

Development and Validation of a Non-Cognitive Character Diagnostic Assessment Instrument for Junior High Graduates

Muchamad Aulia Arifin[✉], Sarwi Sarwi, Djuniadi Djuniadi

Department of Educational Research and Evaluation, Graduate School, Universitas Negeri Semarang, Semarang, Indonesia

Info Articles

History Articles:

Received: October 24, 2024

Accepted: January 19, 2025

Published: March 08, 2025

Keywords:

development, assessment, diagnostics, non-cognitive, character.

Abstract

The characteristics of the independent curriculum are assessments at the beginning, process, and end of learning to understand learning needs and the development of the learning process that students have gone through. One of the objectives of non-cognitive diagnostic assessment is to obtain information about students' initial character. The existing non-cognitive character diagnostic assessment instruments, according to the author, are good but need to be developed. The Graduate Profile Dimension in In-Depth Learning is an indicator in creating statement items for this non-cognitive character diagnostic assessment. This study aims to develop a valid and reliable non-cognitive character diagnostic assessment instrument for junior high school graduates. The design chosen in this study is Research and Development (R & D) with a 4D development model. Content and construct validity were analyzed using Aiken's V and CFA, while reliability was analyzed using the Cronbach Alpha formula. The results of Confirmatory Factor Analysis (CFA) with the help of the Lisrel program. The average content validity value from the experts was 0.96. Therefore, it can be stated that the instrument is valid. The results of the analysis show that the Construct Reliability (CR) value of 0.98 has met and even exceeded the required minimum limit value, which is 0.70. This CR value shows that the indicators in the instrument have a very good level of consistency and show very high internal consistency.

✉ Address correspondence:

Department of Educational Research and Evaluation, Graduate School, Universitas Negeri Semarang, Semarang, Indonesia 50229
E-mail: muchauliaarifin8@students.unnes.ac.id

P-ISSN 2252-6420

E-ISSN 2503-1732

INTRODUCTION

Globalization is the process of cultural, political, economic, and technological encounters and blending of people, ideas, and resources across local, national, and regional boundaries, which is widely considered to have increased in intensity and scale during the late 20th and early 21st centuries. This century of globalization is characterized by advances in science and technology (iptek), increasing materialism, global competition, and free movement. Unfortunately, this revolution brings not only benefits but also dangers. Misleading information, destructive ideas, and immoral content spread rapidly through the internet, influencing the thinking and behavior of the current generation. Values, norms, and lifestyles can change so quickly and become a new order. This order ultimately distances humans from the certainty of values they have held for decades (Abdurakhmonova et al., 2021; Hismanoglu, 2012).

Technological developments, changes in the socio-cultural environment, social interactions, and national identity have now experienced moral degradation or decline. This moral degradation occurring in the current generation is heavily influenced by the presence of gadgets, resulting in students lacking a sense of concern for their surroundings, and a decline in discipline and religiosity. Technological advances have also significantly impacted the level of student misconduct. While previously common offenses included cheating and truancy, today they range into adult delinquency, such as excessive dating, violence, and gadget abuse.

This situation has been further exacerbated by the COVID-19 pandemic, which has significantly disrupted schools worldwide. The COVID-19 pandemic has led to learning loss, a loss of knowledge or skills among students. This can occur due to prolonged school closures and the shift to online learning, creating gaps in access and quality of learning, which alter the substance of student learning (Aziz et al., 2023; Donnelly & Patrinos, 2022).

This certainly requires teachers to be astute in their attitudes and actions as possible so that we can maximize the

positive impacts of the globalization era. Education can be effectively absorbed and act as a brake on life. One characteristic of the independent curriculum is the use of assessments at the beginning, during, and at the end of learning to understand students' learning needs and progress throughout the learning process. Providing diagnostic assessments at the beginning of learning will improve students' learning outcomes (Agustinah & Indriyani, 2019; Supriyadi et al., 2022).

