelitian ini. Responden dalam penelitian ini adalah mahasiswa yang telah mengikuti mata kuliah Fisika Dasar. Data yang diperoleh dianalisis menggunakan metode Grade Response Model (GRM) setelah dibandingkan dengan beberapa metode lain untuk skala polytomous dengan Item Response Theory (IRT) seperti Partial Credit Model (PCM), Rating Scale Model (RSM), dan Generalized Partial Credit Model (GPCM). Hasil penelitian menunjukkan bahwa diantara keempat model pendekatan dan berdasarkan kriteria yang digunakan, model yang dianggap paling sesuai adalah GRM. Penelitian juga menunjukkan bahwa jumlah item yang dinyatakan konsisten dengan model tidak mencakup seluruh item CLASS melainkan beberapa item. Temuan ini menunjukkan bahwa diperlukan eksplorasi lebih lanjut terhadap item instrumen CLASS, khususnya pada versi bahasa Indonesia. Temuan penelitian ini juga menambah kekayaan pengetahuan terkait penilaian kualitas instrumen CLASS melalui pendekatan teori tes modern (IRT). Oleh karena itu, instrumen CLASS dapat dianggap sebagai instrumen standar yang digunakan secara global di berbagai populasi.*

***Kata Kunci**: CLASS versi Bahasa Indonesia, pendidikan fisika, IRT, respon politomus*

Introduction

Students begin their studies with expectations and beliefs that generally have a different perspective from scientists. A perspective aligned with that of physicists is a vital learning outcome(Gray et al., 2008). In the field of Physics Education, there is a specific topic that examines this. Research related to this topic has been vigorously conducted until now. This is considered essential to continue learning because it is believed that most teaching practices can cause a substantial decrease in student grades, the likelihood of students choosing a physics major in correlation with their interests, and for the majority of the student population, men's scores in some categories are very different from women's scores (Adams et al., 2006). Instruments such as the Maryland Expectations about Science Survey (MPEX) (Redish et al., 1998), Epistemological Beliefs Assessment for Physical Science (EBAPS) (Halloun, 1997), Views about Science Survey (VASS) (A. Elby, 2002), and Colorado Learning Attitudes about Science Survey (CLASS) (Adams et al., 2006) have been developed to measure students' attitudes towards physics and compare them with experts (Kontro & Buschhüter, 2020).

CLASS is one of the latest and most widely used instruments to measure attitudes. The CLASS item statements are formulated concisely and can be used in various physics courses (Adams et al., 2006). This instrument consists of 42 statements rated on a five-point Likert scale (ranging from strongly agree to strongly disagree). CLASS has been adapted for use in Biology (Semsar et al., 2011) and Chemistry (CLASS-Chem) (Adams et al., 2008). In addition, CLASS has also been translated into various national languages, such as Spanish (de la Garza et al., 2010), Chinese (Zhang & Ding, 2013), Arabic (Alhadlaq et al., 2009), and others. When translating CLASS, great care has been taken to ensure students understand the questions correctly. For example, the Mandarin translation was validated by bilingual physicists and through student interviews (Zhang & Ding, 2013). Student interviews also validated the English version for a predominantly Hispanic sample, which found one misinterpretation. For the item discussing the situation where students do not remember an equation during an exam, many said they would try to answer the question another way. In contrast, expert reasoning would say that the equation can be derived (Sawtelle et al., 2009).

Data from different populations have also been used in the development of CLASS. Moreover, the CLASS statements are formulated concisely and involve situations that, according to the authors, can arise in all Physics studies (Adams et al., 2006). Adams et al. (2006) explain that CLASS aims to measure physics as a practice and the process of learning physics, and Physics is viewed as a science. All statements in CLASS were developed with validation and testing by Physicists and through interviews with students. This explanation is one of the empirical evidence that CLASS was developed into one of the reference instruments that can be used across various populations.

Research related to the quality testing of the CLASS instrument after being developed by Adams et al. (2006), was followed up in several articles, such as the validity test by Douglas et al. (2014), through Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) suggested that the CLASS construct of eight components be simplified to three components. Kontro et al. [6] followed up on the findings (Douglas et al., 2014) but applied it to the Finnish population and registered findings by Douglas et al. (2014). Martins & Lindsay (2022) evaluated CLASS with K-12 curriculum adjustments. In addition to testing the validity of CLASS through the classical approach, several studies have also been conducted by applying modern test theory in the form of Item Response Theory (IRT) like Jahanifar & Derafsh (2020) through factor analysis and IRT performance with graded response found that Cronbach's alpha of 0.701 – 0.891 for the eight components of CLASS fits the IRT model with the GRM method. Another finding by Christman et al. (Christman et al., 2020) through Exploratory Factor Analysis (EFA) and Graded

IRT obtained results that EFA failed to find the construct described in the original version (eight components) but found things in line with the findings of (Douglas et al., 2014). In addition, through Graded IRT analysis, IRT correlates more strongly with the data obtained than with factor analysis.

The CLASS instrument in current research has been rendered into Bahasa Indonesia and is available online at https://www.physport.org/assessments/assessment.cfm?A=CLASS. Studies within Indonesia's physics education landscape have explored the concept of learning attitude through various research. These studies have delved into the correlation between high school students' motivation and their stance on Physics classes (Astalini et al., 2019), assessed how students perceive Physics education (Kurniawan et al., 2019), and examined the variance in student attitudes when different educational strategies are employed (Sakliressy et al., 2021). However, none of these investigations utilised the CLASS tool to gauge students' attitudes toward Physics learning within the Indonesian context.

This study attempts to convey research findings on student responses to CLASS by presenting them through an IRT approach. The intended approach is the Grade Response Model (GRM), Partial Credit Model (PCM), Rating Scale Model (RSM), and Generalized Partial Credit Model (GPCM). CLASS is an instrument that produces category responses using a Likert scale. In simulation studies for instruments like the Likert scale by Hauck et al. (Hauck Filho et al., 2014), GRM and other methods such as RSM show some parameters consistently higher than those obtained from other models and provide sTable estimates for responses in a Likert scale.

Using four models of the IRT approach, this study aims to report the most suitable modelling of the Indonesian version of CLASS among the four IRT approach models used and to explore whether there are instrument items that are considered to need attention for follow-up based on the findings obtained. This step is taken to make this research one of the reference reports on the quality of the CLASS instrument about testing on various populations, including in Indonesia. Moreover, this step is part of strengthening the Physics Education community internationally.

The CLASS

Over the last ten years, physics education has undergone reforms designed to improve students' attitudes towards physics, and CLASS has become an essential tool for assessing curriculum reforms (Douglas et al., 2014). According to Google Scholar on December 6, 2023, the CLASS article (Adams et al., 2006) published in physics education research (PER) has been cited in 1054 articles. Meanwhile, according to the Physical Review journal site, the CLASS article (Adams et al., 2006) has been cited in 362 articles for research within the scope of PER. CLASS has been used in various studies published in PER (for example, Corsiglia et al., 2023; Freed et al., 2022; Hynninen et al., 2023; Thacker, 2023). In addition, it has been modified for biology and chemistry (Adams et al., 2008; Semsar et al., 2011) and translated for use in several languages (Alhadlaq et al., 2009; de la Garza et al., 2010; Zhang & Ding, 2013).

CLASS was developed by scholars at the University of Colorado and is based on other established attitude and epistemology surveys towards Fishbein's attitude theory (Ajzen & Fishbein, 1977) and science (Adams et al., 2006; Douglas et al., 2014). Adams et al. also reported spending much time interviewing experts and students to understand better students' attitudes and beliefs about physics and physics learning. According to them, this survey investigates students' beliefs about physics and physics learning and distinguishes these beliefs from the views of experts. CLASS is written to make statements as clear and concise as possible and is suitable for use in various physics courses (Adams, 2005).

Adams et al. (2006) developed CLASS by administering it in comprehensive high schools and universities across various physics courses involving 5000 students from different majors. Initially, CLASS consisted of 42 Likert scale items that explored students' attitudes towards physics. Students rated their level of agreement with each item on a five-point scale from strongly disagree to strongly agree. CLASS assessments were determined by participant responses that agreed with opinions predetermined by physicists. Based on factor analysis work, 26 out of 41 items were grouped into eight overlapping attitude factors about physics and physics learning. The eight identified factors are (1) Real-World Connections, (2) Personal Interest, (3) Effort and Sense Making, (4) Conceptual Connections, (5) Applied Conceptual Understanding, (6) General Problem Solving, (7) Problem Solving Confidence, and (8) Problem Solving Sophistication (Adams et al., 2006; Kontro & Buschhüter, 2020; Martins & Lindsay, 2022). In a psychometric re-evaluation, Heredia and Lewis suggested that the psychometric re-evaluation of the chemistry version of CLASS found a three-factor solution from 16 items that most closely fit their data and Fishbein's attitude theory (Ajzen & Fishbein, 1977).

