

DEVELOPMENT OF AUTHENTIC ASSESSMENT INSTRUMENTS FOR SAINTIFICAL LEARNING IN TOURISM VOCATIONAL HIGH SCHOOLS

Sugiyono¹, Badraningsih Lastariwati², Emy Budiastuti³, A Yudianto⁴

^{1,2,3}Department of Fashion Design and Hospitality Education, Universitas Negeri Yogyakarta, Indonesia

⁴Politecnico de Torino, Torino, Italy

Email: giyono@uny.ac.id

ABSTRACT

Teachers of Tourism Vocational High Schools in assessing students' performance generally use an invalid written test. It does not measure what it was designed to measure. Measurement of performance should use authentic assessments. The objectives of this study were (1) identifying aspects required to develop an instrument, (2) developing authentic assessment instruments on scientific learning of Tourism Vocational High Schools, and (3) developing authentic rubrics of scientific learning. This study was Research and Development applying the 4D model consisting of Define, Design, Developed, and Disseminate. The first year study completed stages of Defining and Designing, while the second year completed stages of Developing and Disseminating. The results of the first year study were the prototype of the instruments and the rubric script. This study generated 3 instrument prototypes, namely attitude instrument (8 items), knowledge instrument (9 items), and skills instrument (18 items), along with the rubrics for each instrument. The three instruments qualify face validity with expert judgment, content validity with Aiken formula, construct validity with goodness-of-fit test, and reliability estimate with Intraclass Correlation Coefficient.

Keywords: authentic assessment, instrument development, scientific learning

INTRODUCTION

Authentic assessments conducted in the field are still based on the perceptions of each teacher without using an authentic learning assessment instrument. The implementation of 2013 Curriculum dominantly use portfolio assessments, while teachers of tourism Vocational High Schools have not all mastered the guidelines of portfolio assessments, consequently the assessment results have not reached expected results.

Most teachers are not interested and do not want to use authentic assessments or performance-based assessments. In general teachers argue that doing authentic assessment is a waste of time and energy and too expensive. Moreover, authentic assessments need to be well designed. This opinion is certainly not true. Assessing performance with a written test is certainly not valid, as it does not measure what it is supposed to be measured. Performance needs to be assessed as the activity is in progress. If the performance appraisal is done to a number of students and not designed

first or done carelessly, of course the result cannot be accounted for because it is not consistent. Wiggins (2005) states that designing and executing performance appraisals are very efficient, because it is steady or consistent or reliable, inexpensive and wasted no time. Standards cannot be created without performing performance-based assessments.

In some cases there may be tasks that cannot be done in the classroom, so they should be done outside of school hours even outside the school. How to assess such learning? How can we assess such learning outcomes? People usually mention this type of learning as project-based learning (Wiggins, 2005). Thus, authentic assessment is also used to assess learning outcomes based on assignments or projects

According to the Regulation of Minister of Education and Culture of Indonesia (2014) on the techniques and instruments of the 2013 curriculum, the assessments use authentic assessments to assess students' learning progress of (1) attitude, (2) knowledge, and (3) skills. Skills competency assessments can be done using performance, projects, products,

portfolios, and written tests.

Performance assessment is done by observing the activities of learners in doing something, for example to assess the competence of the students in practices in the laboratory. Budiastuti et al. (2014) suggested that the assessment of practices can be done with self assessment involving the students as the assessor for themselves. Portfolio assessments basically assess the work of individual learners at a given period for a subject. At the end of a period, the work is collected and assessed by teachers and students. Ekawatiningsih (2008) showed that students, who follow the learning of restaurant courses with a portfolio assessment method, achieve better learning outcomes than those who follow the learning by a conventional evaluation method.

In abstract skills, the objectives of the assessment of learning outcomes by educators follow the scoring patterns of scientific learning outcomes in terms of learning ability: observing, asking, trying, associating, and communicating (Ministry of Education and Culture of Indonesia, 2014). Samani (1998) suggest that in science learning, learners do not receive the real benefits of what is learned, less functional teaching materials due to what people face in everyday life and the learning is too theoretical and elusive.