Non-cognitive diagnostic assessments aim to obtain information about students' well-being, psychology, social-emotional well-being, learning styles, character, and interests. By understanding students' cognitive levels and characteristics, teachers will more easily determine learning strategies (Nurjanah, 2021; Puspita Rakhmi et al., 2023).

The author believes that the existing non-cognitive character diagnostic assessment instruments are good but need to be developed. The Graduate Profile Dimensions in In-Depth Learning can be used as indicators in creating non-cognitive character diagnostic assessment items. Based on the background above, the researcher wants to examine the development of a valid and reliable non-cognitive character diagnostic assessment instrument for junior high school graduates. Based on the results of the non-cognitive character diagnostic assessment, teachers can determine appropriate learning strategies and material emphasis for each student. This will ensure that learning runs smoothly and on target.

METHODS

The research design chosen in this study is Research and Development (R&D). This design is used to conduct research, development, and testing of a product. Producing certain products requires research based on needs analysis. This study developed an instrument product using Thiagarajan's 4D development model. This model was chosen because it represents an innovation from previous studies related to the development of non-cognitive diagnostic

assessment instruments or character assessments. This model consists of four stages: definition, design, development, and dissemination (Rachman, 2015).

Content validity was obtained from the results of the validation sheet (assessment instrument review) which was carried out by verifying the validation results, input and suggestions from the validator. Validation of the test items was based on three assessment aspects: material, construction, and language. The validation process of the assessment instrument was carried out by instrument expert lecturers and Pancasila Education teachers. After obtaining the validation data for the assessment instrument from the experts, the data was analyzed using the Aiken's V formula to calculate the content validity coefficient using the Aiken V coefficient formula, namely:

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

Information :

V : Aiken index

s : score given by the assessor minus the lowest score in the category

r : score given by the assessor

lo: lowest assessment score

c : highest assessment score

n : number of validators (raters)

This study also tested construct validity after obtaining data from the field trial phase. The data were analyzed using Lisrel to calculate the CFA.

Confirmatory Factor Analysis (CFA) aims to test whether the indicators forming the latent variables are significant and valid (Ghozali, 2022). CFA measurements are based on questionnaire item validity and loading factor reliability. Validity testing demonstrates how the manifest variables (indicators) reflect the latent variables being measured. To be considered valid, the value of the validity test must have a standardized loading factor >0.50 (Ghozali, 2017).

In this study, reliability testing was conducted using inter-rater reliability and Kuder-Richardson reliability. Inter-rater

reliability was conducted to measure the level of expert agreement on the developed instrument. The type of inter-rater reliability applied was Cronbach's Alpha reliability. This reliability was used after completing the content validity calculation. Reliability is declared reliable if the r11 result is > 0.70. The formula used in Cronbach's Alpha is:

$$r_{11} = \left(\frac{n}{n-1} \right) \left(1 - \frac{\sum S_i^2}{S_t^2} \right)$$

Information :

r11 : Instrument coefficient

n : Number of questions

S_i^2 : Grain variants

S_t^2 : Question variance

The reliability of the instrument was also analyzed using the Cronbach Alpha coefficient obtained through calculations in the Winstep program. The reliability criteria are presented in a table as explained by (Hasanah & Aini, 2025).

Table 1. Reliability Value Criteria

Interval	Information
< 0.67	Weak
0.67 – 0.8	Enough
0.8 – 0.9	Good
0.91– 0.94	Very Good
> 0.94	Excellent

The model reliability test is carried out by calculating the Construct Reliability (CR) and Average Variance Extracted (AVE) values from the Standardized Factor Loading values using the following formula:

$$\text{Composite Reliability} = \sqrt{\frac{(\sum \text{Standar Loading})^2}{(\sum \text{Standar Loading})^2 + \sum E_j}}$$

$$\text{Average Variance Extracted} = \sqrt{\frac{(\sum \text{Standar Loading}^2)}{(\sum \text{Standar Loading}^2) + \sum E_j}}$$

