



Development of project-based historical knowledge assessment instruments

Supianto

Yogyakarta State University, Indonesia

Kumaidi

Muhammadiyah University of Surakarta, Indonesia

Yoyon Suryono

Yogyakarta State University, Indonesia

ABSTRACT

The history teacher is quite often using a project-based learning model to improve students' historical knowledge and skills. Unfortunately, this learning model is not supported by the availability of qualified assessment instruments. The limitations of project assessment instruments make teachers rely more on subjectivity in assessing student projects. Meanwhile, to assess historical knowledge, most teachers still use multiple choice and essay questions. Both types of assessments provide little information about students' abilities and are not sufficient as a basis for teachers to make decisions in the classroom. Therefore, it is necessary to develop an assessment instrument that can be used in assessing students' knowledge based on the project that they produce in project-based learning. This study aims to develop a project-based historical knowledge assessment instrument that has good evidence of validity and reliability that can be used to measure students' historical knowledge more precisely and more accurately. Historical knowledge is assessed based on six indicators: what, who, where, when, why, and how. The six indicators are described in a 12 item Likert scale. Based on these 12 items, students' historical knowledge was classified into four categories: very high, high, acceptable, and weak. The instrument was tested on 426 10th grade high school students in Yogyakarta, Indonesia. The test results data were used to analyze the characteristics of the instrument consisting of validity and reliability. The test results show that the instrument developed is valid and reliable, so it is suitable to be used to assess historical knowledge on a broader scale.

KEYWORDS

Project-based learning, Historical knowledge, Assessment instrument, High school, Indonesia

CITATION

Supianto, Kumaidi, & Suryono, Y. (2021). Development of project-based historical knowledge assessment instruments, *Historical Encounters*, 8(1), 16-29. <https://doi.org/10.52289/hej8.102>

COPYRIGHT

© Copyright retained by Author/s

Published 4 May 2021

Distributed under a CC BY-NC-ND 4.0 License

Introduction

The industrial revolution 4.0 has changed the way humans communicate and interact with each other. This change has a direct impact on the world of education. The internet and web 2.0 development has had a significant influence on what and how learning and assessment should be carried out. History lessons as subjects taught at school also cannot close themselves from these developments. Websites and social media such as Facebook and YouTube contain many representations of history shared by others free of charge (Baron, 2014; Bell, Carland, Fraser, & Thomson, 2016). Historical sources are diverse, rich, and are no longer limited to learning resources provided by teachers and schools. The types of history that can be identified are no longer limited to national history as found in school textbooks but provide opportunities for the introduction of local history and promote it as a national identity.

Besides the historical learning model, the historical evaluation model is also used, depending on the objectives of history learning (Shemilt, 2018). Historical assessment and learning aim to foster national integration, promote awareness of life in history and dynamic socio-cultural contexts, and obtain meaning and historical wisdom from the history that has been studied (Alvén, 2017; Funkenstein, 1989; Lapin & Tomanova, 2016). However, the assessment model used by history teachers has not been able to measure the level of students' historical knowledge accurately.

In the mid-20th century, most of the history lessons at school attempted to transmit factual knowledge about national and regional history. The assessment result is used for selection purposes by companies and universities. Rating is rarely considered problematic. Criticism may be made to ask the objectivity and reliability of ratings, but rarely questioned the significance of assessing students based on their memory of events, names, and dates. Multiple-choice tests, although criticized, are believed to be the most appropriate way to describe student learning outcomes accurately (Black & William, 1998).

In recent years, the relevance and usefulness of the assessment approach used so far in historical education have been questioned. Traditional methods and approaches are said to have led us to assess what is measured rather than encourage measurement of what we value (Biesta, 2009). Even to the extent that many teachers feel depressed because they have to teach for exams (Wineburg, 2006). Such criticism is less driven by deficiencies found in traditional assessment methods and procedures than by new ideas about the purpose of historical education.

The assessment model used by the history teacher is still a paper and pencil test with the multiple-choice format and question matters (Breakstone, Wineburg, Grossman, & Labaree, 2013). These types of assessments provide little information about students' knowledge and are not sufficient as a basis for teachers to make decisions in the classroom (Madaus & Russel, 1989). This is in line with the results of interviews with several high school history teachers in Yogyakarta, Indonesia. According to most history teachers, the learning models used are quite diverse; for example, project-based learning. However, the obstacle faced by teachers is the lack of availability of assessment models that can be used to assess student projects. The limitations of the project assessment model make teachers rely more on subjectivity in assessing students' historical writing project.

Multiple choice questions are not a reasonably accurate tool that can be used to gather information about students' understanding of specific dimensions of the constructs of historical learning, such as historical knowledge, historical thinking, and historical consciousness. Researchers argue that this method is solely to measure students' ability to remember historical facts (Reich, 2009; Wineburg, 2004), and students more often answer the multiple-choice questions without using historical thinking procedures correctly (Breakstone, J & Smith, 2013). Multiple-choice questions also provide limited information to the teacher because they do not express students' thoughts that lead to answers (Madaus & Russel, 1989). The dark circle on the answer sheet is the only indication of student thought.

