International Journal of Innovative Research and Scientific Studies, 8(2) 2025, pages: 1672-1683

IJIRSS
International Journal of Innovative Research & Scientific Studies

ISSN: 2617-6548

URL: www.ijirss.com

Assessing the validity of the Pancasila Student Character Learning Environment Inventory instrument

Achmad Ridwan[1*], Indra Darmawan[2], Riyadi[3], Yuli Rahmawati[4], Soleh[5]

[1,2,3,4]*Universitas Negeri Jakarta, Indonesia.*
[5]*La Tansa Mashiro University Jl. Soekarno - Hatta, Rangkasbitung, Lebak Regency, Banten 42317, Indonesia.*

Corresponding author: Achmad Ridwan (*Email: AchmadRidwan2024@proton.me*)

Abstract

The purpose of this study was to test the Pancasila Student Character Learning Environment Inventory Instrument for high school students using the Multi-Item Response Theory (MIRT) as a measurement model. An analysis was conducted to find proof of validity and reliability, as well as the level of difficulty of the items and any signs of item bias. There was a pilot study with 330 students from 7 high schools (4 public high schools and 3 private high schools) to try the instruments. The MIRT model analysis shows that 6 of the 106 items have a negative total correlation of items, and 1 item has a different item function (DIF) that needs to be calculated and is not part of the second empirical test. The Multidimensional Graded Response Model (MGRM), the Multidimensional Partial Credit Model (MPCM), and the Multidimensional Generalized Partial Credit Model (MGPCM) were all tested. The Pancasila Student Character Learning Environment Inventory tool passed the psychometric testing that was conducted using MIRT model analysis. The tests were meant to show that the scale was accurate and reliable, to check for multidimensionality, to examine how difficult the items were and how well the respondents performed, and to identify any item bias (DIF).

Keywords: Inventory instrument, Learning environment, Pancasila student character.

DOI: 10.53894/ijirss.v8i2.5527
Funding: This study received no specific financial support.
History: Received: 28 January 2025 / **Revised:** 3 March 2025 / **Accepted:** 10 March 2025 / **Published:** 20 March 2025

Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Publisher: Innovative Research Publishing

1. Introduction

Many schools are looking for strategies to help students develop their competence and character in light of the growing demand for people with strong 21st-century skills in society [1]. Structured to accommodate a variety of learning styles and a less prescriptive role for teachers than in more conventional classrooms, an inclusive learning environment promotes the development of students' competence and character [2]. In an ideal classroom, students understand the significance of creative

learning for their personal growth and professional success, the school as a whole has clear goals for student achievement, and teachers work together to achieve these goals [3]. These classrooms connect what students learn in the classroom with what they can apply in the real world through the use of realistic assignments that make learning more meaningful to each student [4].

According to Davies, et al. [5], students gain knowledge and engage in the learning process in a classroom setting. According to Beghetto and Kaufman [6], the development of students' creative potential is significantly influenced by the learning environment. Several studies have shown that the learning environment greatly influences students' motivation, academic abilities, cognition, attitudes, and achievement [7-12]. Assuming the learning process is well-designed and executed, students' attitudes and abilities will improve [13].

In accordance with the principles of Pancasila, Indonesian students are perceived as competent, lifelong learners whose strong character is in harmony with these principles. With this declaration, the Pancasila Student Profile is summed up. Those students who have mastered all six components of this profile will graduate. The following dimensions are easy for teachers and students in Indonesia to remember: 1) having faith, being devoted to God Almighty, and having noble character; 2) encouraging independence; 3) collaborating; 4) celebrating diversity on a global scale; 5) reasoning critically; and 6) encouraging creativity. The Republic of Indonesia's Ministry of Education and Culture has a policy that is founded on the Pancasila framework. All parties involved, including students and educators, can have a clear picture of where they are going and how they are getting there when they follow the guidelines laid out in the Pancasila Student Profile.

Competencies that are pertinent to 21st-century skills include attitudes and abilities. The 'Four Cs' of critical thinking, communication, cooperation, and creativity-innovation are the kinds of 21st-century talents that certain nations' education programs aim to foster [14]. Attitude and skill terminology have long been thought of as a kind of intricate mental operation. Learning environment instruments were initially created by Anderson, et al. [15] and then refined by Fraser, et al. [16] to be used for evaluating learning settings. First established in 1974 as a CES manual, Trickett and Moos [17] used CES to evaluate the atmosphere in middle and high school classrooms. Several other instruments were created by different researchers; these include the following: the Constructivist Learning Environment Survey (CLES) [18], the Science Laboratory Environment Inventory (SLEI) [19], the Geography Classroom Environment Inventory (GCEI) [20], the Chemistry Laboratory Environment Inventory (CLEI) [21, 22], the Clinical Learning Environment Inventory (CLEI) [23], the Chinese Language Classroom Environment Inventory (CLCEI) [24], and the Elementary School Science Classroom Environment Scales (ESSCES) [25].

This study expands upon previous work in the field of learning environments that has shown that students' views are the most important social and psychological elements affecting their academic performance [26]. The study's proponents argue that students' impressions of their classroom social environment can be used to make predictively valid judgments about that environment and that kids are capable of perceiving and assessing inputs. Students' emotional and intellectual outcomes are predicted by their impressions of the classroom learning environment, according to a large body of research spanning decades [16].