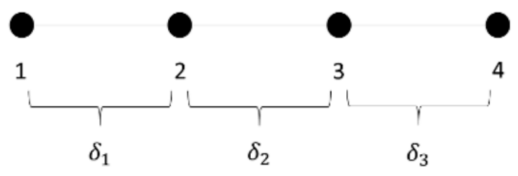
Item Response Theory with Polytomous Item Response

Item response can encompass more than just dichotomous outcomes. Scores for items that are constructed responses with partial credit and those that use a Likert scale response are scored polytomous (Paek & Cole, 2019). In a framework where item responses are scored dichotomously, the item response function (IRF) is the fundamental component for modelling these responses. The primary element for models with polytomous responses is the Category Response Function (CRF), also known as the category probability function. Similarly to how a dichotomous IRF is visually represented through an item characteristic curve (ICC), a polytomous CRF is depicted using a category characteristic curve (CCC) or a category probability curve (Adams, 2005). Several models that can be used if item responses are in the form of polytomous scores include the Partial Credit Model (PCM), Rating Scale Model (RSM), Generalized Partial Credit Model (GPCM), Grade Response Model (GRM), and Nominal Response Model (NRM) (Paek & Cole, 2019; Widhiarso, 2010). Essentially, each approach model has its characteristics according to its intended use.

The following is a brief description related to PCM, RSM, GPCM, and GRM. NRM is not explained in this section because the data analysed in this study are not on a nominal scale. Initially, PCM and RSM will be explained. Both approach models are considered polytomous forms of the Rasch model (Desjardins & Bulut, 2018). PCM was initially developed to analyse test items, which required several steps. For example, calculation problems in physics subjects consist of parts from problem identification to the final solution (Widhiarso, 2010). PCM can also be used to analyse responses on a multi-point personality scale (Embretson & Reise, 2000). PCM was developed from the 1-PL Model and included in the Rasch model (Desjardins & Bulut, 2018; Paek & Cole, 2019; Widhiarso, 2010).In PCM, the probability of obtaining X_i points ($X_i = 0, 1, \dots m_i$) on item i for PCM can be written as:

$$P(X_i|\theta, \delta_{ih}) = \frac{exp\left[\sum_{h=0}^{X_i}(\theta - \delta_{ih})\right]}{\sum_{k=0}^{m_i} exp\left[\sum_{h=0}^{k}(\theta - \delta_i h)\right]},$$

where θ is the latent trait, δ_{ih} is the step parameter (also as step difficulty) that represents the relative difficulty in obtaining h points over $(h-1)$ points (De Ayala, 2013; Desjardins & Bulut, 2018).



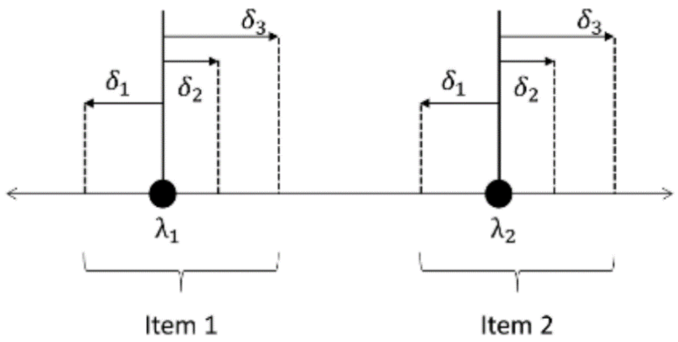
Sources: Desjardins & Bulut (2018)

Figure 1. PCM - Thresholds between the four ordered response categories

Furthermore, the RSM requires that all items have the same number of options or categories, and it assumes that adjacent threshold parameters are equally spaced, i.e., are equidistant, across all items model (Desjardins & Bulut, 2018; Paek & Cole, 2019; Widhiarso, 2010). If items in the scale have different formats, then RSM cannot be applied (Widhiarso, 2010). Each item is denoted by one location parameter (reflecting the λ_i)relative difficulty of a particular item. For the RSM, the probability of selecting category c ($c = 0,1, \ \ldots, m$) for item i may be written as:

$$P(X_{ic}|\theta, \lambda_i, \delta_1, \ldots, \delta_m) = \frac{exp\left[\sum_{j=0}^{c}\left(\theta - (\lambda_i + \delta_j)\right)\right]}{\sum_{h=0}^{m} exp\left[\sum_{j=0}^{h}\left(\theta - (\lambda_i + \delta_j)\right)\right]},$$

where λ_i is the location parameter for item i and $\delta_1, \ldots, \delta_m$ are the category threshold parameters



Sources: Desjardins & Bulut (2018)

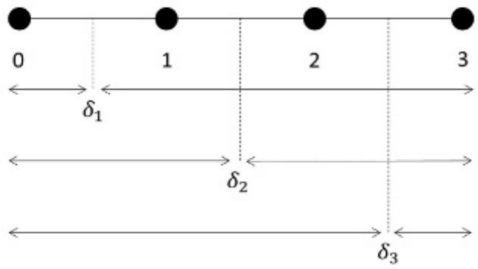
Figure 2. RSM - Location and threshold parameters of two rating scale items.

The next will discuss GPCM and GRM. Both models are categorised as polytomous non-Rasch Models and can be viewed as polytomous forms of the two-parameter (2PL). GPCM is a model developed by Muraki (Muraki, 1992), which is a redevelopment of the PCM model. GPCM allows grains in the scale to have differences in slope parameters (Paek & Cole, 2019). The probability of obtaining X_{ik} points ($X_{ik} = 0, 1, \ldots m_i$) on item i for GPCM can be written as

$$P(X_{ik}|\theta, a_i, \delta_{ik}) = \frac{exp\left[\sum_{h=1}^{X_{ik}} a_i(\theta - \delta_{ih})\right]}{\sum_{c=1}^{m_i} exp\left[\sum_{h=1}^{c} a_i(\theta - \delta_i h)\right]},$$

where a_i is the discrimination parameter for item i. Like PCM, the thresholds (δ_{ik}) are not restricted to the same order as the response categories. Furthermore, GRM was initially introduced as a homogeneous case of the stratified response model because the forms of cumulative category response functions are the same and never cross each other (Adams, 2005). The Common

Response Model (GRM) is appropriate for items with different response continuums. It predicts the likelihood of choosing a particular category of responses or one that ranks higher, in line with the order in which responses are chosen. In GRM, each response category adds some detail to the likelihood of someone choosing a particular response category. When dealing with items with sequential response categories K, GRM generates binary items K − 1 by dividing the response categories cumulatively (Desjardins & Bulut, 2018).



Sources: Desjardins & Bulut (2018)

Figure 3. GRM - Cumulative thresholds between the four response categories

Figure 3 illustrates the cumulative division of the four response categories through three thresholds (δ_1, δ_2, δ_3). Each threshold parameter indicates the latent trait level necessary for a 50% chance of choosing a specific response category or higher. The probability of obtaining X_i points or higher ($X_i = 1, 2, \ldots m_i$) on an item for i the GRM can be written as follows:

$$P^*(X_i|\theta, a_i, \delta_{X_i}) = \frac{e^{a_i(\theta - \delta_{X_i})}}{1 + e^{a_i(\theta - \delta_{X_i})}},$$

where θ is the latent trait, a_i is the discrimination parameter for item i, δ_{Xi} is the category boundary location for the category X_i (similar to the category threshold parameter in the previous models), and $P^*(X_i|\theta, a_i, \ \delta_{Xi})$ is the probability of a person obtaining the score of X_i or higher (Embretson & Reise, 2000). As mentioned, GRM splits a polytomous item into a series of dichotomous items using cumulative probabilities. That is, item i consists of mi dichotomous items that share the same discrimination parameter (a_i) but have unique difficulty parameters (δ_{Xi}).

Method

Participant and Data Collection

A total of 292 undergraduate students were sampled in this study. The students come from four universities in Indonesia, three from central Indonesia and one from western Indonesia. The sampling technique carried out in this study was obtained through the help of lecturers who teach introductory physics courses at the four universities. CLASS is an instrument that measures student attitudes towards Physics and its learning, so this instrument is intended for students who have or are currently programming Physics courses, specifically in Basic Physics courses.

The profile of participants in this study consisted of 21.23% men and 78.77% women. In addition, it consists of 219 students in the first semester, 29 in the third semester, 11 in the fifth semester, and 33 in the seventh and eighth semesters. Regarding the origin of the study program, students consist of 10.62% from Physics, 55.82% from Physics Education, 19.18% from Biology Education, 7.88% from Chemistry Education, 4.79% from Informatics Engineering and 1.71% from other.

Translate CLASS into Indonesian

CLASS was developed by Adams, W. K. et al. (Adams et al., 2006) in the original version in English form. Translating the CLASS into the Indonesian version in this study involved three validators assessing the CLASS instrument independently translated by the researcher. The qualifications used in choosing validators are having decent skills in English and being able to understand Physics terms in context so that the selected validators consist of linguistic experts and researchers in the field of Physics Education who have a track record of English research searched through Google Scholar. The assessment results were analysed using the Aiken Index (Aiken, 1980). The revision process based on validators' input is carried out until the Indonesian version of the CLASS instrument, according to the calculation results with the Aiken Index, is declared valid. The validator assessment results are then uploaded to https://www.physport.org, as presented in Figure 4.

