

Development of Portfolio Assessment to Measure the Student's Skill of Using Primary Source Evidence

DOI: 10.15804/tner.2019.56.2.08

Abstract

The study aims to develop a portfolio assessment to measure the student's skill of using primary source evidence in historical learning. The portfolio assessment was developed based on Seixas' concept of historical thinking. The portfolio assessment consists of essay test and interview guide, which has proved to be valid and reliable. The developed portfolio assessment can be used to depict the student's skills to organize information from historical documents or other written resources, and then reconstruct it as historical narratives. The developed portfolio assessment can be posited as an alternative evaluation mode for history teachers.

Keywords: portfolio assessment, historical thinking, using primary source evidence

Introduction

The main objective of learning history is to enhance students' historical thinking skills. Seixas and Peck (2004) and Ercikan and Seixas (2015) defined historical thinking as a process by which students use procedural knowledge in the discipline of inquiry. Furthermore, Seixas (2006) divided historical thinking into six concepts consisting of constructing historical meaning, using the main sources of evidence, recognizing continuity and change, analyzing causality, understanding historical perspectives, and understanding moral dimensions in historical interpretation. Seixas (1998) then explained that using primary source evidence is an

important skill to reconstruct the past because it teaches students to find evidence from historical resources. Therefore, students should be able to inventively read and contextualize the evidence in its historical context. By these skills, students can provide new evidence of historical events.

Considering its importance, the student's skill in using primary source evidence needs to be measured. Seixas, Gibson, and Ercikan (2015) developed a historical thinking assessment model called One-Hour Test model. The One-Hour Test model consists of three types of assessment, i.e., short answer, multiple choice, and constructed response question. Furthermore, Smith (2017) adopted Seixas' concept of historical thinking and developed a multiple-choice model to measure historical thinking skills. Nevertheless, these instruments have a fundamental weakness if they are analyzed based on some theoretical studies. Reich (2009) revealed that the multiple-choice test contains a striking misalignment between the student's correct answers and the standard of historical analysis claimed by the test. Reisman (2012) confirmed that the reliability of resources in constructing information from the past with students' different knowledge and reading comprehension is difficult to evaluate with the use of a multiple-choice test. The authors believe that the One-Hour Test Model might have a weakness similar to the multiple-choice instrument. It could become inadequate to measure the complexity of cognitive progress in students' historical thinking. Therefore, the existing model of historical thinking assessment should be advanced by holistic approaches based on the six concepts of historical thinking.

Based on the above discussion, this study aimed to develop an instrument to measure students' historical thinking especially the skills of using primary source evidence in the form of portfolio assessment, which consists of essay items and interview guides. The authors try to adopt Seixas, Gibson, and Ercikan's model of assessment (2015) by developing portfolio assessment based on the concept of historical thinking. By using portfolio assessment, the teacher would be able to measure the level of students' higher order thinking skill in learning history.

Research Methodology

The authors used Getsdottir, Boxtel, Drie's (2018) framework as a research method and the Guide to Developing Assessment Tools compiled by the Australian Skills Quality Authority for the Design and Development stages. The framework consists of literature review, expert consultation, first trial, and second trial.

The research participants were 60 senior high school students. Firstly, the authors reviewed the literature related to the concepts of using primary source. Four history teachers were also interviewed and 30 students were asked to fill out questionnaires to find the empirical problems of using primary source evidence. Moreover, the authors designed and developed the outline of portfolio instruments. The developed instrument was then validated by an expert, including the feasibility of assessment instrument, material expert, and text comprehension. The feasibility of the developed instrument was measured using the criteria used by Sutimin, Joebagio, Sariyatun, and Abidin (2018).

Table 1.

Range	Criteria
3.26–4.00	Very Feasible
2.51–3.25	Feasible
1.76–2.50	Quite Feasible
1.00–1.75	Infeasible

The next stage aimed to test the developed instrument. The trial stage was divided into two phases: small group trial and a large group trial. In the second trial, 60 students, divided into the control and experimental groups, were involved in measuring the effectiveness of the developed instruments. Data were analyzed with the Quest program by using the Partial Credit Model (PCM) proposed by Masters and Wright (1982). Validity was determined based on the average of INFIT Mean of Square (INFIT MNSQ) along with the standard deviation or the average value of INFIT Mean of INFIT t. The developed item is declared to be fit if the value of INFIT MNSQ is 0.77 to 1.30 (Adam & Khoo, 1996).