The MIRT MGRM measuring methodology was utilized for the analysis of the Pancasila Student Character Learning Environment Inventory Instrument (PCLEII) for high school students. One up-to-date method for ensuring precise measurements is the MIRT MGRM model, which is based on test theory by Boone and Noltemeyer [27]. Among the many benefits of the MIRT model for measurement are the following: the ability to handle missing data, the provision of tools to detect discrepancies, the ability to measure object parameters independently of measurement instruments, and the provision of latent variable measurement units in logit units [28]. Data from rating scales and Likert scales, which are examples of polytomous test results, can be analyzed using MIRT modeling [29]. Analyzing the compatibility of individual and item data with the MIRT model forms the basis of instrument analysis in the model. This study set out to determine whether or not the PCLEII was valid and reliable when administered to high school students. Item characteristics, including difficulty level and bias, as well as indications of multidimensional adequacy, were also retrieved using instrument analysis.

2. Research Methods

2.1. Research design

In order to assess the learning environment and character of Pancasila pupils, research into the creation of an inventory instrument was conducted in stages. A group of experts in education and related fields reviewed and approved the inventory instrument. Next, we conducted an instrument test to empirically validate the content [30, 31]. Using the MIRT model, we checked the inventory's instrument test results for item bias, difficulty levels, multidimensional indicators, and measurement validity and reliability. The results of the measurements taken using the MIRT model were described using a descriptive quantitative analysis.

2.2. Pilot Study Sample

Seven high schools participated in the instrument trials, and 330 students served as the pilot study's sample. Out of the total number of students, 178 attended public high schools, making up 54% of the total, while 152 attended private high schools, making up 46%.

2.3. Instrument

The PCLEII in the following table was tested and validated by experts and a panel:

Table 1.
Factors, indicators, and number of items (attitude scale).

No	Factors	Indicators	Favorable	Unfavorable	Total
1	Have faith, be devoted to God Almighty, and have a noble character.	The extent to which students believe that religion can regulate life.	1,3	2	3
		The extent to which students accept themselves as a gift from God.	4,6	5	3
		The extent to which students accept other people as God's creation.	7,8,9	10	4
		The extent to which students appreciate the universe as God's creation.	11,12		2
		The extent to which students feel that obeying state regulations is a part of their devotion to Allah.	14,15	13	3
2	Global Diversity	The extent to which students can recognize and appreciate culture.	16,17,18	19	4
		The extent to which students can communicate and interact with various cultures.	20,21		2
		The extent to which students can reflect on and take responsibility for their experiences of diversity.	22,24	23	3
		The extent to which students recognize social justice in everyday life.	25,26	27	3
3	Worked together	The extent to which students become actively involved in completing group assignments.	28,29		2
		The extent to which students are able to share.	30,31,32		3
4	Independent	The extent to which students can control themselves in any situation.	33,34,35		3
		The extent to which students can self-regulate.	36,37		2
5	Critical Reasoning	The extent to which students can process information so that new ideas emerge.	38,39,40		3
		The extent to which students can analyze and evaluate their reasoning.	41,42,44	43	4
		The extent to which students can reflect on and evaluate their own thinking.	45,46		2
6	Creative	The extent to which students can produce original ideas.	48	47	2
		The extent to which students can produce original work and actions.	49,50	51	3
		The extent to which students demonstrate flexibility in thinking to find alternative solutions to problems.	52,53		2
		Total	43	10	53

Table 2.
Factors, indicators, and number of items (behavior scale).

No	Factors	Indicators	Favorable	Unfavorable	Total
1	Have faith, be devoted to God Almighty, and have a noble character.	The extent to which students believe that religion can regulate life.	1,3	2	3
		The extent to which students accept themselves as a gift from God.	4,6	5	3
		The extent to which students accept other people as God's creation.	7,8,9	10	4
		The extent to which students appreciate the universe as God's creation.	11,12		2
		The extent to which students feel that obeying state regulations is a part of their devotion to Allah.	14,15	13	3
2	Global Diversity	The extent to which students can recognize and appreciate culture.	16,17,18	19	4
		The extent to which students can communicate and interact with various cultures.	20,21		2
		The extent to which students can reflect on and take responsibility for their experiences of diversity.	22,24	23	3
		The extent to which students can recognize social justice in everyday life.	25,26	27	3
3	Worked together	The extent to which students become actively involved in completing group assignments.	28,29		2
		The extent to which students are able to share.	30,31,32		3
4	Independent	The extent to which students can control themselves in any situation.	33,34,35		3
		The extent to which students can self-regulate.	36,37		2
5	Critical Reasoning	The extent to which students can process information so that new ideas emerge.	38,39,40		3
		The extent to which students can analyze and evaluate their reasoning.	41,42,44	43	4
		The extent to which students can reflect on and evaluate their own thinking.	45,46		2
6	Creative	The extent to which students can produce original ideas.	48	47	2
		The extent to which students can produce original work and actions.	49,50	51	3
		The extent to which students demonstrate flexibility in thinking to find alternative solutions to problems.	52,53		2
		Total	43	10	53

The PCLEII was designed in the form of a frequency scale type assessment scale [32]. The response options on the frequency scale included five choices for the attitude scale: strongly agree (SA), agree (A), doubtful (D), disagree (DA), and strongly disagree (SDA). The SA attitude scale used five codes: A with code 4, D with code 3, DA with code 2, and SDA with code 1. The behavior scale indicated always (AL), often (OF), sometimes (ST), never (NV), and never at all (NVA). Frequency scale coding was used with the following conditions: AL with code 5, OF with code 4, ST with code 3, NV with code 2, and NVA with code 1. This frequency scale was used by the participants in response to each item of the Pancasila

student character learning environment inventory. Participants responded according to their perceptions and experiences of learning in class.