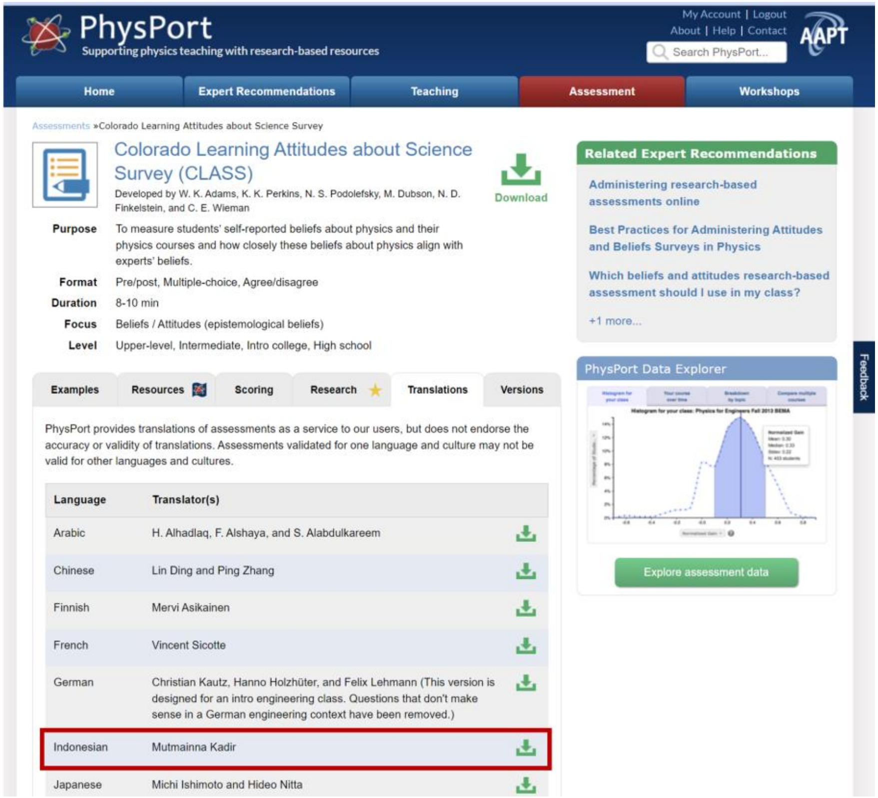


Figure 4. CLASS Indonesian version at Physport

Data Analysis

The CLASS test instrument consists of 42 items that measure eight components, namely (1) Real World Connections, (2) Personal Interests, (3) Sense Making and Effort, (4) Conceptual Connections, (5) Applied Conceptual Understanding, (6) General Problem Solving, (7) Problem-Solving Confidence, and (8) Problem-Solving Sophistication. Of the 42 items, one item (item 31) was not included in the analysis because it is provided to filter the answers of respondents who are considered not to answer thoughtfully. As for what was applied in this study, 292 respondents were analysed. In this case, the role of item 31 can be considered for future research, especially in validating the Indonesian version of the CLASS instrument. The modelling of the Indonesian language version of the CLASS instrument is carried out in several stages until the most suitable Table model is obtained. The steps are based on data from 292 samples for responses to 41 instrument items that were entirely analysed using several instrument analysis methods. The CLASS instrument is a questionnaire with a Likert scale with five options, so this analysis uses IRT with polytomous item response. Some of the methods chosen include the Partial Credit Model (PCM), Rating Scale Model (RSM), Generalized Partial Credit Model (GPCM), and Graded

Responses Model (GRM). Data analysis using R Program version 4.3.0 using the MIRT package. The model is suitable for the first part of the analysis concerning the AIC, SABIC, HQ, and BIC values. The four methods applied were compared using ANOVA. The model best suits the instrument looks at the smallest AIC, SABIC, HQ, and BIC values. The next stage is to explore whether the analysis fits into the item-by-item model. Considerations used in determining an item's quality include looking at the Item Probability Function Graph and the Fit status of an item against a model that refers to the value $X^2 \geq 0{,}05$. The final step is to eliminate items declared unfit for the model from the initial data and then analyse and report back the AIC, SABIC, HQ, and BIC values obtained along with the appropriate items.

Results and Discussions

The results of the comparison of model conformity values with the four approaches used are presented in Table 1 below:

Table 1. The matching value of the assessment model against four unidimensional IRTs with the Polytomous Item Responses method was applied.

Metode	AIC	SABIC	HQ	BIC
GRM	29798.26	29901.89	30100.17	30551.99
GPCM	30084.49	30188.12	30386.40	30838.22
RSM	31466.34	31489.08	31532.61	31631.79
PCM	414113.51	414217.14	414415.42	414867.24

Based on Table 1, the smallest AIC, SABIC, HQ, and BIC values are obtained in the GRM method. At a later stage, digging into more detailed results in the GRM method, other data are presented. Based on the data obtained from Table 2 shows that the distribution of values a, b_1, b_2, b_3, and b_4, each number is generally in the same range, except for data containing items 13, 21 and 23.

Table 2. Graded responses calibration trait and item parameter values (N = 292)

Item	a	b_1	b_2	b_3	b_4
B1	0.897	-3535	-1.811	0.369	3.284
B2	-0.512	1.317	-1.366	-3.722	-6.753
B3	2.053	-2.259	-1.601	-0.353	0.729
B4	1.527	-2.929	-2.029	-0.375	0.869
B5	0.727	-4.493	-2.245	0.170	2.603
B6	-0.692	1.096	-1.323	-3.466	-4.753
B7	0.738	-3.620	-2.317	0.299	2.256
B8	2.221	-2.533	-1.824	-0.808	0.569
B9	1.226	-2.935	-1.894	-0.260	1.307
B10	-0.327	3.861	-1.198	-6.804	-10.717
B11	1.847	-2.755	-1.818	-0.368	0.827
B12	1.389	-2.724	-1.737	-0.668	0.633
B13	-0.053	32.157	1.730	-25.263	-53.855
B14	2.189	-2.538	-1.635	-0.281	0.816
B15	2.043	-2.927	-1.873	-0.395	1.050
B16	1.965	-2.859	-1.884	-0.905	0.337
B17	0.250	-8.507	-1.320	5.891	11.453
B18	0.666	-3.762	-2.012	0.916	3.684
B19	2.057	-2.636	-1.769	-0.687	0.591

B20	0.314	-7.570	-2.970	1.451	7.751
B21	-0.031	60.609	18.820	-26.936	-77.066
B22	0.372	-8.210	-2.967	1.869	6.008
B23	-0.112	16.106	3.861	-9.595	-24.322
B24	2.146	-2.644	-1.850	-0.404	0.721
B25	0.924	-3.497	-2.047	0.593	2.728
B26	2.739	-2.610	-1.671	-0.427	0.805
B27	1.468	-2.943	-1.846	0.201	1.513
B28	2.237	-2.769	-1.680	-0.412	0.667
B29	0.799	-3.561	-1.715	0.369	2.458
B30	2.448	-2.705	-1.585	-0.277	0.798
B32	-0.623	1.428	-0.919	-3.072	-5.813
B33	1.603	-3.376	-1.836	-0.401	1.053
B34	1.142	-4.239	-2.016	0.507	2.522
B35	-0.669	1.628	-0.618	-3.350	-5.318
B36	1.537	-2.887	-1.590	0.064	1.839
B37	1.841	-2.648	-1.623	-0.151	1.372
B38	0.424	-4.656	-1.761	2.182	5.920
B39	2.369	-2.733	-1.760	-0.217	1.174
B40	0.555	-5.289	-1.680	1.518	4.955
B41	1.701	-3.089	-1.913	-0.097	1.333
B42	2.424	-2.530	-1.577	-0.272	0.860

As seen in Table 2, it is observed that items 13, 21, and 23 exhibit distinct relative ranges compared to the rest of the items within the parameters from b_1 to b_4. A high number on the finite range b_1 to b_4 indicates that the difficulty level of these three items is higher than that of others. Next, the following are presented examples of graphs of Probability Functions for items 28 and 41.

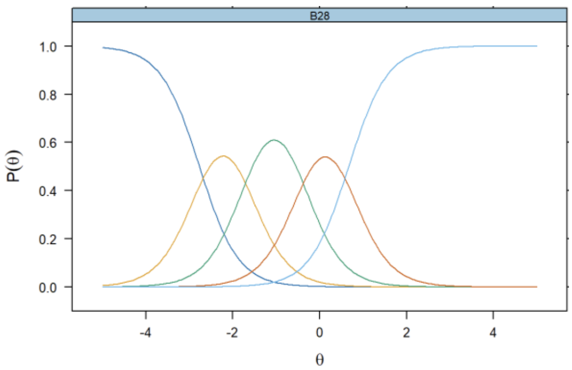


Figure 5. Item 28 a = 2.237; b_1=-2.769; b_2=-1.680; b_3=-0.412; b_4=0.667

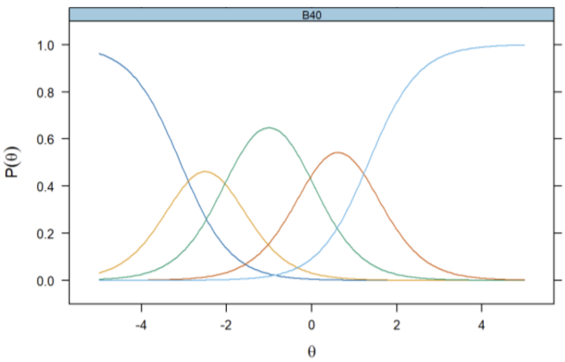


Figure 6. Item 41 a = 1.701; b_1=-3.089; b_2=-1.913; b_3=-0.097; b_4=1.333

Figure 5 and Figure 6 show the distribution of the conversion scale (θ) of student response learning attitudes through the CLASS instrument in an axis inclined to the left. This is in line with the graph of the information function for all items in Figure 7. Furthermore, for the graph display of the Probability Function/CRF, the TIF for the 41 question items and the information function graph and error standards are presented in Figure 7. Based on the data obtained from these two numbers, for per-item analysis of IRF graphs and information functions, some numbers are considered less functional in Learning Attitude for ranges θ = -6 until θ = 6 as applicable to other items. Examples of such items include items 13, 21, and 23. This is in line with the data in Table 2.

