

Integrating Krathwohl's taxonomy into digital self-assessment: Psychometric validation of a civic affective scale for secondary education

Rabiatul Adawiah 

Universitas Lambung Mangkurat, Indonesia
rabiatuladawiah@ulm.ac.id

Dian Agus Ruchliyadi 

Universitas Lambung Mangkurat, Indonesia
dianagus@ulm.ac.id

Nurul Huda * 

* Universitas Lambung Mangkurat, Indonesia
nurul.huda@ulm.ac.id

Yonas Belay Gela 

Dilla University, Dilla, Ethiopia
yonasbelay118@gmail.com

*Corresponding Author

Article Info

Article History

Submitted:

6 March 2026

Revised:

14 Agustus 2026

Accepted:

7 September 2026

Keywords

affective domain; civic education; digital self-assessment; Krathwohl's taxonomy; psychometric validation

Abstract

This study develops and evaluates a digital self-assessment tool to measure students' affective domain in Civics Education at the senior high school level. The research addresses the lack of measurement of affective outcomes in civic and character education. Although the affective dimension is an important part of the curriculum, most teachers lack reliable tools to document students' value internalisation. Using a research and development (R&D) design based on the ADDIE (analysis, design, development, implementation and evaluation) model, the tool was built on Krathwohl's taxonomy of the affective domain and supported by attitude theory, social cognitive theory, and the concepts of global citizenship and character education. Content validity was assessed using Aiken's V with five experts, followed by empirical testing with 182 participants in the pilot test and 498 participants in the field test. Item-total correlation and Cronbach's alpha were used to assess item performance and internal consistency, with reliability coefficients between 0.706 and 0.804. Exploratory factor analysis (EFA) revealed a general affective factor while retaining five theoretically relevant factors: receiving, responding, valuing, organisation, and characterisation. Confirmatory factor analysis (CFA) indicated that the measurement model was acceptable (CMIN/df = 2.545; CFI = 0.927; TLI = 0.916; RMSEA = 0.056); although the adjusted goodness-of-fit index (AGFI = 0.882) fell slightly below the strict 0.90 criterion. This research contributes both a context-relevant yet globally applicable civic affective scale and an illustration of digital self-assessment for evidence-based affective assessment, while addressing issues such as social desirability and cross-cultural validity.

Article Link: <https://journal.uny.ac.id/index.php/civics/article/view/96277>

Introduction

The affective domain is an important element of quality education as it covers issues such as values, attitudes, emotions, motivation, empathy, responsibility and self-regulation. For civic and character education, learning outcomes cannot be limited to cognitive acquisition as students are also expected to internalise values and display responsible behaviour in school, community and even online. The topic is also relevant internationally: Global Citizenship Education (GCED) comprises cognitive, socio-emotional and behavioural domains, while in character education value formation needs to be proven aremainst knowledge (UNESCO, 2015).

However, even thoalt is an important issue globally, the asseassessingective domain continues to be one of the weakest parts of classroom assessment. While teachers acknowledge the importance of attitudes and values, they tend to use informal or unstructured observation, or general impressions. This makes it difficult to obtain valid, reliable and comparable evidence about the development of the affective domain of the students. This is particularly important in Civic Education, whose aim is the cultivation of civic values, ethical responsibility and respect for diversity.

The theoretical underpinning of this study is not confined to Krathwohl's work. The taxonomy developed by Krathwohl et al. (1964) offers a hierarchical progression from receiving and responding through valuing, organisation, and characterisation. However, affective competence is also multidimensional, consisting of cognitive beliefs, affective responses, evaluative commitment, and behavioural tendencies. The theory of attitude states that values can be displayed through cognitive-affective-behavioural components (Fishbein & Ajzen, 2010). Social Cognitive Theory points out that attitudes can be formed through observation, social interaction, self-regulatory and reflective processes (Bandura, 1991). These theories provide justification for measuring civic affective competence as multidimensional and developmental in nature.

Pancasila Education places emphasis on the importance of the affective domain because the expected learning outcomes include not only the acquisition of knowledge related to civic principles, but also value, attitude, and disposition development towards proper conduct as citizens. From this standpoint, Krathwohl's revised taxonomy becomes very pertinent because it views affective development as the development from reception and response to valuing, organisation, and characterisation. As such, affective learning is viewed as a process of development rather than an assessment of whether or not students possess certain civic values. Therefore, assessment in Pancasila education should be able to measure students' internalisation of civic values as well as their reflection on these values.

Self-assessment is relevant for affective measurement because many affective states and value commitments are partly internal and cannot be fully captured through teacher observation alone. By inviting students to reflect on their own attitudes, commitments, and behavioural tendencies, self-assessment provides a mechanism for capturing aspects of affective development that correspond to the progressive internalisation of values described in Krathwohl's taxonomy. It can also support students' reflection and learner autonomy in the development of civic dispositions. However, self-assessment is vulnerable to social desirability bias, especially when measuring moral, civic, or character-related constructs. Students may report what is socially expected rather than what accurately reflects their dispositions (Lavidas et al., 2022; Setiawati et al., 2024). Therefore, the use of self-assessment for civic affective measurement requires a theoretically grounded and psychometrically validated instrument capable of translating the dimensions of affective development into clear, measurable, and contextually relevant indicators.

To reduce this risk, we designed the instrument as a non-punitive digital self-assessment. Students were informed that the instrument was not a high-stakes test and that responses

would be used for learning reflection and instrument development. Although the digital format offers advantages in standardisation, efficiency, and response privacy, the study does not claim that digital delivery prevents response bias; instead, it treats such bias as a methodological limitation to be addressed through clear instructions, anonymity, triangulation, and validity studies.

Previous studies on affective assessment addressed teachers' problems, students' self-assessment, and the measurement of character qualities. More recent literature has focused increasingly on self-assessment and digital assessment in education. Nevertheless, these three research areas seem to have developed separately. Research on affective assessment has generally investigated students' attitudes, values, or character qualities. Self-assessment research mostly focuses on students' reflection in the learning process. Research on digital assessment, in turn, has tended to examine issues such as accessibility, efficiency, and feedback rather than operationalising students' affective development in a civic education setting. Thus, less attention has been paid to integrating a civic affective construct, digital self-assessment, and validation of such an assessment instrument by means of Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) in the case of secondary civic education. Evidence on how to operationalise Krathwohl's affective taxonomy through a digital self-assessment instrument for assessing civic values in Pancasila Education remains scarce. This gap points to the need for instruments that are theoretical, empirical, and practically applicable in classrooms and for clearer evidence on the structure of affective constructs in civic education.

On this basis, the present study develops and validates an online self-assessment instrument of the affective domain in Civic Education. The novelty of this study lies in integrating Krathwohl's affective taxonomy, civic affective assessment, self-assessment, and digital delivery into a single measurement instrument supported by EFA and CFA evidence. The study contributes theoretically by demonstrating the applicability of Krathwohl's taxonomy to measuring civic affective outcomes, and methodologically by providing psychometric evidence for a multidimensional civic affective scale for secondary education. It also contributes to educational assessment by positioning Pancasila Education within the broader international discussion of civic, character, and global citizenship education, while providing an instrument that can support practical classroom assessment.

