Al-Ishlah: Jurnal Pendidikan
Vol.15, 1 (March, 2023), pp. 953-963
ISSN: 2087-9490 EISSN: 2597-940X, DOI: 10.35445/alishlah.v15i1.2980

Development of Summative Assessment Instruments for the Affective Aspects of Learning History

Syarifah
Universitas Negeri Medan, Medan, Indonesia; syarifahahmad6@gmail.com

ARTICLE INFO

Keywords:

instrument development; affective competencies; values history.

Article history:

Received 2022-11-25
Revised 2022-12-29
Accepted 2023-02-09

ABSTRACT

Instruments for measuring both the affective and psychomotor aspects of teaching are required. The study's overarching goal is to create a reliable and valid assessment tool for measuring affective competency (values contained in history). Nine hundred thirty-five students from Medan's XI Social Sciences high school participated as research subjects. An affective aspect assessment sheet, which is a questionnaire with various statements to gauge affective aspects, was used to collect research data. To compare measurement models and make it simpler to analyze model correctness, structural equation modelling (SEM) analysis with first- and second-order components was employed in the data analysis. The findings of the first-order confirmatory factor analysis using the ML (maximum likelihood) technique indicated that all items had a correlation coefficient greater than 0.32, and the results of testing using the model had a significance level of 0.0001, according to the research analysis. With support from second-order structures, the structural equation model (SEM) findings are also estimated with adequate theoretical confidence. The construct reliability (CR) and variance extract (VE) coefficients indicate that the designed instrument complied with the reliability coefficient acceptability limits of 0.70 for CR and 0.50 for VE. The development process resulted in an affective competency assessment tool with 84 valid and reliable items spread across 5 dimensions and 14 indicators. The instruments created were therefore shown to be valid and reliable for assessing affective aspects of learning history. The findings of this study are utilized to diagnose student potential, particularly concerning affective components, so that students are aware of their strengths and weaknesses. The findings of this study are utilized by stakeholders as input for policymaking, particularly in the area of education.

Corresponding Author:
Syarifah
Universitas Negeri Medan, Medan, Indonesia; syarifahahmad6@gmail.com

1. INTRODUCTION

History is a subject that is taught in schools from elementary through high. The purpose of learning history in schools is to foster a feeling of nationality, historical awareness, historical thinking skills, the development of moral and ethical conduct, and the formation of the present- and future-focused attitudes (Henderson, 2019; Mathis & Parkes, 2020; Sakki & Pirttilä-Backman, 2019). According to the Merdeka curriculum, learning history in schools aims to foster historical awareness, nurture nationalism, foster chronological thinking, and build collective memory as a nation so that we know our people and serve as a foundation for building a sense of unity. This is in line with the aforementioned opinion. Based on the aforementioned goals of history education, it can be said that teaching history in schools aims to help students develop their attitudes, actions, skills, and cognitive capacities. To promote the development of these three elements, proper and adequate learning aids are required.

However, history learning in schools is still focused on cognitive components, specifically those that are just connected to historical facts (Wansink et al., 2018; Yildirim et al., 2018). It appears that the affective component associated with the ideals expressed in the historical event itself has been overlooked. This is evident from the assessment of learning outcomes, which exclusively considers cognitive factors. Although the purpose of learning history is primarily cognitive, affective learning also plays a significant role. One of the effective elements for achieving the goal of developing an educated generation that is enthusiastic, ardent, and supportive, as well as having and upholding the nation's noble ideals, is teaching history in schools (Saputra et al., 2018). In this context, the values in issue are those of a person's or society's life as a guide for choosing behaviours and attitudes. They include principles, norms, ideals, and worldviews. The values included in the Indonesian people's history are the conceptual soul and spirit that create the Indonesian people's beliefs, desires, and shared objectives with all the effectiveness that impacts not actions.

It would be impossible to determine if the curriculum's learning objectives had been met without an assessment in learning. Assessment is therefore an essential component of the learning process (P. Black & Wiliam, 2018; Papadakis et al., 2020). The process of assessment in learning is important, among other things, to determine students' mastery of the material being studied, to know students' potential, to increase students' motivation, to diagnose students' learning difficulties, to provide evidence of competence that students have mastered, and as a benchmark for improving future learning programs(P. Black & Wiliam, 2018; Hansem et al., 2021; Ningsih et al., 2019; Ofianto & Suhartono, 2016; Reisman et al., 2019; Sajjad et al., 2018; Sya'bandari et al., 2022). Teachers must be able to create assessment instruments that can measure cognitive, affective, and skill-related factors (Ofianto et al., 2022; Ofianto & Ningsih, 2021). Students' capacities to absorb, respond to, evaluate, organize, internalize, and uphold historical values—both fundamental and operational values—are considered aspects of their affective competence in the study of history.