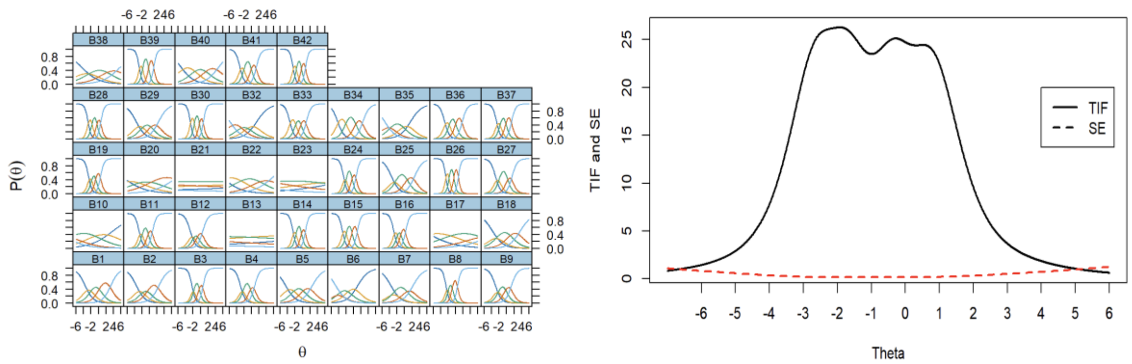


Figure 7. Item Probability Function, Test Information Function and Standard Error of Measurement

The data obtained based on whether each item matches the measurement model and the item number to be dropped in the following analysis is reported in Table 3.

Table 3. Fit test per item based on X^2 value.

Item	S_ X ^{2}	df.S_ X ^{2}	RMSEA.S_ X ^{2}	p.S_X ^{2}	Note	Information
B1	71.880	57	0.030	0.089	Fit	
B2	77.531	68	0.022	0.201	Fit	
B3	90.002	51	0.051	0.001	Not Fit	
B4	96.598	53	0.053	0.000	Not Fit	
B5	86.665	73	0.025	0.131	Fit	
B6	92.081	67	0.036	0.023	Not Fit	
B7	109.908	78	0.037	0.010	Not Fit	
B8	69.828	45	0.044	0.010	Not Fit	
B9	98.984	59	0.048	0.001	Not Fit	
B10	110.455	74	0.041	0.004	Not Fit	
B11	77.311	51	0.042	0.010	Fit	

B12	70.601	63	0.020	0.239	Fit	
B13	100.059	79	0.030	0.055	Not Fit	Drop
B14	71.266	49	0.040	0.021	Not Fit	
B15	69.303	43	0.046	0.007	Not Fit	
B16	93.943	46	0.060	0.000	Not Fit	
B17	102.609	72	0.038	0.010	Not Fit	Drop
B18	55.646	62	0.000	0.702	Fit	
B19	67.906	48	0.038	0.031	Not Fit	
B20	94.380	72	0.033	0.040	Not Fit	
B21	101.912	72	0.038	0.012	Not Fit	Drop
B22	87.119	72	0.027	0.108	Fit	
B23	104.801	84	0.029	0.062	Fit	Drop
B24	78.499	46	0.049	0.002	Not Fit	
B25	52.171	56	0.000	0.621	Fit	
B26	60.420	42	0.039	0.033	Not Fit	
B27	55.866	50	0.020	0.264	Fit	
B28	124.723	48	0.074	0.000	Not Fit	
B29	87.200	78	0.020	0.223	Fit	
B30	93.649	45	0.061	0.000	Not Fit	
B32	116.720	81	0.039	0.006	Not Fit	
B33	63.527	54	0.025	0.176	Fit	
B34	42.588	45	0.000	0.575	Fit	
B35	87.448	75	0.024	0.154	Fit	
B36	54.672	49	0.020	0.268	Fit	
B37	82.640	45	0.054	0.001	Not Fit	
B38	96.918	80	0.027	0.096	Fit	
B39	61.993	37	0.048	0.006	Not Fit	
B40	58.786	64	0.000	0.661	Fit	
B41	64.436	47	0.036	0.046	Not Fit	
B42	78.488	44	0.052	0.001	Not Fit	

Referring to data presented in Table 3, the range of previous items indicated shows a relatively different score range from other items so that a drop is made. The items consist of 13, 17, 21, and 23. After the number of items from 41 items to 37 items, the analysis is carried out again, and the comparison data presented in Table 4 is obtained.

Table 4. The matching value of the assessment model against four unidimensional IRT with Polytomous Item response methods was applied before and after the reduction.

Number of Items	Method	AIC	SABIC	HQ	BIC
41	GRM	29798.26	29901.89	30100.17	30551.99
	GPCM	30084.49	30188.12	30386.40	30838.22
	RSM	31466.34	31489.08	31532.61	31631.79
	PCM	414113.51	414217.14	414415.42	414867.24
37	GRM	26347.96	26441.49	26620.42	27028.16
	GPCM	26636.44	26729.96	26908.90	27316.64
	RSM	27953.35	27974.08	28013.73	28104.10
	PCM	414073.51	414167.03	414345.97	414753.71

Referring to data in Table 4, of the four methods, GRM still shows as the most fit method after subtracting the CLASS question item because it has the smallest AIC, SABIC, HQ, and BIC values. This is also in line with the analysis of the suitability of each item to the model based on Chi-Square values (X^2) presented in Table 5 below. The number of CLASS items declared most suitable for the model is the GRM method.

Table 5. The number of fit and unfit items based on Chi-Square value after dropping four items.

Method	The total number of Fit items	The total number of Unfit items
GRM	24	13
GPCM	23	14
RSM	10	27
PCM	23	14

The processed data after reducing special items, the comparison item suitability test, is presented in Table 6. Based on the data obtained, after the reduction of 4 items, the number of items that fit the model increased by about 30%, from 17 to 24 items. The increase in the number of items that fit goes hand in hand with the decrease in items that do not fit the model. This indicates that the GRM method is the model that best suits the data obtained.

Table 6. Comparison of the number of fit and non-fit items after the reduction

Number of Items	Number of Fit items	Number of items The unfit	Fit Item	Unfit item
41	17	24	1-2, 5, 12-13, 18, 22-23, 25, 27, 29, 33-38,	3-4, 6-11, 14-17, 19-21, 24, 26, 28, 30, 32, 36, 39, 41-42
37	24	13	1-7, 10-12, 15, 20, 22, 24, 26-27, 30, 32-34, 36-40	8,9,14, 16, 18, 19, 25, 28-29,35, 37, 41-42

Discussion

The results of data analysis obtained from the Indonesian version of the CLASS instrument through the application of the four IRT approach models (PCM, RSM, GPCM, and GRM) show that the most fit model is the Grade Response Model (GRM). The GRM suits items with a clear underlying response continuum (Desjardins & Bulut, 2018). GRM models are appropriate for use on items with a category response, such as the Likert scale (Widhiarso, 2010).

If considered in detail, the data generated either in the analysis of 41 items or after eliminating four items so that the CLASS becomes 37 items, the match test results obtained based on the values of AIC, SABIC, HQ, and BIC are obtained in a consistent order from smallest to largest respectively, namely GRM, GPCM, RSM and PCM. Furthermore, there is a relationship between match test data between GRM and GPCM, both through match parameters for all items and match tests for each item. Where the values obtained from the two methods are close to each other (with a tiny difference), GRM and GPCM methods are categorised as polytomous non-Rasch Models. They are polytomous forms of the two-parameter/2PL(Widhiarso, 2010). While the findings obtained in the suitability test analysis of each item appear that PCM and GPCM obtained the same results, especially after reducing 4 CLASS items where 23 items were found to be fit and 14 items were declared not fit with the model even though the value of the AIC, SABIC parameters, HQ, dan BIC These two models differ significantly. GPCM is a model developed by Muraki in 1992, which is a redevelopment of the PCM model (Widhiarso, 2010).

Referring to previous research by Jahanifar et al. [29], applying the polytomous IRT to the Persian version of the CLASS instrument did not specify the method used. However, it reported the study results that eight of the original version of the CLASS instrument conformed to the

Polytomous Item Response Theory model. The research was conducted by Christman et al. [30] through the application of Graded Item Response Theory on the CLASS instrument. The Graded-IRT method allows the exploration of respondents' learning attitudes into three groups, namely "not like an expert", "neutral", and "expert-like", which are all three denoted -1, 0, and 1, which are modelled in two probability functions expressed in θ.

