Assessment Instrument Development for a Chronological Thinking Skills with the Rasch Model

Ofianto [1], Tri Zahra Ningsih[2]

Article Info	Abstract
	Chronological thinking skills are one of the most critical goals in learning history, distinguishing it from other sciences. This skill is not a natural skill that already exists in students but is a skill that needs to be developed in students, especially in studying history. This study aims to develop an instrument for assessing chronological thinking skills with the Rasch model. The development model adopted the Gall & Borg development model by adjusting to the research objectives and needs. The adaptation of the Borg & Gall model resulted in four stages in this study: (1) needs analysis and preliminary investigation, (2) planning and preparation of product development, (3) expert validation, and (4) instrument implementation. This research was conducted in Senior High School (SMA), involving 120 students from three schools. Sampling was carried out employing the proportional sampling technique. The data was collected using validation sheets, tests, and assessment sheets (scoring rubric). Data analysis was then performed with the Quest Program utilizing the Partial Credit Model (PCM). The results showed that the chronological thinking skills assessment instrument with the Rasch model was valid and reliable. The validity value of items that fit the model ranged from 0.77 to 1.30, and the instrument reliability value of 0.76 was in the high category. Therefore, it could be concluded that the assessment instrument developed can measure students' chronological thinking skills.
	Abstrak
	Berpikir kronologis merupakan salah satu tujuan terpenting dalam pembelajaran sejarah yang membedakannya dengan ilmu-ilmu lainnya. keterampilan ini, bukan merupakan keterampilan alami yang sudah ada pada diri siswa, tetapi merupakan keterampilan yang perlu untuk dikembangkan dalam diri siswa khususnya dalam mempelajari sejarah. Penilaian merupakan elemen penting untuk mengembangkan keterampilan berpikir kronologis. Instrumen penilaian yang tepat akan mempengaruhi perkembangan keterampilan secara kronologis. Penelitian ini bertujuan untuk mengembangkan instrumen penilaian keterampilan berpikir kronologis dengan model *Rasch*. Model pengembangan yang digunakan mengadaptasi dari model pengembangan Borg & Gall dengan menyesuaikan pada tujuan dan kebutuhan penelitian. Adaptasi model Borg & Gall menghasilkan 4 tahap dalam penelitian ini yaitu (1) analisis kebutuhan dan pengumpulan infromasi awal, (2) perencanaan dan penyusunan produk pengembangan, (3) validasi ahli, dan (4)

[1] Universitas Negeri Padang, Padang, Indonesia
Email: ofianto.anto@yahoo.com
2 SMP Negeri 46 Kerinci, Jambi, Indonesia
Email: trizahra10019@gmail.com

implementasi instrumen. Penelitian ini dilakukan di Sekolah Menengah Atas (SMA) dengan melibatkan 120 siswa dari 3 sekolah berbeda. Pengambilan sampel dilakukan dengan teknik *proportional sampling*. Pengumpulan data dilakukan dengan menggunakan lembar validasi, tes, dan lembar penilaian (rubrik penskoran). Analisis data dilakukan dengan Program Quest menggunakan *Partial Credit Model* (PCM). Hasil penelitian menunjukkan bahwa instrumen penilaian keterampilan berpikir kronologis dengan model rasch valid dan reliabel dengan nilai validitas item yang *fit* dengan model berkisar antara 0.77 hingga 1.30 dan nilai reliabilitas instrumen 0.76 yang berada pada kategori tinggi. Sehingga dapat disimpulkan bahwa instrumen penilaian yang dikembangkan dapat digunakan sebagai instrumen untuk mengukur keterampilan berpikir kronologis siswa.

INTRODUCTION

Chronological thinking is a characteristic of historical learning that distinguishes it from other sciences (Waring, 2010). It is thinking coherently, regularly, and continuously. Chronological thinking skills are among the most critical goals in learning history because the sequence of events is critical in understanding the past and the present. A series of events would be meaningless without a chronology because past events can only be correctly interpreted if placed on the proper timeline. Besides, chronological thinking is at the core of historical reasoning, is a central element in the education system in many countries, and is the basis for developing historical thinking skills (Stearns, 2010; Wineburg, 2010). It is more than just a concept of sequence of events in a time context; it is a periodization framework that refers to historical duration and succession patterns.

