MULTIDISCIPLINARY SCIENCE JOURNAL

RESEARCH ARTICLE
Published Online: March 22, 2025
https://doi.org/10.31893/multiscience.2025452

Development of formative assessment in history lessons with research-based learning model to measure historical thinking skills

Asyrul Fikri[a] ✉ | **Suarman**[a] | **Isjoni**[a] | **Sumarno**[a] | **Ofianto**[b]

[a]Riau University, Indonesia.
[b]Riau Negeri Padang, Indonesia.

Abstract This study aims to develop a valid, practical, and effective formative assessment framework tailored for history education, utilizing a research-based learning model within the History Education Study Program at the University of Riau. By adopting Plomp's educational development model, the research emphasizes three core criteria: validity, practicality, and effectiveness. The methodology is structured into two main phases: the predevelopment phase and the development phase. The predevelopment phase is further divided into three stages: initial investigation, design phase, and realization phase. These stages focus on understanding the educational context, designing assessment tools, and realizing the initial framework. The development phase encompasses testing, evaluation, and revision stages, ensuring that the assessment tools are refined and optimized for educational use. The study's findings reveal a comprehensive set of assessments specifically designed for history education, utilizing the research-based learning model. The trial results indicate that the formative assessment framework successfully meets the criteria of validity, practicality, and effectiveness. The validity of the formative assessment and its associated tools is categorized as excellent, demonstrating a robust alignment with educational standards and objectives. Furthermore, the practicality and effectiveness of the formative assessment and all implemented tools are rated as very practical and effective, highlighting their ease of use and positive impact on learning outcomes. The research further indicates that the effectiveness of alternative assessments, based on both process and learning outcomes, shows a consistent improvement trend, categorizing them as good. This study not only contributes to the enhancement of formative assessments in history education but also provides a model that can be adapted to other educational contexts. The implications of this research suggest significant potential for improving teaching strategies and student engagement through well-designed assessment tools, ultimately contributing to the advancement of educational practices.

Keywords: formative assessment, history education, research-based learning model, historical thinking skills

1. Introduction

The role of history education is crucial in shaping the understanding and historical thinking skills of college students. According to Cortese (2003), history education significantly contributes to helping students develop continuous assessment, analysis, and integration skills. Yilmaz (2008) emphasized the importance of focusing on critical and analytical thinking skills in history education, as well as the ability to draw conclusions on the basis of evidence (Sherren, 2008). One of the challenges in teaching history is transforming a subject often considered boring into something engaging for 21st-century students (Sebbowa & Ng'ambi, 2020). Higher education in Indonesia emphasizes the importance of high-level thinking skills, particularly historical thinking, for graduates, especially prospective history teachers. Universities play a crucial role in equipping students with relevant skills to meet the demands of the workforce and society (Boelens et al., 2018). The development and strengthening of history education in universities are key to preparing a generation with 21st-century skills essential for success in the future (Chankseliani et al., 2020), such as critical thinking, communication, and collaboration, to meet the demands of modern jobs (Choy & Le, 2023).

The importance of history education in higher education is reinforced by the presence of an independent curriculum program or Freedom to Learn-Campus Freedom (MBKM). The MBKM program, initiated by the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia, allows students the flexibility to design their own curriculum and choose courses outside their main field of study. The MBKM program has proven beneficial, providing real-world work experience and practical learning (Dian et al., 2023), as well as developing 21st-century skills in general (Astuti et al., 2022). In developing historical thinking skills, active learning activities and various methods need to be involved (Keleşzade et al., 2018).

 Received: November 21, 2024 | Accepted: January 21, 2025

Historical thinking involves more than just factual knowledge, including engaging in connecting past events with the present and future. Additionally, the process of deconstruction and reconstruction is crucial in historical thinking, where learners compare information, describe continuity and change, and explain past events (Van Drie & Van Boxtel, 2008). Historical thinking skills facilitate an active and self-reflective process that explores the intentions behind historical narratives and interpretations of the past (Körber & Meyer-Hamme, 2015).

The significance of history learning and students' historical thinking skills lies not only in factual knowledge but also in the ability to analyze, synthesize, and interpret historical information (Boxtel et al., 2021; Nokes, 2010). History learning is not just about memorizing content and narratives of the social past but also about understanding what historical knowledge is, how it is represented, and how it can guide present and future life (Carrasco & Serrano, 2022). Historical thinking is a complex process involving various skills, including interpretation and argumentation via assessment methods such as portfolios, projects, written tests, discussions, and simulations (Körber & Meyer-Hamme, 2015). Assessment of the freedom curriculum has become an integral element in efforts to improve the education system, providing more flexibility and autonomy to universities in designing and implementing their curriculum. However, with this paradigm shift, assessments must be designed to reflect the diversity of students and provide relevant information to plan learning that suits individual needs.

In the freedom curve paradigm, assessment plays a crucial role in measuring student achievements (Brown, 2017; Wilson, 2018), providing essential information for curriculum-related decision-making (Confrey et al., 2019), and ensuring that the learning process focuses on the needs and development of each student (Driscoll & Wood, 2023; Konopasek et al., 2016). In this framework, formative assessment becomes a vital element by providing continuous feedback that helps students understand their own progress and assists instructors in designing learning tailored to students' needs (Irons & Elkington, 2021). The freedom curriculum also encourages active student engagement in the learning process (Rombe et al., 2023). However, amidst efforts to provide more contextual and responsive history education, several challenges arise. The presurvey results at several universities in Region XVII, Riau Province, especially in the history education program, reveal difficulties in assessing students' historical thinking skills effectively. Instructors face challenges in assessing the historical thinking process effectively and integrating research-based learning outcomes well.

The main challenge also arises in the context of team-teaching guidelines that are not adequately available. The lack of these guidelines creates confusion in assessing student learning outcomes relevant to research-based learning. Instructors are often caught up in conventional assessment approaches, with more emphasis on the results than on the learning process. Claravall and Irey (2022) indicated that research-based document instruction can have positive potential in interrogating historical sources and correcting information from various sources. Shemilt (2018) outlines six key dimensions of historical thinking skills that can be assessed, including chronology and arrangement, causality, evidence, significance, perspective, and empathy.

The performance evaluation results of lecturers at the University of Riau show relatively low achievement, especially in terms of teaching management, which is caused by the mismatch of teaching methods and semester learning plans. History Education students also face difficulties in developing critical reading skills for historical sources, identifying perspectives, and generalizing them on the basis of historical evidence. This issue emerges because inappropriate learning models are used to improve historical thinking skills. To address these issues, this research takes an approach to develop formative assessments in history education. By combining research-based learning and the freedom-to-learn-campus-freedom model, this research aims to create a more effective and relevant learning framework.

History education in universities plays a crucial role in shaping students' historical thinking skills. Research-based learning (RBL), a learning approach that emphasizes the use of research as the basis for learning practices, is a recognized solution, although it is not without challenges (Brew & Saunders, 2020; McLaren, 2019). RBL has key concepts such as research use, active engagement, reflection and application, and collaborative learning. While beneficial, its implementation requires time, commitment, research skills, and the ability to translate research findings into learning practices (Brew & Saunders, 2020; Siegel, 2005).

This model is designed around investigative activities, encouraging students to engage in real-life research and experiments (Camacho et al., 2017). In the context of local history, this model can enhance learning productivity by focusing on student-led investigations (Griffiths, 2004). The importance of developing historical thinking skills through research-based learning models has been confirmed by related research (Körber & Meyer-Hamme, 2015). RBL helps students analyze historical sources, evaluate evidence, and build coherent arguments (Seixas & Morton, 2013). Research-based learning models are recognized to significantly improve students' understanding of historical concepts and historical thinking skills (Husna & Syukur, 2020). The effectiveness and long-term impact of research-based learning (RBL) have also been evaluated; for example, Ofianto and Ningsih (2021) reported that RBL improved students' historical thinking skills. Ahdika's (2017) demonstrated its effectiveness through tests, assignments, and research reports. Although successful in improving learning satisfaction, research has yet to explore the long-term impact of RBL beyond the program (Tungkasamit, 2019).