The testing of the CLASS instrument, in addition to using the IRT approach, for example, by Douglas et al. (Douglas et al., 2014) through EFA and CFA analysis, obtained the results of research the CLASS instrument consists of three components, namely (1) Personal Application and Relation to Real World includes six items, (2) Problem Solving/Learning includes five items, and (3) Effort/Sense Making includes four items. Another study explains that some items in the factors identified by Douglas et al. have substantially different parameters of difficulty and discrimination than other items in factors, indicating that the subscale is not unidimensional (Embretson & Reise, 2000). This is in line with the initial design of the CLASS by Adams et al. (Adams et al., 2006). Furthermore, research reported by Kontro et al. (Kontro & Buschhüter, 2020) evaluated the CLASS instrument in the Finnish population with Confirmatory Factor Analysis (CFA) by following the findings of Douglas et al. (Douglas et al., 2014) so that three factors consisting of 14 items were obtained. These results were obtained after following up the findings by conducting expert interviews. One item removed from Douglas et al. (Douglas et al., 2014) was found in point 25. Both findings show that the CLASS instrument validation results are generally shorter than the original version.

As for tracing instrument items that are considered to require attention to be followed up, data items that are declared dropped from the data are obtained to obtain a suitable model, namely points 13, 17, 21, and 23 (redaction of these four items both in the original version and the Indonesian version are included in the appendix). The four items are dropped because they represent different indicators, and others still represent them. As for other items through follow-up of item fit test results through the GRM approach, some items are declared not fit with the model, namely items 8, 9, 14, 16, 18, 19, 25, 28, 29, 35, 37, 41, 42. Therefore, 17 CLASS items in the Indonesian version need to be explored further. The remaining 24 items are 1, 2, 3, 4, 5, 6, 7, 10, 11, 12, 15, 20, 22, 24, 26, 27, 30, 32, 33, 34, 36, 38, 39, and 40. Through factor analysis, Douglas et al. (Douglas et al., 2014) obtained three main components and proposed as many as 15 items from CLASS. In our findings, some items remain consistently considered items expressed according to the model. These items include items 3, 5, 22, 24, 30, 32, 34, 37, and 40. Therefore, we consider that most of the CLASS instrument items in the Indonesian version are conducive to further research of their use in the population in Indonesia. This is also in line with Kontro et al. (Kontro & Buschhüter, 2020), who stated that the results of the study showed that the interpretations made of the CLASS results were broadly usable and that CLASS remained a valuable instrument for various populations (Kontro & Buschhüter, 2020). From the various descriptions above, these studies show that the CLASS instrument could be researched continuously through various approaches, such as factor analysis and IRT approaches.

Conclusion

The Indonesian version of the CLASS instrument was evaluated using four Item Response Theory (IRT) models: Partial Credit Model (PCM), Rating Scale Model (RSM), Generalized Partial Credit Model (GPCM), and Graded Response Model (GRM). The Graded Response Model (GRM) is the most suitable among the four approach models. This provides empirical evidence that the model of each approach model (PCM, RSM, GRM, and GPCM) has its own characteristics. CLASS which is one of the instruments in the form of the Likert scale and was found to fit with GRM model. Furthermore, this study confirmed the consistency of certain items previously validated in other studies, showing strong validity. However, not all items in the

CLASS instrument are considered suitable according to the model, indicating the need for further research to explore alternative approaches in evaluating the Indonesian version of the CLASS instrument.

Acknowledgments

This article is one of the scientific papers compiled as a student (first author) and supported by funding by the government. Therefore we would like to express highest gratitude to The Ministry of Education, Culture, Research, and Technology (KEMENDIKBUDRISTEK), the Center for Higher Education Funding (BPPT), Education Fund Management Institution (LPDP) of the Republic of Indonesia, and Indonesian Education Scholarship (BPI) for funding in the first author's doctoral study and this research. In addition, we would like to thank the Department of Physics Education, University of West Sulawesi, Yogyakarta State University, Tadulako University, and Makassar State University.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

M is responsible for the overall content of the article including instrument preparation, data collection, data analysis, and reporting of research results. MH prepares research instruments, takes data, and provides suggestions and improvements to the article manuscript. S prepares research instruments, takes data, and provides suggestions and improvements to the article manuscript, especially in the research results and grammar in the article manuscript. BS prepares research instruments, takes data, and provides suggestions and improvements to the article manuscript. FAS concepts the content in the article manuscript.

References

A. Elby. (2002). *The Idea Behind EBAPS*. Http://Www2.Physics.Umd.Edu/~elby/EBAPS/Idea.Htm.

Adams, W. K. (2005). The Design and Validation of the Colorado Learning Attitudes about Science Survey. *AIP Conference Proceedings*, 45–48. https://doi.org/10.1063/1.2084697

Adams, W. K., Perkins, K. K., Podolefsky, N. S., Dubson, M., Finkelstein, N. D., & Wieman, C. E. (2006). New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey. *Physical Review Special Topics - Physics Education Research*, *2*(1), 010101. https://doi.org/10.1103/PhysRevSTPER.2.010101

Adams, W. K., Wieman, C. E., Perkins, K. K., & Barbera, J. (2008). Modifying and Validating the Colorado Learning Attitudes about Science Survey for Use in Chemistry. *Journal of Chemical Education*, *85*(10), 1435. https://doi.org/10.1021/ed085p1435

Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, *40*(4), 955–959.

Ajzen, I., & Fishbein, M. (1977). Attitude-behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, *84*(5), 888–918. https://doi.org/10.1037/0033-2909.84.5.888

Alhadlaq, H., Alshaya, F., Alabdulkareem, S., Perkins, K. K., Adams, W. K., Wieman, C. E., Sabella, M., Henderson, C., & Singh, C. (2009). *Measuring Students' Beliefs about Physics in Saudi Arabia*. 69–72. https://doi.org/10.1063/1.3266756

Astalini, A., Darmaji, D., Pathoni, H., Kurniawan, W., Jufrida, J., Kurniawan, D. A., & Perdan, R. (2019). Motivation and Attitude of Students on Physics Subject in the Middle School in Indonesia. *International Education Studies*, *12*(9), 15. https://doi.org/10.5539/ies.v12n9p15

Christman, E., Miller, P. M., & Stewart, J. (2020). Exploring the CLASS with Item Response Theory. *2020 Physics Education Research Conference Proceedings*, 81–86. https://doi.org/10.1119/perc.2020.pr.Christman

Corsiglia, G., Pollock, S., & Passante, G. (2023). Intuition in quantum mechanics: Student perspectives and expectations. *Physical Review Physics Education Research*, *19*(1), 010109. https://doi.org/10.1103/PhysRevPhysEducRes.19.010109

De Ayala, R. J. (2013). *The theory and practice of item response theory*. Guilford Publications.

de la Garza, J., Alarcon, H., Singh, C., Sabella, M., & Rebello, S. (2010). *Assessing Students' Attitudes In A College Physics Course In Mexico*. 129–132. https://doi.org/10.1063/1.3515178

Desjardins, C. D., & Bulut, O. (2018). *Handbook of educational measurement and psychometrics using R*. CRC Press.

Douglas, K. A., Yale, M. S., Bennett, D. E., Haugan, M. P., & Bryan, L. A. (2014). Evaluation of Colorado Learning Attitudes about Science Survey. *Physical Review Special Topics - Physics Education Research*, *10*(2), 020128. https://doi.org/10.1103/PhysRevSTPER.10.020128

Embretson, S. E., & Reise, P. E. (2000). Item response theory for psychologists. Mahwah (NJ): Lawrence Earlbaum Associates. *Inc.[Google Scholar]*.

Freed, R., McKinnon, D., Fitzgerald, M., & Norris, C. M. (2022). Development and validation of an astronomy self-efficacy instrument for understanding and doing. *Physical Review Physics Education Research*, *18*(1), 010117. https://doi.org/10.1103/PhysRevPhysEducRes.18.010117

Gray, K. E., Adams, W. K., Wieman, C. E., & Perkins, K. K. (2008). Students know what physicists believe, but they don't agree: A study using the CLASS survey. *Physical Review Special Topics - Physics Education Research*, *4*(2), 020106. https://doi.org/10.1103/PhysRevSTPER.4.020106

Halloun, I. (1997). Views About Science and physics achievement: The VASS story. *AIP Conference Proceedings*, 605–614. https://doi.org/10.1063/1.53156

Hauck Filho, N., Machado, W. de L., & Damásio, B. F. (2014). Effects of statistical models and items difficulties on making trait-level inferences: A simulation study. *Psicologia: Reflexão e Crítica*, *27*(4), 670–678. https://doi.org/10.1590/1678-7153.201427407

Hynninen, T., Pesonen, H., Lintu, O., & Paturi, P. (2023). Ongoing effects of pandemic-imposed learning environment disruption on student attitudes. *Physical Review Physics Education Research*, *19*(1), 010101. https://doi.org/10.1103/PhysRevPhysEducRes.19.010101

Jahanifar, M., & Derafsh, H. (2020). The Evaluation of Psychometric Characteristics of Persian Version of the Colorado Learning Attitudes about Science Using the Polytomous Item Response Theory Models. *Journal of Cognitive Psychology*, *8*(1), 1–26.