Description:

1. Standard loading = obtained from the standardized loading for each indicator calculated using the Lisrel 8.0 program.
2. ΣE_j = measurement error for each indicator. Measurement error can be obtained from $1 - (\text{standard loading})^2$

Reliability, measured through the construct reliability coefficient, has a critical value limit of ≥ 0.70 (good reliability). However, if the value is in the range of 0.60–0.70, it is still acceptable, provided the validity of the indicators in the model is good. The average variance extracted coefficient has a critical value limit of ≥ 0.50 , which is considered reliable. This coefficient is optional in research (Hair et al., 2021).

RESULTS AND DISCUSSION

Expert judgment validated the non-cognitive diagnostic assessment instrument for the character of junior high school graduates by assigning a score of 1–4 to 15 questions. The scores obtained from the expert judgments were then analyzed using the Aiken V formula, and the results are presented in table 3 below.

Table 2. Content Validity Analysis Results

Item	s1	s2	s3	s4	s5	Σs	n(c-1)	V	
1	3	3	3	3	3	15	15	1	
2	2	3	3	3	3	14	15	0,93	
3	2	3	3	3	2	13	15	0,87	
4	2	3	3	3	2	13	15	0,87	
5	3	3	3	3	3	15	15	1	
6	3	3	3	3	3	15	15	1	
7	3	3	3	3	3	15	15	1	
8	3	3	3	3	3	15	15	1	
9	3	3	3	3	3	15	15	1	
10	3	3	3	3	3	15	15	1	
11	3	2	3	3	3	14	15	0,93	
12	3	2	3	3	3	14	15	0,93	
13	3	3	3	3	3	15	15	1	
14	3	3	3	3	3	15	15	1	
15	3	3	2	3	1	12	15	0,8	
Average value V								0,96	

It can be seen that the 15 questions regarding the non-cognitive diagnostic assessment instrument assessment of the character of junior high school graduates

obtained an Aiken V index ranging from 0.8 to 1 with an average of 0.96. According Aiken, (1985) In the validity coefficient V table, if the expert judgment consists of 5 people using a 4-point assessment scale, the V coefficient value is 0.87. If the V value obtained is ≥ 0.87 , it can be stated that the instrument is valid

Expert validation is a standard step in the instrument development process. Proving content validity plays a role in ensuring the overall validity of the instrument, particularly regarding the correspondence between the grid and indicators, indicators and statement items, scales and scoring, and the ability of indicators and statement items to measure the targeted variables. Therefore, a systematic approach is needed to evaluate content validity, based on evidence from the assessment of the developed instrument (Indiati, 2022; Liu et al., 2024).

The data obtained from expert judgment were analyzed again to estimate its reliability using the Cronbach's Alpha formula with SPSS 26. The results of the reliability between expert judgments can be seen in table 4 below.

Table 3. Results of Reliability Between Expert Judgements

Cronbach Alpha	N of Items
.214	15

The analysis of scores from five expert judgements on the instrument yielded a score of 0.214. This score is below the minimum required limit of 0.70. Therefore, it can be concluded that the expert assessments are unreliable. A low reliability score indicates significant differences in assessments between experts on the same indicator.

This condition likely arose from differing interpretations of several criteria on the validation sheet. However, the validity results using Aiken's V still confirmed that the instrument was suitable for use. This low inter-rater reliability also reinforces the need for a revision phase based on qualitative expert notes to clarify and standardize each component of the

instrument to meet standards before empirical testing (Natarisa, 2025).

After the instrument was revised based on expert input, a field trial was conducted to determine the construct validity of the instrument. The field trial results for each item can be seen in the item validity summary in Table 4 below.