Method

Study design

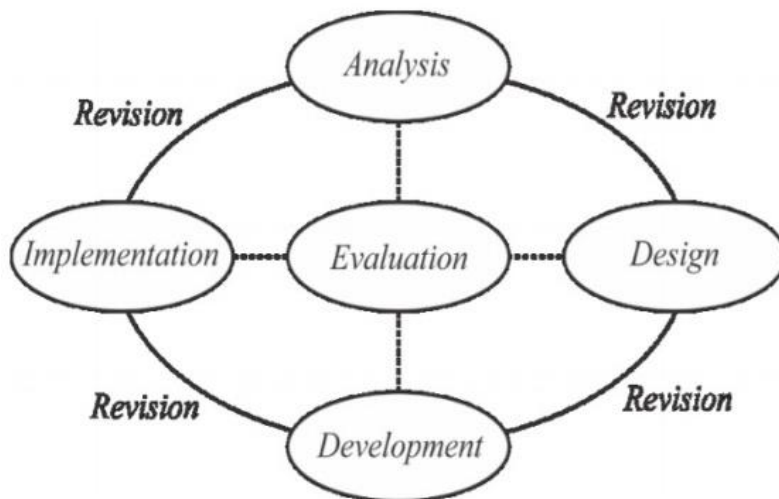
This study used a Research and Development (R&D) design with the ADDIE model to develop and psychometrically validate a digital self-assessment instrument for measuring the affective domain in Pancasila Education at the senior high school level. The ADDIE stages were used not only as a product-development sequence but also as a validity-building framework. The process began with needs analysis and theoretical specification, followed by instrument development, expert validation and formative revision, preliminary testing, large-scale empirical validation, and final evaluation. Thus, the development process combined theoretical specification, expert judgement, and empirical psychometric testing to establish the quality of the instrument.

The Analysis stage started with identifying needs and problems related to affective domain assessment. This activity was conducted through field observations and preliminary studies to understand the actual conditions and obstacles encountered in implementing attitude assessment in schools. Subsequently, a literature review was conducted to identify indicators of attitudes, character, and affective values relevant to the learning objectives. Based on the needs analysis and theoretical review, the researchers prepared a preliminary design of the assessment instrument as a basis for further development.

The Design stage included compiling assessment indicators based on the needs identification results from the previous stage. These indicators were then developed into instrument items that represented the attitudes, character, and affective values to be measured. At this stage, the researchers systematically designed a prototype assessment instrument to include all established indicators.

Figure 1.

Stages of the ADDIE development model



Source: Branch (2009).

The Development stage aimed to create the designed instrument during the process. At this stage, experts validated the instrument (expert judgement) to ensure the suitability of the content and the constructs being measured. Validators were purposively selected based on a minimum academic qualification of a doctorate (S3) in civic education, values education, or educational evaluation, and experience in developing or evaluating assessment instruments. In addition, the validators were actively involved in studies or scientific publications related to educational assessment. According to these criteria, the participating validators were competent experts in their respective fields. The validation procedure included providing the instrument and its outline to the experts for assessment based on the suitability of the indicators, clarity of wording, relevance of the content, and measurability of each statement item. The evaluation used a validation sheet with specific scales, enabling experts to rate items and provide qualitative recommendations to improve the instrument. The expert evaluation findings were analysed to establish the content validity of the instrument. In addition, feedback and comments from the validators were used to modify the instrument, especially items that appeared unclear, ambiguous, or not reflective of the constructs being measured.

Expert validation was not merely a process of approving the prototype but a formative revision process in which the experts provided qualitative and quantitative ratings of item clarity, indicator representation, normative wording, and appropriateness of the items for senior high school students (Table 1).

Table 1.

Expert validators and their roles in the instrument validation process

Validator	Area of expertise	Role in validation
V1	Educational assessment and evaluation	Evaluated construct alignment, measurability, and scoring clarity
V2	Psychometrics / instrument development	Reviewed item representativeness, validity logic, and scale structure
V3	Values education / character education	Assessed relevance to affective and character indicators

Validator	Area of expertise	Role in validation
V4	Pancasila and Civic Education	Assessed curricular relevance and civic value representation
V5	Language and academic writing	Reviewed clarity, readability, ambiguity, and student-level wording

Source: Author (2026).

Following expert assessment, the tool was piloted to obtain empirical evidence of validity. The tool was piloted in one round using a limited number of participants; 182 eleventh-grade students at State Senior High School 3, Banjarmasin, Indonesia, participated in the study. Purposive sampling was used to select the participants because their characteristics fit well with the development goals of the research. The pilot testing involved administering the tool to the respondents in a classroom situation; they were instructed to conduct self-assessments on the questions asked in the tool. Item validity and instrument reliability were assessed using the data collected during the testing phase. The results were used to identify items that failed to meet the required criteria, including poor wording and weak construct validity. The first revision of the tool therefore involved correcting unclear language used in item statements, making the item statements more operational, and discarding invalid items.

The Implementation stage took place through field trials with a larger sample. These trials were meant to measure the validity and reliability of the instrument construct. The developed instrument was then implemented in the Pancasila Education learning process as a tool to assess the affective domain of students.

The final stage was evaluation, which included conducting a final evaluation and revision based on the field trial results to obtain a valid and reliable assessment instrument. Furthermore, the evaluation analysed affective domain assessment results of students by gender and school of origin to ensure that the developed instrument was valid, reliable and practical across various student situations.

Study population and sample

The population comprised all 11th-grade students at public senior high schools in Banjarmasin City, totalling 2,034 students. The sample consisted of two groups, namely a small-scale and a large-scale trial. The small-scale trial was conducted at State Senior High School 3, Banjarmasin, including 182 students, to obtain initial data regarding the quality of the instrument items.

The sample size for the large-scale trial was determined based on the recommended sample size for the instrument test, namely 5–10 respondents per item (Memon et al., 2020). This study selected a purposive sample of 600 students from seven public senior high schools to increase model stability and generalisability. A total of 498 questionnaires were used in the analysis, as the remaining questionnaires were incomplete or did not meet the criteria for valid data. The number of respondents was considered sufficient for testing the construct validity and reliability of the instrument during the field trial phase. These trials were conducted at seven public senior high schools in Banjarmasin City, with the population and sample distribution shown in Table 2.

Table 2.

Distribution of population and sample by school

School Name	Population Size	Sample Size
State Senior High School 1 Banjarmasin	316	67
State Senior High School 2 Banjarmasin	400	45
State Senior High School 4 Banjarmasin	289	77

State Senior High School 5 Banjarmasin	385	56
State Senior High School 10 Banjarmasin	216	83
State Senior High School 11 Banjarmasin	248	90
State Senior High School 12 Banjarmasin	180	80
Total	2,034	498

Source: Author (2026).