Kontro, I., & Buschhüter, D. (2020). Validity of Colorado Learning Attitudes about Science Survey for a high-achieving, Finnish population. *Physical Review Physics Education Research*, *16*(2), 020104. https://doi.org/10.1103/PhysRevPhysEducRes.16.020104

Kurniawan, D. A., Astalini, A., & Sari, D. K. (2019). An evaluation analysis of students' attitude towards physics learning at senior high school. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *23*(1), 26–35. https://doi.org/10.21831/pep.v23i1.20821

Martins, J. S., & Lindsay, W. E. (2022). Evaluation of high school student responses to the Colorado Learning Attitudes about Science Survey. *Physical Review Physics Education Research*, *18*(1), 010132. https://doi.org/10.1103/PhysRevPhysEducRes.18.010132

Muraki, E. (1992). A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement*, *16*(2), 159–176. https://doi.org/10.1177/014662169201600206

Paek, I., & Cole, K. (2019). *Using R for item response theory model applications*. Routledge.

Redish, E. F., Saul, J. M., & Steinberg, R. N. (1998). Student expectations in introductory physics. *American Journal of Physics*, *66*(3), 212–224. https://doi.org/10.1119/1.18847

Sakliressy, M. T., Sunarno, W., & Nurosyid, F. (2021). Students Scientific Attitude in Learning Physics Using Problem Based Learning Model with Experimental and Project Methods. *Jurnal Ilmiah Pendidikan Fisika Al-Biruni*, *10*(1), 59–70. https://doi.org/10.24042/jipfalbiruni.v10i1.8347

Sawtelle, V., Brewe, E., & Kramer, L. (2009). Validation study of the Colorado Learning Attitudes about Science Survey at a Hispanic-serving institution. *Physical Review Special Topics - Physics Education Research*, *5*(2), 023101. https://doi.org/10.1103/PhysRevSTPER.5.023101

Semsar, K., Knight, J. K., Birol, G., & Smith, M. K. (2011). The Colorado Learning Attitudes about Science Survey (CLASS) for Use in Biology. *CBE—Life Sciences Education*, *10*(3), 268–278. https://doi.org/10.1187/cbe.10-10-0133

Thacker, B. (2023). Inquiry-based experimental physics: Twenty years of an evidence-based, laboratory-based physics course for algebra-based physics students. *Physical Review Physics Education Research*, *19*(2), 020116. https://doi.org/10.1103/PhysRevPhysEducRes.19.020116

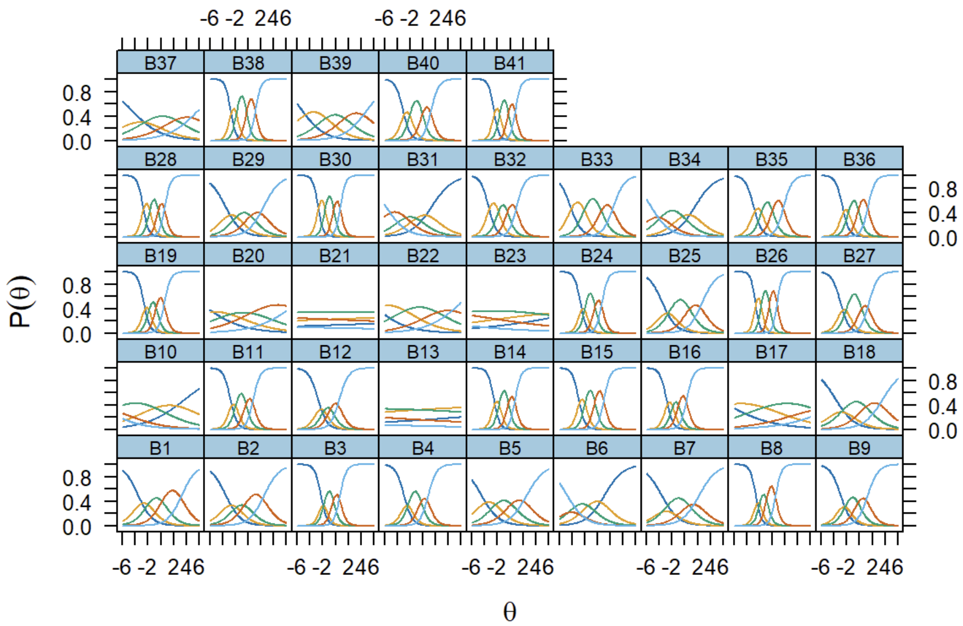
Widhiarso, W. (2010). Model politomi dalam teori respons butir. *Available at SSRN 2593459*.

Zhang, P., & Ding, L. (2013). Large-scale survey of Chinese precollege students' epistemological beliefs about physics: A progression or a regression? *Physical Review Special Topics - Physics Education Research*, *9*(1), 010110. https://doi.org/10.1103/PhysRevSTPER.9.010110

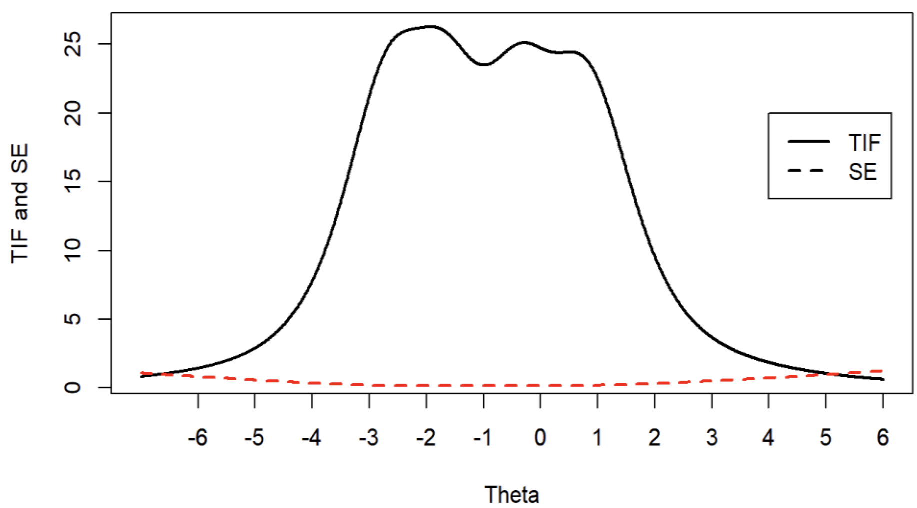
Appendix

In order to access the instrument in the original version (in English version) or in Indonesian form, we recommend that readers open via link:
https://www.physport.org/assessments/assessment.cfm?A=CLASS
Instruments can be opened after creating an account on the https://www.physport.org

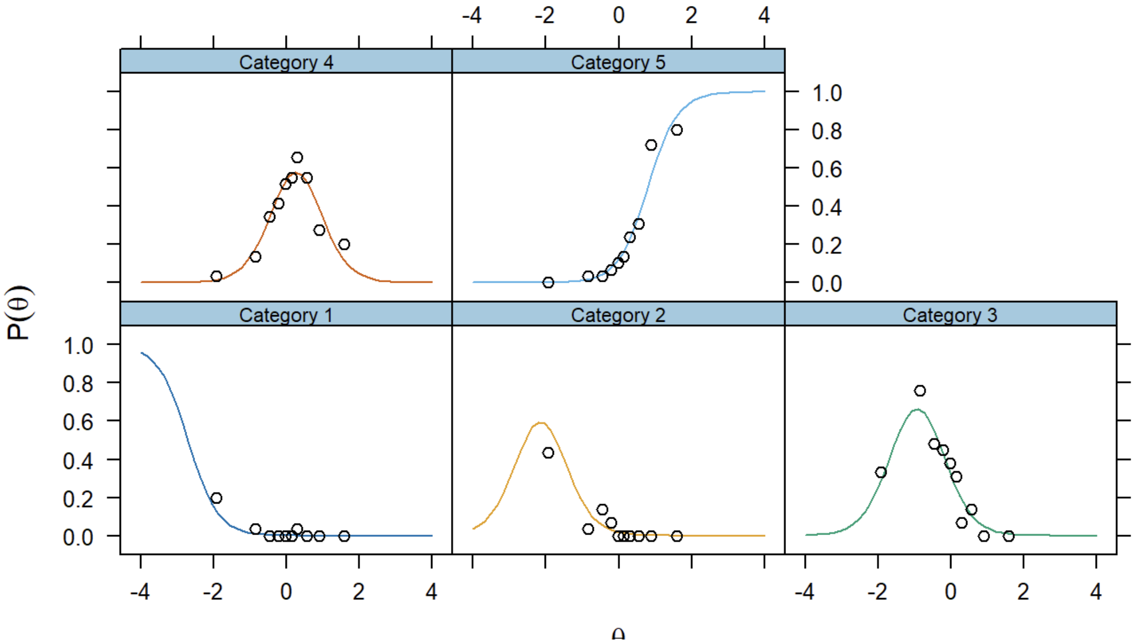
Other analysis results are presented in this section. Graded Response Model analysis data for 41 instrument items
Item Probability Function