Zainul (2001) emphasized the need for performance appraisals to measure other aspects beyond cognitive, i.e. Howard Gardner's six basic skills that cannot be assessed only in the usual ways. The seven basic abilities are (1) visual-spatial, (2) bodily-kinesthetic, (3) musical-rhythmical, (4) interpersonal, (5) intrapersonal, (6) logical mathematical, and (7) verbal linguistic. Only two of the latter abilities are widely measured or assessed by people, while the other five skills have not been much revealed. From the description above it is clear that the assessment process or assessments, especially the performance appraisal becomes the main focus of the assessment.

Mueller (2006) states that an authentic assessment is a form of assessments whose students are required to present tasks in real situations demonstrating the application of essential knowledge and skills. A similar opinion is expressed by Stiggins (1987), who emphasizes specific skills and competencies, to apply skills and knowledge already mastered. Usually an authentic assessment involves a task for students to display and an assessment criterion or rubrics that will be used to assess performance based on the task.

Briefly scoring rubrics consists of several components, namely dimensions, definitions and examples, scales, and standards. Dimensions will serve as the basis for assessing students' performance. Definitions and examples are explanations of each dimension. The scale is set because it will be used to assess the dimension, while the standard is specified for each performance category.

This study was designed for two years. In the first year, the research aimed to develop an authentic assessment prototype on the scientific teaching of Tourism Vocational High School in the Province of Yogyakarta Special Region and the guidelines for the use of the prototype. In the second year, it continued developing the instrument prototype of the first year results into standard instruments, and refine the rubric as a guideline for evaluation using standard instruments. The objectives of the study are formulated in an integrated manner between the first year and the second year, namely (1) to identify the aspects required to develop an authentic assessment instrument on scientific learning, (2) to develop a set of authentic assessment instruments on scientific learning of Vocational High Schools, and (3) to develop assessment criteria (rubrics) of authentic assessments for scientific learning.

METHOD

This study is Research and Development (R & D). This approach was chosen based on

the objectives of the study to be achieved. The development model used in this study is the 4D model proposed by Thiagarajan that consisted of Define, Design, Develop, and Disseminate.

The stages were divided into two stages. The first stage and the second stage were implemented in the first year and the second year respectively. The first year study was to reveal the need assessment based on the 2013 curriculum of tourism Vocational High Schools. The design stage was to arrange a prototype of authentic assessment instruments on scientific learning based on need assessment. The second year of the study was to develop authentic assessment instruments by conducting trials on the students of tourism Vocational High Schools with a limited scale in the province of Yogyakarta Special Region, by using purposive sampling to test the validity and reliability of the instrument.

The data were obtained from the experts, participants and the students, each was described as follows (1) experts consisted of lecturers of practical subjects in the laboratory who had role as information giver in determining need assessment to design the instrument prototype, (2) the experts from teachers of practical subject in the laboratory had role as information giver in determining need assessment to design the instrument prototype, and (3) students of tourism Vocational High Schools play the role as the subject of the tryout of the developed instrument prototype to know construct validity and reliability of the instrument prototype.

Methods of data collection used a Delphi method, Focus Group Discussion (FGD), and questionnaires or instrument filling, each described as follows: (1) needs analysis was used to know the needs in the field related to the development of authentic assessment instruments on scientific learning in tourism Vocational High Schools, (2) focus Group Discussion (FGD) was conducted to obtain input and improvement of the instrument design, (4) workshop to evaluate the instrument by experts and practitioners, and (5) a limited-

scale trial to test the validity and reliability of the instrument prototype. Content validity test was done by analyzing data obtained from expert evaluation using Aiken formula, as follows.

$$V = \frac{\sum s}{n(c-1)} \quad (1)$$

Where

- V = coefficient of validity count
- s = r - lo
- r = value provided by the validator
- lo = lowest validation score is 1
- c = highest validation score is 4
- n = number of experts is 8

The validity of the construct was used to determine the suitability between the measurement result of the measuring instrument and the theoretical construct about the variables studied. Data obtained from the measurement of these indicators were analyzed quantitatively using Confirmatory Factor Analysis (CFA). The extraction was done using principal component analysis method. To make the results of factor analysis was more convincing, rotation process was conducted in this study using a Varimax method.