This research makes a significant contribution to the revitalization of history education in Indonesian universities. By combining research-based learning models and formative assessment, this research attempts to integrate the freedom-to-learn-campus-freedom framework into the development of historical thinking skills. Innovative and relevant, this research not

only fills gaps in the academic literature but also provides practical guidance for curriculum development and teaching strategies in history education programs, which are relevant for dissemination to other history programs in universities.

On the basis of the previously stated issues, the research questions are formulated as follows: 1) How can the development procedure of formative assessment be considered valid, practical, and effective on the basis of history subjects by applying the research-based learning model? 2) To what extent can the results of developing formative assessments based on history subjects by applying the research-based learning model be considered valid, practical, and effective? 3) To what extent can the developed formative assessment contribute to improving students' historical thinking skills by implementing the research-based learning model?

The purpose of this research is to develop a formative assessment framework that is valid, practical, and effective for history education by integrating research-based learning and the freedom-to-learn-campus-freedom model. This framework aims to enhance students' historical thinking skills and provide a model adaptable to other educational contexts.

2. Literature Review

2.1. Development of Historical Thinking Skills in History Education

In recent years, the paradigm of history teaching has undergone significant changes. History is no longer considered merely a narrative requiring students to memorize facts but is now viewed as a discipline involving the understanding of concepts and historical thinking skills. This shift includes the separation of historical content knowledge and an understanding of how history engages in disciplinary practices. Kaposi (2016) explained that the reinterpretation of the historical understanding process forms the basis of this change.

When discussing historical thinking skills, Seixas et al. (2015) knowledge is defined as the ability of students to analyze and apply historical content knowledge when interpreting the past (Reisman, 2015). These skills include constructing historical arguments, identifying patterns over time, and establishing cause–effect relationships. Essentially, historical thinking skills characterize a cognitive process in which students are required to think critically and creatively with the goal of increasing the likelihood of achieving desired outcomes (Fahim & Masouleh, 2012). The five historical thinking skills listed in the history curriculum are understanding chronology, exploring historical evidence, interpreting evidence, imagining, and rationalizing historical events (Talin, 2016). Notably, Vansledright (2013) emphasizes that historical thinking involves the investigative process of evidence, authors, and context, as well as the validation of evidence.

The ability to think historically is recognized as a crucial intellectual skill that needs to be integrated into history teaching (Davis et al., 2001). Hardy and Iwatani (2021) outline the dimensions of historical thinking, including historical argumentation, causality, comparison, contextualization, continuity and change over time, and sourcing (Seixas & Morton, 2013). However, research indicates that history teaching still tends to be traditional (Talin, 2015). Lee (2023) highlight the potential of inquiry-based learning in developing students' critical thinking skills. As an alternative, Laksana's (2020) demonstrated a significant improvement in historical thinking skills by applying the cause-and-effect learning model. Therefore, there is a need for specific training programs for teachers to teach historical thinking skills, as suggested by Guerrero-Romera and Perez-Ortiz (2022), who advocate for collaboration among the government, universities, and professional organizations to meet these needs.

2.2. Formative Assessment in History Education

In the field of education, evaluation and assessment play crucial roles. Evaluation, as a systematic effort, involves the collection, organization, and processing of data to draw conclusions about values, meanings, and the achievement of goals for accountability purposes (Mardapi, 2008). On the other hand, assessment, as described by Brookhart (2011), Brookhart et al (2010), is a process of obtaining information for decision-making related to students, the curriculum, programs, schools, and educational policies.

Two main types of evaluation, formative and summative, are the focus of understanding in the education system (Harlen, 2006; McMillan, 2010). Formative evaluation, as an integral part of the learning process, provides feedback to teachers, students, and educational stakeholders to understand and improve the teaching and learning process (Clark, 2012; Curry et al., 2015; Irons & Elkington, 2021). In line with this, formative assessment involves teachers collecting evidence of student learning to identify strengths and weaknesses and adjust teaching methods (Black & Wiliam, 2018).

Moreover, summative assessment, conducted at the end of the learning period, aims to measure students' progress (Harlen, 2006; Tran et al., 2023; Zeng et al., 2018). The types of summative assessment include tests, exams, and projects (Ishaq et al., 2020). Drake and McBride (2000) emphasized the importance of summative teaching portfolios in helping teachers reflect on teaching practices and improve teaching quality.

In formative assessment, understanding the social and cultural context is key to designing assessment strategies that are appropriate for students' needs and realities. Feedback in formative assessment serves not only as an evaluation of student achievement but also as a key instrument to guide learning and student improvement (Harrison et al., 2017). Thus, the combination of formative and summative evaluation, which considers social and cultural contexts, provides a solid foundation in efforts to improve the quality of education holistically.

2.3. Research-Based Learning: Transforming Learning in the 21st Century.

Research-based learning integrates research methods into the curriculum, requiring educators to reflect on the role of research in the professional development of prospective teachers (Brew & Saunders, 2020). The philosophy of constructivism serves as the foundation, utilizing authentic approaches, problem solving, cooperative learning, hands-on activities, and research discovery (Poonpan & Siriphan, 2001). The procedures for research-based learning, as outlined by Tremp (2010) and Saptuti Susiani et al. (2018), involve formulating general questions, literature review, question determination, research activity planning, method/methodology clarification, investigation, data analysis, interpretation of results, and reporting/presenting findings. Molebash et al. (2019) identify six stages: engage, inquire, plan, seek, analyze, and communicate.

Enhancing critical thinking skills and student learning outcomes are benefits of research-based learning (Usmeldi et al., 2017). Lambert (2009) emphasized the role of students as active intellectual producers who participate in the research culture within academic environments. Although effective at the higher education level (Wessels et al., 2021), the effectiveness of research-based learning in the social sciences has not been fully systematically tested. Integration into the curriculum is recognized to expand the community learning approach (Walkington et al., 2011). Research experiences in child development learning provide skills in data collection, writing, and presentation, reinforcing the connection between research and learning (Kazura & Tuttle, 2010). This model also enhances student creativity (Khwanchai et al., 2017). Research-based learning transforms the learning paradigm, enabling students to become knowledge producers in the 21st century (Srikoon et al., 2014).

3. Materials and methods

3.1. Research Design

This research refers to the Plomp's (1997) educational development model, which consists of the following steps: 1) preliminary investigation; 2) design; 3) realization/construction; 4) testing, evaluation, and revision; and 5) implementation. The initial investigation phase involves analyzing needs; the design phase formulates solutions; the realization/construction phase creates a prototype; and the test, evaluation, and revision phase tests refine the design. This cycle repeats until a design meeting the criteria is obtained. In the implementation phase, the generated design is applied in real situations. The entire development process revolves around two main activities: the predevelopment stage involves initial investigation, design, and realization, whereas the development stage involves testing, evaluation, and revision. The primary goal is to ensure that the development results are theoretically and empirically justifiable, with consistent involvement between the theoretical development stages and field implementation. This cycle is shown in Figure 1.

3.2. Research Subjects

The research subjects in this study are odd-semester students and lecturers with a home basis in the history education study program, specifically lecturers with experience in teaching local history, national history, and history teaching in the academic year 2023–2024. The involvement of these subjects depends on needs at the development stage. The trial location is selected via purposive sampling techniques, namely, universities/educational institutions that teach history via a research-based learning model. The details of the research subjects involved in this study are presented in Table 1.

Table 1 Research Subjects.

Subject	Number of Trial Subjects			Total
	Readability	Limited	Expanded	
Lecturers	14	10	19	43
Students	65	48	173	286

3.3. Development and Testing Procedure of the Product

The procedure for developing formative assessment in this research is divided into two main stages: the predevelopment stage and the development stage. The predevelopment stage consists of three main phases: the initial investigation phase, the design phase, and the realization phase. The development stage includes the test, evaluation, and revision phases.

Product testing consists of readability testing, limited testing, and expanded testing. Readability testing involves lecturers from UNRI, STKIP YAP, and UNIV. PSB, and students at UNRI to obtain perspectives and suggestions on formative assessment. Limited testing is conducted at UNRI, involving the development of homework assignment instruments and overall assessment. The results are used for improvement. Expanded testing is carried out at UNRI, STKIP, YAP, and UNIV. PSB and continues until validation testing. This process aims to ensure the validity, practicality, and effectiveness of the assessment. The implementation design of the product testing is explained in the scheme in Figure 1.