Test Information Function and Standard Error of Measurement

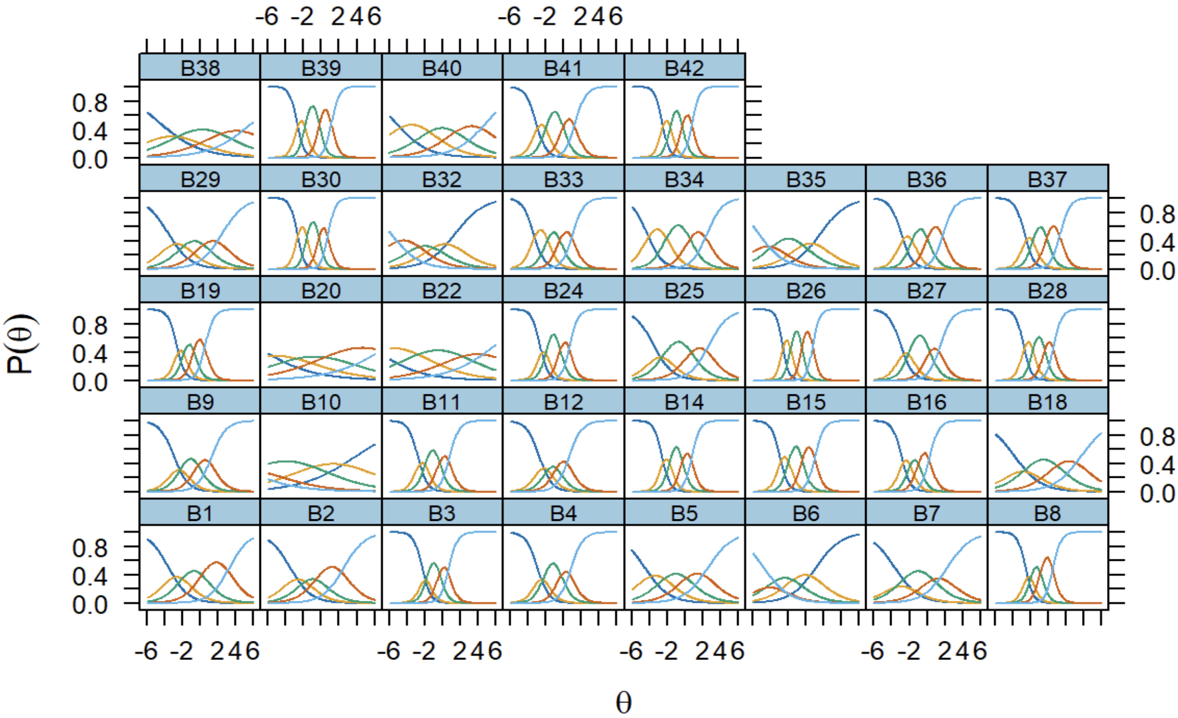


Empirical Plot for item 28

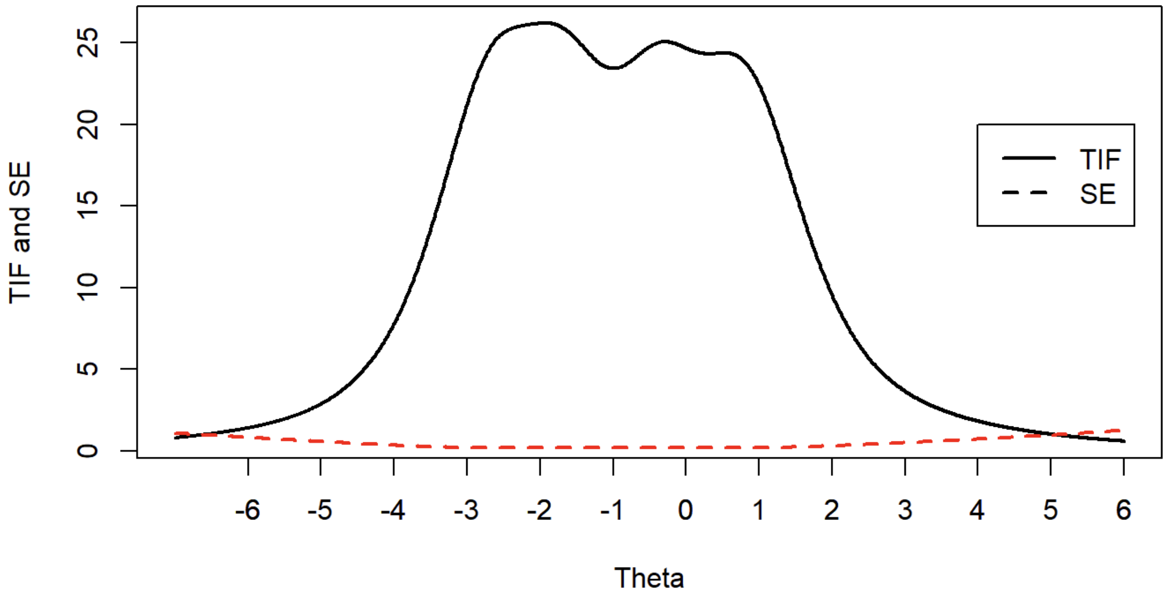


Graded Response Model analysis data for 37 instrument items

Item Probability Function



Test Information Function and Standard Error of Measurement



Empirical Plot for item 38

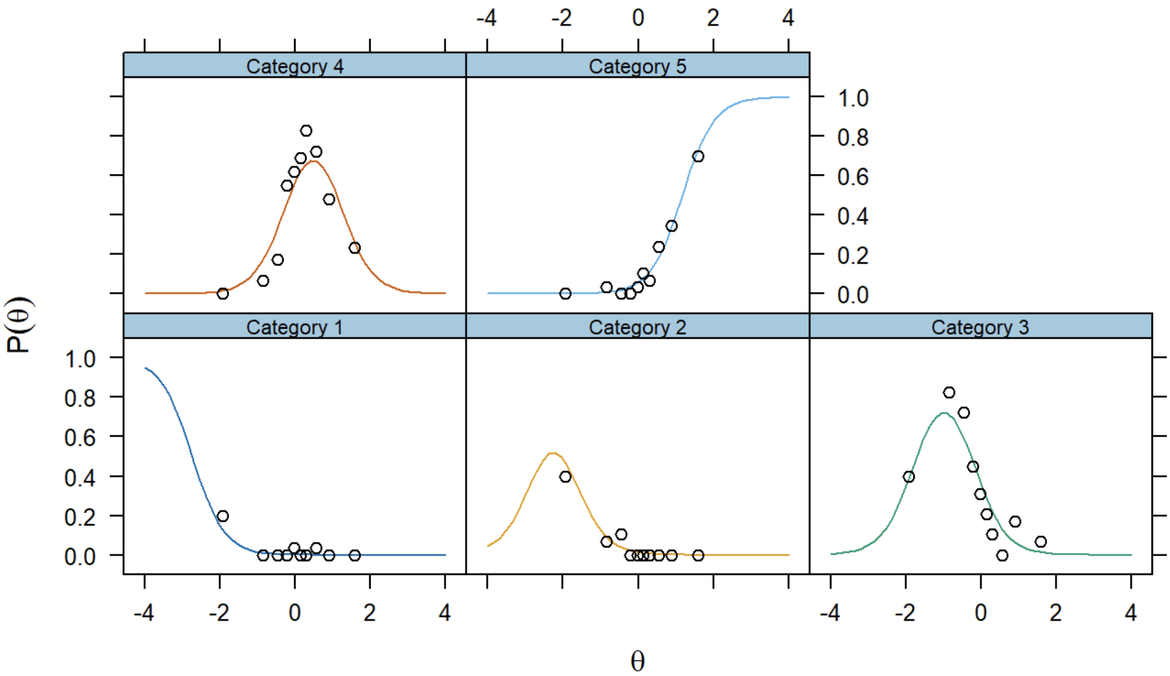


Table 7. Fit test per item based on Chi Square value PCM with 4 numbers dropped.

item <chr>	S_X2 <dbl>	df.S_X2 <dbl>	RMSEA.S_X2 <dbl>	p.S_X2 <dbl>	Note <chr>
B1	73.998	61	0.027	0.123	Fit
B2	93.857	76	0.028	0.081	Fit
B3	64.733	51	0.030	0.094	Fit
B4	64.192	58	0.019	0.269	Fit
B5	80.933	77	0.013	0.357	Fit
B6	88.418	70	0.030	0.068	Fit
B7	94.906	82	0.023	0.156	Fit
B8	72.205	46	0.044	0.008	Not Fit
B9	91.201	64	0.038	0.014	Not Fit
B10	96.557	75	0.031	0.048	Not Fit
B11	55.183	52	0.015	0.355	Fit
B12	69.746	69	0.006	0.452	Fit
B14	71.883	52	0.036	0.035	Not Fit
B15	50.657	50	0.007	0.447	Fit
B16	92.581	48	0.056	0.000	Not Fit
B18	102.885	71	0.039	0.008	Not Fit
B19	79.533	51	0.044	0.006	Not Fit
B20	104.000	73	0.038	0.010	Not Fit
B22	91.135	67	0.035	0.027	Not Fit
B24	45.966	50	0.000	0.636	Fit
B25	76.484	59	0.032	0.063	Fit
B26	46.259	48	0.000	0.544	Fit
B27	62.729	49	0.031	0.090	Fit
B28	77.786	51	0.042	0.009	Not Fit
B29	110.995	81	0.036	0.015	Not Fit
B30	59.825	50	0.026	0.161	Fit
B32	113.644	89	0.031	0.040	Not Fit
B33	48.149	56	0.000	0.763	Fit
B34	47.840	47	0.008	0.438	Fit
B35	93.760	78	0.026	0.108	Fit
B36	57.493	48	0.026	0.164	Fit
B37	67.201	49	0.036	0.043	Not Fit
B38	85.947	77	0.020	0.227	Fit
B39	52.672	43	0.028	0.148	Fit
B40	74.269	67	0.019	0.253	Fit
B41	69.699	49	0.038	0.028	Not Fit
B42	61.443	49	0.030	0.109	Fit