2.4. Data Analysis

The data analysis used the MIRT model for measurement analysis, facilitated by the R program. Validation using the MIRT model was carried out to obtain samples/respondents and inventory items that align with the model's requirements. The validation process commenced with respondent analysis to obtain a suitable sample, followed by an assessment of the suitability of inventory items.

Measurement analysis using the MIRT model was conducted to obtain empirical evidence regarding the validity and reliability of the inventory. Additionally, an assessment of multidimensional adequacy was carried out to demonstrate the instruments' multidimensionality. The analysis process culminated with an examination of item difficulty levels and item biases.

3. Results and Discussion

3.1. Item Validity

The trial results of the Pancasila student character learning environment inventory instrument obtained 99 valid items out of the 99 items tested in the inventory. The analysis of item validity in the MIRT model relies on how well the items fit the model [33]. Below is a compilation of the valid items derived from the item suitability test results using the MIRT MGRM model analysis.

Table 3, reveals that all items have infit values ranging from 0.5 logit to 1.5 logit, indicating conformity with the MIRT MGRM model. The assessment of item fit through MIRT model analysis uses the Infit value measure [29, 34-36] with the use of ZSTD standard scores recommended by Linacre [36] for small sample sizes, typically between 30 and 300 [29]. The standard MNSQ value for testing item suitability falls within the range 0.5<MNSQ<1.5. As per Linacre's classification, MNSQ values within 0.5 – 1.5 signify productive measurement, MNSQ>1.5 suggests poor measurement, while MNSQ<0.5 indicates less productive measurement [37]. Inventory items conforming to the MIRT MGRM model (as shown in Table 3) have MNSQ Infit values ranging from 0.753 to 1.105 logit. A total of 99 inventory items were considered productive or effective in measuring the learning environment of Pancasila student character based on their MNSQ infit values.

Table 3.
Valid items from suitability test results.

items	infit MGRM	items	infit MGRM	items	infit MGRM
B1	1.008	B42	0.944	B50	0.999
B2	0.887	B56	1.015	B51	0.989
B3	0.975	B58	1.015	B52	0.977
B4	0.900	B59	1.020	B53	0.971
B5	1.018	B65	0.973	B54	0.981
B6	0.874	B66	1.004	B60	1.007
B7	0.946	B67	0.997	B64	0.996
B8	0.881	B69	0.987	B70	0.921
B9	0.861	B75	0.997	B71	1.105
B10	0.797	B76	1.004	B72	0.946
B11	1.192	B77	0.987	B73	0.961
B12	0.988	B78	0.988	B74	0.917
B13	0.964	B79	1.020	B84	1.033
B14	0.753	B80	1.047	B85	1.033
B15	0.833	B81	1.002	B86	1.019
B31	1.025	B82	1.020	B87	1.022
B32	0.985	B83	1.002	B88	1.010
B33	1.048	B93	0.997	B89	0.976
B34	0.972	B94	0.965	B90	1.020
B35	1.032	B95	0.982	B91	0.952
B36	1.000	B96	1.000	B92	0.991
B37	1.085	B97	0.989	B100	0.997
B38	1.047	B98	1.017	B101	0.979
B39	1.000	B99	0.999	B102	0.990
B40	1.050	B16	0.914	B103	0.971
B41	0.987	B17	0.932	B104	0.996

3.2. Measurement Reliability

The concept of reliability in the context of the MIRT model is summarized by the stratified alpha coefficient (stratified alpha), as introduced by Cronbach, et al. [38]. This coefficient is a valuable tool for gauging the reliability of instruments comprising several subtests. Similar to the alpha coefficient, the stratified alpha coefficient assesses internal consistency

among the various components of a test. It is particularly well-suited for evaluating the reliability of multidimensional composite scores. Presented below are the outcomes of the stratified alpha calculation:

Table 4.
Descriptive statistics of measurement results.

Factors	Variance	Reliability of each Factors
Attitude	462.78	0.88
Behavior	474.33	0.87
Total score	1733.65	

Based on the calculations, a stratified alpha of 0.93 can be obtained, indicating that the instrument's reliability is highly satisfactory. A reliability coefficient of 0.93 implies that 93% of the variance observed in individuals (students) stems from systematic measurements, while 7% is due to measurement error. This measurement reliability value exceeds the minimum alpha standard of 0.80, which is particularly noteworthy for newly developed instruments [39]. Measurement reliability shows the consistency and dependability of the instrument in providing measurement data. The instruments consistently provide the same or equivalent information. Consequently, judging from the reliability index, the quality of inventory measurement demonstrates relatively strong consistency [40]. When applied to different groups or individuals, the inventory instrument is consistent in measurement reproduction.

3.3. Item Information Function

In the measurement of attitude scales, the item that provided the most information up to 6.5 logits was statement item number 8, which states: "Every person is a noble creature created by God, so I should not insult or look down on anyone." This statement contributed to assessing the extent to which students can accept other people as creations of God. It also relates to the dimension of faith, devotion to God Almighty, and possessing noble character.