The composition of respondents from several schools was intended to increase data heterogeneity and provide a broader picture of the characteristics of affective domain assessment of students. This study was conducted in schools in a single geographic location, and the results were interpreted with consideration of the limitations of the local context. As a result, broad generalisations of the outcomes were approached with caution. Future studies should include more diverse locations and larger sample sizes to ensure the findings are more representative at the regional or national level.

Data collection and analysis

Data analysis was conducted sequentially through item analysis, reliability testing, Exploratory Factor Analysis (EFA), and Confirmatory Factor Analysis (CFA). The item analysis process was used to evaluate the quality and performance of individual items. Reliability testing was done to determine the internal consistency of the instrument and its factors. Third, Exploratory Factor Analysis (EFA) using SPSS version 26 was conducted to investigate the instrument's factor structure. Interpretation of the EFA results was done according to the obtained eigenvalues and factor loadings, along with the expected five-factor structure of the affective domain. The Confirmatory Factor Analysis (CFA) was carried out using AMOS software. Because the available large-scale dataset was analysed sequentially rather than split into fully independent EFA and CFA subsamples, the CFA results were interpreted as confirmatory evidence within the same field-trial context and not as independent cross-validation. This reporting decision was made to avoid overclaiming and to make the boundary of the psychometric evidence explicit.

Results and Discussion

Analysis stage

The initial stage of this study was a needs analysis conducted through field observations, namely the implementation of Pancasila Education learning in schools. Through these observations, several aspects required further analysis, including the school environment, teachers, students, and the education system. Observations and interviews were conducted with Pancasila Education teachers in Banjarmasin City to understand the environmental conditions. Based on these observations, Pancasila Education lessons were generally held during the day. Interviews with teachers in charge showed that Pancasila Education learning was implemented appropriately, both in planning and implementation, as well as in assessment. A total of three aspects were available in the assessment, including cognitive, affective, and psychomotor. However, the affective aspect was not always implemented effectively because it was not assessed with adequate instruments. The interviews also showed that teachers considered affective assessment important for shaping students' character but difficult to implement, particularly in the creation of assessment instruments. Affective assessment was conducted solely through unsystematic observation because no measurable assessment instruments supported the practice. Therefore, developing instruments that facilitated more accurate and sustainable affective assessment was necessary.

A literature review was conducted to identify references related to affective domain assessment. The review showed that teachers lacked appropriate and high-quality instruments

to measure this domain. Self-assessment was a commonly used method for assessing the affective domain because it allowed students to independently express personal perceptions, attitudes, and affective states (Andrade, 2019; Yan & Carless, 2022). The method assumed that individuals could evaluate and reflect on personal internal states, including emotions, motivations, as well as values. In this study, self-assessment was used to measure the affective domain, in which students were trained to evaluate and monitor personal thoughts, feelings, and actions to identify strengths and weaknesses as a basis for achieving expected learning outcomes.

Instrument specification was conducted by first identifying relevant affective indicators based on the affective domain of Bloom's taxonomy, which Krathwohl et al. (1964) organised into five hierarchical levels, namely receiving, responding, valuing, organisation, and characterisation. These indicators served as the basis for constructing the instrument grid and student self-assessment questionnaire items. The response types in this study reflected typical performance, which could not be classified as right or wrong. All responses were considered appropriate based on the individual context of each respondent. During this process, a questionnaire format with a five-point Likert scale (1 = strongly disagree to 5 = strongly agree) was used to accommodate this characteristic when measuring students' affective domain abilities.

Design stage

In the design phase, this study focused on designing indicators and developing assessment instruments for the affective domain. Indicators were developed based on the results of needs identification obtained during the analysis phase. These indicators were designed to represent aspects of attitudes, character, and affective values relevant to learning objectives, particularly in the context of Pancasila Education. Each indicator was broken down into several statements, systematically structured to measure the attitudinal achievement of students comprehensively. Furthermore, the assessment instrument was designed as an initial prototype, as shown in Table 3.

Table 3.
Example of an early prototype instrument grid

No	Affective Aspect	Attitude Indicators	Statement Items	Rating Scale
1	Receiving	Willingness to accept and obey values or rules.	I am willing to accept criticism or input with an open attitude.	1-5
2	Responding	Participate actively and show positive responses.	I actively participate in group discussions or school project activities.	1-5
3	Valuing	Show appreciation for specific values through behaviour.	I respect differences in religion, culture, and opinion in my environment.	1-5
4	Organisation	Integrate various values into the personal system.	I use the value of cooperation as a guideline in all my actions.	1-5
5	Characterisation	Consistently make values a part of personality and behaviour.	I am honest even when no one is paying attention.	1-5

Source: Author (2026).

The instrument prototype was developed by considering the suitability between indicators, assessment objectives, and student characteristics. It was also designed to cover all established indicators of attitudes, character, and affective values, providing a comprehensive picture of the affective development of students. The result of this design phase was a prototype affective assessment instrument ready for validation and further development in the next phase.

Development stage

A total of five experts participated in the validation of the instrument prototype. The panel consisted of two experts in educational assessment and evaluation, one expert in values and character education, one expert in Pancasila and Civic Education, and one expert in language and academic writing. Content validity was examined using Aiken's *V* coefficient to determine the relevance and representativeness of each item with respect to the intended indicators. The experts evaluated each item using a five-point rating scale ranging from 1 (very inappropriate) to 5 (very appropriate). The resulting Aiken's *V* coefficients were interpreted according to the validity categories presented in Table 4.

Table 4.

Instrument validity criteria

$0.80 < V \leq 1.00$	Very high
$0.60 < V \leq 0.80$	High
$0.40 < V \leq 0.60$	Moderate
$0.20 < V \leq 0.40$	Low
$0.00 < V \leq 0.20$	Very low

Source: Author (2026).

The validity classification in Table 4 was based on Aiken's *V* coefficient, which ranged from 0 to 1, with values closer to 1 indicating greater agreement among experts regarding the relevance of instrument items (Aiken, 1985). In this study, the interpretation of Aiken's *V* value followed criteria widely used in the instrument evaluation literature, where $V \geq 0.80$ was categorised as very high, 0.60–0.79 as high, and < 0.60 as requiring revision. Various empirical studies in psychometric measurement supported the use of thresholds, although they were conventional and varied depending on the number of validators and the scale categories used (Polit & Beck, 2006). This study interpreted validity based on the index value and the overall measurement context.

In line with contemporary measurement perspectives, validity interpretation considered the response distribution characteristics and the instrument's sensitivity. The results showed that scores on the affective dimension tended to be high, indicating a possible ceiling effect, namely the concentration of responses in the highest Likert category. This phenomenon could limit the instrument's ability to differentiate respondents at a high construct level, affecting interpretations of validity and reliability.