Estimates of construct reliability were used to determine the reliability of the instrument if the instrument was used at different times or by different assessors. Data obtained from indicator measurements were analyzed quantitatively using the Structural Equation Modeling (SEM) program. Statistical analysis program used is Intaclass Correlation Coefficient (ICC) program. The formula of the correlation coefficient of ICC as follows.

$$r_{xx} = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_o^2 + \sigma_e^2} \quad (2)$$

Where

- r_{xx} : ICC correlation coefficient
- σ_s^2 : subject variants
- σ_o^2 : observer variants (observer, rater)
- σ_e^2 : variants error

Intraclass correlation coefficients are the ratio between inter-group variants and total variants. The total variant comes from three sources, namely: subject, observer, and random error or residual error. The criteria for instrument stability or reliability were adequate stability categories with ICC greater than or equal to 0.50 and high stability categories with ICC greater than or equal to 0.80.

RESULTS AND DISCUSSION

The defining stage was done by collecting information to make the instrument design. Information was obtained through literature studies of the 2013 curriculum of tourism Vocational High Schools, and the regulation of Minister of Education and Culture of Indonesia no 104 of 2014, as well as relevant studies. Based on the needs analysis, at this defining stage, the result was an authentic assessment with a scientific approach.

Form of authentic assessments according to Regulation of Minister of Education and Culture of Indonesia No 104 year 2014 article 2, among others, include an assessment based on laboratory works and performance. While Article 5 mentions the scope of assessment of learning outcomes by educators includes attitude, knowledge, abstract skills and concrete skills. Abstract skills include observing, asking, trying, reasoning and communicating. Concrete skills include imitating, performing, deciphering, composing, modifying, and creating. the Regulation of Minister of Education and Culture of Indonesia No.104 year 2014 states that authentic assessment is a form of assessment that requires learners to display attitudes, using knowledge and skills gained from learning in performing tasks in real situations.

Based on the analysis of the needs mentioned above, it can be made an authentic assessment design instrument on science learning in tourism Vocational High Schools as follows: (1) the design of attitude instruments, (2) the design of knowledge instruments, and

(3) the design of skills instruments.

Instrument validity was used to find out whether the instruments of attitude, knowledge, and skills meet the criteria of good instruments based on the face validity. The result of the validity test was based on the opinion of experts and practitioners through expert judgment. The results show that it met the requirements of face validity in a good category and was ready to be used.

Content validity was calculated using Aiken's formula for data obtained from the evaluation of prototype instruments by experts, i.e. instruments prototypes of attitude, knowledge, and skills. The result of calculation of prototype instrument validation of attitude data by experts using V-Aiken formula is presented in Table1.

Table 1. V-Aiken Value of the Attitude Instrument Prototype

Item		V-Aiken	Criteria	Status
No.	Var.	V-cal.	5%	5%
1	X01	0.79	0.75	Valid
2	X02	0.83	0.75	Valid
3	X03	0.96	0.75	Valid
4	X04	0.83	0.75	Valid
5	X05	0.92	0.75	Valid
6	X06	0.88	0.75	Valid
7	X07	0.79	0.75	Valid
8	X08	0.83	0.75	Valid

The variables of student attention, attitude of tolerance, cooperative attitude, attitude of care, attitude of responsibility, discipline attitude, attitude of neatness, and attitude of honesty are defined as X01, X02, X03, X04, X05, X06, X07, and X08 respectively. Table 1 provides information that all indicators are valid based on the rating of experts and practitioners. Therefore the prototype of attitudes instruments in the opinion of experts and practitioners qualifies the validity of the contents and is prepared by the corresponding indicators of the theory. The calculation of validation data prototype of knowledge instruments by experts using the V-Aiken formula is presented in Table2.

Table 2. V-Aiken Value of the Knowledge Instrument Prototype

Item No.	Var.	V-Aiken V-cal.	Criteria 5%	Status 5%
1	X09	0.88	0.75	Valid
2	X10	0.83	0.75	Valid
3	X11	0.92	0.75	Valid
4	X12	0.79	0.75	Valid
5	X13	0.79	0.75	Valid
6	X14	0.79	0.75	Valid
7	X15	0.92	0.75	Valid
8	X16	0.88	0.75	Valid
9.	X17	0.79	0.75	Valid

The variables of material characteristics, material type, tool type, working step, change during processing, safety requirements, environmental sanitation requirements, personal hygiene, and order are defined as X09, X10, X11, X12, X13, X14, X15, X16, and X17 respectively.