3.4. Data collection instruments

In this research, two types of data collection instruments are used: one for the validation of assessment tools and the other for the development of assessment tools. The assessment tools include activity assessment instruments such as observation sheets and activity assessment guidelines; self-assessment tools such as self-assessment sheet formats and self-assessment guidelines; and assessment tools such as case-making guidelines, historical thinking skill scoring guidelines, and historical thinking skill development value guidelines. In addition to assessment tools, various research instruments, such as formative assessment rating sheets, are used to assess the validity of assessments. The feasibility observation sheet records the feasibility of assessments in the classroom, whereas the formative effectiveness questionnaire is used to assess the effectiveness of assessments by lecturers. Other instruments include student activity observation sheets to obtain information on student activities, lecturer activity observation sheets to observe lecturer activities, student response questionnaires to obtain student feedback, homework assignments such as historical thinking questions, semester learning plans such as lecturer guidelines, and validation sheets used by experts/practitioners to assess the validity of the developed instruments. All these instruments are designed carefully to ensure the validity, effectiveness, and practicality of formative assessments in research-based learning.

3.5. Validity and Reliability of Instruments

The validity and reliability of homework assignment instruments are evaluated from the dimensions of reliability, difficulty level, and item discrimination, measured by following the development procedures that refer to the views of Cohen et al. (1996). The test reliability coefficient is calculated via the Cronbach's alpha formula (Allen & Yen, 2001).

$$a = \frac{k}{k-1}\left(1 - \frac{1}{s_x^2}\sum s_i^2\right) \quad (1)$$

Explanation:
k = number of test items
s_i^2= variance of the score of item i
s_x^2= variance of total scores
The difficulty level of essay questions is calculated via the following formula:

$$P_i = \frac{\sum x_i}{Sm_i N} \quad (2)$$

Explanation:
P_i= difficulty level of item i or proportion of correctly answered item i.
$\sum x_i$= sum of scores for item i that are correctly answered by the testee.
Sm_i= maximum score for item i
N = number of subjects.

The developed homework assignment instruments meet the criteria for reliability, difficulty level, and item discrimination, as shown in Table 2. Allen & Yen's (2001) indicates that a difficulty level in the interval of 0.3–0.7 and item discrimination greater than or equal to 0.2 are considered good. To measure the validity of the instruments, theoretical evaluation is carried out through expert/practitioner assessments, including homework assignment instruments. Experts/practitioners provide assessments of instructions, coverage, and language instruments with validity categories on the basis of the criteria for categorizing the quality of instruments. Furthermore, the level of agreement between assessors (interrater reliability) was measured via Cohen's kappa coefficient and percentage of agreement statistics. The lower limit of the reliability coefficient used for good instruments is 0.70 (Wilkerson & Lang, 2007).

3.6. Data analysis

Data analysis in this research is performed descriptively and qualitatively with a logical narrative, focusing on whether formative assessments and the developed assessment tools are valid, practical, and effective. The validation results of experts are used to assess the validity of instruments, whereas trial data are used to evaluate practical and effective criteria. The level of agreement between assessors is measured via Cohen's kappa (Cohen et al., 1996) and percentage of agreements (Unrau et al., 2007), which are calculated via predetermined formulas. The kappa coefficient is obtained with the help of SPSS 15, making data analysis an integral part of the overall evaluation of the validity, practicality, and effectiveness of formative assessments and their tools.

$$K = \frac{f_0 - \sum f_e}{N - \sum f_e} \quad (3)$$

Explanation:
K: agreement level of assessors (reliability coefficient between assessors)
f_0: frequency of observed results

f_e: expected frequency.
N: number of items classified or evaluated.

Table 2 Homework assignment test trial results.

Aspek	Item	Homework Assignment Questions				
		VI	VII	VIII	IX	X
Reliability	α	0,87	0,91	0,71	0,88	0,91
	(each task) Description	Reliable	Reliable	Reliable	Reliable	Reliable
	α			0,97 Reliable		
Difficulty Level	1	0,58	0,52	0,64	0.70	0,54
	2	0,39	0,61	0,48	0,48	0,70
	3	0,72	0,70	0,58	0,59	0,53
	Category	Medium	Medium	Medium	Medium	Medium
		Medium	Medium	Medium	Medium	Medium
		Easy	Medium	Medium	Medium	Medium
Item Discrimination	1	0,49	0,60	0,31	0,40	0,47
	2	0,50	0,56	0,43	0,56	0.46
	3	0,54	0,47	0,49	0,63	0,47
	Category	V. Good	V. Good	V. Good	V. Good	V. Good
		V. Good	V. Good	V. Good	V. Good	V. Good
		V. Good	V. Good	V. Good	V. Good	V. Good

Note: V. Good (very good).

$$Percentage\ of\ Agreements\ = \frac{Agreements}{Disagreements + Agreements}\ x100 \qquad (4)$$

Note:
Agreements: the frequency of matches between the data of two validators/observers.
Disagreements: the frequency of mismatches between the data of validators/observers.

Reference criteria for determining the level of validity, practicality, and effectiveness of the developed formative assessments are determined on the basis of quality level criteria for products adopting assessment criteria from Azwar (2007), as presented in Table 3.

Table 3 Reference for Determining Product Quality, Validity, Practicality, and Effectiveness.

Score	Average Score Interval	Category			
		Product Quality	Validity	Practicality	Effectiveness
4	$3{,}25 < M \leq 4{,}00$	Very Good	Very Valid	Very Practical	Very Effective
3	$2{,}50 < M \leq 3{,}25$	Good	Valid	Practical	Effective
2	$1{,}75 < M \leq 2{,}50$	Less Good	Sufficiently Valid	Sufficiently Practical	Sufficiently Effective
1	$0{,}00 < M \leq 1{,}75$	Not Good	Less Valid	Less Practical	Less Effective

4. Results

4.1. Data and Data Analysis Results

4.1.1. Pre-Development Stage

The predevelopment stage comprises three main phases: initial assessment investigation, assessment tool investigation, and early research instrument investigation. In the initial assessment investigation phase, research was conducted to establish a solid foundation for the development of formative assessments for history lessons via a research-based learning model. Lecturers faced difficulties in assessment because of the lack of relevant guidelines for the learning process in the classroom, such as local and national history. A presurvey at universities provided confidence in developing formative assessments as a solution. The assessment tool investigation phase revealed that the assessment tools used by lecturers were not effective. Lecturers struggled to assess the learning process and outcomes, including poorly evaluated homework assignments. Therefore, the development of new assessment tools, such as Student Activity Assessment Devices, Student Self-Assessment Devices, and Historical Thinking Skills Assessment Devices, is necessary. The early research instrument investigation involved the development of instruments to assess the validity, practicality, and effectiveness of the developed assessments. In the design phase, formative design, assessment tools, and research instruments were subsequently created considering syntax, principles, impact, and implementation in line with the research-based learning model. Assessment tool prototypes were designed to support history lessons in courses such as local history, national history, and history teaching practices. Finally, in the realization phase, the results of the initial and design phases were reflected to compose formative assessments in history

learning via the research-based learning model. The final product of this phase is the initial prototype of formative assessments, including assessment tools and research instruments.