A categorical upper-bound analysis was performed to examine the proportion of category selections for each item and evaluate this outcome. This analysis aimed to determine whether high scores reflected true affective ability or resulted from limited instrument sensitivity. During the process, the function of the Likert scale categories was analysed through the frequency distribution of responses and the tendency to select each category to ensure adequate discriminatory power. The analysis showed that, although the scale categories remained generally functional, responses tended to accumulate in the upper categories, signifying decreased discrimination at higher levels of the construct. The findings suggest that certain categories were not ideal in measuring the differences in affective domain abilities among students.

The results show that the instrument has limitations in capturing differences among students at high levels of the affective domain. The next step would be to adjust item difficulty, expand the range of indicators, and introduce more discriminating items to improve the instrument.

Based on the expert validation results (Table 5), the instrument received scores in the high and very high categories, indicating that the instrument's content validity was accepted. However, high Aiken's *V* values should not be taken as proof that the initial items were flawless. The qualitative validation results indicated that some items still needed modification, especially

regarding ambiguous phrasing, normative statements, and imbalance among the character dimensions. There was no item deletion based on expert validation results, but item wording was modified, and indicators were realigned before the preliminary empirical trial.

Table 5.

Expert validation test results

Validity Criteria	Instrument Number	Total
Very high	2, 3, 4, 9, 10, 13, 15, 18, 22	9
High	1, 5, 7, 8, 11, 17, 21	7
Moderate	6, 12, 14, 16, 19, 20, 23	7
Low	-	0
Very low	-	0

Source: Author (2026).

Following the modifications based on expert validation, the instrument was tested through a preliminary trial (limited testing) to get empirical validity of the instrument. The preliminary test was carried out in one testing cycle involving 182 eleventh-grade students from State Senior High School 3, Banjarmasin, Indonesia. The preliminary test involved giving the instrument to the respondents under normal learning conditions, and the data were analysed through an item validity test using item-total correlation in statistical software. The validity criterion was $p < 0.05$.

In the preliminary trial, the students performed the self-assessment independently after receiving standardised instructions in order to assess the readability of the items, consistency of the responses, and empirical validity of the items. No item was discarded in this process because all the items satisfied the significance criterion; nevertheless, the wording of several items was modified in order to make them less ambiguous and also less normative in their nature.

All items satisfied the validity criteria because none of the significance values exceeded 0.05. Although no item was discarded, we carried out another revision process to make the sentences more readable and representative of the affective indicators being assessed. This revision phase was conducted to improve the clarity and practicality of the instrument before proceeding to the main field trial. The revised instrument was used at the field implementation stage to test its construct validity and reliability on a larger and more heterogeneous sample.

Implementation stage

The revised instrument was integrated into a web-based application through Google Forms and used in a large-scale trial including 498 students from State Senior High Schools 1, 2, 4, 5, 10, 11, and 12 in Banjarmasin. Affective domain abilities of students were measured using five aspects proposed by Krathwohl, namely receiving, responding, valuing, organisation, and characterisation. Table 6 presents the results of the descriptive analysis and correlations among the five aspects of the affective domain, with reference to Krathwohl's taxonomy.

Table 6.

Descriptive statistics and correlations between aspects of students' affective domain

	N	Mean	SD	Rec	Res	Val	Org	Cha
Receiving (Rec)	498	21.40	2.23	1				
Responding (Res)	498	21.30	2.35	0.714	1			
Valuing (Val)	498	22.22	2.11	0.698	0.740	1		
Organisation (Org)	498	16.58	2.12	0.605	0.655	0.637	1	
Characterisation (Cha)	498	17.48	1.98	0.627	0.670	0.721	0.585	1

Source: Author (2026).

The results showed that all aspects had high means and a consistent distribution. Significant positive correlations were observed among the aspects, with the highest being Responding-Valuing ($r = 0.740$) and Valuing-Characterisation ($r = 0.721$), indicating a

progressive relationship among the affective dimensions. All correlations were < 0.85 , signifying no multicollinearity. This pattern supported Krathwohl's affective domain hierarchy, from Receiving $\rightarrow$ Responding $\rightarrow$ Valuing $\rightarrow$ Organisation $\rightarrow$ Characterisation, which reflected the development of attitudes among students from acceptance to internalisation of values and character.

The outcomes of the large-scale field trial, including 498 students from seven secondary schools, indicated that the instrument had good potential for generalisability. The diversity of respondent characteristics and different school contexts supported external validity. The instrument was relevant in a single study setting and also had the potential to be applied to broader educational contexts. However, the generalisability of these results was still limited to the context of public schools with specific characteristics. As a result, further testing would be needed in other types of schools, such as private schools or settings with different social conditions.

Reliability and validity tests were conducted on each dimension of the affective domain to ensure that the instruments used in this study were of adequate quality. Item validity was tested using the item-total correlation coefficient for each statement. An item was considered valid when its correlation coefficient exceeded 0.60, indicating a substantial correlation between the item and the dimension total. Reliability, in turn, was determined using Cronbach's alpha. The results are presented in Table 7.

Table 7.

Validity and reliability testing

Dimensions	Item	Item-total correlation	Reliability coefficient
Receiving	Rec1	0.664	0.731
	Rec2	0.682	
	Rec3	0.675	
	Rec4	0.727	
	Rec5	0.728	
Responding	Res1	0.714	0.783
	Res2	0.715	
	Res3	0.731	
	Res4	0.696	
	Res5	0.776	
Valuing	Val1	0.752	0.804
	Val2	0.702	
	Val3	0.775	
	Val4	0.789	
	Val5	0.733	
Organisation	Org1	0.776	0.719
	Org2	0.748	
	Org3	0.797	
	Org4	0.704	
Characterisation	Cha1	0.738	0.706
	Cha2	0.771	
	Cha3	0.725	
	Cha4	0.724	

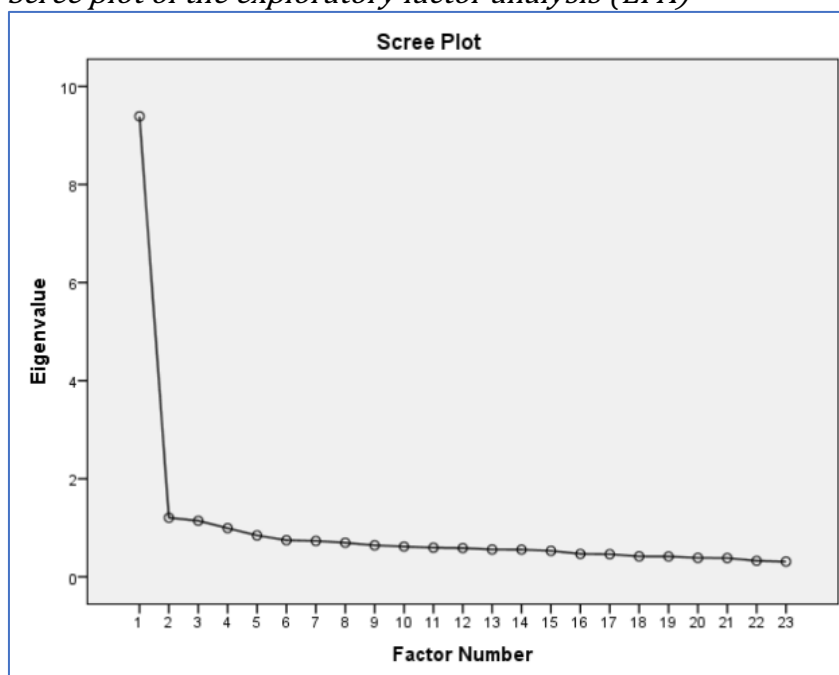
Source: Author (2026).