Table 8. Fit test per item based on Chi Square value RSM with 4 numbers dropped.

item <chr>	S_X2 <dbl>	df.S_X2 <dbl>	RMSEA.S_X2 <dbl>	p.S_X2 <dbl>	Note <chr>
B1	75.763	69	0.018	0.270	Fit
B2	92.457	69	0.034	0.031	Not Fit
B3	90.955	60	0.042	0.006	Not Fit
B4	68.650	59	0.024	0.183	Fit
B5	74.768	64	0.024	0.168	Fit
B6	169.350	64	0.075	0.000	Not Fit
B7	103.218	71	0.039	0.008	Not Fit
B8	111.364	61	0.053	0.000	Not Fit
B9	92.713	61	0.042	0.005	Not Fit
B10	147.857	74	0.059	0.000	Not Fit
B11	71.788	59	0.027	0.123	Fit
B12	82.303	60	0.036	0.030	Not Fit
B14	108.568	59	0.054	0.000	Not Fit
B15	68.738	59	0.024	0.181	Fit
B16	105.908	61	0.050	0.000	Not Fit
B18	86.821	62	0.037	0.020	Not Fit
B19	114.396	61	0.055	0.000	Not Fit
B20	108.935	63	0.050	0.000	Not Fit
B22	81.828	62	0.033	0.047	Not Fit
B24	71.681	59	0.027	0.124	Fit
B25	101.690	68	0.041	0.005	Not Fit
B26	82.508	59	0.037	0.023	Not Fit
B27	95.174	64	0.041	0.007	Not Fit
B28	104.277	59	0.051	0.000	Not Fit
B29	104.146	69	0.042	0.004	Not Fit
B30	91.480	59	0.043	0.004	Not Fit
B32	163.383	66	0.071	0.000	Not Fit
B33	68.129	59	0.023	0.195	Fit
B34	104.922	71	0.041	0.005	Not Fit
B35	168.528	71	0.069	0.000	Not Fit
B36	83.297	69	0.027	0.116	Fit
B37	90.582	62	0.040	0.010	Not Fit
B38	76.159	68	0.020	0.233	Fit
B39	87.464	60	0.040	0.012	Not Fit
B40	73.099	61	0.026	0.138	Fit
B41	80.656	61	0.033	0.047	Not Fit
B42	84.972	59	0.039	0.015	Not Fit

Table 9. Fit test per item based on Chi Square value GPCM with 4 items dropped.

item <chr>	S_X2 <dbl>	df.S_X2 <dbl>	RMSEA.S_X2 <dbl>	p.S_X2 <dbl>	Note <chr>
B1	73.998	61	0.027	0.123	Fit
B2	93.857	76	0.028	0.081	Fit
B3	64.733	51	0.030	0.094	Fit
B4	64.192	58	0.019	0.269	Fit
B5	80.933	77	0.013	0.357	Fit
B6	88.418	70	0.030	0.068	Fit
B7	94.906	82	0.023	0.156	Fit
B8	72.205	46	0.044	0.008	Not Fit
B9	91.201	64	0.038	0.014	Not Fit
B10	96.557	75	0.031	0.048	Not Fit
B11	55.183	52	0.015	0.355	Fit
B12	69.746	69	0.006	0.452	Fit
B14	71.883	52	0.036	0.035	Not Fit
B15	50.657	50	0.007	0.447	Fit
B16	92.581	48	0.056	0.000	Not Fit
B18	102.885	71	0.039	0.008	Not Fit
B19	79.533	51	0.044	0.006	Not Fit
B20	104.000	73	0.038	0.010	Not Fit
B22	91.135	67	0.035	0.027	Not Fit
B24	45.966	50	0.000	0.636	Fit
B25	76.484	59	0.032	0.063	Fit
B26	46.259	48	0.000	0.544	Fit
B27	62.729	49	0.031	0.090	Fit
B28	77.786	51	0.042	0.009	Not Fit
B29	110.995	81	0.036	0.015	Not Fit
B30	59.825	50	0.026	0.161	Fit
B32	113.644	89	0.031	0.040	Not Fit
B33	48.149	56	0.000	0.763	Fit
B34	47.840	47	0.008	0.438	Fit
B35	93.760	78	0.026	0.108	Fit
B36	57.493	48	0.026	0.164	Fit
B37	67.201	49	0.036	0.043	Not Fit
B38	85.947	77	0.020	0.227	Fit
B39	52.672	43	0.028	0.148	Fit
B40	74.269	67	0.019	0.253	Fit
B41	69.699	49	0.038	0.028	Not Fit
B42	61.443	49	0.030	0.109	Fit

Table 10. Fit test per item based on Chi Square value GRM with 4 items dropped.

item	S_X2	df.S_X2	RMSEA.S_X2	p.S_X2	Note
<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>
B1	73.767	61	0.027	0.126	Fit
B2	91.631	72	0.031	0.059	Fit
B3	68.523	51	0.034	0.051	Fit
B4	67.102	55	0.027	0.127	Fit
B5	80.781	78	0.011	0.392	Fit
B6	88.509	69	0.031	0.057	Fit
B7	101.645	81	0.030	0.060	Fit
B8	68.370	42	0.046	0.006	Not Fit
B9	81.117	60	0.035	0.036	Not Fit
B10	74.888	67	0.020	0.238	Fit
B11	57.922	51	0.022	0.235	Fit
B12	71.895	64	0.021	0.233	Fit
B14	73.249	50	0.040	0.018	Not Fit
B15	54.377	47	0.023	0.214	Fit
B16	84.784	43	0.058	0.000	Not Fit
B18	89.616	65	0.036	0.023	Not Fit
B19	71.294	49	0.040	0.020	Not Fit
B20	88.795	70	0.030	0.064	Fit
B22	92.355	73	0.030	0.063	Fit
B24	48.020	50	0.000	0.553	Fit
B25	74.994	52	0.039	0.020	Not Fit
B26	44.810	45	0.000	0.480	Fit
B27	60.435	48	0.030	0.107	Fit
B28	80.048	49	0.047	0.003	Not Fit
B29	108.335	80	0.035	0.019	Not Fit
B30	60.353	48	0.030	0.109	Fit
B32	101.181	86	0.025	0.126	Fit
B33	43.658	54	0.000	0.842	Fit
B34	46.998	45	0.012	0.391	Fit
B35	97.517	73	0.034	0.029	Not Fit
B36	54.581	46	0.025	0.181	Fit
B37	64.095	44	0.040	0.026	Not Fit
B38	88.919	79	0.021	0.209	Fit
B39	52.600	42	0.029	0.127	Fit
B40	71.060	66	0.016	0.313	Fit
B41	69.392	47	0.040	0.018	Not Fit
B42	65.771	48	0.036	0.045	Not Fit

The numbers was dropped from analysis

Item 13

Original Version

13. I do not expect physics equations to help my understanding of the ideas; they are just for doing calculations.

Strongly Disagree	1	2	3	4	5	Strongly Agree

Indonesian version

13. Saya merasa rumus-rumus tidak dapat membantu saya memahami konsep fisika. Rumus-rumus tersebut hanyalah alat untuk menghitung.

Sangat Tidak Setuju	1	2	3	4	5	Sangat Setuju

Item 17

Original Version

17. Understanding physics basically means being able to recall something you've read or been shown.

Strongly Disagree	1	2	3	4	5	Strongly Agree

Indonesian version

17. Memahami fisika hanyalah sekadar untuk mengingat materi yang telah kita pelajari sebelumnya.

Sangat Tidak Setuju	1	2	3	4	5	Sangat Setuju

Item 21

Original Version

21. If I don't remember a particular equation needed to solve a problem on an exam, there's nothing much I can do (legally!) to come up with it.

Strongly Disagree	1	2	3	4	5	Strongly Agree

Indonesian version

21. Jika saya tidak ingat rumus saat mengerjakan soal ujian fisika, saya tidak dapat berbuat apa pun untuk menyelesaikannya.

Sangat Tidak Setuju	1	2	3	4	5	Sangat Setuju

Item 23

Original Version

23. In doing a physics problem, if my calculation gives a result very different from what I'd expect, I'd trust the calculation rather than going back through the problem.

Strongly Disagree	1	2	3	4	5	Strongly Agree

Indonesian version

23. Dalam mengerjakan setiap soal Fisika, jika perhitungan saya memberikan hasil yang sangat berbeda dari apa yang saya harapkan, saya akan mempercayai perhitungannya daripada kembali meninjau soal.

Sangat Tidak Setuju	1	2	3	4	5	Sangat Setuju