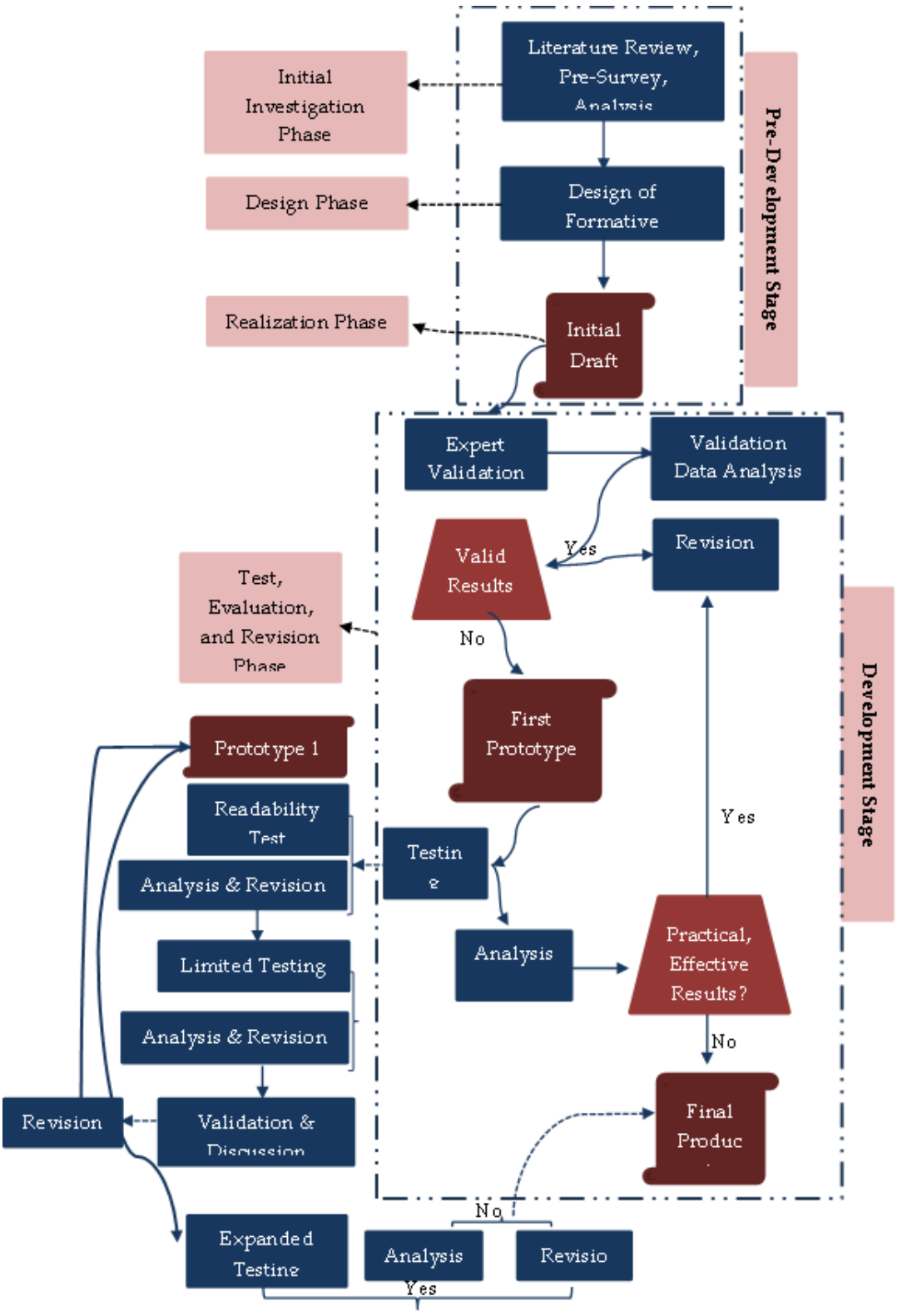

Figure 1 Development and Testing Procedure of the Product.

4.1.2. Development Stage

Table 4 summarizes the data analysis results concerning the feasibility of the research instruments based on expert validation.

Table 4 Results of Data Analysis of the Feasibility of Research Instruments.

No	Instrument	Analysis Results by Validator						Cohen's Kappa Statistic	
		1	2	3	4	TAS	Category	Average	Category
		Average Score by Each Validator							
1	Formative Assessment Validation Sheet	3,47	3,40	3,47	3,40	3,44	Very Valid	0,85	Reliable
2	Validation Sheet for Student Activity Assessment Tools	3,31	3,42	3,49	3,49	3,43	Very Valid	0,79	Reliable
3	Self-Assessment Tool Validation Sheet	3,36	3,44	3,36	3,44	3,40	Very Valid	0,86	Reliable
4	Validation Sheet for Formative Historical Thinking Skills Devices	3,25	3,30	3,30	3,35	3,30	Very Valid	0,76	Reliable
5	Observance of Feasibility Validation Sheet	3,56	3,56	3,67	3,67	3,62	Very Valid	0,80	Reliable
6	Effectiveness Questionnaire Validation Sheet	3,87	3,87	3,80	3,80	3,84	Very Valid	0,82	Reliable
7	Observation Sheet for Lecturer Activities	3,58	3,58	3,50	3,58	3,56	Very Valid	0,89	Reliable
8	Student Response Questionnaire Validation Sheet	3,56	3,67	3,56	3,67	3,62	Very Valid	0.83	Reliable
9	Homework Assignment Sheet Validation	3,39	3,50	3,39	3,50	3,45	Very Valid	0,83	Reliable
10	RPS (Syllabus) Validation Sheet	3,81	3,81	3,81	3,75	3,80	Very Valid	0,90	Reliable

Note: Total average score (TAS).

Analysis of the assessment/validation data to be presented will be obtained by determining the average value of all the components on the validation/assessment sheet of formative assessments given by the four validators. The results of the analysis of the assessment/validation of formative assessments are presented in Table 5.

Table 5 Results of the formative assessment validation analysis.

No	Assessment Aspect	Average Score by Validator				Cohen's Kappa Statistic
		1	2	3	4	
1	Supporting Theory	3,00	3,00	3,00	3,00	
2	Syntax	3,70	3,70	3,51	3,51	
3	Assessment Principles	3,65	3,65	3,69	3,69	
4	Impact of Assessment Implementation	3,55	3,50	3,30	3,31	
5	Assessment Implementation	3,25	3,50	3,50	3,00	
Average Score by Each Validator		3,43	3,47	3,40	3,31	
Total Average Score		3,40				0,78
Category		Very Valid				Reliable

The results of the analysis of the assessment by validators indicate that all instruments generally obtained a very valid category on the basis of high average scores. Furthermore, the reliability of these instruments was also measured via Cohen's kappa coefficient to measure consistency among the validators. The results show interrater reliability coefficients exceeding the minimum criterion of 0.70 for each instrument. Thus, the developed instruments can be considered reliable overall.

The findings of the practicality assessment of formative assessments, as evaluated by experts, are delineated in Table 6. This table provides a comprehensive overview of the expert evaluation, underscoring the efficacy and practicality of the assessments.

Table 6 Analysis of Practicality Assessment Data of Formative Assessments According to Experts.

No	Assessment Indicator	Expert Rating			
		1	2	3	4
1	The lecturer theoretically possesses the ability to follow the technical steps of assessment.	3	3	3	3
2	The lecturer has theoretical proficiency in applying fundamental principles in assessment.	4	4	3	4
3	Theoretically, the lecturer can carry out the assessment process in the classroom.	3	3	3	3
Average Score		3,33	3,33	3,00	3,33
Level of Practicality		High			

Table 7 indicates that the developed formative assessment has been highly rated by experts/practitioners and meets effectiveness criteria overall. However, revisions are deemed necessary on the basis of validator feedback, particularly

concerning the assessment of student activities, faculty activities, and the execution of the formative assessment. Instrument revisions involve adding a scale to the student observation sheet and changing assessment criteria on the faculty observation sheet as well as the implementation of the assessment. Prototype-1, resulting from these revisions, underwent limited testing (Trial I) to assess its validity, effectiveness, and practicality. The results of the revisions and Prototype-1 are further discussed in the product revision section.

Table 7 Analysis of Effectiveness Assessment Data of Formative Assessments According to Experts.

No	Assessment Indicator	Expert Rating			
		1	2	3	4
1	Improvement of the Quality of the Learning Process	4	4	4	4
2	Improvement of Learning Outcomes	3	3	3	3
3	Positive response from students and positive feedback from lecturers regarding the implementation of formative assessment.	4	4	4	4
4	The lecturer is theoretically capable of carrying out assessment syntax.	3	3	3	3
5	Theoretically, there is a possibility that the lecturer can apply assessment principles.	4	4	4	3
6	There is a likelihood that the lecturer can apply assessment principles.	3	3	3	3
Average Score		3,50	3,50	3,50	3,33
Level of Effectiveness		Very High			

4.2. Results of the Formative Assessment and Research Instrument Trial Data Analysis

4.2.1. Limited Trial Data Analysis Results

Table 8 presents a summary of the limited trial analysis results for formative assessment. The Implementation Assessment received a score of 2.46, with an agreement level of 0.89, indicating an unfavorable yet reliable assessment. Student activities and observer agreement were 2.41 points, with an agreement level of 0.78. In contrast, lecturer activities and observer agreement achieved a score of 2.60, with an agreement level of 0.89. Historical thinking skills on the basis of student self-assessment score 1.09, whereas historical thinking skills on the basis of homework assignment score 1.37. Student responses to components and activities of formative assessments achieved a high positive score of 91.39, and lecturer responses to the effectiveness of formative assessments received a score of 3.15, indicating positive assessments. Therefore, the limited trial results indicate strengths and potential improvements in each aspect of formative assessment.