According to the findings of the researchers' observations, teachers in schools frequently exclusively assess students' cognitive abilities through examinations, leaving out other factors. According to the findings of the researchers' observations and interviews with several history teachers in Medan, it was discovered that teachers found it challenging to assess the affective components of history learning outcomes tests because these components included students' appreciation and practice. It is challenging for written examinations, which are frequently used, to accurately depict this affective element. Based on the above issues, this research aims to develop a summative assessment instrument to measure the affective components of learning history. The urgency of this research is to provide other techniques for assessing students' potential, particularly in affective aspects. This is due to the fact that the affective component is a crucial component that the teacher must take into account and develop while helping students learn about history.

2. METHODS

This research utilizes a development research technique, which includes several stages of research, including (i) issue analysis through literature reviews and field observations; (ii) instrument design for affective assessment; (iii) validity and reliability testing; and (iv) instrument trials. According to theory, testing the validity and reliability of the instrument includes evaluating the fit between variables and dimensions, the fit between indicators and dimensions, and the fit between items and indicators by 20 individuals statistically and five individuals qualitatively. Using a Likert scale of 1 to 5, perform quantitative analysis using the formula below (Zhou, 2013).

$$V = \frac{\sum n_i \left| i - r \right|}{N(t-1)},$$

with $i = (r+1)$ to $(r+t-1)$, $N = \sum n_i$

Information :
r = Floor value
t = Ceiling Value
n_i = number of expert choices on value i

Reliability interrater

$$r_{11} = \frac{RK_b - RK_e}{RK_b}$$

Information:
RK_b = Average Squared Rows (rater)
RK_e = Mean Squared Error

If the r_{xx} value is greater than 0.7, Hoyt's intermediate reliability with reliable criteria is used, and Aiken's validity with valid criteria is used (Zhou, 2013).

Nine hundred thirty-five students in Social Senior High School's class XI participated as research participants. Affective aspect test sheets in questionnaires with several stations according to the indicators to be measured were used to collect research data. In total, 515 respondents participated in trial 1, and 420 respondents in stage 2. This was done after the theoretical validation stage. The number of instrument items tested—102 in the first stage and 84 in the second—was used to calculate the number of responders. In response to a statement on the outcomes of studying history, respondents self-assessed how much they agreed with the statement based on the condition or affective competency of learning history. A Likert scale with four options and categories—strongly agree (SA), agree (A), disagree (D), and strongly disagree (SD)—was employed as the scale model for this research instrument.

SEM (structural equation modelling) is used to analyze research data. This was followed by second-order confirmatory factor analysis (CFA) using the standard loading factor (SLF) or load factor 0.32 with T-values of 1.96 or 2.00. While the construct dependability (CR) and variance extracted (VE) are used by the following formula in the SEM reliability test:

$$CR = Construct\ Reliability = \frac{(\sum \lambda)^2}{(\sum \lambda)^2 + (\sum \varepsilon)}$$

$$= \frac{(\sum standardized\ loading)^2}{(\sum standardized\ loading)^2 + (\sum error)}$$

or

$$VR = VarianceExtracted = \frac{(\sum \lambda^2)}{(\sum \lambda^2) + (\sum \varepsilon)}$$
$$= \frac{(\sum s\tan dardized\ loading^2)}{(\sum st\,an\,dardized\ loading^2) + (\sum error)}$$

A construct has strong reliability, according to Hair et al. (1995), if the values of CR and VE are both above 0.70. Additionally, tests are done to determine how well the measurement model and the entire model fit together. Different goodness of fit (GOF) measurements are used to test the model's overall fit.

As indicated in table 1, the results of a qualitative assessment of students' affective competence in history may be categorized into four categories: very excellent, good, sufficient, and bad (Azwar, 2013).