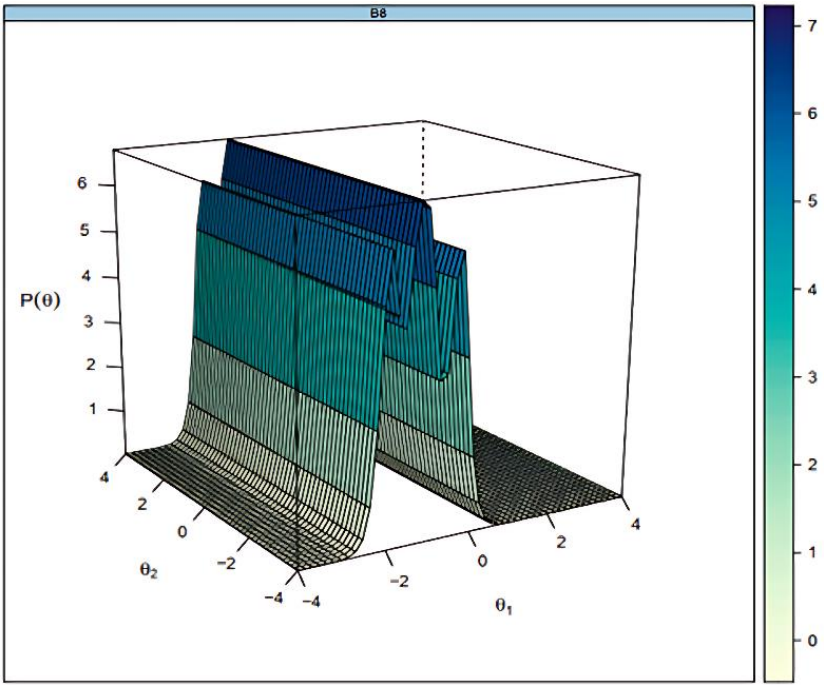

Figure 1.
Information function item no 8.

Figure 1 illustrates the information function of item number 8 on the Pancasila student character learning environment instrument for the attitude scale. The scaling model used is the MGRM polytomous. The maximum information function value attained is 6.5 logits, occurring at an ability (Theta) -1. A higher information value of an item indicates greater reliability. Item number 8 is considered suitable for respondents with abilities ranging between -3 and 0.5 logits.

In the measurement of behavioral scales, the item providing the highest information up to 0.9 logit was statement item number 49, which states: "I really love domestic products." This statement contributes to assessing the extent to which students can reflect on and take responsibility for their experiences. It is also part of the global diversity dimension.

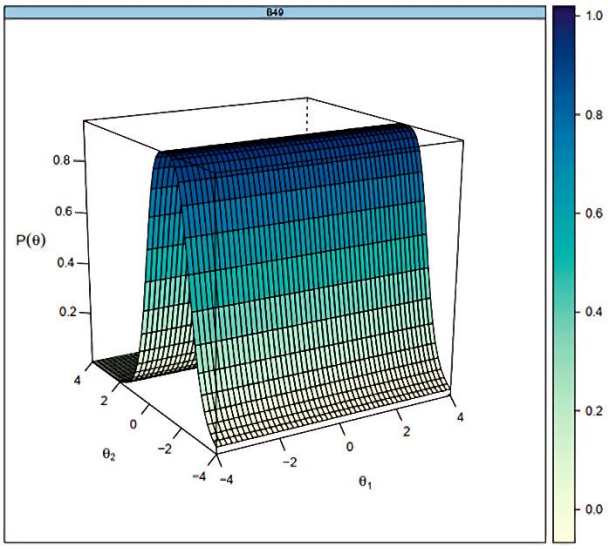

Figure 2.
Information function item no 49.

Figure 2 displays the information function of item number 49 in the Pancasila student character learning environment instrument for the behavior scale. The scaling model used is the MGRM polytomous. The maximum value of the information function obtained is 0.9 logit, occurring at an ability (Theta) of 0. The higher the value of item information, the more reliable the item is considered. Item 49 is deemed appropriate for respondents with abilities ranging between -3 to 1.5 logits.

3.4. Instrument Information Function

From the information function depicted in Figure 3, it can be interpreted that the item information level is highest when both latent traits are around zero. Conversely, the item information level is lowest when both latent traits are either very low or very high (i.e., at 0 -4 and 0 4).

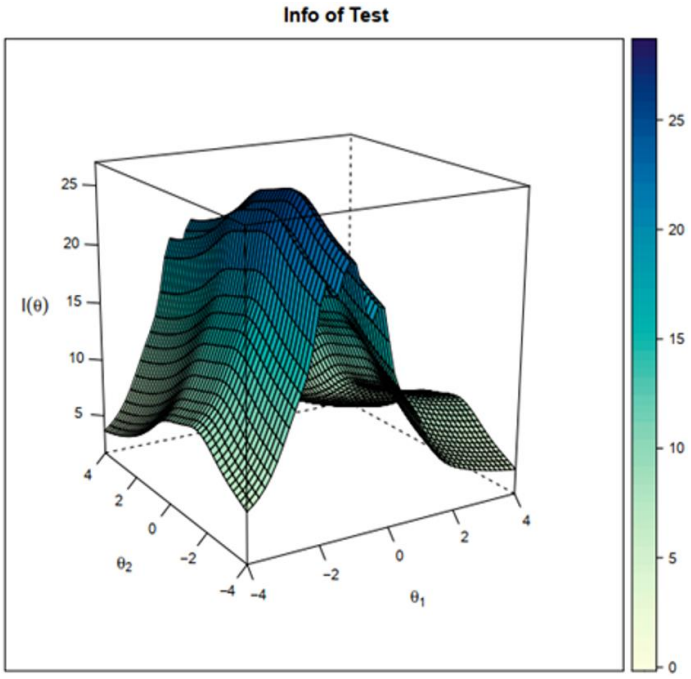

Figure 3.
Instrument information function.

3.5. Standard Error

From the standard error depicted in Figure 4, it illustrates the relationship between the respondent's ability and the available instruments. Visually, it is evident that respondents with abilities between -2 and 2 produced the lowest standard error in both dimension 1 and dimension 2 (indicated by ɵ1 and ɵ2). This indicates that this instrument is suitable for application with respondents with abilities ranging from -2 to 2, representing a medium level of ability.