Table 8 Comprehensive Analysis of Limited Trial Assessment Results for Formative Assessment.

Component	Assessment		Percentage of Agreement	
	TAS	Category	TAS	Category
Implementation Assessment	2,46	Less	0,89	Reliable
Student Activities and Observer Agreement	2,41	Less	0,78	Reliable
Lecturer Activities and Observer Agreement	2,60	Good	0,89	Reliable
Historical Thinking Skills Based on Student Self-Assessment	1,09	Simply		
Historical Thinking Skills Based on Homework Assignment Results	1,37	Good		
Student Responses to Components and Activities of Formative Assessments	91,39	Positive		
Lecturer Responses to the Effectiveness of Formative Assessments	3,15	Positive		

Note: Total average score (TAS).

4.3. Expanded Trial Data Analysis Results

Expanding the trial activity serves as a follow-up to the recommendations derived from the limited trial results, as the developed prototype-2 of the formative assessment is deemed not to meet the criteria of being effective and practical. This expanded trial functions as a form of validation for the formative assessment prototype that has been evaluated and revised in the preceding trial. The objective is to determine whether the developed assessment tool fulfills the expected criteria of being valid, effective, and practical.

On the basis of Table 9, formative assessment meets practical criteria with a high level of feasibility. The expanded trial results indicate smooth and efficient implementation of the assessment. The observer consistency in assessing feasibility is very high, with the average percentage of agreement reaching 92%. These data confirm a high level of precision in the execution of formative assessment in the classroom.

Table 10 indicates a positive assessment of student activities, meeting established standards. The formative assessment, which is evaluated on the basis of observations of student activities, has proven to be effective, with an improvement in the quality of learning over time. Despite some variation in percentage agreement between the two observers, the average coefficient PA reached a high level at 0.85. The findings from Table 11 suggest that both observers share a consistent perception regarding the quality of student learning with a high level of precision. Thus, the observation format can be relied upon for data collection in subsequent trials.

Table 9 Feasibility and Observer Agreement on Formative Assessments in Expanded Trials.

Assessment	Average Score of the Assessment Results			Percentage of Agreement		
	UNRI	STKIP YAP	UNIV. PSB	UNRI	STKIP YAP	UNIV. PSB
1	2,54	2,11	2,21	0,86	0,79	0,86
2	3,21	2,43	2,46	0,93	0,86	0,86
3	3,36	3,00	3,07	1,00	0,93	1,00
4	3,57	3,07	3,14	1,00	1,00	1,00
Average Score	3,17	2,65	2,58	0,95	0,89	0,93
Total Average Score		2,72			0,92	
Level of Implementation		High			Reliable	

Table 10 Student Activities and Observer Agreement in Extended Trials.

Assessment	Student Activities		Percentage of Agreement	
	Average Score	Category	Average Score	Category
1	2.96	Good	0.63	Unreliable
2	3.00	Good	0.75	Reliable
3	3.04	Good	1.00	Reliable
4	3.10	Good	1.00	Reliable
Total Average Score	3.03	Good	0.85	Reliable

Table 11 Lecturer Activities and Agreement Percentage in the Learning Process—Expanded Trial.

Assessment	Average Score of the Assessment Results			Percentage of Agreement		
	UNRI	STKIP YAP	UNIV. PSB	UNRI	STKIP YAP	UNIV. PSB
1	3,02	2,30	2,26	0,96	0,88	0,88
2	3,36	2,92	2,96	1,00	0,92	0,92
3	3,44	3,16	3,18	1,00	0,96	1,00
4	3,58	3,38	3,36	1,00	1,00	1,00
Average Score	3,35	2,94	2,95	0,99	0,94	0,95
Total Average Score		3,08			0,96	
Category		Good			Reliable	

Table 11 presents a positive assessment of lecturer activities during the learning process, meeting the established criteria. The evaluation of formative assessment effectiveness, which is based on the observation of lecturer activities, also meets the effectiveness criteria. Observations by two observers revealed an improvement in the quality of learning over time. The average percentage agreement reaches 0.96, confirming that lecturer activities run well as expected, with a high level of precision. The observation format for Lecturer activities proves reliable for data collection in the subsequent trial. Thus, the developed formative assessment has proven effective in enhancing learning activities after two trials.

Table 12 illustrates student learning outcomes, the development of historical thinking skills, student responses to formative assessment, and lecturer responses to formative assessment. All findings demonstrate a high level of effectiveness, meeting the established criteria. Thus, the developed formative assessment has proven to be effective in enhancing student learning outcomes and fostering the development of historical thinking skills. Positive feedback from both the students and the lecturers after two trials further confirmed the success of the formative assessment.

Table 12 Extended Trial Assessment Results for Historical Thinking Skills Development.

Component	Assessment	
	Total Average Score	Category
Self-Assessment of Historical Thinking Skills	1.51	Good
Historical Thinking Skills Development Based on Homework Assignment	1.51	Good
Student Responses to Components and Activities of Formative Assessments	93.84	Positive
Lecturer Responses to the Effectiveness of Formative Assessments	3.32	Very Effective

The results obtained from this expanded trial, as previously outlined, indicate that the developed formative assessment meets the criteria of validity, practicality, and effectiveness summarized in Table 13. In other words, after two rounds of testing, a formative assessment is produced that is valid, practical, and effective in assessing the process and outcomes of learning in a research-based learning context within the history program.

4.4. Product Revision

Formative assessment instruments and tools have undergone a series of revisions aimed at enhancing their quality in terms of validity, practicality, and effectiveness. Several improvements have been focused on various components of these instruments. First, in the effectiveness questionnaire for formative assessment, special attention was given to typos and punctuation errors in the instructions, which were then corrected to ensure clearer and easily understandable instructions.

Second, in the observation sheet for the implementation of formative assessment, assessment criteria were expanded using a more comprehensive rating scale. This is intended to provide more detailed and varied information about the implementation of observations, resulting in more accurate outcomes.

Table 13 Levels of Validity, Practicality, and Effectiveness of Formative Assessments.

Aspect	Average Score	Category
Validity	3.44	Very Valid
Practicality	3.37	Very Practical
Effectiveness	3.47	Very Effective

Note: For the average scores of the assessed aspects, their scales were first equated by converting them to a four-point scale.

Third, in the observation sheet of the lecturer's activities during the learning process, the rating scale was expanded to cover a more varied level of activities, providing deeper insights into the quality of learning. Fourth, in the student response questionnaire, language usage was simplified and clarified to make it easier for students to understand the questions. Finally, some homework assignment questions were revised to improve clarity and avoid potential misinterpretations. Although the developed formative assessment meets the criteria of being valid, practical, and effective, its implementation needs to consider its characteristics, strengths, and limitations.

The characteristics of this assessment include development in line with a realistic mathematics learning approach, emphasis on procedural and process assessment with quantified assessment results, integrated assessment implementation during the learning process, continuous assessment, use of free-response formats, and emphasis on contextual tasks. In terms of strengths, formative assessment proves to be relevant for mathematics learning with a realistic approach, providing information according to actual conditions, assessing students' problem-solving abilities, obtaining as much information as possible using various assessment tools, encouraging improvement and enhancing the quality of learning, building good communication between lecturers and students through student self-assessment, ensuring fairness through feedback, avoiding lecturer subjectivity, not being tied to specific content, and being open to further development.