Table 2 provides information that all indicators are valid based on the rating of experts and practitioners. Therefore the prototype of the knowledge instrument in the opinion of experts and practitioners qualifies the content validity and is prepared by the corresponding indicators of theory.

Table 3 is the calculation of validation data of skill instrument prototype by experts using V-Aiken formula. The variables of material requirements, choosing practice technique, determining work step, selecting work tool, patience student, concept of practice type, practice procedure, practice materials, equipment practice, teacher information, interpretation of practice results, developing interpretation, conclusion of practice result, report of practice result, multimedia, stages planning, practice, and results of practice are defined as X18, X19, X20, X21, X22, X23, X24, X25, X26, X27, X28, X29, X30, X31, X32, X33, X34, and X35 respectively.

Table 3 provides information that all indicators are valid based on the rating of experts and practitioners. Therefore the prototype of the skill instrument in the opinion of the experts and practitioners qualifies the content validity and is prepared by the appropriate indicators of the theory.

Table 3. V-Aiken Value of the Skills Instrument Prototype

Item No.	Var.	V-Aiken V-cal.	Criteria 5%	Status 5%
1	X18	0.79	0.75	Valid
2	X19	0.79	0.75	Valid
3	X20	0.79	0.75	Valid
4	X21	0.83	0.75	Valid
5	X22	0.96	0.75	Valid
6	X23	0.83	0.75	Valid
7	X24	0.92	0.75	Valid
8	X25	0.88	0.75	Valid
9.	X26	0.79	0.75	Valid
10.	X27	0.83	0.75	Valid
11.	X28	0.83	0.75	Valid
12.	X29	0.92	0.75	Valid
13.	X30	0.83	0.75	Valid
14.	X31	0.88	0.75	Valid
15.	X32	0.83	0.75	Valid
16.	X33	0.83	0.75	Valid
17.	X34	0.83	0.75	Valid
18.	X35	0.79	0.75	Valid

Constructs validity, its determination begins with factor analysis. The purpose of factor analysis is to classify the indicators into the appropriate factors using the Maximum Likelihood extraction method with Varimax rotation method. The determination of the indicator into the factor is based on the largest loading factor value among the loading factor values on the same item. The validity of the construct is established by Goodness-of-fit Test resulting from the extraction process in factor analysis. An experimental prototype test was conducted on 90 respondents from SMK N Godean, SMK N 4 Yogyakarta and SMK N Kalasan. The number of respondents from each school was 30 respondents.

The prototype of the attitude instrument consists of 8 items, grouped into the appropriate constructs using factor analysis. The KMO and Bartlett's Test of the Attitude Instrument prototype is presented in Table 4.

Table 4. KMO and Bartlett's Test of the Attitude Instrument Prototype

Kiser-Meyer-Olkin Measure of Sampling	of	.846
Bartlett's Test of Sphenricity	Approx. Chi Square	225.671
	df	28
	Sig.	.000

The KMO and Bartlett's Test table gives KMO-MSA values of 0.846 and Chi-Square values of 225.671 with the degree of freedom (df) of 28 and the significance level of 0.000. The value of KMO is 0.846 which is higher than 0.50. It means the instrument developed in either category. Chi-Square value of 225.671 and significance level of 0.0 mean that the correlation matrix is not an identity matrix so it can be used for factor analysis. The rotated factor matrix is presented in Table 5.

Table 5. Rotated Factor Matrix of the Attitude Instrument Prototype

	Factor	
	1	2
X01	.295	.514
X02	.168	.795
X03	.200	.502
X04	.607	.479
X05	.681	.347
X06	.616	.270
X07	.628	.142
X08	.637	.192

Extraction Method: Maximum Likelihood
 Rotation Method: Varimax with Kaiser Normalization
 a. Rotation converged in 3 iterations

The result of the rotated factor shows that the items are grouped into two factors, namely: factor 1 consists of items of X04, X05, X06, X07, and X08; while factor 2 consists of items of X01, X02, and X03. The result of the validity analysis of prototype construct of attitude instrument with Goodness-of-fit Test using Maximum Likelihood method is presented in Table 6.