Results and Development

The framework of portfolio assessment

The frame of the portfolio assessment refers to Seixas' (2006) definition of using primary source evidence consisting of an ability to use historical sources. Seixas, Gibson, and Ercikan (2015) defined these skills as a way to learn about the past. In order to make a valid historical conclusion from the traces of the past, students need to read and analyse primary sources. The skills of using primary source evidence do not merely relate to an ability to read books to find information, but

they lead to the meaning of what is conveyed by historical sources. In terms of using primary source evidences, students should be able to examine the historical sources, formulate questions related to historical sources, analyse historical sources to find out their value and perspectives, compare the perspectives and uses of several historical sources, assess what historical sources cannot answer and use historical sources to construct an argument or narrative. The outline of the developed instruments is presented in Table 2.

Table 2. The outline of assessment of using primary source evidence

No.	Framework of using primary source evidence	Indicators of using primary source evidence	Type of instruments
1.	Using primary source to explain a particular past event	• Determining the types of primary sources	Essay Test
		• Identifying information included in primary sources.	Essay Test
		• Arranging historical narrative from primary sources.	Essay Test
2.	Using primary source to reconstruct past event.	• Determining the primary sources to reconstruct past event.	Essay Test
		• Identifying information included in primary sources.	Essay Test
		• Arranging a historical narrative from primary sources.	Essay Test
3.	Implementing the knowledge of a concept of primary sources in oral data collection.	• Using the knowledge of a concept of historical resources in oral data collection.	Interview

The development of portfolio assessment

The expert then assessed the feasibility of the outline. The results of expert judgement are presented in Table 3.

Table 3. Feasibility of developed outline

Product	Indicators	Scores	Interpretation
Essay Test	Materials	3.5	Very Feasible
	Construction	3.2	Very Feasible
	Language	3.25	Very Feasible
	Mean	3.31	Very Feasible

Product	Indicators	Scores	Interpretation
Interview Sheet	Materials	3	Feasible
	Construction	3.17	Very Feasible
	Language	3.33	Very Feasible
	Mean	3.13	Very Feasible

The levels of validity and reliability of the instruments were measured by Quest Program. The validity of the essay questions and interview guides is shown in Figure 1.

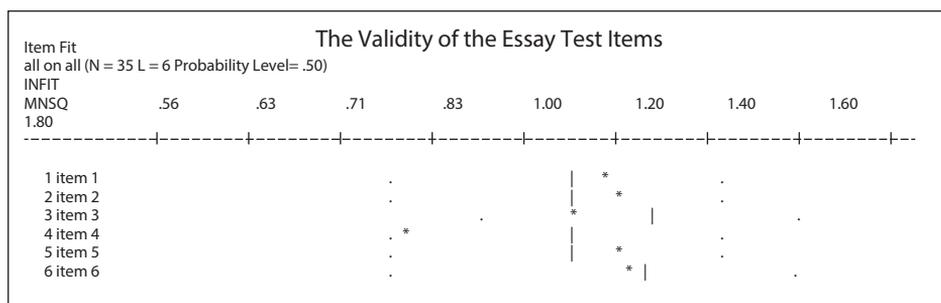


Figure 1. The results of the validity test using the Quest program

Figure 1 shows that all the items on the essay test are valid because they locate between the lower limit of 0.77 and the upper limit of 1.30. The validity value of each item is presented in Table 4.

Table 4. The validity of the essay test items

Product	Items	INFIT MNSQ (Validity Value)
Essay Test	1	0.108
	2	0.109
	3	0.86
	4	0.80
	5	0.125
	6	0.95

The reliability value of the developed item was determined based on internal consistency value. The essay test instrument is said to be reliable because it has a value of 0.63. The reliability of the developed items is shown in Figure 2.

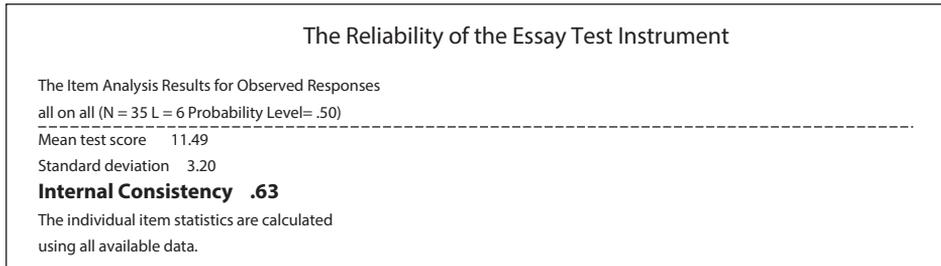


Figure 2. The results of the reliability test of the essay test items

The developed interview guides proved to be valid. The values of calculation locate between the lower limit of 0.77 and the upper limit of 1.30. The result of the validity of the interview sheet is presented in Figure 3.