Moreover, chronological thinking skills are limited to knowledge of the time context. They include the skills to clearly understand the causes of change or even the causes of continuity and the effects of continuity or change (Williams, 2016). Therefore, it is vital to develop chronological thinking skills in students because it allows students to (1) distinguish past, present, and future; (2) identify historical narratives and construct historical narratives based on time and causal relationships; (3) measure, calculate, and interpret time; (4) identify and reconstruct patterns; (5) understand the various modes of periodization; (6) create and interpret data on a timeline; (7) describe change and continuity (Lorenc, 2013).

Chronological thinking skills are essential for activating historical thinking. They are a mental scaffolding (scaffolding tool) in organizing the mind not to mess up the time sequence. Without having chronological thinking skills, individuals will find it difficult to trace the relationships between events in a structured manner and explain their cause and effect. Chronological thinking skills are also necessary to construct historical narratives and give meaning to a series of events. However, these skills are not natural skills possessed by students but need to be trained and developed. In this case, the teacher must help develop these skills (Martin, 2015).

Lorenc (2013) reported that countries such as Finland, France, the Netherlands, Russia, the United States, and the United Kingdom assessed students' chronological thinking skills by providing written assignments and open-ended questions on various issues. Students were required to collect information from various sources to answer the questions given and presented in descriptive explanations. Such assessments have been shown to improve students' analytical skills and lead students to think chronologically. In Indonesia, especially West Sumatra, based on the interview and observation results by the author of high school history teachers in West Sumatra, an assessment of the history learning outcomes was generally carried out using a multiple-choice test. Teachers utilized multiple-choice tests to assess student learning outcomes based on the material studied. Nevertheless, this kind of assessment will not be able to develop the students' thinking skills. The test items are arranged based not on the demands of history learning objectives but on the subject matter studied. Thus, students' thinking skills are often neglected, including chronological thinking skills.

On the other hand, multiple-choice tests have various limitations, including discrimination against students' answers, whether the answers are chosen result from student knowledge or just guessing answers; hence, tests like this cannot describe the students' actual abilities (Samritin & Suryanto, 2016). Also, multiple-choice items produce students who cannot describe answers or arguments in solving problems (Norris & Ennis, 1989). Smith (2017) dan VanSledright (2015) show that multiple-choice items used to measure the achievement of historical learning outcomes tend to measure factual memory but are less able to measure historical thinking skills include measuring chronological thinking skills. These limitations indicate that multiple-choice tests are less appropriate as an instrument to measure chronological thinking skills.

Based on the above problems, this study aimed to develop an instrument for assessing chronological thinking skills through the Rasch model. The Rasch model was chosen in developing a chronological thinking skill assessment instrument because this model could provide accurate information about students' thinking skills. The information generated by the Rasch model could be used to monitor progress, diagnose students' strengths, weaknesses, and needs in history learning in general and chronological thinking skills in particular. The assessment instrument developed was essay questions that produced a polytomous response (more than two alternative answers, 1, 2, and 3) and was equipped with a scoring rubric to measure chronological thinking skills.

METHODS

This study adapted the Gall & Borg (1989) development research model. The Gall, Borg & Gall model adaptation was adjusted to the research objectives and interests, consisting of four stages: needs analysis and preliminary investigation, planning and preparation of development products, expert validation, and instrument implementation.

The needs analysis and preliminary investigation stage were carried out through direct observation/surveys and literature studies. This activity's results were used as the basis for compiling an initial draft of the chronological thinking skills assessment instrument. This study's planning and modeling stage was screening and formulating instruments based on all the information obtained in the first stage. At this stage, the development of chronological thinking skills assessment instrument was carried out. The expert validation stage was the instruments' validation that involved experts, including an assessment instrument expert, a material expert, and a linguist. The prototype feasibility was then measured utilizing the criteria used by Sutimin, Joebagio, Sariyatun, Hum, & Abidin (2018) as follows:

Table 1. Feasibility criteria for prototypes

Range	Criteria
3,26-4,00	Very feasible
2,51-3,25	Feasible
1,76-2,50	Quite feasible
1,00-1,75	Not feasible