History education curriculum in Indonesia

Indonesia has made changes to the education curriculum three times in the last twenty years. The first change was made in 2004, where history education was given a tiny portion of time, only 40 minutes a week in high school (Depdiknas, 2003). The main focus of history lessons in this curriculum is to provide as much material as possible in the shortest possible time. The time to do the assessment is neglected, so the teacher uses the simplest, easiest, and fastest method, namely multiple choices.

The second change was made in 2006. The 2006 curriculum eliminates the position of history lessons in junior high schools as an independent subject, where history and other social science subjects are integrated into the field of social science studies (Depdiknas, 2009). The 2006 curriculum was implemented for a long time until a third curriculum change was made in 2013.

In the 2013 Curriculum, history subjects get quite essential positions, both from the allocation of time and from the lesson content. In terms of time, history lessons are taught for 80 minutes in grades 10 and 11, and 120 minutes in grade 12 (Kemendikbud, 2018). In terms of content, history lessons are no longer focused only on the transmission of historical knowledge from teachers and books to students. Still, the content of the lessons has also included material on historical research methods that demand the competency development of historical skills.

The 2013 curriculum provides more flexibility for history teachers to use various learning and assessment models that emphasize learning by doing (Kemendikbud, 2017). This was emphasized by Minister of Education and Culture, Nadiem Makarim, with the idea *Merdeka Belajar*, where learning and assessment can be carried out anytime, anywhere, in any way, even by anyone (Kemendikbud, 2019).

For example, in grade 10 on historical methods material, the teacher can integrate project-based learning, project-based assessment, and class discussions. In this material, students can be given a series of project assignments to apply historical methods in simple research practice. Through this project assignment, the teacher can assess students' historical knowledge regarding the historical project they are working on, as well as assessing students' historical research skills. Unfortunately, as previously explained, these kinds of learning models have not been widely supported by the availability of appropriate assessment models in the literature (Hellström, Nilsson, & Olsson, 2009; Touimi, Faddouli, Bennani, & Idrissi, 2013).

Project assessment in history learning

Assessment and learning are linked (Hargreaves, 1997; Rust, 2002). What and how students learn depends mainly on how they think they will be assessed (Ludvigsson, 2003). This implies that assessment is a learning tool that can be strategically used in a learning environment to obtain better learning outcomes.

Project assignments challenge students to think outside the classroom boundaries, helping them develop the skills, behaviours, and beliefs necessary for success in the 21st century. Designing learning environments that help students to question, analyze, evaluate, and forecast their possible plans, conclusions, and ideas, leading them to higher-order thinking, requiring feedback and evaluation that goes beyond letters or numbers.

Project assessment is an assessment of performance and products. It can be carried out in a variety of ways, from conventional written tests to more innovative assessment models such as self-assessments, peer assessments, joint assessments, portfolio assessments, and reflective journals (Van den Bergh et al., 2006). Teachers can assess students' cognitive abilities based not only on student work, reflection, evidence of progress and performance, but also their attitudes and learning progress. Through project assessment, teachers can simultaneously assess both the learning process and student project outcomes.

Project-based learning emphasizes the integration of students' knowledge and problem-solving skills. Project assessment focuses on a variety of multidimensional processes. Project material that students work on focuses on the application of knowledge that students have in the

form of real work and the products they produce. Project-based learning emphasizes these products as outputs and as the primary outcome in learning. Dennis et al. (2006) concluded that reflection on project-based learning could be done at any time by observing the progress of the project done by students. This allows the teacher to be able to guide students until the project they are working on is completed. Therefore, both project-based learning and project assessment can be carried out simultaneously.

According to some literature, conventional assessment methods do not support students' understanding and skills gained from project-based learning (Frank and Barzilia, 2004). This is what motivates researchers to develop an alternative assessment model consisting of a combination of assessment methods.

Historical knowledge

Historical knowledge is knowledge of what has happened in the past in human history, or knowledge of historical facts and processes (Topolski, 1973). The essence of historical knowledge is the disclosure of events with all the facts that include *what, who, when, where, why, and how* (Grant, 2003). Historical knowledge lies not in what can tell about the future, but in what can tell about the past (Elliott, 2003). Historical knowledge can be measured through learnable facts (Grant, 2003).

Intellectual curiosity about the past is a reason why people study and read history (Tosh, 2002). Historical knowledge is one element of historical understanding (Grant, 2003). Historical understanding is seen in terms of substantive and procedural knowledge of historical disciplines (Husbands, Kitson & Pendry, 2005). Students' knowledge of history includes an understanding of causality (Kitson and Husbands, 2011). Students consume and produce information through a wide variety of texts and develop skills concerning the interpretation and construction of historical knowledge and ideas. Students can be positioned as consumers as well as producers of historical knowledge when they study history.