Table 1. The Range of Affective Competency Values in High School History

Score Interval	Criteria
Mi + 1,5 SDi ≤ $\bar{M}$ ≤ Mi + 3,0 SDi	Very excellent
Mi + 0 SDi ≤ $\bar{M}$ ≤ Mi + 1,5 SDi	Good
Mi - 1,5 SDi ≤ $\bar{M}$ ≤ Mi + 0 SDi	Sufficient
Mi - 3 SDi ≤ $\bar{M}$ ≤ Mi – 1,5 SDi	Bad

The research strategy is shown in Figure 1 below:

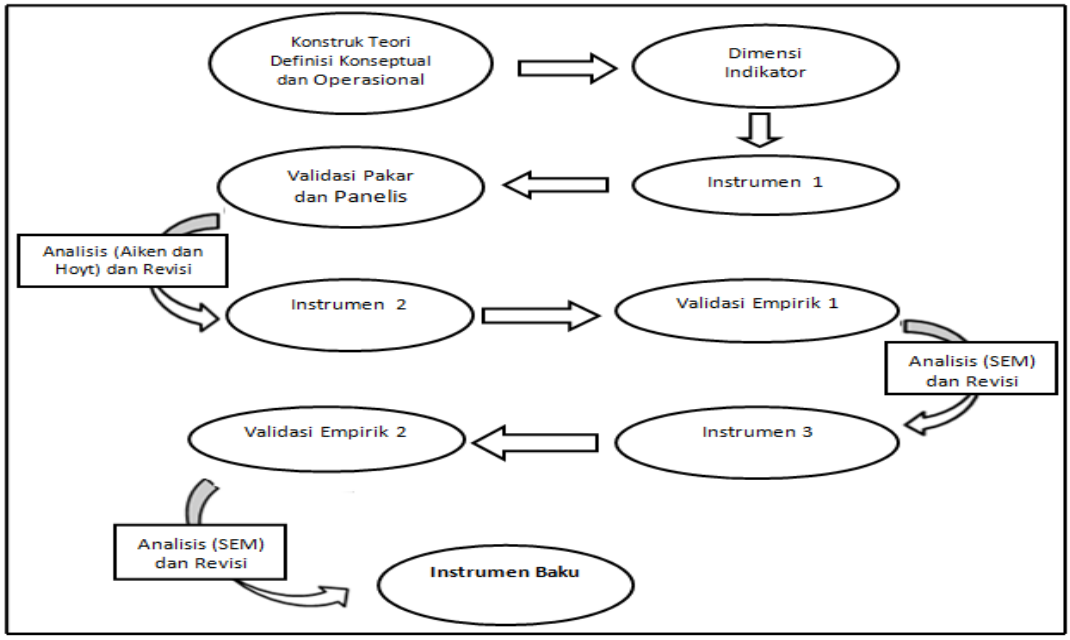

Figure 1. Design of Instrument Development Research

3. FINDINGS AND DISCUSSION

3.1 Findings

The affective assessment instrument developed consists of 5 dimensions, namely: (i) awareness of the significance of time and place in history (R); (ii) appreciation of historical heritage (S); (iii) awareness

of unity, togetherness, and solidarity in the face of the threat of national disintegration (T); (iv) pride in Indonesia as a nation and love for the motherland (U); and (v) awareness of heroic values (V). From these five dimensions, 110 items and 14 indicators are created. The validity of each item in this affective aspect assessment instrument is examined to determine whether it can be used to measure various aspects of affective competency.

The instrument's findings from the theoretical construct were verified by experts using qualitative analysis, followed by panellists using quantitative analysis, which led to the identification of 8 invalid items: items 1, 10, 28, 68, 70, 85, 86, and 97. This is due to the fact that the Aiken value is less than 0.2, making it possible to declare that all of the instrument's components are invalid. As a result, there are 102 legitimate things and 8 invalid items out of a total of 110 items. A reliability value of $r_x = 0.992$ was acquired from further testing to determine the interrater reliability (r_{11}) using the Hoyt formula. All of the questions in the affective competency assessment instrument on values connected to high school history learning are considered trustworthy because the findings of r_{xx} are greater than 0.7.

Validity and reliability tests were conducted to determine how consistently the manifest variable measured its latent components. In order to compare measurement models and make it easier to investigate model correctness, structural equation modeling (SEM) analysis was performed to determine to construct validity in the trial 1 and trial 2 stages. Valid indicator criteria (items) are tested using a goodness-of-fit model after being experimentally established using the standard loading factor (SLF) value or load factor 0.32 with a T-value of 1.96 or 2.00 for the outcomes of analysis with SEM.