The findings from the validity and reliability analysis revealed that all the instrument items had high validity (item-total correlation > 0.60) and good internal consistency (Cronbach's alpha 0.706–0.804). The dimension with the highest level of reliability was Valuing followed by Responding, Receiving, Organisation, and Characterisation. This finding supports the reliability of the instrument in measuring various aspects of the affective domain of the students.

The eigenvalues, scree plot (Figure 2) and EFA results were analysed to determine the best factor structure of the instrument. Three factors had eigenvalues greater than 1, namely 9.389, 1.205, and 1.143. However, factor extraction was not based on the eigenvalue criterion alone but on this criterion in combination with the scree plot and factor loadings.

Figure 2.

Scree plot of the exploratory factor analysis (EFA)



Source: Author (2026).

The scree plot revealed an elbow at the first factor, which was then followed by a sharp drop, with relative stability of subsequent factors. This pattern suggested one dominant factor accounting for most of the variability within the data. The discrepancy between the two criteria reflects a common difficulty: when a scale possesses a general factor alongside several meaningful subdimensions, the number of factors is hard to determine. The factor structure was therefore interpreted in light of both the statistical results of the EFA along with the theoretically specified five-dimensional structure of the affective construct.

In order to examine the adequacy of the data prior to carrying out factor analysis, the Kaiser-Meyer-Olkin (KMO) test and Bartlett's Test of Sphericity were carried out. In addition, the Measures of Sampling Adequacy (MSA) were examined. The factor solution was determined based on eigenvalues, the scree plot and factor loading pattern. The results of this test are presented in Table 8.

Table 8.

Data adequacy and exploratory factor analysis

	Result
Kaiser-Meyer-Olkin (KMO)	0.957
Measure of Sampling Adequacy.	
Bartlett's Test of Sphericity	0.000
MSA value	0.962; 0.954; 0.970; 0.950; 0.937; 0.963; 0.967; 0.957; 0.962; 0.940; 0.966; 0.930; 0.958; 0.975; 0.966; 0.926; 0.959; 0.935; 0.974; 0.954; 0.962; 0.929; 0.965
Eigenvalues	9.389; 1.205; 1.143; 0.992; 0.845

Source: Author (2026).

The output shows that the data were suitable for the next step of analysis (KMO = 0.957; $p < 0.05$). All MSA scores were greater than 0.90, indicating adequate sampling properties. The

data could therefore be analysed further using EFA. EFA was conducted to analyse the data's underlying factor structure. The factor matrix (Table 9) shows how strongly each item loads on the retained factors; higher loadings indicate a stronger relationship between the item and the factor.

Table 9.

Exploratory factor analysis matrix

Dimensions and Items	Factors				
	1	2	3	4	5
Receiving: Rec1	0.581				
Rec2	0.497				
Rec3	0.605				
Rec4	0.558				
Rec5	0.597				
Responding: Res1		0.602			
Res2		0.633			
Res3		0.662			
Res4		0.640			
Res5		0.644			
Valuing: Val1			0.663		
Val2			0.566		
Val3			0.677		
Val4			0.720		
Val5			0.674		
Organisation: Org1				0.363	
Org2				0.668	
Org3				0.680	
Org4				0.681	
Characterisation: Cha1					0.659
Cha2					0.689
Cha3					0.467
Cha4					0.613

Source: Author (2026).

Based on the EFA results, many items had high factor loadings (above 0.50), indicating strong relationships between items and factors. Although some items had low factor loadings, they were retained because of their theoretical relevance and the absence of cross-loading. The next step was CFA, used to confirm the hypothesised five-factor structure based on the EFA results and Krathwohl's affective taxonomy. This hypothesis was confirmed by evaluating various fit indices: absolute, incremental, and parsimonious fit. The analysis indicated an acceptable fit with a CMIN/df ratio of 2.545, which falls within the acceptable range (2–3). Most other indices (GFI, IFI, TLI, CFI, RMSEA and RMR) also met the model criteria (Table 10).

Table 10.

Evaluation of confirmatory factor analysis model fit

Fit Index	Value	Interpretation
CMIN/df	2.545	Acceptable Fit
GFI	0.906	Acceptable Fit
AGFI	0.882	Acceptable Fit
IFI	0.927	Acceptable Fit
TLI	0.916	Acceptable Fit
CFI	0.927	Acceptable Fit
RMR	0.018	Good Fit
RMSEA	0.056	Acceptable Fit

Source: Author (2026).

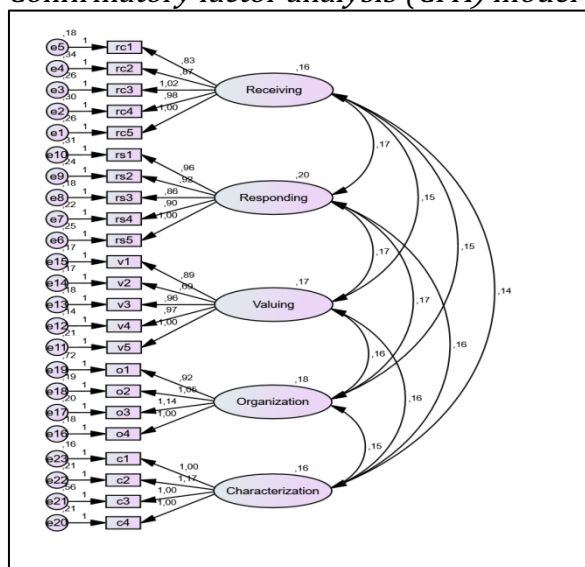
CFA results indicated a reasonable fit between the model and the data, though not perfect. CMIN/df (2.545), GFI (0.906), IFI (0.927), TLI (0.916), CFI (0.927), RMR (0.018), and RMSEA

(0.056) supported adequate model fit. However, AGFI (0.882) was slightly below the stricter 0.90 criterion. We retained this value and reported it transparently rather than modifying it mechanically because the model represented a theoretically grounded affective construct with five related dimensions. Modification indices were not used to force a better fit unless theoretically justified, in order to avoid overfitting the model to the present sample.

Based on the model-fit indices, the five-dimensional model showed an acceptable fit with the empirical data. The findings supported representing the affective domain through five dimensions (Receiving, Responding, Valuing, Organisation, and Characterisation) and supported the multidimensional structure of the measurement model.

Figure 3.

Confirmatory factor analysis (CFA) model of the affective domain structure



Source: Author (2026).