Based on the definition described above, the historical knowledge construct is described using six indicators, each of which is broken down into several items. The six indicators are *what, who, when, where, why, and how*. *What* is the ability of students to know whether an event belongs to the category of history or not history (for example, myths, legends, fables). *Who*, the ability of students to identify who are the actors involved in historical events. *When* the ability to analyze precisely when an event occurs. *Where*, the ability of students to identify places where historical events occur and how they are related between areas within the same time frame and events. *Why*, the ability of students to analyze the factors that cause events to occur. *How*, the ability of students to explain how the historical events occur chronologically.

The sixth indicator of historical knowledge that has been described above will be assessed through the project of writing history. Through this assessment, the teacher cannot just simply assess students' knowledge of history, but at the same time, assess the critical thinking ability of students to use the knowledge they have.

Methodology

This research is development research adopting the R&D model proposed by Plomp (Plomp, 2010), which consists of five phases: the preliminary investigation phase, design phase, construction phase, test, evaluation and revision phase, and implementation phase. The product developed in this study is a project-based historical knowledge assessment instrument.

Trials were conducted on 426 10th grade students in five high schools in Yogyakarta, Indonesia. The instrument was validated by four experts in the field of assessment and three experts in history. The results of the validation were analyzed with the Aiken validity index. The construct validity was analyzed using Confirmatory Factor Analysis of Second Order with Lisrel 8.50. Instrument reliability was estimated using Cronbach's Alpha. Item characteristics were analyzed

using IRT Partial Credit Model. The following is an instrument used to assess a student's project-based historical knowledge.

Skills	On this assignment, student used the skill on...			
	All (4)	Most (3)	Some (2)	None (1)
Determine the theme of the article				
Explain the historical site of the event				
Mention offenders in the same direction				
Identify the role of each actor				
Identify the location of the event				
Analyzing the relationship between regions				
Mention the time of the event				
Analyze the relationship between events at different times				
Analyze the causes of events				
Analyzing can be of an event				
Explain the process of occurrence of events				
Chronology history writing				

Table 1. Historical knowledge assessment instrument

In this study, the learning design and project assessment to measure the level of student knowledge are briefly described as follows.

1. The learning process in the classroom applies project-based learning methods.
2. Teacher gives project assignments to students.
3. The project assignment referred to in point 2 is a project to write a popular history article on a topic determined by the teacher.
4. The popular article writing project was done for two weeks.
5. Popular history articles that have been completed by students are posted to the www.historista.id website that has been prepared. Historical knowledge assessment instruments developed have been integrated into the website.
6. The teacher assesses the students' historical knowledge.
7. All project assessment results can be downloaded on the www.historista.id for further analysis.
8. The polytomous IRT PCM model then analyzes the raw data that has been downloaded.

The capability parameters obtained through the IRT analysis of the PCM model polytomous are converted into scores on a scale of 1-100 so that a final score is achieved. This final score will describe the overall profile of students' historical knowledge. The results of historical knowledge assessment are presented in the categories of Very High, High, Acceptable, and Weak. The range of values used for each aspect differs from one another. This is based on the average value and standard deviation of each element. The provisions used to determine the range of scores for each category are as follows (see Table 2).

Score	Category
$X \geq \tilde{x} + 1. s_x$	Very High
$\tilde{x} \leq X < \tilde{x} + 1. s_x$	High
$\tilde{x} - 1. s_x \leq X < \tilde{x}$	Acceptable
$X \leq \tilde{x} - 1. s_x$	Weak

Table 2. Score category

Results

Instrument Development

The development of the instrument was carried out after the researchers conducted a preliminary investigation, which included curriculum studies, pre-surveys, literature studies, and relevant research studies. Curriculum study is intended to review government regulations and policies related to history lessons in the curriculum. The pre-survey was conducted to obtain information on current conditions of learning and historical assessment, the real problem of historical learning and assessment, and identification of the needs of history teachers regarding the assessment of historical knowledge. Information about these matters was obtained through direct interviews with history teachers.

Based on an interview with a history teacher in Yogyakarta, Indonesia, information was obtained about the historical knowledge assessments that teachers do through daily tests, mid-term assessments, and assessment of the final semester. The assessment techniques are mostly in the form of paper and pencil test by giving a series of multiple-choice questions and filling in related material that has been taught. Peer-to-peer assessments and self-assessments are also carried out to assess aspects of attitudes. As for the component of historical research methods, the teacher prefers to test students' conceptual knowledge of historical research methods through multiple-choice tests and descriptions. A small portion provides the assignment of history book resumes.

As support to learning, history teachers have used the internet as a source of learning history. However, the availability of websites that contain quality historical content is not widely available. While for assessment, history teachers have not used the internet at all as an assessment medium. This is because there are no website-based historical assessment instruments available to be used by teachers. Google Forms can be used as an alternative internet-based assessment medium, but due to the limited knowledge and skills of teachers in using it, its utilization is also minimal.