Table 4. Validity of Field Trial Statement Items

No	Statement	SLF	Desc
1	I am grateful for the blessings I receive from God every day	0.75	Valid
2	I maintain personal hygiene by bathing regularly and washing my hands before eating.	0.68	Valid
3	Saya nyaman berteman dengan siapa saja tanpa memandang latar belakangnya. I feel comfortable making friends with anyone regardless of their background	0.81	Valid
4	I dispose of waste properly in designated places	0.73	Valid
5	I follow the school rules and regulations	0.66	Valid
6	I participate in cultural activities in my local community	0.60	Valid
7	I feel proud of Indonesia's cultural diversity	0.77	Valid
8	I do not laugh when a friend is mocked or humiliated	0.81	Valid
9	I am willing to help friends or others in need without discriminating against their background	0.82	Valid
10	I support cooperation among countries to promote world peace	0.64	Valid
11	Saya aktif dalam pekerjaan kelompok tanpa bergantung pada orang lain.	0.88	Valid
12	Saya aktif dalam kegiatankegiatan sosial	0.83	Valid
14	Saya belajar melalui internet selain belajar dari buku paket.	0.99	Valid
15	Saya tidak menunda-nunda waktu untuk mengerjakan tugas dan belajar.	0.81	Valid
16	Saya bertanya kepada guru atau orang tua untuk mengetahui kebenaran sebuah informasi.	0.88	Valid
17	Saya mempertimbangkan fakta yang ada sebelum mengambil sebuah keputusan.	0.84	Valid
18	Saya belajar dari kesalahan yang pernah saya buat agar tidak mengulangnya lagi di masa yang akan datang.	0.90	Valid
19	Saya berusaha mencari cara baru untuk menyelesaikan masalah yang ada.	0.86	Valid
20	Saya senang ketika menghasilkan karya yang didorong oleh minat saya sendiri.	0.85	Valid
21	Saya menganalisis kelebihan dan kekurangan dari setiap pilihan sebelum menentukan solusi untuk suatu masalah.	0.88	Valid
22	Saya menggunakan HP dengan bijak agar kesehatan tetap terjaga.	0.99	Valid
23	Saya rutin berolahraga untuk menjaga kesehatan tubuh.	0.76	Valid
24	Saya percaya diri saat berbicara di depan teman atau guru.	0.99	Valid
25	Saya menggunakan kata-kata yang sopan saat berbicara kepada siapapun.	0.78	Valid

The analysis results show that all statements in this dimension have a SLF value greater than 0.50, ranging from 0.78 to 0.99. This value meets the construct validity criteria, thus concluding that all statements in the Communication dimension are valid and capable of measuring the intended construct.

The statements in this dimension reflect students' ability to communicate confidently and use polite language. A good SLF value for each statement indicates that these indicators truly reflect the students' communication characteristics. Therefore, the non-cognitive diagnostic assessment instrument used is capable of accurately measuring the Communication dimension. This demonstrates the instrument's effectiveness in assessing the character of junior high school graduates.

SLF values that meet these criteria indicate that each indicator has a strong relationship with the construct being measured. This is in line with the opinion of Hair (2022) who stated that an indicator is said to be construct valid if it has a factor loading value ≥ 0.50 . Substantially, each measured character dimension also reflects students' actual behavior in daily life, such as the ability to work together, think critically, be independent, maintain health, and communicate politely and confidently. This shows that the non-cognitive character constructs developed are not only statistically valid but also contextually relevant to the characteristics of junior high school students. Based on the results obtained, according to Saputra et al., (2022), if a statement has a high level of validity, it is reliable and appropriate for use.

Based on the reliability test results of the non-cognitive diagnostic assessment constructs for junior high school graduates' character, presented in Table 5, a reliability analysis was conducted using the Construct Reliability (CR) and Average Variance Extracted (AVE) values obtained from Confirmatory Factor Analysis (CFA) using the Lisrel program. The analysis showed a Construct Reliability (CR) value of 0.98, which meets and even exceeds the minimum required value of 0.70. This CR value indicates that the indicators in the instrument have a very good level of consistency in measuring the non-cognitive diagnostic assessment constructs for junior high school graduates' character.

