

CHARACTERISTICS OF ASSESSMENT INSTRUMENT
OF COMPETENCE TEST ABOUT FASHION SEWING
ON VOCATIONAL HIGH SCHOOL (SMK)
IN YOGYAKARTA SPECIAL TERRITORY

Emy Budiastuti
Lecturer of Cuisin and Fashion Education
Faculty of Engineering
Yogyakarta State University
emy.budiastuti@yahoo.com

Abstract

The purpose of this research was to identify validity and reliability of assessment instrument about competence test of Fashion Sewing on Vocational High School's student in Yogyakarta Special Territory.

This research was conducted by a Research and Development (R & D) design by using Borg and Gall's model. Procedure of this design were: preliminary study; developing of instrument by instrument expert validation and revision, by instrument tryout and revision; instrument field testing and final product revision. Samples as the information source were third-class students of Vocational High School (SMK) of Fashion Program in Yogyakarta Special territory. The sampling method was taken by purposive sampling. The rater of competence test was conducted by school and industry persons. Research-focused variable were development of test instrument, evaluation sheet, and evaluation guide. The validation of evaluation instrument was determined by expert judgement and reliability coefficient was calculated by using *Genova* dan *Cohen's Kappa* program.

Result of the research shows that: (1) validity of assessment instrument based on expert judgement about competence test of fashion program were classified as good category, (2) reliability coefficient of the instrument tested by using *Genova* program was 0.79 and *K (Kappa)* reliability coefficient was 0.77 more than the standard (0,70), so the reliability coefficient was significant. It was concluded that the assessment instrument can be used for competence test of Fashion Sewing on Vocational High School's student in Yogyakarta Special Territory.

Key words: assessment instrument, fashion sewing

Introduction

Fashion Design Program in vocational school is a part of secondary vocational school which aims at preparing the graduates to face work field. Education in vocational school must keep on being continued in order to the graduates have the competences needed in work field and they are ready to work.

Considering competence aspect so the substance of Vocational School Curriculum is selected and packaged with *competency based curriculum*. The learning with competence-based approach is especially aimed to make the learning materials really reflecting the necessity to achieve the required competences in work field. The competence achieved can be known through assessment. The assessment of fashion sewing competence can be done through performance or tasks forming competences. To find out the competence of each learner, it can be done by doing the test called competence test. The assessment in competence test on fashion sewing is in the form of performance of the test's participants to demonstrate knowledge and skills according to the expected criteria and can be applied actually.

The result of observation shows that just a few of teachers developing assessment instrument, however it is still limited and the validity and reliability has not been tested in the field. That is supported by data stating that 50% of state vocational school teachers and 25 % of private vocational school teachers developed and applied assessment instrument in competence test in fashion sewing.

The purposes of this research are: (1) to find out the instrument validity of competence test in sewing women coat in Vocational School in DIY, (2) to find out the instrument's reliability of competence test assessment in sewing women coat in Vocational School in Yogyakarta Special Region.

Competence-based assessment is the way a teacher evaluating the student's performance, which aims at placing and planning of professional development (Yorkovich, 2008:1). Djemari Mardapi (2004:7) states that the competence-based assessment principle are: (1) describing the characteristics of someone's behavior effectively; (2) the assessment which does not depend on others but based on someone's behavior; (3) helping

to gain the most effective way to think about behavior; (4) the assessment system used based on the principles of accuracy, economic, and learning quality improvement.

Competence-based assessment assesses someone's capability or success based on criteria, not compares based someone's capability with others in the classroom (Yoyoh Jubaedah, 2007:9). According to Moore (2002: 314-316), competency refers to aspect of working behavior supporting a job. Meanwhile based on KBBI (2005), competency is capability and ability which are sufficient in something or in a job.

Depdiknas (Ministry of National Education, 2007:III-1) states that the assessment of competency mastering in skill or psychomotoric aspect of students in Vocational School is conducted by *performance assessment*. *Performance assessment* is various tasks and situations where the test participants are asked to demonstrate their understanding and to apply their deep knowledge as well as skills in various contexts according to the expected criteria). According to Berk (1986: i), performance assessment is a process of collecting data by conducting observation systematically to make a decision about someone.

The important characteristics in education measuring are validity and reliability. Djemari Mardapi (2004:14) proposes that the validity of a measuring instrument can be assessed by construct validity of the instrument. Construct validity will show that the measurement is suitable with the plan. Through the grid of the measurement of the instrument, it will be found the validity of a measurement instrument. The grid contains of the tested materials, the form of the problems, hierarchial thinking level, the quality of the problems, and scoring procedure. In assessment using inter-rater reliability, it obliges the tester to assess or measure some aspects of student's behavior. To achieve high inter-rater reliability, it needs to design and apply rubric. Rubric is arranged to avoid tester's subjectivity and to obtain inter-rater reliability level (Bresciani, 2009: 2-3). Johnson (2009:22) states that:

The consistency of examinees' scores across such facets as occasions, tasks, and raters. In other words, reliability addresses whether an examinee's score would be the same if she were to take the exam on a different occasion, complete different tasks, or be scored by different raters.