Through literature studies and relevant research conducted at the preliminary investigation stage, researchers have collected research support materials in the form of literature relating to the concept of historical knowledge and previous studies related to the assessment of historical knowledge. Then at the design stage, the researcher sought to formulate a conceptual model of historical knowledge consisting of the following six indicators:

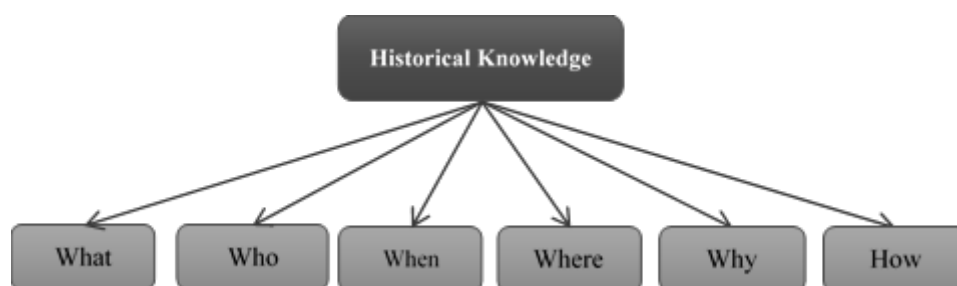


Figure 1. Conceptual model of historical knowledge assessment

The six indicators that make up historical knowledge (as outlined previously) are *what*, *who*, *when*, *where*, *why*, and *how*. *What* is knowledge about historical events, including the ability to distinguish between historical events or legends. *Who* is knowledge about historical actors and all people involved in the event, including knowledge about who are the main actors for the occurrence of a historical event. *When* is the ability to remember and mention time accurately, including date, day, hour, month, and year. *Where* is the ability to identify, find, show, and mention the exact place where an event occurred.

Besides, *where* also includes the ability to connect events that occur in one place with events that occur in other places. *Why* is the ability to find the causal relationship of an event. Which factors cause the occurrence of events and how the consequences of these events, both short-term consequences and long-term consequences. *How*, the ability to describe, describe, and recount past events chronologically and periodically so that the story conveyed has meaning and can be enjoyed by others.

After the constructs and components of historical knowledge are successfully designed, the next stage is the preparation of historical knowledge assessment instruments. The following is a grid and instrument for the assessment of historical knowledge.

Variable	Indicator	Item
Historical Knowledge	1. What	1. Determine the theme of the article 2. Explain the historical site of the event
	2. Who	3. Mention offenders in the same direction 4. Identify the role of each actor
	3. Where	5. Identify the location of the event 6. Analyzing the relationship between regions
	4. When	7. Mention the time of the event 8. Analyze the relationship between events at different times
	5. Why	9. Analyze the causes of events 10. Analyzing can be of an event
	6. How	11. Explain the process of occurrence of events 12. Chronology history writing

Table 3. Indicators and items of historical knowledge instruments

Instrument validity and reliability

The quality of the historical knowledge assessment instrument that has been developed was determined through content validity, construct validity, instrument reliability, and instrument characteristics. Content validation was carried out by experts to assess the appropriateness of the instrument before being tested.

There are seven experts involved in validating the contents of this instrument consisting of three measurement experts as well as historians, one measurement expert, one historian, and two outstanding history teachers.

The expert conducted validation based on aspects: (a) conformity of items with indicators and components, (b) language used, (c) statement of items that are not biased, (d) clarity of statement, (e) clarity of instructions for filling instruments. Then the expert gives a score in the form of a rating scale with a range of 1-4, where 1 = Weak, 2 = Acceptable, 3 = Good, and 4 = Very Good.

While the V value of the table is the minimum value of this validity index based on the number of rater in V_{table} , the following are the results of the experts' evaluation of historical knowledge assessment instruments.

Items	Content Validity					Construct Validity		Exp
	$\sum s$	n	c-1	V_{table}	V_{count}	t-value	SLF	
x1	19	7	3	0.76	0.9	**	0.739	Valid
x2	18	7	3	0.76	0.86	14,415	0.791	Valid
x3	20	7	3	0.76	0.95	**	0.762	Valid
x4	18	7	3	0.76	0.86	15,299	0.790	Valid
x5	19	7	3	0.76	0.9	**	0.654	Valid
x6	20	7	3	0.76	0.95	12,629	0.764	Valid
x7	19	7	3	0.76	0.9	**	0.809	Valid
x8	18	7	3	0.76	0.86	10,628	0.562	Valid
x9	19	7	3	0.76	0.9	**	0.374	Valid
x10	17	7	3	0.76	0.81	5,941	0.521	Valid
x11	20	7	3	0.76	0.95	**	0.561	Valid
x12	19	7	3	0.76	0.9	10,072	0.744	Valid

Table 4. Content and construct validity

Results of the analysis showed that the content validity of s Retained Earnings 12 item instrument for measuring knowledge of history meets good validity because of the average content validity index above the minimum limit of 0.76, which amounted to 0.89. In other words, the unidirectional knowledge assessment instrument that has been developed deserves further testing.