Construct reliability testing shows that the non-cognitive diagnostic assessment instrument for junior high school graduates' character has a very good level of reliability. A Construct Reliability (CR) value of 0.98 indicates very high internal consistency, while an Average Variance Extracted (AVE) value of 0.68 indicates that most of the indicator variance can be explained by the measured construct.

The results of this study met both criteria. Furthermore, the SLF values for the indicators, which ranged from 0.60 to 0.99 with relatively low error, confirmed that each indicator contributed significantly to the formation of the construct. These

findings demonstrate that the developed instrument is reliable and consistent in measuring the non-cognitive characteristics of junior high school graduates. This instrument can be used repeatedly with relatively stable measurement results.

Table 5. Reliability Test Results

Indi-cator	SLF	value	SLF ²	CR	AVE
P1	0.75	0.43	0.56		
P2	0.68	0.53	0.46		
P3	0.81	0.35	0.66		
P4	0.73	0.46	0.53		
P5	0.66	0.57	0.44		
P6	0.60	0.64	0.36		
P7	0.77	0.41	0.59		
P8	0.81	0.34	0.66		
P9	0.82	0.32	0.67		
P10	0.64	0.59	0.41		
P11	0.88	0.23	0.77		
P12	0.83	0.31	0.69		
P13	0.90	0.19	0.81		
P14	0.99	0.02	0.98	0.98	0.68
P15	0.81	0.34	0.66		
P16	0.88	0.22	0.77		
P17	0.84	0.29	0.71		
P18	0.90	0.19	0.81		
P19	0.86	0.27	0.74		
P20	0.85	0.27	0.72		
P21	0.88	0.23	0.77		
P22	0.99	0.02	0.98		
P23	0.76	0.42	0.58		
P24	0.99	0.02	0.98		
P25	0.78	0.40	0.61		

CONCLUSION

Development and validation of the non-cognitive character diagnostic assessment instrument for junior high graduates demonstrate strong evidence of content validity, construct validity, and internal reliability. Expert judgment confirmed high content validity, while Confirmatory Factor Analysis supported

the adequacy of construct validity across all indicators representing the targeted character dimensions. Reliability analysis further indicated excellent internal consistency, demonstrating that the instrument provides stable and dependable measurement results for identifying students' non-cognitive character profiles. Availability of a validated diagnostic assessment instrument strengthens evidence-based assessment practices within the implementation of the Independent Curriculum, particularly in supporting teachers to design more responsive and student-centered learning strategies. Integration of graduate profile dimensions within the instrument also represents an important contribution to character-based assessment development in lower secondary education contexts. Limitations of the study include the relatively restricted sample scope and the focus on selected character dimensions aligned with the Graduate Profile framework. Future research may extend validation across broader regional contexts and incorporate additional non-cognitive constructs to enhance the applicability of diagnostic assessment instruments in diverse educational settings.