Table 6. Goodness-of-fit Test of the Attitude Instrument Prototype

Chi-Square	df	Sig.
13.482	13	.411

The calculated Chi-Square value is 13.482 at degrees of freedom (df) of 13 with the significance level of p is 0.411. The value of the analysis results of p is 0.411 which is much larger than α of 0.05, meaning that there is no difference between the construct constructed from the theory with the constructs generated

from the analysis of empirical data. Based on the description, it can be concluded that the grouping of items into the factor or the construct is valid based on the validity of the construct. So the prototype of a valid attitude instrument is based on the validity of the construct with and in the Goodness-of-fit Test analysis.

The prototype of the knowledge instrument consists of 9 items, grouped into appropriate constructs using factor analysis. The KMO and Bartlett's Test of the knowledge instrument prototype is presented in Table 7.

Table 7. KMO and Bartlett's Test of the Knowledge Instrument Prototype

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.845
Bartlett's Test of Sphenricity	Approx. Chi Square	256.187
	df	36
	Sig.	.000

The KMO and Bartlett's Test table gives KMO-MSA values of 0.845 and Chi-Square values of 256.187 with degrees of freedom (df) of 36 and significance levels of 0.000. The value of KMO is 0.845 which is higher than 0.50 meaning that the instrument developed in either category. Chi-Square value of 256.187 and the significance level of 0.0 mean that the correlation matrix is not an identity matrix so it can be used for factor analysis.

Table 8. Rotated Factor Matrix of the Knowledge Instrument Prototype

	Factor	
	1	2
X09	.328	.498
X10	.176	.802
X11	.194	.500
X12	.635	.463
X13	.683	.330
X14	.623	.256
X15	.633	.131
X16	.612	.190
X17	.474	.218

Extraction Method: Maximum Likelihood
 Rotation Method: Varimax with Kaiser Normalization
 a. Rotation converged in 3 iterations

The result of the rotated factor presented in Table 8 shows that the items are grouped into two factors, namely: factor-1 consists of items X12, X13, X14, X15, X16, and X17; while factor-2 consists of items X09, X10, and X11. The result of the validity analysis of the prototype construct of knowledge instrument with Goodness-of-fit Test using Maximum Likelihood method is presented in Table 9.

Table 9. Goodness-of-fit Test of the Knowledge Instrument Prototype

Chi-Square	df	Sig.
22.582	19	.256

The calculated Chi-Square value is 22.582 on the degrees of freedom (df) of 19 with the significance level of Sig. $p = 0.256$. Value of analysis result: $p = 0.256$ is far bigger than α of 0.05, meaning there is no difference between construction constructed from theory with construct result from empirical data analysis. Based on the description can be concluded that the grouping of items into the factor or the construct is valid based on the validity of the construct. So the prototype of valid knowledge instrument is based on the construct validity with $\chi^2 = 22.582$ and $p = 0.256$ in the Goodness-of-fit Test analysis.

The prototype of skill instruments consists of 18 items, grouped into appropriate constructs using factor analysis. The KMO and Bartlett's Test of the prototype of the Skills Instrument is shown in Table 9.

Table 9. KMO and Bartlett's Test of the the Skills Instrument Prototype

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.927
Bartlett's Test of Sphericity	Approx. Chi Square	780.325
	df	153
	Sig.	.000

The KMO and Bartlett's Test table gives KMO-MSA values of 0.927 and Chi-Square values of 780.325 with degrees of freedom (df) of 153 and significance levels of 0.0. The value of KMO is 0.927 which is higher than 0.50 means the instrument developed in either

category. Chi-Square value of 780.325 and Sig significance level of 0.000 mean that the correlation matrix is not an identity matrix so it can be used for factor analysis. The rotated factor analysis is presented in Table 10.

Table 10. Rotated Factor Matrix of the Skills Instrument Prototype

	Factor		
	1	2	3
X18	.261	.238	.488
X19	.128	.420	.513
X20	.168	.129	.892
X21	.387	.157	.718
X22	.420	.155	.583
X23	.611	.339	.128
X24	.574	.204	.179
X25	.396	.376	.151
X26	.502	.335	-.012
X27	.737	.164	.201
X28	.523	.485	.307
X29	.627	.338	.209
X30	.543	.278	.371
X31	.554	.396	.068
X32	.554	.421	.117
X33	.320	.781	.134
X34	.419	.509	.121
X35	.289	.564	.255

Extraction Method: Maximum Likelihood
Rotation Method: Varimax with Kaiser Normalization

b. Rotation converged in 7 iterations

The result of the rotated factor shows that the items are grouped into three factors, namely factor 1 consists of items X23 up to X32; whereas factor 2 consists of items X33, X34, X35, and factor 3 consists of X18 up to X22. The result of validity analysis of skills instrument prototype construct with Goodness-of-fit Test using Maximum Likelihood method is presented in Table 11.

