The instruments development assessment of the affective competence: values contained in history

Syarifah1, Hidayat1, Deny Setiawan2, Lukitaningsih1 and Arfan Diansyah1

1Department of History Education, Universitas Negeri Medan, Medan, Indonesia
2Department of Pancasila and Civic Education, Universitas Negeri Medan, Medan, Indonesia
*syarifah@unimed.ac.id

Abstract. This study aims to develop valid and reliable instruments of the affective competency assessment (historical values). The instrument trial involved 935 students of high school grade XI IPS in Medan. Respondents were selected by random sampling and quota sampling. Data collected is analyzed using Structural Equation Modeling (SEM) and second order in Confirmatory Factor Analysis (CFA) with the standard loading factor (SLF) value criteria. From CFA analysis with the Maximum Likelihood method on the first order, it is known that all instrument items have a factor load of > 0.32 and the results of the model match test with a significance level of 0.0001. The results of reliability testing with Construct Reliability (CR) and Variance Extracted (VE) show that the instruments developed have met the acceptance limit of the reliability coefficient of $r_{xx} \geq 0.70$ for CR and $r_{xx}^2 \geq 0.50$ for VE. In conclusion, the instrument of affective competency assessment on historical values developed includes 5 dimensions, 14 indicators and 84 items including valid and reliable categories.

1. Introduction
Assessment is an integral part of the learning process and the achievement of learning objectives in accordance with the established curriculum. This is in line with the opinion of Ducan and Criss assessment in learning is intended to determine the level of mastery and competence of students towards the material being studied, obtain information on the potential and motivation of students, materials to diagnose student learning difficulties [1].

Assessment of learning history that takes place in schools has tended to be oriented only to the cognitive aspects of historical facts. Affective aspects related to the values contained in historical events are ignored and the assessment of the affective aspects tends to be done in writing so as not to give a factual picture of the affective dimension. This condition requires an improvement in the affective aspiration assessment system, so that the objectives and process of learning history are more meaningful. Efforts to improve the affective assessment system of history learning, require the development of instruments of affective assessment instruments in harmony. Departing from this thought, the aim of this study was to develop an instrument for evaluating valid and reliable affective competency learning outcomes that teachers can use to measure students' appreciation of the values contained in history. Instrument is a tool used to measure an object measuring Scriven, both natural phenomena and social phenomena, including measuring student achievement or other factors that have a relationship between...
student learning outcomes [2]. Valid instruments will determine the quality of education [3]. The steps for
developing instruments include: 1) theoretical study to strengthen the concept of variables and their
operational definitions, 2) compile dimensions and indicators, 3) make grilles and write instrument items,
4) theoretical validation for experts and panelists, 5) empirical trials, 6) validity analysis and empirical
data reliability, 7) perfecting instruments with their completeness so that they can be used as standard
instruments for affective competency assessment in history learning.[4]

The development of affective values encompasses historical aspects of the concept, that history
contains knowledge about past events, has a mission to give birth to educated generations who are soulful,
passionate and live the noble values of the nation[5][6][7][8][9][10]. So that the historical value becomes,
aspirations, views of life and references in acting and behaving or becoming the basic values and
operational value of the National Daily Council of Force ’45 and the basis of personal formation and
mental attitude of Meulen[11][12]. In the formulation of Bloom and Krathwohl the affective dimensions
of history learning include receiving, responding, valuing, organization, and characterization[13]

2. Methods
The focus of this study is the development of affective instruments and methods used in instrument
development research methods, beginning with theoretical and theoretical validation studies involving
experts and panelists to assess the suitability of dimensions with variables, suitability of indicators with
dimensions, suitability of items with indicators. Quantitative analysis uses a likert scale of 1 to 5 with the
formula:

a. $\sum_{i} n_i \mid i - r \mid$

2. $V = \frac{\sum_{i} (r + 1) to (r + t-1)}{N = \sum n_i}$
a. $N (t - 1)$ Description: $r = floor value, r = ceiling value = number of values i$

Aiken validity with valid criteria if the value of $rxx 2 0.2$ and Hoyt interrater reliability with reliability
criteria if above $r_{xx} 7 0.7$. Furthermore, instrument testing was carried out in two stages (102 items in the
first phase and 84 points in the second stage involving 935 respondents from the 11th grade social high
school students (515 respondents tested 1 and 420 respondents in stage 2).