In this study, the implementation stage was testing the instrument in small and large groups. The test results were then analyzed to determine how far the product developed met the criteria as expected. This research was conducted in Senior High School (SMA), involving 120 students from three schools. The selection of research subjects was based on school with the categories favorite, medium, and not favorite. Thus, the maximum variability was obtained from the measurement results. Sampling was done using the proportional sampling technique, while the data were collected employing validation sheets, tests, and assessment sheets (scoring rubric). Data analysis was then performed with the Quest Program utilizing the Partial Credit Model. The determination of fit items and test as a whole with the model in the Quest program (Adams, R.J & Khoo, 1996) was based on the magnitude of the INFIT Mean of Square (INFIT MNSQ) values, along with their standard deviation or the INFIT Mean of INFIT t value. An item or test is declared fit with the model if the

INFIT MNSQ range limits from 0.77 to 1.30 (Adams, R.J & Khoo, 1996). Meanwhile, the determination of instrument reliability was based on the internal consistency value. The difficulty level of the items had a range of ±2 (between -2 to +2), with the interpretation that the value of -2 indicates that the items were very easy, and +2 signifies that the items were very difficult (Swaminathan, Hambleton, & Rogers, 2006)(Swaminathan et al., 2006).

FINDINGS AND DISCUSSION

The Chronological Thinking Skills Framework

The theoretical framework of chronological thinking skills follows Lorenc's (2013) chronological thinking model. Lorenc (2013) stated that chronological thinking skills are indispensable in compiling historical narratives and giving meaning to a series of events. The ability to match events to historical times and periods by placing them in the order of events is a fundamental skill in history learning. Chronological thinking skills provide a framework or mind mapping that gives meaning and coherence to historical learning. Reaching this skill level requires a long developmental process involving elaborating, restructuring, and synthesizing student knowledge.

Through chronological thinking skills, students are expected to develop the ability to explain the time sequence of events, understand cycle patterns that occur repeatedly, and relate the temporal system to the concept of numbers to understand duration challenges. Gibson & Peck (2020) argued that without understanding the concept of time, there would be no fundamental understanding of change, development, continuity, and regeneration. Percival (2020)asserted that the ability to sort is a fundamental feature of historical understanding and that the past is the chaos of sequencing. The ability to sort events is one way students can understand the past correctly and determine the relationship between the past, present, and future. Chronological thinking skills enable students to (1) differentiate between past, present, and future; (2) identify historical narratives and construct historical narratives based on time and causal relationships; (3) measure, calculate, and interpret time; (4) identify and reconstruct patterns; (5) understand the various modes of periodization; (6) create and interpret data on a timeline; (7) describe change and continuity (Lorenc, 2013). An outline of the assessment instruments developed is presented in Table 2.

Table 2. Framework for the chronological thinking skills assessment

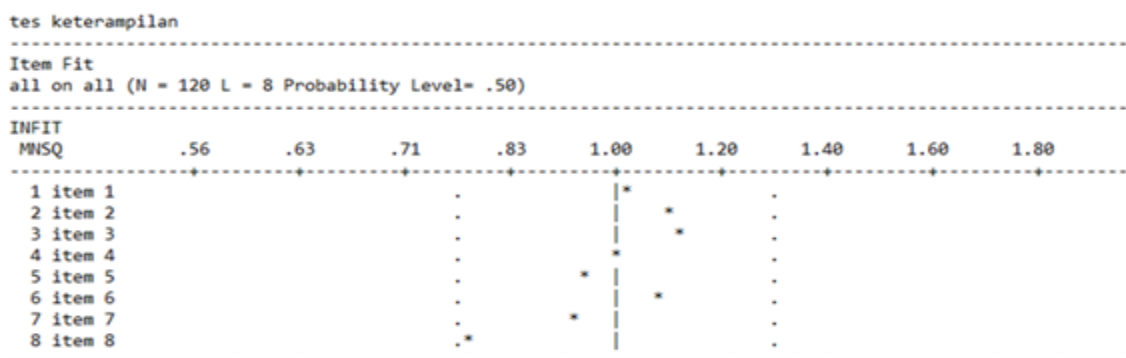| No	Chronological Thinking Skills Framework	Chronological Thinking Indicators
1.	Identify historical periodization models	• Clearly define historical periodization models • Distinguish each of the historical periodization models
2.	Construct temporal structures in constructing historical narratives	• Identify the temporal structure of a historical narrative or historical narrative • Compile historical narratives based on the identification of historical narrative structures
3.	Interpret data presented in a timeline	• Interpret the information presented in a timeline • Describe information obtained from a timeline in the form of historical narratives
4.	Compare alternative models for historical periodization	• Identify alternative models for historical periodization • Compare each alternative model of historical periodization

The result of the development process

The expert validation

The assessment instrument developed was then tested for expert validity. The expert validity test involved two material experts, two assessment experts, and two linguists. The instrument validity test results by experts are shown in Table 3.