Table 11. Goodness-of-fit Test of the Skills Instrument Prototype

Chi-Square	df	Sig.
88.667	102	.824

The calculated Chi-Square value is 88.667 in degrees of freedom (df) of 102 with the level of significance of Sig. $p = 0.824$. The value of the analysis results: $p = 0.824$ is much larger than α of 0.05, meaning that there is no

difference between the construct constructed from the theory with the constructs generated from the analysis of empirical data. Based on the description, it can be concluded that the grouping of items into the factor or the construct is valid based on the validity of the construct. So the prototype of the skill instrument is valid based on the construct validity with χ^2 of 88.667 and p of 0.824 in the Goodness-of-fit Test analysis.

Reliability estimation is done by using intraclass correlation coefficients (ICC) with type of consistency definition test. ICC analysis is used to determine the stability of the instrument, based on instrument test by rater.

The attitude instrument prototype consists of 8 items, tested to 24 respondents. The results of the analysis and calculation are presented in Table 12 and Table 13.

Table 12. Reliability Statistics of the Attitude Instrument Prototype

Cronbach's Alpha	N of Items
.988	24

The result of reliability statistics analysis above shows that prototype attitude instrument has Cronbach Alpha reliability coefficient of 0.988 means that the instrument has good reliability.

Table 13. Intraclass Correlation Coefficient of the Attitude Instrument Prototype

	Intraclass Correlation ^a	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.777 ^b	.592	.936	84.643	7.0	161	.000
Average Measures	.988 ^c	.972	.997	84.643	7.0	161	.000

Two-way mixed effects model where people effects are random and measures effects are fixed

- a. Type C intraclass correlation using a consistency definition-the-between-measure variance is excluded from the denominator variance
- b. The estimator is the same, whether the interaction effect is present or not
- c. This estimate is computed assuming the interaction effect is absent because it is not estimable otherwise

The results of the above analysis provide information that the prototype of attitude instruments tested against 24 respondents resulted in a single measurement ICC coefficient of 0.777 and the average ICC average measurement coefficient of 0.988 which is higher than criterion coefficient minimum of 0.70 required ICC, meaning that the instrument prototype attitude has consistency and stability in category is adequate.

The prototype of the knowledge instrument consists of 9 items, tested to 24

respondents. The results of the analysis and calculation are presented in Table 14 and Table 15.

Table 14. Reliability Statistics

Cronbach's Alpha	N of Items
.988	24

The reliability analysis statistics above shows that the prototype of knowledge instrument has a Cronbach Alpha reliability coefficient of 0.988 meaning that the instrument has a good reliability.

Table 15. Intraclass Correlation Coefficient of the Knowledge Instrument Prototype

	Intraclass Correlation ^a	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.779 ^b	.605	.929	85.384	8.0	184	.000
Average Measures	.988 ^c	.974	.997	85.384	8.0	161	.000

Two-way mixed effects model where people effects are random and measures effects are fixed

- a. Type C intraclass correlation using a consistency definition-the-between-measure variance is excluded from the denominator variance
- b. The estimator is the same, whether the interaction effect is present or not
- c. This estimate is computed assuming the interaction effect is absent because it is not estimable otherwise

The results of the above analysis informed that the prototype of knowledge instruments tested on 24 respondents resulted in a single measurement ICC coefficient of 0.779 and the ICC average measurement coefficient of 0.988 which is higher than ICC coefficient criteria of at least 0.70 required, meaning that the prototype of knowledge instrument has consistency and stability in category is adequate.

The skills instrument prototype consists of 18 items, tested to 24 respondents. The

results of the analysis and calculation are presented in Table 16 and Table 17.

Table. Reliability Statistics of the Skills Instrument Prototype

Cronbach's Alpha	N of Items
.988	24

The result of reliability statistics analysis above shows that the prototype of skill instrument has a Cronbach Alpha reliability coefficient of 0.988 means that the instrument has good category reliability.