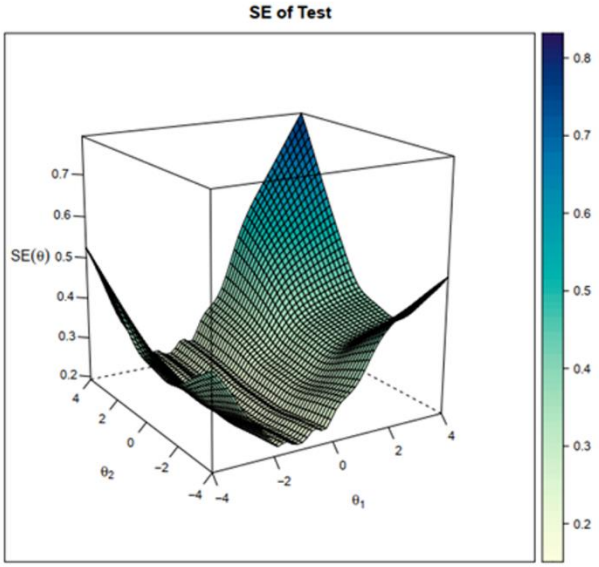

Figure 4.
Standard error.

3.6. Dimensionality

Multidimensional adequacy is a fundamental necessity when measuring latent variables or constructs within the MIRT model [41]. The test for Multidimensionality was based on the MIRT model. If the test for dimensionality indicates that if more than one dimension is involved, then the instrument's parameter estimates will be analyzed using the Multidimensional IRT (MIRT) approach. Testing the dimensions of the instrument device employs principal component analysis, taking into consideration the eigenvalue and total variance as depicted in Table 5.

Table 5.
Dimensionality Testing Trials.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative%	Total	% of Variance	Cumulative %
1	17.964	16.947	16.947	17.964	16,947	16.947
2	14.277	10.639	27.586	14.277	10.639	27.586
4	9.043	8.531	36.117	9.043	8.531	36.117
4	7.396	6.977	43.094	7.396	6.977	43.094
5	6.478	6.111	49.205	6.478	6.111	49.205
6	$0.23	4.934	54,139	5.230	4.934	54,139
7	4.389	4.141	58.26	4.389	4.141	58.2680
a	4.351	4.104	62.384	4.351	4.104	2.384
9	4.136	3.901	66.286	4.136	3.901	66.286
10	3.669	3.462	69.746	3.669	3.462	69.746
17	3.02	2.0508	72.604	3.020	2.0508	72.604
12	2.836	2.676	75.260	2.836	2.676	75.280
13	2.684	2.532	77.812	2.684	2.532	77.812
14	2.668	2.423	80.236	2.668	2.423	80.2365
15	2.465	2.311	82.646	2.4650	2.391	82.646
16	2.199	2.075	84.620	2.199	2.075	84.620
7	2.047	1.931	86.551	2.047	1.931	86.551
18	1.787	1.685	88.237	1.787	1.685	88.237
19	1.718	1.62	89.867	1.718	1.620	89.857
20	1.698	1.564	91.421	1.698	1.664	91.421
21	1.57	1.401	92.903	1.570	1.401	92.903
22	1.446	1.364	94.267	1.446	1.364	94.267
23	1.282	1.209	95.476	1.282	1.209	95.476
a4	1.193	1.426	96.601	4.193	1.426	96.601
25	991	936	97.6236			
26	842	794	98.330			
27	759	716	99.046			
28	510	481	99.527			

According to the data presented in Table 5, the ratio of eigenvalues between the first factor and second factors is quite small, indicating a lack of relative difference between them. Additionally, the total variance is less than 20%. These two indicators suggest that the instrument is multidimensional with several dominant factors present. Consequently, researchers opted to use the MIRT model to estimate the instrument parameters, allocating items to their respective domains according to the instrument specifications.

Within the MIRT framework, the analysis of dimensionality aims to ascertain whether the inventory items measure distinct dimensions or a single construct [39]. A sufficient level of dimensionality is essential for an instrument, as items failing to measure the same dimension or construct render it inadequate [40]. Raw variance explained and eigenvalue (λ) serve as indicators of unidimensional adequacy. If the eigenvalue is closely aligned with the size predicted by the MIRT model, then it suggests multidimensional sufficiency in the data [36].

3.7. Multidimensional Difficulty

The polytomous scoring system not only produces a difficulty index for each item but also enables an estimate of the level of difficulty for each category within each item. The item difficulty index is calculated as the average of the category thresholds, while the category difficulty index represents the threshold value for each score [42].

The analysis of the multidimensional difficulty levels of the Pancasila student character learning environment inventory has an average difficulty level index of -2.07, indicating an easy level of difficulty overall. In item response theory (IRT), item difficulty ranges from, $-\infty \leq b \leq \infty$. Research findings suggest that items with difficulty levels falling within the range of $-2 \leq b \leq +2$ are considered good items [43].

Furthermore, the average difficulty level index for measuring attitudes is -2.72 while for measuring behavior it is -1.40. This indicates that the statements assessing attitudes are easier to agree with compared to those measuring behavior.

3.8. Person Ability

In practical terms, a person's ability is typically assessed on a scale ranging from -3 to 3. The closer the score is to -3, the lower the ability is considered, whereas, the closer it is to 3, the higher the ability is classified [42]. According to the IRT approach, the ability or ability item parameter (θ) lies within the interval $-\infty \leq \theta \leq \infty$, and is standardized to approximate normal distribution with a mean of 0 and a standard deviation of 1 [43]. Research findings indicate that an individual's abilities generally range from $-3 \leq \theta \leq 3$ [43].