However, limitations have also been identified, such as the restriction of assessment to the process and outcomes of history learning with a research-based learning model in higher education, limitations in students' historical thinking skills that are assessed, the requirement for mature preparation by lecturers, potential unfairness in student self-assessment, and the need for explanations, exercises, and guidance for students, especially those in the history education program, in applying these assessment tools. To date, peer assessment among students has not been implemented; thus, it needs to be considered for further development.

5. Discussion

This research aims to evaluate the implementation of research-based learning (RBL), with a focus on assessment aspects, which is expected to enhance the quality of history learning. RBL is a systematic approach that integrates research and practice in teacher education to improve knowledge, skills, and the quality of learning (Brew & Saunders, 2020). Lecturers, as primary assessors, need adequate skills to assess students' historical thinking objectively and fairly. The implementation of RBL requires time, resources, and significant support from the university (Brew & Saunders, 2020). Similarly, Brew and Mantai (2017) reported that despite RBL having many benefits, some challenges, such as the lack of training and support for teaching staff, need to be overcome for effective implementation. Therefore, continuous assessment is needed to monitor the development of students' historical thinking skills during learning.

The importance of active student involvement in assessment, including self-assessment, provides the basis for lecturers to assign final grades. Formative assessment, involving various methods such as observation, self-assessment, and performance, is key to providing comprehensive information and maintaining the quality of assessment. Lecturers need professional development to enhance their knowledge of formative and summative assessment, which will improve their assessment practices (Sulaiman et al., 2020). This aligns with the emphasis of Martínez-Huamán et al. (2022) that lecturers play a crucial role in conducting formative assessments. They actively participate in the student participation and reflection process, aiming to improve their skills through correction and positive reinforcement.

Through the development process using the Plomp procedure, this research successfully produced valid, practical, and effective formative assessments. The implementation phase is expected to be disseminated to students and lecturers for application in history learning across all history-related courses and humanities fields. Constraints that arise during the trial need to be noted and corrected to ensure the validity, practicality, and effectiveness of formative assessments. The findings of this research are consistent with those of Kaya-Capocci et al. (2022), who studied a framework to support the implementation of digital formative assessments in higher education. The analysis revealed that formative assessment can enhance the quality of the learning process and develop students' historical thinking skills. The use of cultural historical activity theory can be a useful framework for understanding how students learn and developing effective formative assessment tools (Asghar, 2013).

Talin (2016) explained that to integrate thinking skills effectively into history teaching and learning, collaboration is needed to prepare and equip lecturers with the necessary knowledge and skills.

Student involvement in assessment has positive impacts, including increased social interaction, confidence, responsibility, and learning motivation. Mendez Lozano and Tirado Segura (2016) highlight the development of six historical thinking elements as a strategy that includes formative assessment. The systematic and continuous feedback process from history lecturers using a ranked rubric helps students identify areas for improvement and develop more effective learning strategies (Tirado-Olivares et al., 2024). Feedback is crucial for reflection, analysis, communication, and adjustment, leading to professional development and improvement in teaching practices (Martínez-Huamán et al., 2022). Pillay and Balele, (2022) suggested that the feedback literacy of lecturers and students has the potential to facilitate research-based feedback processes in the future. Thus, it can be concluded that this formative assessment can be an effective tool to improve the quality of history learning through the research-based learning model, providing information as needed and contributing to the improvement of history education specifically.

6. Conclusions

The research conclusion indicates that formative assessment based on a mathematical learning approach with a realistic perspective, through predevelopment and development stages, meets the criteria of validity, practicality, and effectiveness. The assessment product comprises the Student Activity Assessment Tool, Student Self-Assessment Tool, and Historical Thinking Skills Assessment Tool. The overall implementation of the assessment can enhance the quality of the history learning process and outcomes, reflected in observations of student activities, faculty activities, the development of learning outcomes on the basis of student self-assessment, and the enhancement of historical thinking skills through homework assignments. Both students and faculty responded positively to the formative assessment.

This research has several limitations. First, the study was conducted at only three universities, so the results cannot be generalized to populations with different characteristics. Second, development products have not been widely implemented, so their effectiveness can only be known by lecturers or teams involved in the research. Third, the assessed historical skills are focused only on historical thinking skills. Fourth, not all formative assessment techniques were the subject of development; only a few were considered relevant for history learning with a research-based learning model.

To address the limitations of this research, it is recommended that lecturers utilize this assessment product as one of the assessment approaches in history teaching, especially for those facing challenges in evaluation. The dissemination of the development product can be carried out through community service FGDs with lecturers specializing in history or social sciences at universities. To measure the effectiveness of the assessment tool, implementation on a broader scale is desired. Researchers are also advised to conduct further in-depth and comprehensive studies using a combination of various types of formative assessments in history teaching, especially in the era of the independent curriculum.

Ethical Considerations

This study complies with ethical writing guidelines**.**

Conflict of Interest

The authors declare no conflicts of interest.

Funding

This research did not receive any financial support.

References

Ahdika, A. (2017). Improvement of quality, interest, critical, and analytical thinking ability of students through the application of research based learning (RBL) in introduction to stochastic processes subject. *International Electronic Journal of Mathematics Education*, *12*(2), 167-191. https://doi.org/10.29333/iejme/608

Allen, M. J., & Yen, W. M. (2001). *Introduction to measurement theory*. Waveland Press.

Asghar, M. (2013). Exploring formative assessment using cultural historical activity theory. *Turkish Online Journal of Qualitative Inquiry*, *4*(2), 18–32. https://dergipark.org.tr/en/pub/tojqi/issue/21399/229389

Astuti, S. Y., Rosiawan, R. W., & No, S. (2022). Student Responses to the Implementation of MBKM (Study on FEB Students who follow International Course Asia University). *International Journal of Educational Research & Social Sciences*, *3*(1), 176–185. https://doi.org/10.51601/ijersc.v3i1.249

Azwar, S. (2007). *Sikap Manusia: Teori dan Pengukurannya [Human Attitudes: Theory and Measurement]*. Pustaka pelajar.

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

Boelens, R., Voet, M., & De Wever, B. (2018). The design of blended learning in response to student diversity in higher education: Instructors' views and use of differentiated instruction in blended learning. *Computers & Education*, *120*, 197–212. https://doi.org/10.1016/j.compedu.2018.02.009

Boxtel, C. van, Voet, M., & Stoel, G. (2021). Inquiry Learning in History. In *International Handbook of Inquiry and Learning* (pp. 296–310). Routledge. https://doi.org/10.4324/9781315685779-20

Brew, A., & Mantai, L. (2017). Academics' perceptions of the challenges and barriers to implementing research-based experiences for undergraduates. *Teaching in Higher Education*, *22*(5), 551–568. https://doi.org/10.1080/13562517.2016.1273216

Brew, A., & Saunders, C. (2020). Making sense of research-based learning in teacher education. *Teaching and Teacher Education*, *87*(January), 102935. https://doi.org/10.1016/j.tate.2019.102935

Brookhart, S. M. (2011). Teacher Feedback in Formative Classroom Assessment. In *Leading Student Assessment* (pp. 225–239). Springer Netherlands. https://doi.org/10.1007/978-94-007-1727-5_11

Brookhart, S. M., Moss, C. M., & Long, B. A. (2010). Teacher inquiry into formative assessment practices in remedial reading classrooms. *Assessment in Education: Principles, Policy & Practice*, *17*(1), 41–58. https://doi.org/10.1080/09695940903565545

Brown, G. T. L. (2017). *Assessment of Student Achievement*. Routledge. https://doi.org/10.4324/9781315162058

Camacho, M. H., Valcke, M., & Chiluiza, K. (2017). Research based learning in higher education: a review of literature. In *INTED proceedings*. IATED. https://doi.org/10.21125/inted.2017.1004

Carrasco, C. J. G., & Serrano, J. S. (2022). The origin and development of research into historical thinking. In *Re-imagining the Teaching of European History* (pp. 25–41). Routledge. https://doi.org/10.4324/9781003289470-4

Chankseliani, M., Qoraboyev, I., & Gimranova, D. (2020). Higher education contributing to local, national, and global development: new empirical and conceptual insights. *Higher Education*, *81*(1), 109–127. https://doi.org/10.1007/s10734-020-00565-8

Choy, S., & Le, A. H. (2023). Worklife Learning: Contributions of Tertiary Education. In *Sustaining Employability Through Work-life Learning* (pp. 261–283). Springer Nature Singapore. https://doi.org/10.1007/978-981-99-3959-6_12

Claravall, E. B., & Irey, R. (2022). Fostering historical thinking: The use of document based instruction for students with learning differences. *The Journal of Social Studies Research*, *46*(3), 249–264. https://doi.org/10.1016/j.jssr.2021.08.001

Clark, I. (2012). Formative Assessment: Assessment Is for Self-regulated Learning. *Educational Psychology Review*, *24*(2), 205–249. https://doi.org/10.1007/s10648-011-9191-6

Cohen, R. J., Swerdlik, M. E., & Phillips, S. M. (1996). *Psychological testing and assessment: An introduction to tests and measurement*. Mayfield Publishing Co.