The scale model used is a Likert scale with 4 (four) options, namely: strongly agree (SS), Agree (S),
disagree (TS) and strongly disagree (TS) with the reason the respondent avoided choosing the middle
option. Determination of validity in test 1 and test 2 uses Structural Equation Modeling (SEM) analysis
with first order and second order is continued on Confirmatory Factor Analysis (CFA) with Standard
Loading Factor (SLF) / load factor $\geq 0.32$ criteria with $T$-Value 96 1.96 or $\geq 2.00$. While reliability testing
in SEM used Construct Reliability (CR) and Variance Extracted (VE) through the following formula:

$$CR = \frac{(\sum \lambda^2) \cdot (\sum \text{standardized loading}^2)}{(\sum \lambda^2) + (\sum \text{error})}$$

Or

$$VR = \frac{(\sum i^2)}{(\sum i^2) + (\sum e)} = \frac{(\sum \text{standardized loading}^2)}{(\sum \text{standardized loading}^2) + (\sum \text{error})}$$

A construct has good reliability is if the CR value $\geq 0.70$ and the value of VE $\geq 0.50$ [14].
Furthermore, testing the suitability of the measurement model was carried out using various measures of
Goodness Of Fit (GOF). The development of instruments used by the guideline includes (1) How to Use;
(2) Standard Time and Place (3) Reporting Guidelines and (4) Interpretation of the results of students’ affective competency assessment with the criteria listed in table 1.

| Score Range          | Criteria         |
|----------------------|------------------|
| Mi + 1,5 SDi ≤ M ≤ Mi + 3,0 SDi | Very Good       |
| Mi + 0 SDi ≤ M ≤ Mi + 1,5 SDi     | Good             |
| Mi - 1,5 SDi ≤ M ≤ Mi + 0 SDi     | Enough           |
| Mi - 3 SDi ≤ M ≤ Mi - 1,5 SDi     | Less             |

The research method used is presented in Figure 1 below:

![Diagram of Research Method](image)

**Figure 1.** Design of Instrument Development Research

### 3. Result and Discussion

#### 3.1. Research result

The results of expert instrument validation and panelists were generated from 110 items 102 valid items and 8 invalid items, aiken values below 0.2. R11 reliability interrater test, using the Hoyt formula obtained reliability $r_{xx}$ 0.992, where the results of $r_{xx}$ were more than 0.7, all instrument items were reliable. Based on the first order test phase 1, 84 items meet the valid requirements, have a loading factor load above 0.32; 18 items dropped because it has a loading factor below 0.32. For the first order the second stage is all valid $\geq 0.32$. From the model suitability testing using goodness of fit (GOF), the values presented in Table 2 are subtracted:
Table 2. Recapitulation of Validity and Reliability of Affective Competency Dimensions

| Stage (s) | Dimensions | Confirmatory Factor Analysis(CFA) | Goodness of Fit Test |
|-----------|------------|----------------------------------|----------------------|
|           |            | Factor load, t-value, CR, VE     | Chi-Square, Df, p-value |
| Try out I | R          | 1.000, 28.239, 0.911, 0.837     | 1618.95, 73, 0.00000 |
|           | S          | 0.977, 24.782, 0.806, 0.677     | 1618.95, 73, 0.00000 |
|           | T          | 0.859, 21.833, 0.941, 0.841     | 1618.95, 73, 0.00000 |
|           | U          | 0.997, 19.204, 0.939, 0.796     | 1618.95, 73, 0.00000 |
|           | V          | 0.986, 28.589, 0.857, 0.676     | 1618.95, 73, 0.00000 |
| Try out II| R          | 1.000, 28.188, 0.975, 0.951     | 1758.00, 74, 0.00000 |
|           | S          | 0.903, 18.813, 0.825, 0.702     | 1758.00, 74, 0.00000 |
|           | T          | 0.821, 20.098, 0.962, 0.894     | 1758.00, 74, 0.00000 |
|           | U          | 1.000, 25.504, 0.979, 0.922     | 1758.00, 74, 0.00000 |
|           | V          | 1.000, 28.326, 0.969, 0.915     | 1758.00, 74, 0.00000 |

Based on factor analysis shows that the value of Standard Loading Factor (SLF) / factor load ≥ 0.32 with T-Value ≥ 1.96 or ≥ 2.00 for each dimension. This means that the dimensions of the test results in step 1 of the phase 2 trial are valid. To test the reliability of the empirical trial phase 1 and trial 2, reliable for all dimensions, the reliability coefficient index is above 0.70 for CR and VE is 0.5. Recapitulation of the validity and reliability of affective competency variables is presented in Table 3.