One advantage of employing MIRT in item analysis lies in the accuracy of estimating abilities by analyzing each ability for each ability being measured separately. In this case, two distinct abilities were being assessed: attitudes and behavior [42]. The analysis revealed that, on average, the test participants demonstrated moderate abilities, falling within the range of -3 and 3. None of the participants exhibited abilities exceeding 3 or falling below -3.

The results indicate that the ability estimates for each skill fall within the medium category. However, it is worth noting that the lowest ability score in the attitude domain was -1.55, which is slightly lower than the lowest score in the behavioral domain, which stands at -1.49. Despite the average attitude ability being lower, there are respondents who exhibited notably high abilities with a score as high as 2.42.

Based on the estimated proficiency levels for each ability, the respondent's abilities were derived from the average proficiency of the two abilities. The distribution of respondents' abilities is considered relatively ideal, as the majority exhibit moderate proficiency. These findings suggest that respondents possess favorable ability characteristics as depicted in Figure 5:

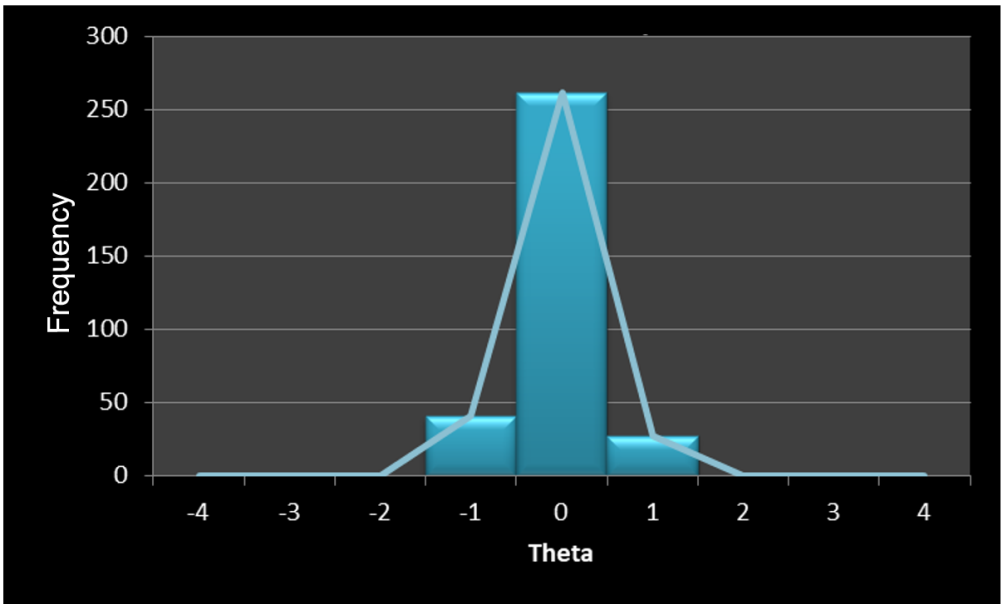

Figure 5.
Distribution of respondents' theta.

The distribution of respondents' abilities is depicted in the histogram graph, which exhibits a relatively bell-shaped curve without significant skewness to the left or right. This indicates that the distribution of respondents' abilities follows a normal distribution range. From the histogram graph, it is visually evident that most respondents have abilities centered around -1 and +1.

3.9. Person Fit

Based on the individual's fit within the model, we can infer that a low Zh person fit (for example < -2) suggests a likelihood that a deviant response pattern will be exhibited. Similarly, a high Zh person fit (for example > 2) also indicates a potential for a deviant response pattern. Participants falling into the category of Zh < -2 and Zh > 2 are those diverge significantly from the expected pattern within the the MGRM model.

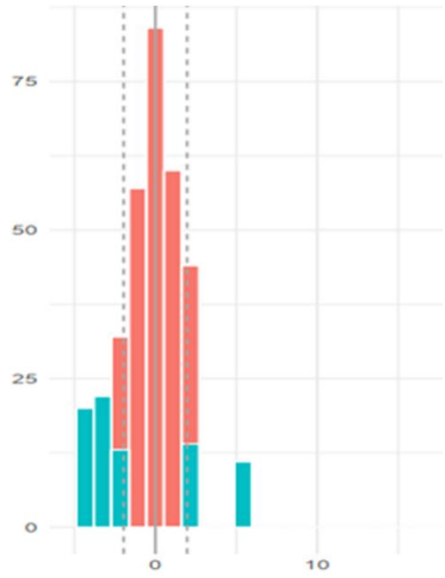

Figure 6.
Person Fit.

4. Conclusion

This research has successfully conducted psychometric testing on the Pancasila student character learning environment inventory instrument using MIRT model analysis. The testing aimed to establish evidence of the validity and reliability, assess multidimensionality, analyze item difficulty, and respondent ability, and identify item bias (DIF). Through matching tests, 99 items were confirmed to be valid based on MNSQ scores. The obtained MNSQ Infit and MNSQ Outfit values fell within the productive range of 0.5 – 1.5 logits, indicating satisfactory measurement characteristics. Additionally, the stratified alpha reliability value was calculated to be 0.93, indicating a high level of consistency in the measurement results and falling into the 'good' category, surpassing the minimum stratified alpha size of 0.80.