Confrey, J., McGowan, W., Shah, M., Belcher, M., Hennessey, M., & Maloney, A. (2019). Using Digital Diagnostic Classroom Assessments Based on Learning Trajectories to Drive Instruction. In *Researching and Using Progressions (Trajectories) in Mathematics Education* (pp. 75–100). BRILL. https://doi.org/10.1163/9789004396449_004

Cortese, A. D. (2003). The critical role of higher education in creating a sustainable future. *Planning for Higher Education*, *31*(3), 15–22.

Curry, K. A., Mwavita, M., Holter, A., & Harris, E. (2015). Getting assessment right at the classroom level: using formative assessment for decision making. *Educational Assessment, Evaluation and Accountability*, *28*(1), 89–104. https://doi.org/10.1007/s11092-015-9226-5

Davis, O. L., Yeager, E. A., & Foster, S. J. (2001). *Historical empathy and perspective taking in the social studies*. Rowman & Littlefield.

Dian, D., Ahmad, C. F., Arsal, F. R., & Mahmudah, S. (2023). Implication And Application MBKM's Curriculum In Education (Madrasah And Universities). *At-Ta'dib*, *18*(1), 106–122. https://doi.org/10.21111/attadib.v18i1.9910

Drake, F. D., & McBride, L. W. (2000). The Summative Teaching Portfolio and the Reflective Practitioner of History. *The History Teacher*, *34*(1), 41. https://doi.org/10.2307/3054374

Driscoll, A., & Wood, S. (2023). *Developing Outcomes-Based Assessment for Learner-Centered Education*. Routledge. https://doi.org/10.4324/9781003444176

Fahim, M., & Masouleh, N. S. (2012). Critical Thinking in Higher Education: A Pedagogical Look. *Theory and Practice in Language Studies*, *2*(7). https://doi.org/10.4304/tpls.2.7.1370-1375

Griffiths *, R. (2004). Knowledge production and the research–teaching nexus: the case of the built environment disciplines. *Studies in Higher Education*, *29*(6), 709–726. https://doi.org/10.1080/0307507042000287212

Guerrero-Romera, C., & Perez-Ortiz, A. L. (2022). The Training Needs of In-Service Teachers for the Teaching of Historical Thinking Skills in Compulsory Secondary Education and the Baccalaureate Level. *Frontiers in Education*, *7*. https://doi.org/10.3389/feduc.2022.934646

Hardy, A., & Iwatani, E. (2021). *Rubrics for Examining Historical Thinking Skills in High School World History Activities and Student Work : Construct Validity Evidence from the Literature*. Digital Promise.

Harlen, W. (2006). On the relationship between assessment for formative and summative purposes. *Assessment and Learning*, *2*, 95–110.

Harrison, C. J., Könings, K. D., Schuwirth, L. W. T., Wass, V., & van der Vleuten, C. P. M. (2017). Changing the culture of assessment: the dominance of the summative assessment paradigm. *BMC Medical Education*, *17*(1), 73. https://doi.org/10.1186/s12909-017-0912-5

History: A Time to Be Mindful: Gaea Leinhardt. (2012). In *Teaching and Learning in History* (pp. 216–262). Routledge. https://doi.org/10.4324/9780203052662-11

Husna, L. A., & Syukur, A. (2020). Meningkatkan Higher Order Thinking Skills Melalui Penerapan Kemampuan Berpikir Historis pada Mahasiswa Pendidikan Sejarah [Enhancing Higher Order Thinking Skills Through the Application of Historical Thinking Abilities in History Education Students]. *Prosiding Seminar Nasional Pascasarjana*, *1*(01), 22–31.

Irons, A., & Elkington, S. (2021). *Enhancing learning through formative assessment and feedback*. books.google.com.

Ishaq, K., Rana, A. M. K., & Zin, N. A. M. (2020). Exploring Summative Assessment and Effects: Primary to Higher Education. *Bulletin of Education and Research*, *42*(3), 23–50.

Kaposi, J. (2016). Skills development Tasks and the development of Historical Thinking. *Fehérvári Anikó (Szerk.), Curriculum, Effectiveness, Equity. Budapest: Hungarian Institute for Educational Research and Development*, 9–22.

Kaya-Capocci, S., O'Leary, M., & Costello, E. (2022). Towards a Framework to Support the Implementation of Digital Formative Assessment in Higher Education. *Education Sciences*, *12*(11), 823. https://doi.org/10.3390/educsci12110823

Kazura, K., & Tuttle, H. (2010). Research Based Learning Approach: Students Perspective of Skills Obtained. *Journal of Instructional Psychology*, *37*(3).

Keleşzade, G., Güneyli, A., & Özkul, A. E. (2018). Effectiveness of History Teaching Based on Social Constructivist Learning and Development of Historical Thinking Skills. *TED EĞİTİM VE BİLİM*. https://doi.org/10.15390/eb.2018.7479

Khwanchai, K., Tanthip, K., & Toansakul, S. (2017). An instructional design model with the cultivating research-based learning strategies for fostering teacher

students creative thinking abilities. *Educational Research and Reviews*, *12*(15), 712–724. https://doi.org/10.5897/err2017.3239

Konopasek, L., Norcini, J., & Krupat, E. (2016). Focusing on the Formative: Building an Assessment System Aimed at Student Growth and Development. *Academic Medicine*, *91*(11), 1492–1497. https://doi.org/10.1097/acm.0000000000001171

Körber, A., & Meyer-Hamme, J. (2015). Historical thinking, competencies, and their measurement: Challenges and approaches. *New Directions in Assessing Historical Thinking*, 89–101.

Laksana, K. (2020). Promoting historical thinking for pre-service social studies teachers: A case study from Thailand. *International Journal of Curriculum and Instruction*, *12*(2), 12–24.

Lambert, C. (2009). Pedagogies of participation in higher education: a case for research-based learning. *Pedagogy, Culture & Society*, *17*(3), 295–309. https://doi.org/10.1080/14681360903194327

LEE, B. I. H. N. I. (2023). Exploring the relationship between history learning and thinking skills: a synthesis of research. *Muallim Journal of Social Science and Humanities*, 18–30. https://doi.org/10.33306/mjssh/233

Levstik, L. S. (n.d.). Learning History. In *Handbook of Research on Learning and Instruction*. Routledge. https://doi.org/10.4324/9780203839089.ch6

Mardapi, D. (2008). *Teknik Penyusunan Instrumen Tes dan Nontes [Techniques for Developing Test and Non-Test Instruments]*. Mitra Cendikia.

Martínez-Huamán, E. L., Landa, J. P. A., Mendoza, V. E. L., & Plata, C. E. G. R. (2022). Formative Assessment: Component of Teaching-Learning Process in the University Context in Post-COVID-19. *International Journal of Learning, Teaching and Educational Research*, *21*(11), 300–314. https://doi.org/10.26803/ijlter.21.11.17

McLaren, P. G. (2019). Stop Blaming Gordon and Howell: Unpacking the Complex History Behind the Research-Based Model of Education. *Academy of Management Learning & Education*, *18*(1), 43–58. https://doi.org/10.5465/amle.2017.0311

McMillan, J. H. (2010). The practical implications of educational aims and contexts for formative assessment. *Handbook of Formative Assessment*, 41–58.