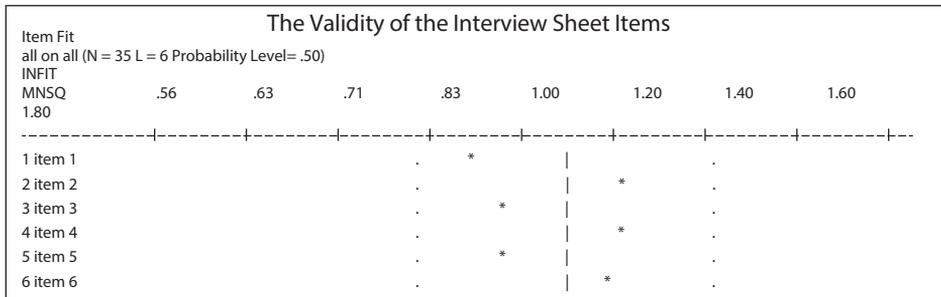


Figure 3. The validity of the interview sheet items

The detailed of values of each item value are shown in Table 5.

The reliability of the developed items is determined based on the internal consistency value. The interview sheet instrument is said to be reliable because the value of internal consistency is 0.66. The result of the reliability test is shown in Figure 4.

Table 5. The validity values of the interview sheet items

Product	Items	INFIT MNSQ (Validity Value)
Interview sheet	1	0.86
	2	0.113
	3	0.90
	4	0.13
	5	0.90
	6	0.108

Interview Sheet Reliability	
The Item Analysis	Results for Observed Responses
all on all (N = 35 L = 6 Probability Level= .50)	
Mean test score 5.00	
Standard deviation 2.07	
Internal Consistency	.66

Figure 4. The reliability of the interview sheet

Based on the above calculation, it can be concluded that the developed instrument of essays (cf., Appendix 1) and interview guides (cf., Appendix 2) proved to be valid and reliable. By considering its validity and reliability, it can be assumed that the developed portfolio instrument can be used to measure students' skill of using primary source evidence.

Discussion

The developed portfolio assessment is an alternative mode to assess students' historical thinking. In comparison to 'The Case of a One-Hour Test (2015)', the developed instrument provides a scope to analyze students' skill in using primary source evidence through essay test and interview guides. Seixas, Gibson, and Ercikan developed their instrument in the form of multiple-choice test, short essays, and constructed response questions. According to Romberg's (1993) critique of a multiple-choice test, it can be assumed that Seixas, Gibson, and Ercikan's instrument theoretically could only assess the student's lower order thinking skills such as memorization skills or historical understanding. Moreover, multiple-choice tests practically cannot reveal the process as well as construction of students'

answers as the representation of the student's level of thinking (Madaus, Russell, & Higgins, 2009).

In the context of learning history, multiple choice only covers the skills of reading texts or documents. The student only requires factual memory to answer a multiple-choice test (Wineburg, 2004). Therefore, they will use unhistorical strategies to answer the questions (Smith, 2017). Furthermore, Reisman's study (2012) shows that multiple-choice tests are unable to evaluate the reliability of sources in building a record of the past with the student's background knowledge and different reading abilities. It means that a multiple-choice test could only measure simple or low-level thinking. In other words, students have not been invited to get involved in finding or discovering their own historical sources.

The developed instrument of portfolio assessment is designed to assess students' historical thinking skills, particularly in using and analysing historical primary sources by considering the framework of higher order thinking skills, including problem-solving skill, critical thinking, historical reasoning, the skill of using primary source evidence, and historical interpretation. Portfolio assessment can overcome the limitations of multiple-choice assessment because it encourages students to develop some of the abilities needed to become independent learners (Ersoy, 2006). Portfolio assessment engages them in complex activities such as discussion, analysis, collaboration with co-workers and demands the students to provide evidence-based answers, improves their critical-thinking and problem-solving skills, and creates the potential to gain student autonomy (Chabelli, 2002). Therefore, portfolio assessment can help teachers to get to know students' abilities such as critical thinking assessment (Micklo, 1997). It can be used to assess students' ability to use historical sources including the ability to find, select, interpret, and contextualize primary sources to build knowledge about the past.