Based on the results of the first order test stage 1, it was determined that 84 items were valid since they had factor loading factors over 0.32, whereas 18 items—items X12, X15, X21, X23, X35, X37, X43, X48, X49, X50, X61, X72, X88, X93, X97, X99, X101, and X102—were discarded. This leads to the conclusion that only 84 of the 102 things in the first order were valid for the first test, but all 84 items were valid for the second test. Second-order confirmatory factor analysis is the next stage. Only the 84 item factor scores will be used because the 84 items in the first factor CFA analysis were already determined to be legitimate. The item value that is free of measurement error is indicated by the score factor.

Additionally, a variety of goodness-of-fit (GOF) measurements were used to further evaluate the factor analysis results in terms of the model's suitability. The second step of the test involved doing the same. Table 2 below shows a general recapitulation of trial models 1 and 2:

Table 2. Result of Validity and Reliability

	Dimensions	**Confirmatory Faktor Analysis(CFA)**				**Uji Goodness of Fit Test**		
		Validity		**Reliability**				
		Factor	**t-value**	**CR**	**VE**	**Chi-Square**	**Df**	**p-value**
Phase of Trial		**loading (λ)**						
1	R	1,000	28,239	0.911	0,837	1618,95	73	0,00000
	S	0,977	24,782	0.806	0,677	1618,95	73	0,00000
	T	0,859	21,833	0,941	0,841	1618,95	73	0,00000
	U	0,997	19,204	0,939	0,796	1618,95	73	0,00000
	V	0,986	28,589	0,857	0,676	1618,95	73	0,00000
2	R	1,000	28,188	0.975	0,951	1758,00	74	0,00000
	S	0,903	18,813	0.825	0,702	1758,00	74	0,00000
	T	0,821	20,098	0,962	0,894	1758,00	74	0,00000
	U	1,000	25,504	0,979	0,922	1758,00	74	0,00000
	V	1,000	28,326	0,969	0,915	1758,00	74	0,00000

Table 2 shows that the test specifications for validity demonstrate that each of these dimensions has a valid standard loading factor (SLF) or factor loading value of 0.32 with a T-value of 1.96 or 2.00. The dimensions of the trial findings, both at trial stage 1 and at trial stage 2, may be determined to be valid. The reliability index coefficient was above the theoretical acceptance limit, namely 0.70 for CR and 0.5 for VE, indicating that the reliability test in the empirical trials in the first and second phases was also reliable for all dimensions.

Table 3. Recapitulation of Variable Reliability and Validity of Affective Competence

No	Trial	Competency	Validity	Reliability		Determination
				CR	VE	
			(Factor Loading)			
1	I	Affective	0,869-1,000	0,986	0,932	Valid and reliable
2	II	Affective	0,821-1,00	0,886	0,650	Valid and reliable

Additionally, trial stages 1 and 2 were conducted to assess the model's overall applicability using goodness of fit (GOF) metrics for the outcomes of the SEM analysis. The output analysis of measurement findings using Lisrel on the empirical data of trials 1 and 2 using second-order CFA factor analysis and the results of different goodness of fit (GOF) measurements are summarized in Table 4.

Table 4. Results of the Overall Fit Test Model Trial 1 and Trial 2

Goodness of Fit	*Cut of value*	Result of Trial 1	Criteria	Result of Trial 2	Criteria
NFI	≥ 0,9	0.937	*Good*	0.929	*Good*
NNFI	≥ 0,9	0.925	*Good*	0,916	*Good*
CFI	≥ 0,9	0.940	*Good*	0.931	*Good*
IFI	≥0.9	0.940	*Good*	0.931	*Good*
RFI	≥0.9	0.922	*Good*	0,912	*Good*

According to Table 4 above, five out of the model's 18 GOF measures meet the criteria for good fit, as evidenced by the fulfilment of a number of fit model sizes, including the Normed Fit Index (NFI), Non-Normed Fit Index (NNFI), Comparative Fit Index (CFI), Incremental Fit Index (IFI), and Relative Fit Index (RFI) for trials 1 and 2.