Table 3. Recapitulation of Validity and Reliability of Affective Competency Variables

| No | Try out | Competency | Validity (Factor Loading) | Reliability (CR, VE) | Information |
|----|---------|------------|---------------------------|----------------------|-------------|
| 1  | I       | Affective  | 0.869-1.000               | 0.986, 0.932         | Valid and reliable |
| 2  | II      | Affective  | 0.821-1.000               | 0.886, 0.650         | Valid and reliable |

The test of the suitability of the model, the results of the goodness of fit (GOF) measure from the output analysis of the measurement results with Lisrel on empirical data of test 1 and test 2 with fact analysis of non-order CFA are shown in Table 4.

Table 4. Test Results of Overall Match Model Try Out 1 and 2

| Goodness of Fit | Cut of value | Try out 1 result | Evaluations | Try out 2 result | Evaluations |
|-----------------|--------------|------------------|-------------|------------------|-------------|
| NFI             | ≥ 0.9        | 0.937            | Good        | 0.929            | Good        |
| NNFI            | ≥ 0.9        | 0.925            | Good        | 0.916            | Good        |
| CFI             | ≥ 0.9        | 0.940            | Good        | 0.931            | Good        |
|IFI              | ≥ 0.9        | 0.940            | Good        | 0.931            | Good        |
|RFI              | ≥ 0.9        | 0.922            | Good        | 0.912            | Good        |

This means that 5 (five) of 18 (eighteen) GOF sizes show good fit criteria, namely: Normed Fit Index (NFI), Non-Normed Fit Index (NNFI), Comparative Fit Index (CFI), Incremental Fit Index (IFI), and Relative Fit Index (RFI) for trial 1 and trial 2. The range and interpretation of the affective competency assessment qualitatively are presented in table 5 below:
Table 5. Qualitative Range and Interpretation of Competitive Competency Assessment.

| Score and Range Assessment | Interpretation of assessment results |
|---------------------------|--------------------------------------|
| 273 ≤ X ≤ 336             | Students have affective competence towards historical values |
| 210 ≤ X < 273             | Students have affective competence towards historical values |
| 147 ≤ X < 210             | Students have affective competence towards historical values |
| 84 ≤ X < 147              | Students have affective competencies toward historical values |

The assessment using a criterion referenced test (CRT) based on Gronlund and Linn obtained the highest score of 336 and the lowest value of 84, as listed in Table 6.[15]

Table 6. Conversion Scores and Range of Affective Assessments from Qualitative to Quantitative

| Score and Range Qualitative Assessment | Score and Range Quantitative Assessment |
|---------------------------------------|-----------------------------------------|
| 273 ≤ X ≤ 336                         | 81,25 ≤ X ≤ 100                         |
| 210 ≤ X < 273                         | 62,50 ≤ X < 81,25                       |
| 147 ≤ X < 210                         | 43,75 ≤ X < 62,50                       |
| 84 ≤ X < 147                          | 25 ≤ X < 43,75                          |

3.2 Discussion

Affective competency instruments about historical values developed after being validated by experts and panelists by using the value of Standard Loading Factor (SLF) / load value and reliability coefficients are declared valid and reliable. The results of the model suitability test used goodness of fit (GOF), test 1 and test 2 for SEM analysis results, 5 of 18 eighteen GOF sizes included the criteria for good fit, because some fit models were met: Normed Fit Index (NFI), Non-Normed Fit Index (NNFI), Comparative Fit Index (CFI), Incremental Fit Index (IFI), and Relative Fit Index (RFI). 13 GOF measures that have not been met, the criteria for good fit models are: Degree of Freedom, Chi-Square, Probability, Root Mean Square Error Approximation (RMSEA), Expected Cross Validation Index (ECVI), Akaike information Index (AIC), Consistent Akaike Information Index (CAIC), Parsimonious Normed Fit Index (PNFI), Root Mean Square Residual (RMR), Parsimonious Goodness of Fit Index (PGFI), Normed Chi-Square (NC), Estimated Non-centrality Parameters (NCP), Parsimonious Normed Fit Index (PNFI), Critical N (CN), Goodness of Fit Index (GFI), and Adjusted Goodness of Fit Index (AGFI).