Figure 3 shows the affective domain's structural model, consisting of five main constructs: Receiving, Responding, Valuing, Organisation, and Characterisation. Each construct is measured by several indicators, each with a loading that indicates its contribution to the latent construct. The constructs are also correlated, indicating strong relationships among affective dimensions and forming an interrelated structure.

Evaluation stage

The evaluation stage in the ADDIE development model was conducted continuously at each phase through formative evaluation, with revisions. This evaluation aimed to ensure that the developed instrument met the criteria of validity, reliability, and practicality. The evaluation process involved experts in assessment, linguistics, value education, and civics education. Following this process, experts assessed the instrument's content and construction, language, and practicality. Expert input was used to revise instrument items, making them clearer, more operational and better aligned with the constructs being measured.

Evaluation was conducted through field trials to assess the performance of the instrument in measuring affective domains of students under various conditions, including gender and school of origin. Based on the entire evaluation and revision process, the instrument was refined in terms of content, construction, and language. The final revision led to a validated self-assessment instrument specifically designed to measure affective competencies of students.

Evaluation results: Students' affective domain skills by gender

The respondents in this study were 498 students, comprising 285 females (57.2%) and 213 males (42.8%). Figure 4 shows students' affective domain abilities in Pancasila Education learning by gender.

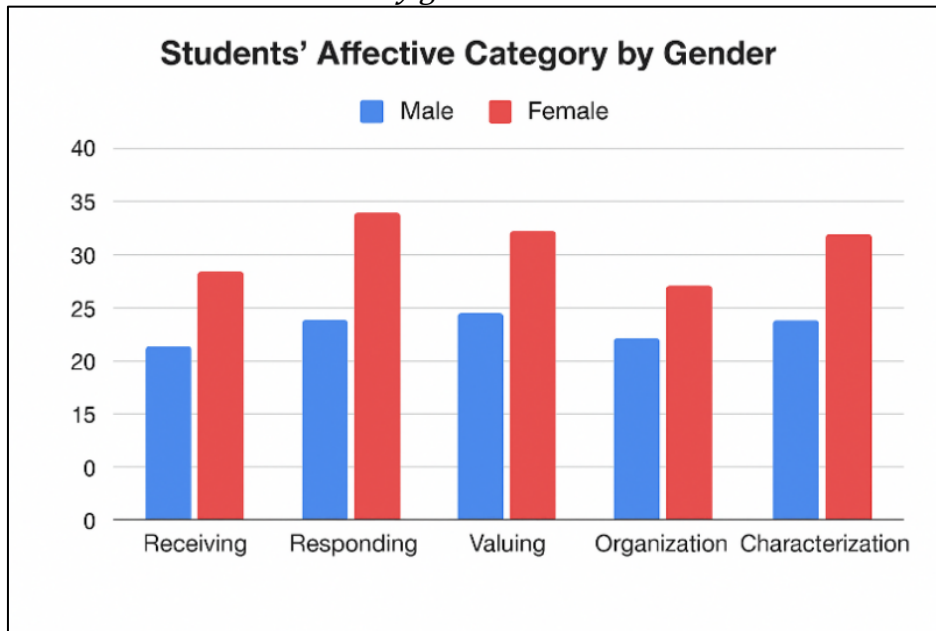
Figure 4.*Students' affective abilities by gender**Source: Author (2026).*

Figure 4 shows that female students demonstrated stronger affective domain abilities across all dimensions, including Receiving, Responding, Valuing, Organisation, and Characterisation. This finding is consistent with previous international studies reporting that female adolescents tend to exhibit higher levels of empathy, emotional awareness, and socio-emotional competencies than their male counterparts. According to Trentini et al. (2022), adolescent girls performed better than their male counterparts when evaluated in terms of different aspects of empathy, especially empathy concerns and emotional responsiveness. Additionally, females generally exhibit stronger social and emotional skills in areas such as interpersonal sensitivity, emotional regulation, and prosocial behaviours, which are linked to affective domain development (Huang et al., 2024). This suggests that gender differences in affective domain competencies may be attributable to differences in emotional socialisation, empathy development and value internalisation processes.

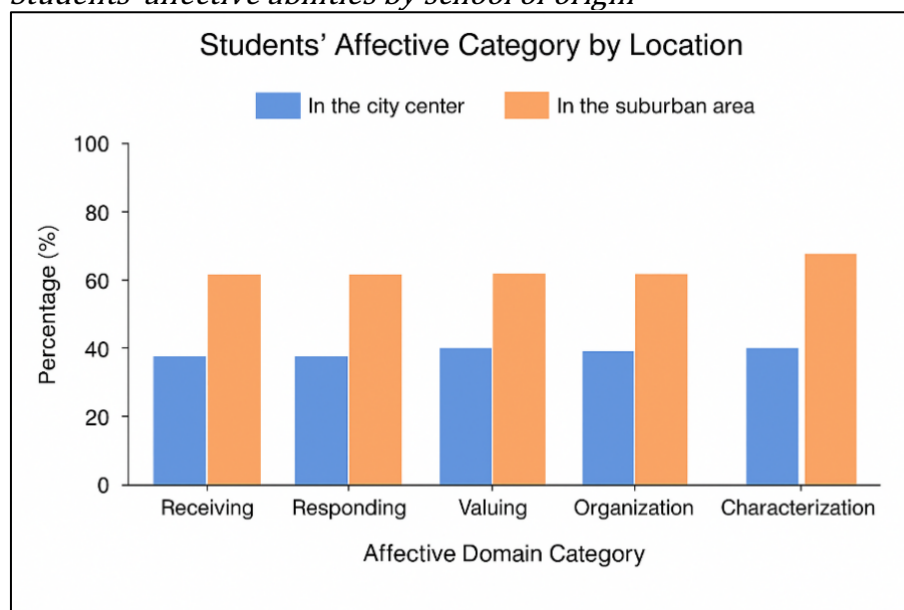
Evaluation results: Students' affective domain abilities by school of origin

The schools used in this study were in the city centre and on the outskirts of Banjarmasin. The results indicated that affective domain abilities of students differed by school of origin, as shown in Figure 5. Figure 5 shows that students from suburban schools demonstrated higher affective-domain scores than those from urban schools. However, the present study was not designed to identify the causal mechanisms underlying these differences. These differences should therefore be interpreted with caution, as they may result from varying school conditions rather than location per se.

Many contextual factors influence affective development. Nonetheless, the observed differences align with the literature indicating that students' social environment influences their socio-emotional and value development. In particular, research shows that students living in less urbanised areas are characterised by high levels of social connectedness, strong community relations, and higher feelings of attachment to schools that contribute to their positive affective development (Allen et al., 2017; Jose et al., 2012). On the other hand, students living in highly urbanised areas face more academic pressure, social competition, and psychological stress. These factors influence emotional wellbeing and social adjustment of adolescents (OECD, 2021). Although these factors may partly explain the differences in the current study, additional research is needed.

Figure 5.

Students' affective abilities by school of origin



Source: Author (2026).