The next step is to know the construct validity. But first, the suitability of the measurement model was tested. Based on the test analysis of the quality of fit model. The value of the index matches the following model: p-value of 0.23 (≥ 0.05), RMSEA 0:02 (≤ 0.08), CFI and GFI amounted to 0.99 and 0.97 (≥ 0.90). With such, these model fit test results indicate that the measurement model good knowledge of the history category.

Next is the analysis of the construct validity of the historical knowledge assessment instruments. The construct validity of the instrument can be seen from the t-value (≥ 1.96), and the loading factor value (≥ 0.3). Based on the results of confirmatory factor analysis (CFA), t-values and factor loading values are obtained in the table below.

Table 3 shows that the loading factor is ranging from 0.374 to 0.809, exceeds 0.3. T-value ranges from 5.941 to 15.299, exceeded the target >1.96 . Thus, it can be said that the instrument of the assessment of historical knowledge has good and significant construct validity. Meanwhile, when viewed from the reliability of the instrument, the SPSS output results show that the Cronbach's Alpha coefficient is 0.903. In other words, if seen from the construct, the scale to measure historical knowledge can be said to be reliable or consistent.

Item characteristics

Figure 2 shows that there is one factor that is more dominant than the others. That means that the unidimensional assumptions have been fulfilled. The second assumption is local independence, which can be interpreted that when the ability to influence test performance is constant, student responses to pairs of items are statistically independent (Hambleton, 1991, p 10). Local independence is equivalent to unidimensional assumptions. That is when the unidimensional assumptions are met then automatically; the assumption of local autonomy is also fulfilled.

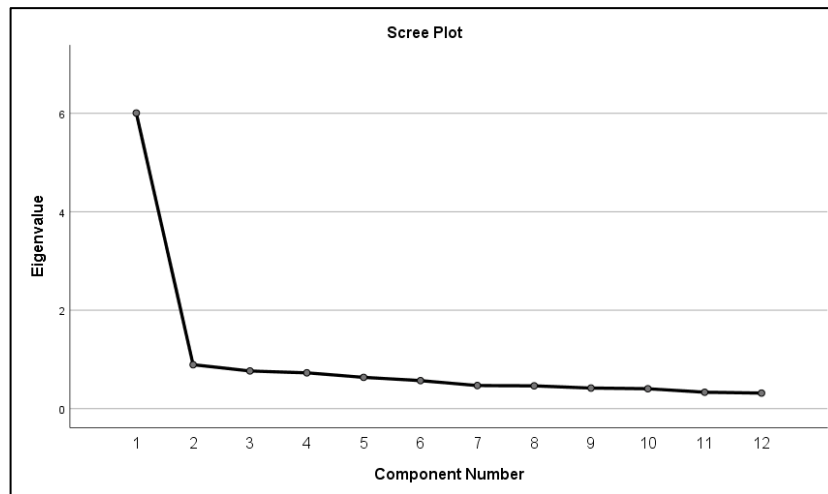


Figure 2. Scree plot unidimension

The third assumption of IRT is parameter invariance, which can be proven by comparing parameter estimates in different groups. In this study, groups were distinguished according to odd and even order numbers. After students are divided into groups, researchers estimate the difficulty parameters of the items in each group. The ability level parameters of students were also classified based on groups of odd and even items. The parameters of each group are correlated, then displayed in the form of a scatter plot. The following is a picture of a scatter plot comparing the parameters of odd and even items as well as the ability parameters in the first half and second half students.