This means that the research on the development of affective instruments that includes 5 dimensions, 14 indicators and 84 items to measure the affective competency study in history learning at the high school level is categorized as close to the requirements of the Goodness of Fit model, because there are 5 GOF criteria that are met. This result according to Hair et.al, instrument development research that produces 5 criteria of goodness of fit is considered to be sufficient to assess the feasibility of a valid and reliable instrument model [14].

4. Conclusions

The results of the development of students' affective competency assessment instruments in the history learning of high school level can be concluded that: the results of the construct validity testing of the instruments developed show all stages of testing, both the theoretical testing stage (construct) and the empirical 1 and 2 testing stages have met the criteria for validity and reliability significant. This can be seen from the results of theoretical validation analysis using Aiken validity, obtained validity values above 0.2 and inter-rater reliability using a reliable Hoyt coefficient Aikan above 0.7. For the stages of empirical testing phase 1 and 2 that are valid that meet the criteria have a factor load of > 0.32 and indicated by the achievement of the reliability coefficient in each test, Construct Reliability (CR) has met the established standards of $r_{\alpha'} \geq 0.70$ and Variance Extracted (VE) $r_{\alpha'} \geq 0.50$. Based on the results of the model suitability test using goodness of fit test, five criteria were produced which contained good value
because they had criteria above the cut value. So that the instrument of affective assessment developed covers 5 dimensions, 14 indicators and 84 items meet the criteria of valid and reliable instruments.

Awards and Acknowledgments
On this occasion the research team thanked the Head of Universitas Negeri Medan, Prof. Syawal Gultom, who has provided enormous support for the completion and smoothness of this research. Thanks are extended to family and colleagues who have provided critical input in the completion of this research.

References

[1] H. Ducan dan C. Bell, Evaluating and Assessing for Learning, New Jersey: Nichols Publishing Company, 1994.
[2] M. Scriven, Evaluation Thesaurus, California: Edgepress Inverness, 1981.
[3] P. M. Djaali dan R. Ramly, Pengukuran dalam Bidang Pendidikan, Jakarta: Pascasarjana Universitas Negeri Jakarta, 2000.
[4] S. Suryabrata, Pengembangan Alat Ukur Psikologi, Yogyakarta: Penerbit Andi, 2005.
[5] N. Notosusanto, Teori Sejarah Peranannya dalam Studi Sejarah, Majalah Ilmu-Ilmu Sastra Indonesia., Jakarta: Fakultas Sastra Universitas Indonesia, 1962.
[6] M. Ali, Pengantar Ilmu Sejarah Indonesia, Jakarta: Bhatara, 1961.
[7] S. Ghazalba, Pengantar Sejarah Sebagai Ilmu, Jakarta: Bhatara Karya Aksara, 1981.
[8] E. C. Hallet, What is History, New York: Alfred A. Knoph, 1965.
[9] T. Abdullah dan A. Sumihardjo, Ilmu Sejarah dan Historiografi Arah dan Perspektif, Jakarta: Gramedia, 1985.
[10] S. Kartodirjo, “Fungsi Sejarah dalam Pembangunan Nasional,” Historika, vol. 1, no. 1, 1989.
[11] W. V. Meulen, Pengajaran Sejarah Moderen, Basis XXI, Yogyakarta: Yayasan BP.Basis, 1972.
[12] L. L. Tyler dan B. W. Walsh, Test and Measurement, Englewood Cliffs N.J: Prentice Hall,Inc, 1979.
[13] B. S. Bloom, Taxonomy of Educational Objectives Book I. Cognitive domain, London: Longman Grup Ltd, 1979.
[14] J. F. Hair, F. A. R, L. T. R dan C. B. W, Multivariate Data Analysis with Readings, 4th Edition., Jersey: Prentice Hall Englewood Cliffs, 1995.
[15] N. E. Gronlund dan R. L. Linn, Measurement and Evaluation in Teaching. Sixth Edition., New York: MacMillan Publishing Company, 1990.
[16] Depdiknas, Pedoman Khusus Pengembangan Silabus dan Penilaian, Mata Pelajaran Sejarah, Jakarta: Dikmenum, 2006.
[17] Dewan Harian Angkatan 1945, Pedoman Umum Pelestarian Jiwa, semngat dan Nilai-Nilai 1945, Jakarta: Dewan Harian Nasional 1945, 1988.
[18] Sukardi, Metodologi Penelitian Pendidikan, Kompetensi dan Prakteknya, Jakarta: Bumi Aksara, 2004.