3.2 Discussion

The study's findings lead to the conclusion that the affective aspects of studying history may be evaluated by history teachers using the affective competency assessment tool, which has been created. Its validity and reliability scores are in the high category. The measures developed in this study give teachers another way to assess students' affective competencies in relation to historical values while they learn history in high school, allowing for the measurement of all student actions, including affective competencies. The findings of this study are consistent with several studies that demonstrate the necessity for an accurate and trustworthy affective competency assessment instrument to measure attitude components of history learning (Efendi et al., 2021; Hensen & Barbera, 2019; A. O. Wulandari et al., 2019). The curriculum's affective assessment instruments are quite complex, and evaluating students with them takes a very long time and procedure. In order to provide answers for history teachers in schools, it was very appropriate to construct this affective competency assessment

instrument. According to the study by Kuntoro and Wardani (2020), a precise assessment tool is required to measure the cognitive components of learning history.

In total, 102 out of 110 developed items were made after receiving some form of conceptual validation from experts and panellists. Then, 102 items that have already been shown reliable and valid are used in a pair of empirical validation experiments. The purpose of the expert validation test was to assess the quality of the manufactured instrument and establish which parts could be used in practise and which could not. Items that fail to garner a high enough average rating from our panel of experts will be eliminated. (Boulden et al., 2021; Eltaybani et al., 2020; Hardesty & Bearden, 2004). While the empirical validation validity and reliability tests are performed using index values Standard Loading Factor (SLF)/load and reliability coefficient above the theoretical acceptance limit set, namely above 0.32 for load factor with T-Value 1.96 or 2.00 and above 0.70 for CR and above 0.5 for VE, with valid and reliable results because it is above the predetermined acceptance.

As a whole, the affective assessment instruments proved to be in a state of the goodness of fit based on the results of stage 1 and stage 2 trials, which were tested with five fit models, namely: the Normed Fit Index (NFI), the Non-Normed Fit Index (NNFI), the Comparative Fit Index (CFI), the Incremental Fit Index (IFI), and the Relative Fit Index (RFI). This implies that the instrument can be used to assess the affective component of history learning.

Because there are already five very good criteria, some of the requirements have been met. According to Black & Babin (2019), using fit criteria 4–5 is considered sufficient to evaluate the feasibility of the model. Therefore, the assessment of affective competence in the values found in high school history subjects is valid and reliable across all 5 dimensions, 14 indicators, and 84 questions. Five indicators make up the resulting instrument: (i) awareness of the significance of time and place in history; (ii) appreciation of historical heritage; (iii) awareness of unity, togetherness, and solidarity in the face of the threat of national disintegration; (iv) pride in Indonesia as a nation and love for the motherland; and (v) awareness of heroic values. The five aforementioned metrics will assess how well high school students understand the goals of studying history and the values that are presented in the course material.

First, indicators of understanding the significance of time and place in history are part of the dimension of awareness of the importance of time and place, which is a process of the past, present, and future (R1), recognizing that history has a temporal dimension (R2). Because historical events took place in both space and time, the concept of time and place is essential to understanding history (Kessler, 2021). Humans can grasp that historical events actually happened and took place in space in the past. While time refers to the historical event's time of occurrence, historians and the general public can use this concept of time to determine how the progression of historical events is linked to the present (Campos, 2020).

Second, the dimension of appreciation for historical heritage as proof of Indonesia's previous civilization (S) comprises indications of maintaining historical heritage artefacts (S1) and visiting historical heritage locations (S2). Because historical events are past occurrences that can only be known by the evidence left behind, it is crucial for students who study history to be aware of and appreciate historical artefacts. Because we do not dwell in the past, neither a historian nor a historical scholar or student will be able to know the past except through its remnants. Appreciation for historical heritage is beneficial for (i) enriching the cultural diversity of the country, (ii) boosting state revenue through historical tourism, (iii) fostering a sense of belonging and community, and (iv) advancing knowledge (Zhang et al., 2021). Therefore, it is crucial for every one of us to do our part to preserve historical heritage in the best way possible, protect historical items from destruction, keep them clean, and abide by any rules that may be in place at a particular historical heritage site.

Third, indications of retaining unity and integrity (T1), respect for differences (T2), deliberation and mutual cooperation (T3), as well as knowledge of unity, fraternity, and solidarity in the face of the threat of national disintegration (T), are included (T3). It is crucial to build and nurture in students an

awareness of togetherness, fraternity, and solidarity in the face of the threat of national disintegration, particularly in the present day and age, where such challenges might originate from anywhere. Students who understand the importance of brotherhood and togetherness will not be as easily impacted by conflict and division (Farouq et al., 2022; Posibi & Canele, 2020; Wulandari et al., 2022). Therefore, students must start developing this indication at a young age.