Table 3. The instrument validity test results by expert teams

Indicators	Score	Interpretation
Material	3.67	Very feasible
Assessment instrument	3.83	Very feasible
Language	3.67	Very feasible
Mean	3.72	Very feasible

Based on Table 3, it is known that the chronological thinking skills assessment instrument developed was very feasible, with a mean value of 3.72 for the expert team validation. Furthermore, a test was carried out using the Quest program to see the instrument's validity and reliability level. The test results are displayed in the figure below.

Figure 1. The validity test results of the chronological thinking skills assessment instrument

Figure 1 above reveals that all items were fit with the model. The MNSQ INFIT value indicated it between 0.77 and 1.30. The values of validity for each item are exhibited in Table 4.

Table 4. The validity values of the chronological thinking skills items

No	Chronological Thinking Skills	Validity Value (INFIT Mean of Square)
1.	Clearly define historical periodization models	1.02
2.	Distinguish each of the historical periodization models	1.09
3.	Identify the characteristics of historical narratives/stories	1.12
4.	Identify the composition of historical narratives/stories	0.99
5.	Identify the information presented in a timeline	0.94
6.	Describe information obtained from a timeline in the form of historical narratives	0.108
7.	Identify alternative models for historical periodization	0.93
8.	Compare each alternative model of historical periodization	0.78

Furthermore, the reliability value of the chronological thinking skills assessment instrument is presented in Figure 2.

Item Analysis Results for Observed Responses
all on all (N = 120 L = 8 Probability Level= .50)
Mean test score   7.99
Standard deviation   2.16
Internal consistency   .76

Figure 2. The reliability value of the instrument

Based on figure 2 above, the reliability value is indicated by the internal consistency value of 0.76 in the high category. Based on this value, it is concluded that the chronological thinking skills assessment instrument developed can be used to measure students' skills.

The study findings indicated that the chronological thinking skills assessment instrument with the Rasch model was valid and reliable to measure students' chronological thinking skills. It was confirmed by the items' validity values that were fit with the model, ranging from 0.77 to 1.30 and the instrument's reliability value in the high category. This study supports Lorenc's (2013) research that a written test with an essay test could improve students' ability to analyze, compare, identify, and compile historical narratives. Thus, this instrument with the Rasch model can improve higher-order thinking skills. The assessment instrument with the Rasch model also guides students to think chronologically. Chronological thinking in history learning will provide a complete picture of events or historical travel from reviewing specific aspects. The relationships' benefits and meanings between events that occur can easily be drawn. Chronological thinking forms the basis for understanding history. If the basic understanding of history is good, it will help to understand history at a higher level of thinking.

The use of essay tests in the chronological thinking skills assessment instrument aimed to overcome the limitations of multiple-choice tests. Essay tests can be used to measure higher-order cognitive processes (Saeed, Hassan, Al Omary, & Alawad, 2018). The ability to describe, explain, analyze, synthesize, evaluate, and construct a historical narrative requires explanation and description. Thus, it is not easy to be measured using short-answer items and multiple-choice questions. Through essay tests, teachers also obtained more information to assess students 'skills and abilities than multiple-choice answers, which could not describe students' abilities comprehensively. However, essay tests also have drawbacks. It is not easy to assess a reliable test response, and essay test responses take a long time to obtain a test result score. Therefore, to overcome the essay test's shortcomings, the Rasch model was employed in developing this assessment instrument. The Rasch model provides an evident and easy-to-understand output score, and this model can identify the suitability of students' unexpected answers.

Rasch (Rasch Measurement Model) is believed to be a beneficial analytical tool to test the validity and reliability of instruments, test takers/person, and items at once. It makes this model unrivaled by other analytical tools (Wolins, Wright, & Masters, 1983). The Rasch model also fulfills the five principles of the measurement model (Rasch, 1977; Sumintono & Widhiarso, 2015). The five principles of measurement are providing a linear scale at the same interval, predicting missing data, providing a more precise estimate, detecting model inaccuracies, and measuring replicable measurements. Based on the above discussion, it could be concluded that the skills assessment instrument with the Rasch model was *appropriate* for use by teachers in measuring students' chronological thinking skills.