Based on the above discussion, the developed instrument of portfolio assessment can be used to assess students' historical thinking skills particularly in using primary source evidence. The developed instrument can overcome the limitation of multiple-choice test by providing space for deep analysis of students' skill in using primary source evidence, which covers complex analysis of thinking, understanding, and reasoning of history. However, the developed instrument also has limitations related to the time consuming aspect in the assessment process. The limitation mainly locates on its ability to assess students' skills in a concise mode. The portfolio assessment requires time to assess students' skills, unlike a multiple-choice test. The portfolio assessment should be developed in further research by considering the needs of teachers and the time consuming aspect of its practicality.

Conclusion and recommendation

The developed portfolio assessment proved to be valid and reliable, thus, it can be used to measure the student's skills to organize information from historical documents, artefacts, or other written resources, and then reconstruct it as historical narration. It means the developed portfolio assessment can be posited as an alternative assessment tool for history teachers, who commonly use multiple-choice tests to assess their students' historical thinking skills of using primary source evidence. However, it also has a limitation related to its time consuming aspect, i.e., it will potentially take a long time to assess students' historical thinking skills. Therefore, the portfolio assessment should be developed by further research.

Acknowledgement

Tri Zahra Ningsih would like thank to the Indonesia Endowment Fund for Education (LPDP) for their support on his study.

References

- Adam, R., & Khoo (1996). *Quest: Interactive Item Analysis Program*. Melbourne: The Australian Council for Educational Research.
- Chabelli, M.M. (2002). Portfolio Assessment and Evaluation: Implication and Guidelines for Clinical Nursing Education. *Curationis* , 25 (3),4–9.
- Ercikan, K & Seixas, P. 2015. Issues in Designing Assessments of Historical Thinking. *Journal Theory Into Practice* , 54 (3), 255–262.
- Ersoy, F. (2006). Opinions of Teacher Candidates as to the Portfolio Assessment. *Elementary Education Online*, 5 (1), 85–89.
- Madaus, G., Russell, M., & Higgins, (2009). *The Paradoxes of High Stakes Testing*. Charlotte, NC: Information Age Publishing.
- Micklo, S.J. (1997). Math Portfolio in the Primary Grades. *Childhood Education*, 97 (Summer), 194–199.
- Reich, G.A. (2009). Testing Historical Knowledge: Standards, Multiple-Choice Questions and Student Reasoning. *Theory & Research in Social Education* , 37 (3), 325–360.
- Reisman, A. (2012). Reading like a Historian: A Document Based History Curriculum Intervention in Urban High Schools. *Cognition and Instruction* , 30 (1), 86–112.
- Romberg, T.A. (1993). Investigations Into Assessment in Mathematics Education. Dalam M. Niss, *Investigations Into Assessment in Mathematics Education: An ICMI Study* (hal. 97–111). Denmark: Springer Netherlands.
- Peter Seixas (1998) Student Teachers Thinking Historically, *Theory & Research in Social Education*, 26:3, 310–341, DOI: 10.1080/00933104.1998.10505854

- Seixas, P et al. (2015). A Design Process for Assessing Historical Thinking: The Case of a One- Hour Test. In P.S. Kadriye Ercikan, *New Directions in Assessing Historical Thinking* (pp. 102–116). New York: Routledge.
- Seixas, P. & Peck, C. (2004). Teaching Historical Thinking. in Sears, A & Wright I, *Challenge and Prospects for Canadian Social Studies* (pp 109–117). Vancouver: Pasific Educational Press.
- Seixas, P. (2006). Benchmarks of historical thinking: A framework for assessment in Canada. *The Centre for the Study of Historical Consciousness. Recuperado el, 12.*
- Smith, Mark D. (2017). New Multiple-Choice Measures of Historical Thinking: An Investigation of Cognitive Validity. *Journal Theory and Research in Social Education* , 46 (1), 1–34.
- Súsanna Margrét Gestsdóttir, C.V. (2018). Teaching Historical Thinking and Reasoning: Construction of an Observation Instrument. *British Educational Research Journal* , 44 (6), 960–981.
- Sutimin, Joebagio, Sariyatun, Abidin. (2018). The Development of Deconstructive of Learning History Model to Promote The Higher Order Thinking Skill of University Students. *The New Educational Review* , 51 (1), 19–29.
- Wineburg, S. (2004). Crazy for history. *Journal of American History*, 90(4), 1401–1414.
- Wright, B.D., and Masters, G.N. (1982). *Rating Scale Analysis: Rasch Measurement*. Chicago: MESA Press.