Fourth, the dimension of having national pride and love for the country, which may be expressed in a variety of spheres of life (U), comprises expressions of national pride (U1), giving national interests first priority (U2), valuing national culture (U3), and protecting the environment (U4). Everyone must adopt a mindset and behaviour that reflect a love for the nation. It is crucial to defend, preserve, and safeguard the nation. The value of instilling a sense of love for the homeland may also inspire love for and preservation of this nation's traditions, culture, and natural beauty (Toirovich & Baxromovich, 2022). It is this devotion to one's motherland that motivates people to work extremely hard to better their nation. Therefore, in order to attain the aim of coexisting, a sense of love for the homeland must be fostered in the spirit of each person who identifies as a citizen of a country or nation.

Fifth, the component of heroic values awareness, which underpins the process of developing pupils' personalities and characters (V), comprises signs, Valuing the contributions of heroes (V1), being prepared to make sacrifices (V2), and adopting heroic behaviour (V3). It is crucial to instil heroic qualities in the next generation so they can identify anyone who is prepared to make a sacrifice for the Indonesian people during the conflict (Salsabila et al., 2021; Subaryana, 2022). Students must model heroic virtues such as a high sense of patriotism and nationalism, comradery and duty, and unselfish sacrifice (Fomina, 2019).

Based on the discussion above, it was concluded that the affective competence assessment instrument in learning history consists of 5 indicators, namely: (i) awareness of the importance of time and place in history; (ii) awareness of the need to appreciate historical heritage; (iii) awareness of unity and brotherhood and solidarity in facing the threat of national disintegration; (iv) pride as a nation of Indonesia and love for the motherland; and (v) awareness of valid and reliable heroic values. Teachers use these indicators to measure students' affective aspects in learning history.

4. CONCLUSIONS

The goal of this study was to create an affective aspect assessment instrument that can be used to evaluate learning history and is valid, reliable, and accurate. Based on the findings of the development of the instruments for measuring students' affective competency in high school history courses, it can be concluded that the construct validity testing of the developed instruments demonstrates that at all testing stages, both the theoretical testing phase and the empirical testing stages 1 and 2 have met the valid criteria quite meaningfully. This is evident from the findings of the theoretical validation study using Aiken, which only meets the requirements for validity if it is above 0.2 and for inter-rater reliability using Hoyt if it is above 0.7. Construct Reliability (CR) has achieved the established standard of 0.70, while Variance Extracted (VE) is determined at 0.50, also based on the results of the model fit test using the goodness of fit test, resulting in five criteria that contain good values because they have the criteria above in the valid empirical trials stage 1 and stage 2. Furthermore, the 5 dimensions, 14 indicators, and 84 items met the criteria for valid and reliable instruments in the valid empirical trials, stages 1 and 2, which meet the criteria of having a factor loading > 0.32 and demonstrated by the achievement of the reliability coefficient in each. Teachers may use research findings to assess student learning outcomes, particularly their affective components. The findings of this study are utilized to diagnose student potential, particularly in relation to affective components, so that students are aware of their strengths and weaknesses. The findings of this study are utilized by stakeholders as input for policymaking, particularly in the area of education. This research is currently limited to trial participants from a single school; thus, more study that is more thorough in terms of subject and

research location, as well as the metrics used, is required. This will enable more advanced historical education.

REFERENCES

Azwar, S. (2013). Kualitas Tes Potensi Akademik Versi 07a. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *12*(2), 231-250. https://doi.org/10.21831/pep.v12i2.1429

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy and Practice*, *25*(6), 551–575. https://doi.org/10.1080/0969594X.2018.1441807

Black, W., & Babin, B. J. (2019). Multivariate Data Analysis: Its Approach, Evolution, and Impact. In B.J. Babin & M. Sarstedt (Eds), S *The Great Facilitator* (pp. 121–130). Springer, Cham.

Boulden, D. C., Rachmatullah, A., Oliver, K. M., & Wiebe, E. (2021). Measuring in-service teacher self-efficacy for teaching computational thinking: development and validation of the T-STEM CT. *Education and Information technologies*, *26*(4), 4663-4689.