Multi IRT can recognize this instrument because it contains more than one factor, as evidenced by the ratio of the eigenvalue of the first factor to the second factor being very small; in fact, it is not relatively much different. The total value of the first factor's variance is also less than 20%. Of the 99 items analyzed, all had infit values between 0.5 logit and 1.5 logit. It also has a stratified Alpha coefficient reliability value of 0.94, indicating very high reliability.

In the first empirical test, the results of the item analysis revealed that the highest information function value in the attitude scale measurement is found in item 8, with a value of 6.5 logits, while in the behavior scale measurement, it is item 49 with a value of 0.9 logits.

The inventory instrument meets the requirements for multidimensionality adequacy. It is proven that the eigenvalue ratio of the first factor to the second factor is very small; in fact, it is not relatively different. The total value of the first factor's variance is also less than 20%. Based on these two indicators, it can be said that there are several dominant factors in measuring the instrument, making it multi-dimensional.

Furthermore, the results of the MIRT model analysis also show that the inventory instruments are contaminated with DIF. The number of items indicated by DIF was only one item, namely item 68.

References

[1] A. Ovbiagbonhia, B. Kollöffel, and P. d. Brok, "Educating for innovation: Students' perceptions of the learning environment and of their own innovation competence," *Learning Environments Research*, vol. 22, pp. 387-407, 2019. https://doi.org/10.1007/s10984-019-09280-3

[2] D. Fisher, P. Den Brok, B. Waldrip, and J. Dorman, "Interpersonal behaviour styles of primary education teachers during science lessons," *Learning Environments Research*, vol. 14, pp. 187-204, 2011. https://doi.org/10.1007/s10984-011-9093-z

[3] Z. Misbah, J. Gulikers, W. Widhiarso, and M. Mulder, "Exploring connections between teacher interpersonal behaviour, student motivation and competency level in competence-based learning environments," *Learning Environments Research*, vol. 25, no. 3, pp. 641–661, 2022. https://doi.org/10.1007/s10984-021-09395-6

[4] S. Telli, R. Maulana, and M. Helms-Lorenz, "Students' perceptions of teaching behaviour in Turkish secondary education: A Mokken scaling of my teacher questionnaire," *Learning Environments Research*, vol. 24, no. 2, pp. 315-337, 2021. https://doi.org/10.1007/s10984-020-09329-8

[5] D. Davies, D. Jindal-Snape, C. Collier, R. Digby, P. Hay, and A. Howe, "Creative learning environments in education—A systematic literature review," *Thinking Skills and Creativity*, vol. 8, pp. 80-91, 2013. https://doi.org/10.1016/j.tsc.2012.07.004

[6] R. A. Beghetto and J. C. Kaufman, "Classroom contexts for creativity," *High Ability Studies*, vol. 25, no. 1, pp. 53-69, 2014. https://doi.org/10.1080/13598139.2014.905247

[7] C. J. McRobbie and B. J. Fraser, "Associations between student outcomes and psychosocial science environment," *The Journal of Educational Research*, vol. 87, no. 2, pp. 78-85, 1993. https://doi.org/10.1080/00220671.1993.9941170

[8] C. Pierce, "Importance of classroom climate for at-risk learners," *The Journal of Educational Research*, vol. 88, no. 1, pp. 37-42, 1994. https://doi.org/10.1080/00220671.1994.9944832

[9] A. J. Reynolds and H. J. Walberg, "A structural model of high school mathematics outcomes," *The Journal of Educational Research*, vol. 85, no. 3, pp. 150-158, 1992. https://doi.org/10.1080/00220671.1992.9944431

[10] H. C. Waxman and S.-Y. L. Huang, "Motivation and learning environment differences in inner-city middle school students," *The Journal of Educational Research*, vol. 90, no. 2, pp. 93-102, 1996. https://doi.org/10.1080/00220671.1996.9944450

[11] D. L. Ching-Tse, "Learning environments in English classrooms in Singapore: Determinants and effects," (Doctoral Dissertation. Curtin University, 2013.

[12] B. J. Fraser and D. L. Fisher, "Predicting students' outcomes from their perceptions of classroom psychosocial environment," *American Educational Research Journal*, vol. 19, no. 4, pp. 498-518, 1982. https://doi.org/https://doi.org/10.3102/00028312019004498

[13] R. J. Stiggins, M. M. Griswold, and K. R. Wikelund, "Measuring thinking skills through classroom assessment," *Journal of Educational Measurement*, vol. 26, no. 3, pp. 233-246, 1989. https://doi.org/10.1111/j.1745-3984.1989.tb00329.x

[14] P21, *21st century readiness for every student: A policymaker's guide*. Washington, DC: Partnership for 21st Century Skills, 2010.

[15] G. J. Anderson, H. J. Walberg, and W. W. Welch, "Curriculum effects on the social climate of learning: A new representation of discriminant functions," *American Educational Research Journal*, vol. 6, no. 3, pp. 315–328, 1969.

[16] B. J. Fraser, G. J. Anderson, and H. J. Walberg, *Assessment of learning environments: Manual for learning environment inventory (LEI) and my class inventory (MCI)*, 3rd ed. Perth, Australia: Western Australian Institute of Technology, 1982.