Discussion

The results show that the assessment of affective competence requires both theoretical grounding and psychometric robustness. Taken together, the findings provide evidence of the content validity, construct validity and internal consistency of the instrument. Generally speaking, the results confirm the validity of the conceptual underpinning of Krathwohl's taxonomy but, at the same time, they point out that civic affective competence can be seen as a multidimensional construct. Furthermore, the reliability and model fit of the scale provide some support for the proposed multidimensional model of measuring students' affective development in Civic Education.

A more detailed analysis of the factor structure helps to better understand how the empirical data corresponds to the Krathwohl taxonomy. The descriptive statistics and inter-factor correlations confirm the connections between the five affective dimensions: Receiving, Responding, Valuing, Organisation, and Characterisation. The strongest correlations were found between Responding and Valuing, as well as Valuing and Characterisation, indicating that active engagement with civic values is related to their higher level of internalisation. These results are consistent with Krathwohl's taxonomy, which postulates that affective development goes from awareness and participation to commitment and value-directed action (Krathwohl et al., 1964). More recent studies suggest that people progress from value awareness to the integration and self-regulated enactment of values through reflection and social experience (Berkowitz & Bier, 2014; Mahoney et al., 2021).

These findings extend the theory to Pancasila Education, showing that the affective elements of civic education can be measured using a multidimensional framework of assessment. The digital self-assessment tool can therefore give students an opportunity to evaluate themselves in terms of their civic values and their development along the affective dimensions, while giving teachers a form of evaluation beyond the cognitive domain. This is because learning outcomes in Pancasila Education include both the conceptual knowledge and the values of civics.

Factor loadings are also in line with the theoretical expectations. In most cases, items in the Receiving, Responding, Valuing, and Characterisation dimensions showed good to excellent loadings, and this means that these measures effectively measured their constructs. The presence of distinct groups of items in each of the five theoretical dimensions supports the construct of Krathwohl. Nevertheless, the results did not show five independent factors but rather correlations among the dimensions, implying that they reflect a broader civic affective

disposition. This finding supports a hierarchical perspective where the five dimensions are stages of affective development rather than traits. Such results are expected because, from current perspectives, affective competence is a multidimensional yet integrated construct, which involves emotional regulation, value commitment, self-reflection, and behaviour (Immordino-Yang & Damasio, 2007; Mahoney et al., 2018).

One notable finding concerns the Organisation dimension, whose loadings were considerably lower compared to several other dimensions. Theoretically, this can be explained by the fact that Organisation represents a more mature level of affective development. At this level, students are expected not only to adopt individual values but also to synthesise several different values in order to create a value system of their own. This requires higher-order thinking and self-control, and such processes have not been fully developed in secondary school students. Research in the field of developmental and educational psychology indicates that such synthesis of values into a consistent personal framework is a process that occurs gradually through social interaction, moral reasoning and reflective practice (Bandura, 1991; Schwartz, 2012). In this sense, responses to the Organisation dimension will likely exhibit more variance and inconsistency compared to the responses to dimensions like Receiving, Responding, and Valuing.

Lower loadings for Organisation may indicate that value integration is more context-bound than other affective processes. Students can support civic values as ideals without problems, but they can experience difficulties in combining various values in social contexts. This interpretation is supported by research showing that value internalisation takes place gradually by engaging in meaningful social experiences, contemplating moral problems, and performing civic acts in real life (Nucci et al., 2014; Mahoney et al., 2021). For this reason, the lower loadings need not be interpreted as a sign of poor measurement but rather as an illustration of developmental difficulty at this level in Krathwohl's scheme. The result emphasises the point that higher-level affective dimensions are inherently harder to measure since they imply both attitude formation and value integration.

The benefits of the digital self-assessment format included standardised administration, efficient gathering of data, and increased privacy for students in reporting their attitudes and value orientations. However, digital administration does not guarantee honest answers. Because civic and character constructs are socially desirable, students may respond according to what is expected of them rather than what they actually believe. The instrument should thus be regarded as one way of obtaining information about students' affective development.

In practical terms, the instrument allows teachers to transition from subjective assessment of attitudes to informed guidance of education. Students with low Receiving scores may need activities that develop awareness and open-mindedness, while those with low Responding scores may require activities that foster involvement. Low Valuing scores could signal the necessity of reflection activities linking civic values to daily life, and low scores on Organisation and Characterisation may call for mentoring, habituation, and reinforcement through cooperation of schools and parents. Consequently, the purpose of the instrument is not classification but diagnosis of character and values internalisation.

Findings also contribute to the wider global conversation about affective, civic, and character education assessment. Although Pancasila Education is specific to the Indonesian context, the problem of measuring affective competencies remains a global challenge. Educational institutions around the world are increasingly focused on teaching their students about ethics, citizenship, socio-emotional competencies, and character education; however, many countries still do not have psychometrically proven tools to measure such outcomes. Thus, this study provides an example of a civic affective scale that might be useful for further research related to Global Citizenship Education and character assessment. The validity evidence should be considered critically. The rather high mean scores for all the dimensions indicate the risk of a ceiling effect, meaning that the tool may not be able to distinguish between

different levels of affective competence in students who have already achieved quite high results. Further work on developing the measuring instrument should include creating more complex behavioural items, especially in the Organisation and Characterisation dimensions.

In general, the results provide empirical justification for the multidimensional nature of civic affective competence and illustrate the promise of digital self-assessment as an instrument for educational evaluation. Although further validation is still needed in different populations and cultures, the instrument exhibits good psychometric qualities and practicality in terms of systematic assessment of the development of students' affective skills.

Limitations and Future Research

This study has several limitations. First, the participants were drawn from public senior high schools in Banjarmasin, South Kalimantan only. Thus, the results cannot yet be generalised to private schools, vocational schools, other provinces, or different cultural contexts without further testing. Second, the instrument adopted a self-report approach that is susceptible to social desirability bias, particularly because civic and character values are socially desirable. Although the digital format and non-punitive instructions were intended to encourage honest responses, further studies should combine self-reporting with teacher reporting, peer reporting, observations, and portfolios. Third, EFA and CFA were conducted consecutively on a single field-trial dataset, which provided some valuable information about construct validity. However, the combination is less powerful than cross-validation. Future studies should use split-sample validation, a new independent sample and multi-group CFA to test measurement invariance across gender, school type, region, and culture. Finally, some items showed a strong tendency towards high responses, suggesting a ceiling effect. Further development of the tool should include more discriminating items on the organisation and characterisation level.

Conclusion

This research developed and validated a digital self-assessment instrument to measure students' affective domain in Pancasila Education. The instrument demonstrated adequate content validity, empirical item validity, and internal consistency with Cronbach's alpha coefficients from 0.706 to 0.804. The results of EFA and CFA were consistent with the hypothesis of hierarchical multidimensionality, revealing one general factor and five meaningful factors corresponding to the categories of Krathwohl's taxonomy: receiving, responding, valuing, organisation, and characterisation. The key contribution of this study lies in the combination of Krathwohl's affective taxonomy, attitude/social cognitive approaches, digital self-assessment, and psychometric evaluation in the field of civic education. The instrument can help teachers obtain systematic data on students' affective development and use this information for designing reflection, character building, and value internalisation activities. The findings should be interpreted in light of the study's methodological limitations. Self-assessment may introduce social desirability bias, the participants were geographically limited, and CFA was performed on the same sample as the exploratory analysis. Further research should focus on cross-cultural validation, split-sample/independent-sample CFA, measurement invariance tests, and item refinement.