Campos, D. D. (2020). Campos, I. D., & Bernardo, L. F. (2020). Architecture and Steel. Reflection and Analysis on the Use of Steel Structures (in Sight) as a Concept in the History of Architecture. *Designs*, *4*(3), 1–13.

Efendi, I., Prawitasari, M., & Susanto, H. (2021). Implementasi Penilaian Pembelajaran Pada Kurikulum 2013 Mata Pelajaran Sejarah. *Prabayaksa: Journal of History Education*, *1*(1), 21-25. https://doi.org/10.20527/prb.v1i1.3081

Eltaybani, S., Igarashi, A., & Yamamoto-mitani, N. (2020). Assessing the palliative and end-of-life care education-practice-competence triad in intensive care units: Content validity, feasibility, and reliability of a new tool. *Journal of Palliative Care, 36(4), 234-242.* https://doi.org/10.1177/0825859720948972

Farouq, U., Kertawidana, I., & Simatupang, H. (2022). Pancasila as the Foundation of the Defense of the State in the Face of the Threat of Distegration of the Indonesian Nation. *Randwick International of Social Sciences (RISS) Journal*, *3*(1), 134–141.

Fomina, V. (2019). Dead Heroes and Living Saints: Orthodoxy, Nationalism, and Militarism in Contemporary Russia and Cyprus. *Dissertation*, Central European University.

Hardesty, D. M., & Bearden, W. O. (2004). *The use of expert judges in scale development Implications for improving face validity of measures of unobservable constructs*. *57* (2), 98–107. https://doi.org/10.1016/S0148-2963(01)00295-8

Hashem M. Mehany, M. S., & Gebken, R. (2021). Assessing the importance and cognition level of ACCE's student learning outcomes: industry, educator, and student perceptions. *International Journal of Construction Education and Research*, *17*(4), 333-351.

Henderson, D. (2019). The nature of values and why they matter in the teaching and learning of history. In T. Allender (Eds), *Historical Thinking for History Teacher* (1st ed., pp. 102–116). Routledge.

Hensen, C., & Barbera, J. (2019). Assessing Affective Differences between a Virtual General Chemistry Experiment and a Similar Hands-On Experiment. *Journal of Chemical Education*, *96*(10), 2097–2108. https://doi.org/10.1021/acs.jchemed.9b00561

Kessler, O. (2021). Conceptual History in International Relations. In B. de Carvalho., J. C, Lopez., & H. Leira (Eds), *Routledge Handbook of Historical International Relations* (pp. 551–561). Routledge.

Kuntoro, B. T., & Wardani, N. S. (2020). Pengembangan Instrumen Penilaian Sikap Sosial Pembelajaran Tematik Kelas III SD. *Jurnal Ilmiah Wahana Pendidikan*, *6*(2), 295–307. https://doi.org/10.5281/zenodo.3752471

Mathis, C., & Parkes, R. (2020). Historical thinking, epistemic cognition, and history teacher education. In C.W. Berg & T.M. Christou (Eds), *The Palgrave Handbook of History and Social Studies Education* (pp. 189–212). Palgrave Macmillan, Cham.

Ningsih, T. Z., Sariyatun, & Sutimin, L. A. (2019). Development of portfolio assessment to measure the student's skill of using primary source evidence. *The New Educational Review, 56* (2) 101-113. https://doi.org/10.15804/tner.2019.56.2.08

Ofianto, Aman, Ningsih, T. Z., & Abidin, N. F. (2022). The Development of Historical Thinking Assessment to Examine Students' Skills in Analyzing the Causality of Historical Events. *European Journal of Educational Research*, *11*(2), 609–619.

Ofianto, & Ningsih, T. Z. (2021). *Assesmen Keterampilan Berpikir Historis (Historical Thinking)*. Duta Media.

Ofianto, O., & Suhartono, S. (2016). An assessment model of historical thinking skills by means of the RASCH model. *Research and Evaluation in Education 1*(1), 73-83. https://doi.org/10.21831/reid.v1i1.4899

Papadakis, S., Vaiopoulou, J., Kalogiannakis, M., & Stamovlasis, D. (2020). Developing and exploring an evaluation tool for educational apps (E.T.E.A.) targeting kindergarten children. *Sustainability (Switzerland)*, *12*(10), 1–10. https://doi.org/10.3390/su12104201

Posibi, A. ., & Canele, A. (2020). Historical Analysis of the Position of African Countries in the Nigerian Civil War, 1967–1970. *Vestnik of Saint Petersburg University. Asian and African Studies*, *12*(2), 302–311.