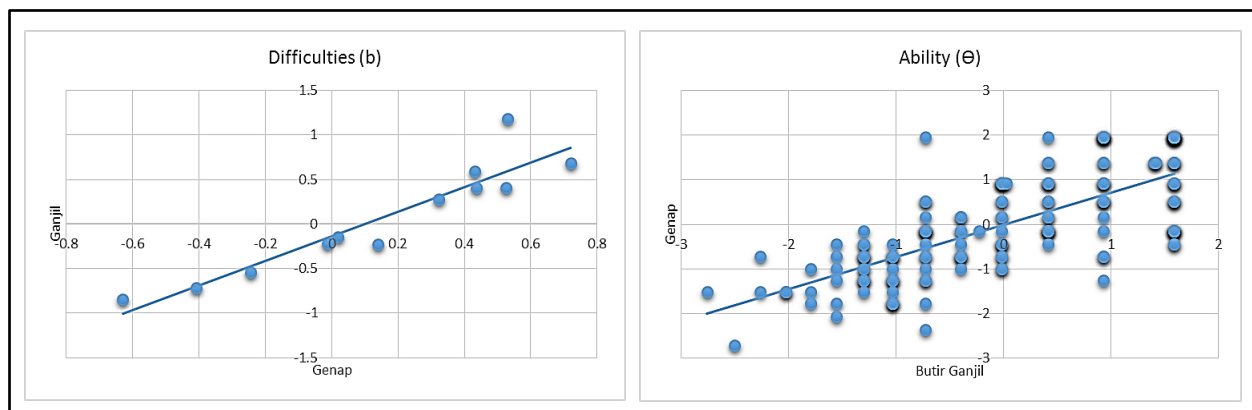


Figure 3. Parameter invariance

Based on figure 3 above, the difficulty parameter (b) and the ability parameter (θ) can be seen that most of the points spread around the slash. So, it can be said that there is no variation in the estimated parameters of two group students and two groups of items. Thus, the assumption of parameter invariance can be fulfilled. Based on the results of the PCM model PCM polytomous IRT analysis using the R Studio program, overall information about the estimated grain parameters in the table below is obtained.

No.	b	b1	b2	b3
x1	-0.01248	-0.62433	-0.89842	1,48530
x2	0.30306	-0.17568	-1.55987	2.64473
x3	-0.07516	-0.40965	-1,09837	1.28253
x4	-0.53874	-0.59124	-1,02417	-0,00081
x5	0.47296	0.16784	-0.50942	1.76047
x6	0.50938	-1.25256	0.36536	2,41533
x7	-0.73595	-1.67067	-1,04864	0.51146
x8	-0.36152	-0.96606	-0.97167	0.85316
x9	0.41285	0.02855	-1.05663	2,26662
x10	0.72109	-0.15756	0.05451	2.26631
x11	0.69573	0.09852	-0.67565	2,66431
x12	-0.10553	-1.24344	0.37729	0.54955

Table 5. Item parameters

Table 5 shows the level of item difficulty is in the range from -0.376 to 0.721, and the step difficulty from -1.670 to 2.664. Thus, all items can be categorized as useful. The items are good if the difficulty level parameter is in the range -3.00 and 3.00. Based on the table above, the 12 instrument items have suitable grain parameters and match the model so that the curve clearly shows the boundaries of each category used.

The next step is to test the suitability of the model. Items that fit the model are items with a significant chi-square probability value, items that have a chi-square probability ≥ 0.05 . The following are the results of the analysis of the compatibility of the model with RStudio.

Figure 4. Output RStudio item fit

	item	G2	df.G2	RMSEA.G2	p.G2	S_X2	df.S_X2	RMSEA.S_X2	p.S_X2
1	x1	33.361	14	0.099	0.003	26.973	24	0.030	0.306
2	x2	16.676	9	0.078	0.054	24.507	20	0.040	0.221
3	x3	21.195	13	0.067	0.069	29.881	25	0.037	0.229
4	x4	41.375	15	0.112	0.000	25.748	19	0.050	0.137
5	x5	27.226	16	0.039	0.284	16.545	18	0.000	0.555
6	x6	23.220	13	0.075	0.039	30.662	23	0.049	0.131
7	x7	18.026	12	0.060	0.115	28.764	21	0.051	0.120
8	x8	26.243	13	0.085	0.016	26.629	27	0.000	0.484
9	x9	25.026	12	0.086	0.026	19.823	20	0.080	0.469
10	x10	16.678	15	0.028	0.338	19.876	26	0.000	0.797
11	x11	33.683	15	0.094	0.004	20.172	20	0.008	0.447
12	x12	37.424	15	0.103	0.001	28.781	27	0.022	0.372

In the output (see Figure 4), column p.S_X2 indicates a p-value where the p-value of 12 items is all above 0.05. Thus, it can be said that the items in the historical knowledge assessment instrument fit into the PCM model.

Student historical knowledge profile

Historical knowledge in this study includes indicators of what, who, when, where, why, and how. The five indicators represent the historical knowledge possessed by students. The ability level parameters obtained by analyzing abilities with the IRT polytomous PCM model are converted into scores on a scale of 10-100. The histogram below shows the distribution of scores of students overall.

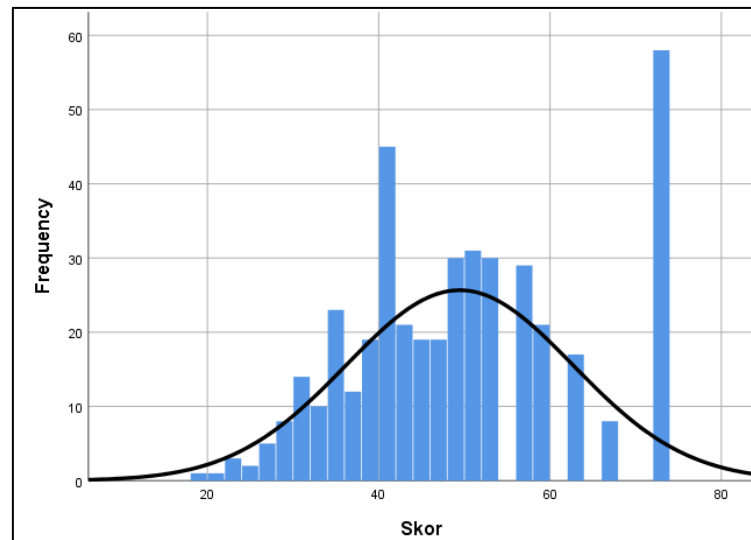


Figure 5. Students' ability of historical knowledge

The histogram above shows that the average score obtained by students is 49.44, with a standard deviation of 13. Based on the average value and the standard deviation, a range of categorical values is obtained as follows: >61 is Very High, 52-61 is High, 42-51 is Acceptable, and <42 Weak. Visually, the results of achieving students' historical knowledge in the form of categories can be seen in Figure 6.