Table 15. Intraclass Correlation Coefficient of the Skills Instrument Prototype

	Intraclass Correlation ^a	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.779 ^b	.658	.889	85.384	17.0	391	.000
Average Measures	.988 ^c	.979	.995	85.384	17.0	391	.000

Two-way mixed effects model where people effects are random and measures effects are fixed

- Type C intraclass correlation using a consistency definition-the-between-measure variance is excluded from the denominator variance
- The estimator is the same, whether the interaction effect is present or not
- This estimate is computed assuming the interaction effect is absent because it is not estimable otherwise

The results of the above analysis informed that the prototype of skill instruments tested on 24 respondents resulted in a single measurement ICC coefficient of 0.779 and the ICC average measurement coefficient of 0.988 exceeded the required minimum 0.70 ICC coefficient criterion, meaning that the skill instrument prototype has consistency and stability in category is adequate.

CONCLUSION

The prototype of the attitude instrument consists of 8 items which qualified the face validity based on expert judgment, fulfilled the validity of the content according to the V-Aiken calculation, fulfilled the validity of the constructs based on goodness-of-fit test grouped into 2 factors, and had alpha reliability of 0.988 with ICC coefficient of 0.777 so the attitude instrument prototype is categorized as adequate. The prototype of knowledge instrument consists of 9 items which qualified

the face validity based on expert judgment, fulfilled the validity of contents according to V-Aiken calculation, fulfilled the validity of constructs based on goodness-of-fit test grouped into 2 factors, and had alpha reliability 0.988 with ICC coefficient of 0.779 so the knowledge instruments prototype is categorized as adequate. The prototype of skill instrument consists of 18 items which qualified the face validity based on expert judgment, fulfilled the validity of contents according to V-Aiken calculation, fulfilled construct validity based on goodness-of-fit test grouped into 3 factors, and had alpha reliability of 0.988 with ICC coefficient of 0.779 so the prototype of skills instruments is categorized as adequate.

REFERENCES

- Aiken, L. R. 1985. Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*. 45. SAGE Social Science Collections

- Borg, W.R. & Gall, M.D. 1983. *Educational Research an Introduction (4th ed.)*. New York: Longman Inc.
- Budiasuti, E., Karomah, P.& Fatmawati, D. 2014. Pengembangan Instrumen Self Assessment pada Praktik Menjahit Rok Berfuring. *Jurnal Pendidikan Teknologi dan Kejuruan*.22.1, 29-35
- Ministry of Education and Culture of Indonesia .2013. *Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 70 Tahun 2013, tentang Kerangka Dasar dan Struktur Kurikulum Sekolah Menengah Kejuruan/Madrasah Aliyah Kejuruan*. Jakarta: Berita Negara R.I. tahun 2013
- Ministry of Education and Culture of Indonesia. 2014. *Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 104 Tahun 2014, tentang Penilaian Hasil Belajar oleh Pendidik pada Pendidikan Dasar dan Pendidikan Menengah*. Jakarta: Berita Negara R.I. tahun 2014 Number 1507
- Samani, M. 1998. Model Pembelajaran Sains dan Teknologi untuk Kejar Paket. *Jurnal Pendidikan Teknologi dan Kejuruan*. 11. 3, 13-26
- Mueller, J. 2006. *Authentic Assessment*. Illinois: North Central College
- Ekawatiningsih, P. 2008. Penerapan Metode Portofolio (Protfolios Based Assesment) untuk Meningkatkan Kualitas Pembelajaran Mata Kuliah Restoran. *Jurnal Pendidikan Teknologi dan Kejuruan*. 17. 5, 257-278
- Stiggins, R. J. 1987. Design and Development of Performance Assessments. *Educational Measurement: Issues and Practice*. 6.3
- Wiggins, G. 2005. Grant Wiggins on Assessment. Edutopia. The George Lucas Educational Foundation (online)
- Wiggins, G. 2006. *Healthier Testing Made Easy: The Idea of Authentic Assessment*. Edutopia. The George Lucas Educational Foundation (online)
- Zainul, A. 2001. *Alternative Assessment. Applied Approach Mengajar di Perguruan Tinggi*. Jakarta: Pusat Antar Universitas untuk Peningkatan Dan Pengembangan Aktivitas Instruksional. Department of National Education of Indonesia