CONCLUSION

The research findings concluded that the chronological thinking skills assessment instrument with the Rasch model was appropriate for teachers to measure students' higher-order thinking skills in learning. The developed instruments provide innovations in history learning, especially for the assessment or evaluation aspects. The assessment instrument with the Rasch model is believed to provide accurate and comprehensive assessments and describe students' abilities appropriately. However, the instrument for assessing chronological thinking skills with the Rasch model also has several limitations. In its application, the assessment instrument with the Rasch model requires the ability to understand basic mathematics and a high understanding level of the software; the Rasch model requires a large sample size; the instrument developed is limited to measuring chronological thinking skills. To overcome the weaknesses in this study, further research is needed on the instrument for assessing chronological thinking skills and other similar studies.

REFERENCES

Adams, R.J & Khoo, S. . (1996). *Quest: Interactive item analysis program*. Melbourne: The Australian Council for Educational Research.

Gall, M. ., & Borg, W. . (1989). Educational Research an Introduction fourth edition. In *Longman Inc*.

Lorenc, J. et al. (2013). How is chronological thinking tested? *Edukacja*.

Martin, G. (2015). Historical Inquiry for Teaching Year 11 & 12 History Students. *Learning and Teaching*. https://doi.org/10.7459/lt/4.2.03

Norris, S. P., & Ennis, R. H. (1989). Evaluating critical thinking. The practitioners' guide to teaching thinking series. In *Critical Thinking Press and Software*.

Percival, J. (2020). *Understanding and Teaching Primary History*. Sage.

Rasch, G. (1977). On Specific Objectivity. An Attempt at Formalizing the Request for Generality and Validity of Scientific Statements in Symposium on Scientific Objectivity. *Danish Year-Book of Philosophy Kobenhavn*, *14*, 58–94.

Saeed, G. T., Hassan, A. E., Al Omary, H. L., & Alawad, Z. M. (2018). Multiple choice questions and essay questions in assessment of success rate in medical physiology. *Journal of the Faculty of Medicine Baghdad*. https://doi.org/10.32007/jfacmedbagdad.59481

Samritin, S., & Suryanto, S. (2016). Developing an assessment instrument of junior high school students' higher-order thinking skills in mathematics. *Research and Evaluation in Education*. https://doi.org/10.21831/reid.v2i1.8268

Smith, M. D. (2017). Cognitive Validity: Can Multiple-Choice Items Tap Historical Thinking Processes? *American Educational Research Journal*. https://doi.org/10.3102/0002831217717949

Stearns, P. N. (2010). World history: The basics. In *World History: The Basics*. https://doi.org/10.4324/9780203839942

Sumintono, B., & Widhiarso, W. (2015). Aplikasi Pemodelan Rasch Pada Assessment Pendidikan [Applications of Rasch Modeling in Educational Assessments]. In *Aplikasi Permodelan Rasch Pada Assesment Pendidikan*.

Sutimin, L. A., Joebagio, H., Sariyatun, Hum, M., & Abidin, N. F. (2018). The development of deconstructive learning history model to promote the higher-order thinking skill of university students. *New Educational Review*. https://doi.org/10.15804/tner.2018.51.1.01

Swaminathan, H., Hambleton, R. K., & Rogers, H. J. (2006). 21 Assessing the Fit of Item Response Theory Models. *Handbook of Statistics*. https://doi.org/10.1016/S0169-7161(06)26021-8

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

Waring, S. M. (2010). Escaping myopia: Teaching students about historical causality. *The History Teacher*, *43*(2), 283–288.

Williams, S. (2016). *Historical thinking and other unnatural acts: charting the future of teaching the past*. DORDT University.

Wineburg, S. (2010). Historical thinking and other unnatural acts. *Phi Delta Kappan*. https://doi.org/10.1177/003172171009200420

Wolins, L., Wright, B. D., & Masters, G. N. (1983). Rating Scale Analysis: Rasch Measurement. *Journal of the American Statistical Association*. https://doi.org/10.2307/2288670