Reisman, A., Brimsek, E., & Hollywood, C. (2019). Assessment of Historical Analysis and Argumentation (AHAA): A New Measure of Document-Based Historical Thinking. *Cognition and Instruction 37(4),* 534-561. https://doi.org/10.1080/07370008.2019.1632861

Sajjad, M., Khan, R. A., & Yasmeen, R. (2018). Measuring assessment standards in undergraduate medical programs: Development and validation of AIM tool. *Pakistan Journal of Medical Sciences*, *34*(1), 164–169. https://doi.org/10.12669/pjms.341.14354

Sakki, I., & Pirttilä-Backman, A. M. (2019). Aims in teaching history and their epistemic correlates: a study of history teachers in ten countries. *Pedagogy, Culture and Society*, *27*(1), 65–85. https://doi.org/10.1080/14681366.2019.1566166

Salsabila, S. R., Dewi, D. A., & Furnamasari, Y. F. (2021). Peranan Perilaku Cinta Tanah Air melalui Pendidikan Kewarganegaraan di Sekolah Dasar. *Jurnal Pendidikan Tambusai*, *5*(3), 7792–7800. https://jptam.org/index.php/jptam/article/view/2244

Saputra, I. putu A., Sariyatun, S., & Musadad, A. A. (2018). The Reinforcement of Dharma Gita Bali Values in Historical Learning to Improve Student's Tolerace Attitude. *International Journal of Multicultural and Multireligious Understanding*, *5*(2), 247-254. https://doi.org/10.18415/ijmmu.v5i2.326

Subaryana, S. (2022). Menata Ulang Patriotisme dan Nasionalisme Indonesia. *Mukadimah: Jurnal Pendidikan, Sejarah, Dan Ilmu-Ilmu Sosial*, *6*(2), 127–136.

Sya'bandari, Y., Meilani-Fadillah, S., Nurlaelasari-Rusmana, A., Qurota-Aini, R., & Ha, M. (2022). Assessing Cognitive Bias in Korean and Indonesian Scientists: Considering Sociocultural Factors in Judgment and Choice. *Asia-Pacific Science Education*, *8*(February), 1–34. https://doi.org/10.1163/23641177-bja10045

Toirovich, S. X., & Baxromovich, T. S. (2022). National Values are the Main Factor that Includes National and Civil Responsibility for Today's Youth. *Spectrum Journal of Innovation, Reforms and Development*, *2*(3), 136–138.

Wansink, B., Akkerman, S., Zuiker, I., & Wubbels, T. (2018). Where Does Teaching Multiperspectivity in History Education Begin and End? An Analysis of the Uses of Temporality. *Theory and Research in Social Education*, *46*(4), 495–527. https://doi.org/10.1080/00933104.2018.1480439

Wulandari, A. O., Utomo, C. B., & Suryadi, A. (2019). Pelaksanaan Penilaian Afektif Pada Pembelajaran Sejarah Kurikulum 2013 Kelas XI IPS di SMA Negeri 3 Temanggung Tahun Ajaran 2018 / 2019. *Indonesian Journal of History Education*, *7*(1), 9–21.

Wulandari, Y. S., Rachmadani, N. T., Wahdi, A. K., Cahyani, A., Gusman, R., & Asbari, M. (2022). Growing Historical Awareness among the Young Generation of the Indonesian Nation. *Journal of Information Systems and Management*, *1*(5), 43–52.

Yildirim, G., Elban, M., & Yildirim, S. (2018). Analysis of Use of Virtual Reality Technologies in History Education: A Case Study. *Asian Journal of Education and Training*, *4*(2), 62–69. https://doi.org/10.20448/journal.522.2018.42.62.69

Zhang, X., Li, Y., & Lin, J. (2021). The Construction of Placeness in Traditional Handicraft Heritage Sites : A Case Study of Suzhou Embroidery. *Sustainability*, *13*(16), 9176-9200.

Zhou, H. (2013). Dragon under the Blue Helmet: a quantitative analysis of China's motivation for participation in UN peacekeeping operations. *African East-Asian Affairs*, September (3), 31-61. https://doi.org/10.7552/0-3-110