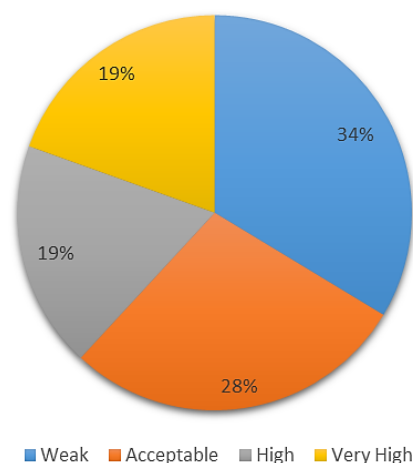


Figure 6. Percentage of historical knowledge categories

Based on figure 6, it can be seen that from 426 students in five research schools, students who have a historical knowledge score of Very High is 19%, High is 19%, Acceptable is 28%, and Weak is 34%. It appears that the results of historical knowledge assessment are dominated by students who have weak or acceptable abilities, with both cumulative percentages of 62%; as many as 263 students.

Conclusion

A valid and reliable historical knowledge assessment instrument is an inevitable need for teachers to be able to determine students' abilities accurately. Project-based learning and assessment is an integral unit. The lack of project assessment instruments in history teaching can be overcome by instruments like those that have been developed in this study. The instruments designed in this study have passed the most important stages in the instrument development process, especially in the validity and reliability tests. This instrument is considered valid, reliable, and fit for use.

The validity of this instrument is proven by the value of content validity and construct validity. The content validity is intended to determine the suitability of items with indicators and historical knowledge variables. The content validity value based on the expert judgment has been fulfilled, likewise with construct validity, which is intended to ensure that the instrument has been able to reveal the thinking constructs that need to be assessed, in this case, historical knowledge. The result of construct validity analysis also exceeds the minimum limit of the agreed terms. The results of the consistency analysis of the instruments to measure the constructs were also proven. So, the instrument can be considered valid and reliable. The fulfillment of the validity and reliability requirements in the development of this instrument makes it feasible to be used on a broader scale. The products produced in this study can be directly used by history teachers in high schools to assess students' historical knowledge. The product implementation provided in this study can be used in formative and summative assessments. If used in formative assessment, then the goal is to track the development of understanding and mastery of historical knowledge. If used in the summative assessment, the goal is to measure students' overall historical knowledge. As an example, this may be achieved by giving students the task of evaluating history articles as final semester assignments.

The assessment model developed in this study is Project-Based Assessment (PBA), which collaborated with Project-Based Learning (PBL). Both the assessment and learning models could be improved if they were implemented with the class discussion method. This class discussion is intended to facilitate students to deliver oral presentations about the topics of history they have written and have been posted on the www.historista.id website. Thus, the assessment process will not only stop at the site but have useful implications for the learning process in the classroom.

References

- Alvén, F. (2017). Making democrats while developing their historical consciousness: A complex task. *Historical Encounters*, 4(1), 52–67.
- Baron, C. (2014). Structuring historic site-based history laboratories for teacher education. *Journal of Museum Education* 39(1), 10–19. <https://doi.org/10.1080/10598650.2014.11510791>
- Bell, J., Carland, R., Fraser, P., & Thomson, A. (2016). 'History is a Conversation': Teaching student historians through making digital histories. *History Australia* 13(3), 415–30. <https://doi.org/10.1080/14490854.2016.1202373>
- Biesta, G. (2009). Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation, and Accountability* 21(1), 33–46. <https://doi.org/10.1007/s11092-008-9064-9>
- Black, P. & William, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-148.
- Breakstone, J., & Smith, M. (2013). Assessing historical thinking skills using Library of Congress primary sources. *The TPS Journal* 6(1), 1–8.