Appendix 1. The Essay Test

ESSAY TEST

1. Write your name and class on your answer sheet
 2. Time allocation in doing the test is 60 minutes
-
1. If you reconstruct a historical event, there will be some kinds of historical resources that can be used, one of those is the written resource. Determine a written resource that can be used to reconstruct a historical event.
 2. From the written source you choose, identify data that can be used to write a historical narration.
 3. Arrange a historical narration from identification result of data toward the written resource used.
 4. A historical event will definitely leave a mark as an evidence of a particular event that happened in the past. It should be in the form of an object resource called an artefact. Determine one artefact in your surrounding that can be used to reconstruct a historical event.
 5. From an object resource you choose, make data identification that can be used to explain a historical narration.
 6. Arrange a historical narration from the primary source you choose.

ESSAY TEST SCORING RUBRIC

NO.	ANSWER CRITERIA	SCORING
1.	(1) Document, (2) Book, (3) Essay, (4) Newspaper, (5) File, (6) Article, (7) Map, (8) Atlas, (9) Autobiography, (10) Biography	Score 1 if the resources chosen by the student is a true written resource. Score 0 if the resources chosen by the student is not a written resource.
2.	a. determine the writer of document b. when the written resources were made. c. where the written resources were published.	Score 0 to 4 depending on the number of items that can be answered by the student.
3.	a. Make one topic b. Compability between reconstructed resources and the topic made. c. Introduction section consists of background, research questions, and research objectives d. Content section e. Closing section includes conclusion and suggestion f. Consists of references consisting of written resources and other related resources	Score 0 to 10 depending on the number of items that can be answered by the student.
4.	a. Work equipment b. Household appliances c. Buildings d. Statue or sculpture e. Weapon f. Jewelry g. Currency	Score 1 if the source chosen is a true artifact resource. Score 0 if the source chosen is not an artifact resource.
5.	a. Identifying types of artefacts. b. Identifying the artefact site c. Determining the age of the artefact. d. Determining the creator of the artefact. e. Identifying the time of making the artefact.	Score 0 to 5 depending on the number of items that can be answered by the student.
6.	a. Proposing a topic b. Compability between the reconstructed resources and the theme made. c. Introduction section consists of background, research questions, and research objectives d. Content section e. Closing section consists of conclusion and suggestion h. References consist of written resources and other related resources	Score 0 to 10 depending on the number of items that can be answered by the student.

$$\text{Score} = \frac{\text{Total obtained score}}{\text{Total score}} \times 100$$

Appendix 2. The interview guide

INTERVIEW GUIDE

1. Decide on one theme related to the learning material.
2. Obtain information through an interview with perpetrators/witnesses/people who know about the actual historical events, documents, oral sources, and other relevant sources.
3. Collect data based on the 5W+1H formula.
4. Categorize data with historical sources concepts from the nature of historical sources, types of historical sources, and the link of historical sources to the selected theme.
5. Report on the interview in form of a written report which contains introduction, content, and conclusion.
6. Write down the bibliography, which consists of sources and other relevant references.

THE SCORING RUBRIC OF THE INTERVIEW

Give a tick (√) on the assessment scores.

No.	Interview results (aspects)	Score		
		1	2	3
1.	Selected theme relates to the learning material			
2.	Understanding of the selected theme based on the interview result.			
3.	The uses of primary sources.			
4.	The obtained information is in line with the criteria.			
5.	Collaboration and activeness among students			
6.	The report fulfils the requirements.			

Notes: 3 = fulfils 2 criteria 2 = fulfils 1 criterion 1 = Does not fulfil criteria

Aspects	Criteria
1	There is a time limit
2	Recount the information using their own sentences and conclude heterogeneous information.
3	Use of oral sources and documents.
4	The presented information is based on data from primary sources.
5.	Students' collaboration.
6.	The report is supported by literature.

$$\text{Score} = \frac{\text{Total obtained score}}{\text{Total score}} \times 100$$