[17] E. J. Trickett and R. H. Moos, "Social environment of junior high and high school classrooms," *Journal of Educational Psychology*, vol. 65, no. 1, pp. 93-120, 1973. https://doi.org/10.1037/h0034823

[18] P. C. Taylor, B. J. Fraser, and D. L. Fisher, "Monitoring constructivist classroom learning environments," *International Journal of Educational Research*, vol. 27, no. 4, pp. 293-302, 1997. https://doi.org/10.1016/S0883-0355(97)90011-2

[19] B. J. Fraser and C. J. McRobbie, "Science laboratory classroom environments at schools and universities: A cross-national study∗," *Educational Research and Evaluation*, vol. 1, no. 4, pp. 289-317, 1995. https://doi.org/10.1080/1380361950010401

[20] G. P. Teh and B. J. Fraser, "An evaluation of computer assisted learning in geography in Singapore," *Australasian Journal of Educational Technology*, vol. 10, no. 1, pp. 55–68, 1994. https://doi.org/10.14742/ajet.2084

[21] Y. Rahmawati, "Fieldwork, co-teaching and co-generative dialogue: Transforming teachers' pedagogy in a cultural context," *Issues in Educational Research* vol. 26, no. 1, pp. 185–204, 2016.

[22] A. F. Wong and B. J. Fraser, "Science laboratory classroom environments and student attitudes in chemistry classes in Singapore," presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, April 5–8, 1994.

[23] D. Chan, "Development of the clinical learning environment inventory: Using the theoretical framework of learning environment studies to assess nursing students' perceptions of the hospital as a learning environment," *The Journal of Nursing Education*, vol. 41, no. 2, pp. 69-75, 2002. https://doi.org/10.3928/0148-4834-20020201-06

[24] S. L. Chua, A. F. Wong, and C. Der-Thanq, "An instrument for investigating Chinese language learning environments in Singapore secondary schools," *Issues in Educational Research*, vol. 19, no. 2, p. 100, 2009.

[25] S. M. Peoples, L. M. O'Dwyer, Y. Wang, J. J. Brown, and C. V. Rosca, "Development and application of the elementary school science classroom environment scale (ESSCES): Measuring student perceptions of constructivism within the science classroom," *Learning Environments Research*, vol. 17, pp. 49-73, 2014. https://doi.org/10.1007/s10984-013-9152-8

[26] R. B. Koul, B. J. Fraser, N. Maynard, and M. Tade, "Evaluation of engineering and technology activities in primary schools in terms of learning environment, attitudes and understanding," *Learning Environments Research*, vol. 21, pp. 285-300, 2018. https://doi.org/10.1007/s10984-017-9255-8

[27] W. J. Boone and A. Noltemeyer, "Rasch analysis: A primer for school psychology researchers and practitioners," *Cogent Education*, vol. 4, no. 1, p. 1416898, 2017. https://doi.org/10.1080/2331186X.2017.1416898

[28] B. D. Wright and M. M. Mok, *An overview of the family of Rasch measurement models. In E. V. Smith Jr. & R. M. Smith (Eds.), Introduction to Rasch measurement: Theory, models and applications*. Maple Grove, MN: JAM Press, 2004.

[29] T. G. Bond, Z. Yan, and M. Heeme, *Applying the Rasch model: Fundamental measurement in the human sciences*, 4th ed. Routledge. https://doi.org/10.4324/9781410614575, 2021.

[30] S. Azwar, *Compilation of psychological scales*, 2nd ed. Yogyakarta, Indonesia: Pustaka Pelajar, 2012.

[31] H. Bisri, *Authentic assessment: Practical and easy*, 1st ed. Bogor, Indonesia: UNIDA Press, 2017.

[32] D. S. Naga, *Score theory on mental measurement*, 2nd ed. Jakarta, Indonesia: PT Nagarani Citrayasa, 2013.

[33] B. D. Wright and M. H. Stone, *Measurement essentials*, 2nd ed. Wilmington, DE: Wide Range, Inc, 1999.

[34] W. J. Boone, J. R. Staver, and M. S. Yale, *Rasch analysis in the human sciences*. Springer. https://doi.org/10.1007/978-94-007-6857-4, 2014.

[35] M. S. Khine, "Objective measurement in psychometric analysis," *Rasch Measurement: Applications in Quantitative Educational Research*, pp. 3-7, 2020. https://doi.org/10.1007/978-981-15-1800-3_1

[36] J. M. Linacre, *A user's guide to Winsteps (Version 3.68)*. Chicago, IL: Winsteps.com, 2008.

[37] B. D. Wright and J. M. Linacre, "Reasonable mean-square fit values," *Rasch Measurement Transactions*, vol. 8, no. 3, pp. 370–371, 1994.

[38] L. J. Cronbach, P. Schönemann, and D. McKie, "Alpha coefficients for stratified-parallel tests," *Educational and Psychological Measurement*, vol. 25, no. 2, pp. 291–312, 1965. https://doi.org/10.1177/001316446502500201

[39] C. Hulin, R. Cudeck, R. Netemeyer, W. R. Dillon, R. McDonald, and W. Bearden, "Cronbach's alpha on two-item scales," *Journal of Consumer Psychology,* vol. 10, no. 1/2, pp. 55–69, 2001.

[40] P. W. Duncan, R. K. Bode, S. M. Lai, S. Perera, and G. A. i. N. A. Investigators, "Rasch analysis of a new stroke-specific outcome scale: The Stroke Impact Scale," *Archives of Physical Medicine and Rehabilitation,* vol. 84, no. 7, pp. 950-963, 2003. https://doi.org/10.1016/S0003-9993(03)00035-2

[41] E. Brentani and S. Golia, "Unidimensionality in the Rasch model: how to detect and interpret," *Statistica,* vol. 67, no. 3, pp. 253-261, 2007. https://doi.org/10.6092/issn.1973-2201/3508

[42] M. D. Reckase, *Multidimensional item response theory*. Springer. https://doi.org/10.1007/978-0-387-89976-3, 2009.

[43] R. L. Brennan, *Educational measurement*, 4th ed. Westport, CT: Praeger Publishers, 2006.