Authors' contributions

Rabiatul Adawiah formulated the research concept, developed the research design, coordinated the research process, contributed to developing the digital self-assessment instrument, carried out data analysis, and prepared the draft manuscript. Dian Agus Ruchliyadi contributed to the research methodology, instrument validation, data collection, and critical revision of the manuscript. Nurul Huda contributed to data collection, the literature review, and interpreting the research findings. Yonas Belay Gela contributed to developing the theoretical framework, interpreting the research results, and editing the manuscript.

Competing interests

The authors declare no conflicts of interest.

Funding Statement

This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Use of artificial intelligence (AI)

OpenAI ChatGPT was used during the revision of this manuscript solely as an assistive tool for language editing, consistency checking, reviewing statistical code and outputs, and supporting the drafting of revisions. The authors independently verified all source data, calculations, citations, statistical results, interpretations, and revised text. All scientific judgments and decisions were made by the authors, who take full responsibility for the accuracy, integrity, and content of the published

Ethics Approval

The research was conducted in accordance with the ethical principles of social and educational research. The Research Ethics Committee of the Institute for Research and Community Service (LPPM), Lambung Mangkurat University, Indonesia granted ethical approval (Approval Number 2890/UN8.2/PG/2025, dated 8 August 2025, valid until 8 August 2026).

References

- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131–142. <https://doi.org/10.1177/0013164485451012>
- Allen, K., Vella-Brodrick, D., & Waters, L. (2017). School belonging and the role of social and emotional competencies in fostering an adolescent's sense of connectedness to their school. In E. Frydenberg, A. J. Martin, & R. J. Collie (Eds.), *Social and emotional learning in Australia and the Asia-Pacific: Perspectives, programs and approaches* (pp. 83–99). Springer. https://doi.org/10.1007/978-981-10-3394-0_5
- Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, 4, Article 87. <https://doi.org/10.3389/feduc.2019.00087>
- Bandura, A. (1991). Social cognitive theory of self-regulation. *Organizational Behavior and Human Decision Processes*, 50(2), 248–287. [https://doi.org/10.1016/0749-5978\(91\)90022-L](https://doi.org/10.1016/0749-5978(91)90022-L)
- Berkowitz, M. W., & Bier, M. C. (2014). Research-based fundamentals of the effective promotion of character development in schools. In L. Nucci, D. Narvaez, & T. Krettenauer (Eds.), *Handbook of moral and character education* (2nd ed., pp. 248–260). Routledge. <https://doi.org/10.4324/9780203114896.ch14>
- Branch, R. M. (2009). *Instructional design: The ADDIE approach*. Springer. <https://doi.org/10.1007/978-0-387-09506-6>
- Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: The reasoned action approach*. Psychology Press. <https://doi.org/10.4324/9780203838020>
- Huang, Z., Zhang, J., & Yuan, Z. (2024). What matters for students' learning? The roles of social and emotional skills and the factors affecting these skills. *ECNU Review of Education*. Advance online publication. <https://doi.org/10.1177/20965311241227453>
- Immordino-Yang, M. H., & Damasio, A. (2007). We feel, therefore we learn: The relevance of affective and social neuroscience to education. *Mind, Brain, and Education*, 1(1), 3–10. <https://doi.org/10.1111/j.1751-228X.2007.00004.x>

- Jose, P. E., Ryan, N., & Pryor, J. (2012). Does social connectedness promote a greater sense of well-being in adolescence over time? *Journal of Research on Adolescence*, 22(2), 235–251. <https://doi.org/10.1111/j.1532-7795.2012.00783.x>
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of educational objectives: The classification of educational goals. Handbook II: Affective domain*. David McKay.
- Lavidas, K., Papadakis, S., Manesis, D., Grigoriadou, A. S., & Gialamas, V. (2022). The effects of social desirability on students' self-reports in two social contexts: Lectures vs. lectures and lab classes. *Information*, 13(10), Article 491. <https://doi.org/10.3390/info13100491>
- Mahoney, J. L., Durlak, J. A., & Weissberg, R. P. (2018). An update on social and emotional learning outcome research. *Phi Delta Kappan*, 100(4), 18–23. <https://doi.org/10.1177/0031721718815668>
- Mahoney, J. L., Weissberg, R. P., Greenberg, M. T., Dusenbury, L., Jagers, R. J., Niemi, K., Schlinger, M., Schlund, J., Shriver, T. P., VanAusdal, K., & Yoder, N. (2021). Systemic social and emotional learning: Promoting educational success for all preschool to high school students. *American Psychologist*, 76(7), 1128–1142. <https://doi.org/10.1037/amp0000701>
- Memon, M. A., Ting, H., Cheah, J.-H., Ramayah, T., Chuah, F., & Cham, T. H. (2020). Sample size for survey research: Review and recommendations. *Journal of Applied Structural Equation Modeling*, 4(2), i–xx. [https://doi.org/10.47263/JASEM.4\(2\)01](https://doi.org/10.47263/JASEM.4(2)01)
- Nucci, L., Narvaez, D., & Krettenauer, T. (Eds.). (2014). *Handbook of moral and character education* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203114896>
- OECD. (2021). *Beyond academic learning: First results from the Survey of Social and Emotional Skills*. OECD Publishing. <https://doi.org/10.1787/92a11084-en>
- Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29(5), 489–497. <https://doi.org/10.1002/nur.20147>
- Schwartz, S. H. (2012). An overview of the Schwartz theory of basic values. *Online Readings in Psychology and Culture*, 2(1), Article 11. <https://doi.org/10.9707/2307-0919.1116>
- Setiawati, F. A., Widyastuti, T., Fathiyah, K. N., & Nabila, T. S. (2024). Minimizing social desirability in questionnaires of non-cognitive measurements. *European Journal of Psychology and Educational Research*, 7(1), 33–43. <https://doi.org/10.12973/ejper.7.1.33>
- Trentini, C., Tambelli, R., Maiorani, S., & Lauriola, M. (2022). Gender differences in empathy during adolescence: Does emotional self-awareness matter? *Psychological Reports*, 125(2), 913–936. <https://doi.org/10.1177/0033294120976631>
- UNESCO. (2015). *Global citizenship education: Topics and learning objectives*. <https://unesdoc.unesco.org/ark:/48223/pf0000232993>
- Yan, Z., & Carless, D. (2022). Self-assessment is about more than self: The enabling role of feedback literacy. *Assessment & Evaluation in Higher Education*, 47(7), 1118–1128. <https://doi.org/10.1080/02602938.2021.2001431>