REFERENCES

- Abdurakhmonova, M. M., Mirzayev, M. A. ugli, Karimov, U. U., & Karimova, G. Y. (2021). Information Culture And Ethical Education In The Globalization Century. *The American Journal of Social Science and Education Innovations*, 03(03), 384–388. <https://doi.org/10.37547/tajssei/volu me03issue03-58>
- Agustinah, S. W., & Indriyani, D. (2019). Dampak Globalisasi Terhadap Perilaku Belajar Siswa Di Smk Negeri 1 Cianjur. *Integralistik*, 30(1), 53–62. <https://doi.org/10.15294/integralistik .v30i1.20767>
- Aiken, L. R. (1985). Analyzing Reliability and Validity Ratings Aiken. *Educational and Psychological Measurement*, 131–142.
- Aziz, R., Susilowati, M., Masturin, M., & Prasajo, Z. H. (2023). Teacher-parent collaboration for developing student character in online learning. *International Journal of Evaluation and Research in Education*, 12(3), 1477–1485. <https://doi.org/10.11591/ijere.v12i3.24456>
- Donnelly, R., & Patrinos, H. A. (2022). Learning loss during Covid-19: An early systematic review. *Prospects*, 51(4), 601–609. <https://doi.org/10.1007/s11125-021-09582-6>
- Ghozali. (2017). *Aplikasi Analisis Multivariate Dengan Program SPSS*. UNDIP.
- Ghozali. (2022). *Aplikasi Analisis Multivariate Dengan Program IBM SPSS 26* (10th ed.). Badan Penerbit Universitas Diponegoro.
- Hair. (2022). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (3e). Thousand Oaks: CA: Sage.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., Danks, N. P., & Ray, S. (2021). *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R*. <https://doi.org/10.1007/978-3-030-80519-7>
- Hasanah, H., & Aini, F. Q. (2025). Pengembangan Instrumen Tes Keterampilan Proses Sains pada Topik Kesetimbangan Kimia: Analisis dengan Model Rasch. *JagoMIPA: Jurnal Pendidikan Matematika Dan IPA*, 5(1), 68–82. <https://doi.org/10.53299/jagomipa.v5i1.1043>
- Hismanoglu, M. (2012). The impact of globalization and information technology on language education policy in Turkey. *Procedia - Social and Behavioral Sciences*, 31(2011), 629–633. <https://doi.org/10.1016/j.sbspro.2011.12.115>
- Indiati, I. (2022). *The Validity Analysis of Learning Kit for Ethnoscience-Oriented Learning Program Operationalization to Improve Teacher Candidates ' s Technological Pedagogical Content Knowledge (TPACK)*. 18. <https://doi.org/10.37394/232015.2022.18.28>

- Liu, Y., Wang, J., Zhang, Z., Wang, J., Luo, T., Lin, S., Li, J., & Xu, S. (2024). Development and validation of an instrument for measuring civic scientific literacy. *Disciplinary and Interdisciplinary Science Education Research*.
<https://doi.org/10.1186/s43031-023-00092-3>
- Natarisa, R. (2025). *Oleh Rahma Natarisa*. Universitas Negeri Semarang.
- Nurjanah, S. (2021). Objek Asesmen Dalam Perspektif Kurikulum 2013. *Jurnal Pendidikan Dan Pengajaran Guru Sekolah Dasar (JPPGuseda)*, 4(2), 85–91.
<https://doi.org/10.55215/jppguseda.v4i2.3607>
- Puspita Rakhmi, M., Purwo Yudi Utomo, A., Ashfiya Adiratna Salim Putri, ah, & Ghufon, W. (2023). Pemanfaatan Google Form dalam Asesmen Diagnostik di SMA Negeri 11 Semarang. *Concept: Journal of Social Humanities and Education*, 2(1), 115–126.
- Rachman, M. (2015). *5 Pendekatan Penelitian Kuantitatif, Kualitatif, Mixed, PTK, R&D* (1st ed.). Magnum Pustaka Utama.
- Saputra, H. D., Purwanto, W., Setiawan, D., Fernandez, D., & Putra, R. (2022). Hasil Belajar Mahasiswa: Analisis Butir Soal Tes. *Edukasi: Jurnal Pendidikan*, 20(1), 15–27.
<https://doi.org/10.31571/edukasi.v20i1.3432>
- Supriyadi, S., Lia, R. M., Rusilowati, A., Isnaeni, W., Susilaningsih, E., & Suraji, S. (2022). Penyusunan Instrumen Asesmen Diagnostik untuk Persiapan Kurikulum Merdeka. *Journal of Community Empowerment*, 2(2), 67–73.
<https://doi.org/10.15294/jce.v2i2.61886>