- Breakstone, J. (2013). *History assessments of thinking: Design, interpretation, and implementation* (Doctoral Dissertation). Stanford University.
- Dennis, C. & Hardy, J. (2006). Development of a model to advance the uptake of e-portfolios for undergraduates in teacher education and registered nurse preparation: An exemplar of best practice. In E. Pearson & P. Bohman (Eds.), *Proceedings of ED-MEDIA 2006--World Conference on Educational Multimedia, Hypermedia & Telecommunications* (pp. 248-253). Orlando, FL USA: Association for the Advancement of Computing in Education.
- Depdiknas. (2003). *Kurikulum berbasis kompetensi*. Balitbang: Jakarta.
- Depdiknas. (2009). *Buku Saku Kurikulum tingkat satuan pendidikan (KTSP) Sekolah menengah pertama*. Direktorat Pembinaan Sekolah Menengah Pertama: Jakarta.
- Elliott, J., (2003). The limits of historical knowledge. *European Review* 11(1), 21–25. <https://doi.org/10.1093/analys/31.1-2.58>.
- Frank, M., & Barzilai, A. (2004). Integrating alternative assessment in a project-based learning course for pre-service science and technology teachers. *Assessment and Evaluation in Higher Education* 29(1), 41–61. <https://doi.org/10.1080/0260293042000160401>
- Funkenstein, A. (1989). Collective memory and historical consciousness. *History and Memory* 1(1), 5–26.
- Garg, B. (2007). *The teaching of history*. New Delhi: Rajat Publications.
- Grant, S. G. (2003). *History lessons: Teaching, learning, and testing in the US. High school classrooms*. New York: Lawrence Erlbaum Associates, Inc.
- Hargreaves, D. J. (1997). Student learning and assessment are inextricably linked. *European Journal of Engineering Education* 22(4), 401–9. <https://doi.org/10.108003043799708923471>
- Kemendikbud. (2017). *Panduan Penilaian Kurikulum 2013*. Balitbang: Jakarta.
- Kemendikbud. (2018). *Manajemen Implementasi Kurikulum 2013*. Direktorat Pembinaan Tenaga Kependidikan Pendidikan Dasar dan Menengah: Balitbang, Jakarta.
- Kemendikbud (2019). *Kebijakan Merdeka Belajar 1: Empat Pokok Kebijakan Merdeka Belajar 2019*. Available online: www.kemdikbud.go.id/main/blog/2019/12/empat-pokok-kebijakan-merdeka-belajar.
- Kitson, A., & Husbands, C. (2011). *Teaching & learning history 11-18: Understanding the past*. New York: Open University Press.
- Lapin, N. S., & Tomanova, M. S. (2016). Historical consciousness and perception of national history by Kazakhstan youth in the 20th century. *Indian Journal of Science and Technology* 9(48). <https://doi.org/10.17485/ijst/2016/v9i47/108366>
- Madaus, G., & Russel, M. (1989). Paradoxes of high-stakes testing. *Journal of Education* 190(1), 20–30.

- Plomp, T. (2010). Educational design research: An introduction. In T. Plomp & N. Nieveen (Eds), *An introduction to educational design research* (pp. 9–36). Netherlands Institute for Curriculum Development: Netherlands.
- Reich, G. A. (2009). Testing historical knowledge: Standards, multiple-choice questions, and student reasoning. *Theory and Research in Social Education* 37(3), 325–60. <https://doi.org/10.1080/00933104.2009.10473401>.
- Rust, C. (2002). The impact of assessment on student learning: How can the research literature practically help to inform the development of departmental assessment strategies and learner-centred assessment practices? *Active Learning in Higher Education*, 3(2), 145–58. <https://doi.org/10.1177/1469787402003002004>.
- Shemilt, D. (2018). Assessment of learning in history education: Past, present, and possible futures. In S. A. Metzger, & L. M. Harris (Eds), *The Wiley international handbook of history teaching and learning* (pp. 449-471). John Wiley & Sons Inc.: Hoboken, New Jersey.
- Topolski, J. (1973). *The methodology of history: Vol. s1-10*. Boston: D. Reidel Publishing Company. <https://doi.org/10.1093/philmat/s1-10.1.111>
- Tosh, J. (2002). *The pursuit of history: Aims, methods, and new directions in the study of modern history* (3rd Ed.). London: Pearson Education Limited.
- Van den Bergh, V., Mortelmans, D., Spooren, P., Van Petegem, P., Gijbels, D., & Vanthournout, G. (2006). New assessment modes within project-based education - the stakeholders. *Studies in Educational Evaluation* 32(4), 345–68. <https://doi.org/10.1016/j.stueduc.2006.10.005>
- Wineburg, S. (2004). Crazy for history. *Journal of American History* 90(4), 1401-1414. <https://doi.org/10.2307/3660360>
- Wineburg, S. (2006). A sobering big idea. *Phi Delta Kappan* 87(5): 401–2. <https://doi.org/10.1177/003172170608700514>

Acknowledgements

The author would like to thank the Indonesia Endowment Fund for Education (LPDP) for funding this research.

About the Authors

Supianto is a doctor in educational research and evaluation, teacher of social science education at junior high school, and compiler of the apparatus development program at BKPSDM, Sumenep Regency.

Kumaidi is a Professor in measurement and classroom assessment at Muhammadiyah University of Surakarta.

Yoyon Suryono is a Professor in educational research and evaluation, Head of Master Program in Nonformal Education, Yogyakarta State University.