Mendez Lozano, S. M., & Tirado Segura, F. (2016). Enhancing Historical Reasoning: A Strategy Including Formative Assessment with Systematic Continuous Feedback. *International Journal of Educational Psychology*, *5*(2), 187–219. https://doi.org/10.17583/ijep.2016.1639

Molebash, P. E., Lee, J. K., & Heinecke, W. F. (2019). Teaching and Learning Inquiry Framework. *Journal of Curriculum and Teaching*, *8*(1), 20. https://doi.org/10.5430/jct.v8n1p20

Nokes, J. D. (2010). Observing Literacy Practices in History Classrooms. *Theory & Research in Social Education*, *38*(4), 515–544. https://doi.org/10.1080/00933104.2010.10473438

Ofianto, & Ningsih, T. Z. (2021a). *Assesmen keterampilan berpikir historis (historical thinking)*. Duta Media.

Ofianto, O., & Ningsih, T. Z. (2021b). The Importance of a Problem-Based Learning Model in History Education for Improving Historical Thinking. *International Journal of Multicultural and Multireligious Understanding*, *8*(10), 503–509. https://ijmmu.com/index.php/ijmmu/article/view/3136/2702

Pillay, P., & Balele, R. (2022). Exploring Learners' Experiences of Receiving Formative Written Assessment Feedback in Business Studies as a Subject in South Africa. *International Journal of Learning, Teaching and Educational Research*, *21*(10), 228–248. https://doi.org/10.26803/ijlter.21.10.12

Plomp, T. (1997). *Development research on/in educational development, Enschede*.

Poonpan, S., & Siriphan, S. (2001). Indocator of Research-Based Learning Instructional Process: A Case Study of Best Practice in a Primary School Faculty of Education, Chulalongkorn University Phaaya Thai Bangkok. *Bangkok. Thailand*.

Reisman, A. (2015). The difficulty of assessing disciplinary historical reading. In *New directions in assessing historical thinking* (pp. 29–39). Routledge.

Rombe, R., Rani, R., Nurlita, N., & Parinding, J. F. (2023). Pembelajaran Berdiferensiasi Dalam Kurikulum Merdeka Belajar Menurut Ki Hajar Dewantara Pada Mata Pelajaran Pendidikan Agama Kristen. *Jurnal Pendidikan Dan Keguruan*, *1*(6), 541–554.

Rosy Talin. (2016). Why Historical Thinking Skills was not there? *International Journal of Learning, Teaching and Educational Research*, *15*(7), 110–126.

Saptuti Susiani, T., Salimi, M., & Hidayah, R. (2018). Research Based Learning (RBL): How to Improve Critical Thinking Skills? *SHS Web of Conferences*, *42*, 42. https://doi.org/10.1051/shsconf/20184200042

Sebbowa, D. K., & Ng'ambi, D. (2020). Teaching History in Ways C21st Students Learn – A Design-Based Research Perspective. *International Journal of Learning, Teaching and Educational Research*, *19*(9), 259–280. https://doi.org/10.26803/ijlter.19.9.14

Seixas, P., Gibson, L., & Ercikan, K. (2015). A Design Process For Assessing Historical Thinking: The Case of a One-Hour Test. In K. Ercikan & P. Seixas (Eds.), *New Directions in Assessing Historical Thinking* (pp. 124–138). Routled.

Seixas, P., & Morton, T. (2013). The Big Six Historical Thinking Concepts. In *The Big Six Historical Thinking Concepts*. https://bit.ly/3FX9lB8

Shemilt, D. (2018). Assessment of Learning in History Education. In *The Wiley International Handbook of History Teaching and Learning* (pp. 449–471). Wiley. https://doi.org/10.1002/9781119100812.ch17

Sherren, K. (2008). A history of the future of higher education for sustainable development. *Environmental Education Research*, *14*(3), 238–256. https://doi.org/10.1080/13504620802148873

Siegel, C. (2005). Implementing a Research-Based Model of Cooperative Learning. *The Journal of Educational Research*, *98*(6), 339–349. https://doi.org/10.3200/joer.98.6.339-349

Srikoon, S., Bunterm, T., Samranjai, J., & Wattanathorn, J. (2014). Research Synthesis of Research-based Learning for Education in Thailand. *Procedia - Social and Behavioral Sciences*, *116*, 913–917. https://doi.org/10.1016/j.sbspro.2014.01.319

Sulaiman, T., Kotamjani, S. S., Rahim, S. S. A., & Hakim, M. N. (2020). Malaysian Public University Lecturers' Perceptions and Practices of Formative and Alternative Assessments. *International Journal of Learning, Teaching and Educational Research*, *19*(5), 379–394. https://doi.org/10.26803/ijlter.19.5.23

Talin, R. (2015). Historical Thinking Skills - The Forgotten Skills? *International Journal of Learning and Teaching*, 15–23. https://doi.org/10.18844/ijlt.v7i1.3

Tirado-Olivares, S., López-Fernández, C., González-Calero, J. A., & Cózar-Gutiérrez, R. (2024). Enhancing historical thinking through learning analytics in Primary Education: A bridge to formative assessment. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-12425-w

Tran, H., Vu-Van, T., Bang, T., Le, T.-V., Pham, H.-A., & Huynh-Tuong, N. (2023). Data Mining of Formative and Summative Assessments for Improving Teaching Materials towards Adaptive Learning: A Case Study of Programming Courses at the University Level. *Electronics*, *12*(14), 3135. https://doi.org/10.3390/electronics12143135

Tremp, P. (2010). Research-based Teaching and Learning, A LERU project. *Universitat Zurich*.

Tungkasamit, A. (2019). The Effect of Using Research - based Learning Model in History Practicum in School Course. *Pedagogia : Jurnal Pendidikan*, *8*(1). https://doi.org/10.21070/pedagogia.v8i1.1770

Unrau, Y. A., Gabor, P. A., & Grinnell, R. M. (2007). *Evaluation in social work: The art and science of practice*. Oxford University Press.

Usmeldi, U., Amini, R., & Trisna, S. (2017). The Development of Research-Based Learning Model with Science, Environment, Technology, and Society Approaches to Improve Critical Thinking of Students. *Jurnal Pendidikan IPA Indonesia*, *6*(2), 318. https://doi.org/10.15294/jpii.v6i2.10680

Van Drie, J., & Van Boxtel, C. (2008). Historical reasoning: Towards a framework for analyzing students' reasoning about the past. *Educational Psychology Review*. https://doi.org/10.1007/s10648-007-9056-1

VanSledright, B. (2015). What does it mean to think historically . . . and how do you teach it? In *Social Studies Today: Research and Practice: Second Edition*. https://doi.org/10.4324/9781315726885-24

Vansledright, B. A. (2013). Assessing historical thinking and understanding: Innovative designs for new standards. In *Assessing Historical Thinking and Understanding: Innovative Designs for New Standards*. Routledge. https://doi.org/10.4324/9780203464632

Walkington, H., Griffin, A. L., Keys-Mathews, L., Metoyer, S. K., Miller, W. E., Baker, R., & France, D. (2011). Embedding Research-Based Learning Early in the Undergraduate Geography Curriculum. *Journal of Geography in Higher Education*, *35*(3), 315–330. https://doi.org/10.1080/03098265.2011.563377

Wessels, I., Rueß, J., Gess, C., Deicke, W., & Ziegler, M. (2021). Is research-based learning effective? Evidence from a pre–post analysis in the social sciences. *Studies in Higher Education*, *46*(12), 2595–2609. https://doi.org/10.1080/03075079.2020.1739014

Wilkerson, J. R., & Lang, W. S. (2007). *Assessing teacher competency: Five standards-based steps to valid measurement using the CAATS Model*. Corwin Press.

Wilson, M. (2018). Making measurement important for education: The crucial role of classroom assessment. *Educational Measurement: Issues and Practice*, *37*(1), 5–20.

Yilmaz, K. (2008). A Vision of History Teaching and Learning: Thoughts on History Education in Secondary Schools. *The High School Journal*, *92*(2), 37–46. https://doi.org/10.1353/hsj.0.0017

Zeng, W., Huang, F., Yu, L., & Chen, S. (2018). Towards a learning-oriented assessment to improve students' learning—a critical review of literature. *Educational Assessment, Evaluation and Accountability*, *30*(3), 211–250. https://doi.org/10.1007/s11092-018-9281-9