Reliability related to the consistency inter-raters in getting the same scores in an assessment. Smith (2007:2) states that in a performance test, *inter-rater reliability* is used to make assessment system meaningful and consistent. The importance of training for tester is to maintain the appropriate level for objective and valid assessment to assess competency (Lichtenberg, 2007: 477). In performance test, it is necessary to develop rubric which is used as measurement basis. Rubric makes the response scale and the difference among levels is similar in item set. Assessment rubric design needs some specifications from the criteria to assess performance quality and assessment procedure choices (Brenan, 2006:394).

In order to give valid and reliable information and to avoid a lot of mistakes in measuring, the assessment is done by more than one person as a team, each of them assesses the same matter. The results of assessment from each tester or *rater* are compared to find out their consistency.

Sewing clothing is done by uniting parts of cloth which have been cut based on certain pattern and design. Sewing is a very important process in making clothing. Sewing technique used must be suitable with the design and materials because if the technique is not appropriate, the result does not have good quality (Ernawati, 2008: 358).

Research Method

The type of this research is Research and Development. The selected model of development is Borg and Gall's model, with the following procedures: preliminary study; materials development, expert judgment and instrument revising; try out and instrument revising; field try out and final product of the instrument. The sample as the source of information is the vocational school's students in the third grade of Clothing Design Program in Yogyakarta Special Region which is ten students. Sampling technique used is *Purposive Sampling*. *Raters* of competence test were from school and industry. Variables which become the focus of this research are problem development, assessment sheet, and rubric (assessment guideline). The validation of the instrument was done through expert

judgment and reliability coefficient was calculated using two approaches, those are Genova program and Cohen's Kappa.

The reliability of the instrument was analyzed by using *Genova* (Generalizability of Variance), which consists of G study and D study. G study was used to find out reliability index of the instrument of assessment and to estimate the components of error variance caused by various sources of variance; those are source of student's variance or Person (P), Rater (T), and Item (I). Furthermore, the analysis of D study was used to find out to what extent the assessment's instrument is accepted by wider facet, especially in the matter of the consistency of measurement of the rater (Tri Hartiti Retnowati, 2009:184).

Research Result and Discussion

Based on the judgments of the expert in clothing field, the developed test in competency test of sewing women's coat has been suitable with the materials in curriculum of clothing design. The developed items have been appropriate with the profile of capability in clothing design study program, especially on sewing women's coat. The criteria of the assessment which was arranged has reflected the mastery of student's competency so that the validity of the developed instrument of competency test assessment is categorized as good.

1. Analysis Result of *Genova* for the Estimation of Variance Components

Analysis Result of *G Study* covers the estimation of variance on the aspect of preparation, process, result, attitude, and time allocation. Based on the result in table 1, the rater is a source of measuring mistake variance which its variance is small that is by 0.004 or 1.19 %, where $0.001 \leq \sigma^2 T \leq 0.242$. (Brenan, 1983:104). Therefore, the role of the rater (T) shows high consistency. It means that the raters have good level of understanding and agreement.

Table 1. Estimation of Component Variance of Student, Rater, and Item on the Assessment of Sewing Preparation

Sources of Variance	SS ₁	SS ₂	df	MS	Variance	% Total Variance
Student (P)	744.00000	9.00000	9	1.00000	0.158	47.02
Rater (T)	735.10000	0.10000	2	0.05000	0.004	1.19
I:T	735.40000	0.30000	3	0.10000	0.000	0.00
PT (Interaction between Student and Rater)	745.00000	0.90000	18	0.05000	0.000	0.00
PI:T (Interaction between Student and Item <i>Nested</i> on the Rater)	750.00000	4.70000	27	0.17407	0.174	51.79
Total	3709.50000	15.00000	59	1.37407	0.336	100.00

Table 2. Estimation of Variance Components of Student, Rater, Item On the Assessment of Sewing Process

Sources of Variance	SS ₁	SS ₁	df	MS	Variance	% Total Variance
Student (P)	3587.05556	40.11111	9	4.45679	0.12052	17.09
Rater (T)	3553.43333	6.48889	2	3.24444	0.00166	0.23
I:T	3659.60000	106.16667	33	3.21717	0.29270	41.57
PT (Interaction between Student and Rater)	3595.66667	2.12222	18	0.11790	0.00000	0.000
PI:T (Interaction between student and Item nested on the Rater)	3788.00000	86.16667	297	0.29012	0.29012	41.151
Total	18183.75556	241.05556	359	11.32642	0.70500	100.0 0

Rater is a source of measurement mistake variance which its variance is small, that is by 0.00166 or 0.23%. Therefore, the role of the raters (T) shows high consistency, meaning that the raters have good understanding level.

Based on the results in table 3, the rater is a source of measuring mistake variance which its variance is small, that is by 0.004 or 1.19 %, where $0.001 \leq \sigma^2 T \leq 0.242$. (Brenan, 1983: 104). Therefore, the role of the rater (T) shows the high consistency, meaning that the raters have good level of understanding and agreement.

Table 3. Estimation of Variance Components of Student, Rater, and Item
On the Assessment on Sewing Result

Sources of Variance	SS ₁	SS ₂	df	MS	Variance	% Total Variance
Students (P)	3202.46667	27.21333	9	3.02370	0.097	22.61
Tester (T)	3175.34000	0.08667	2	0.04333	0.000	0.00
I:T	3198.60000	19.61000	27	0.86148	0.059	13.75
PT (Interaction between Students and Tester)	3204.40000	1.84667	18	0.10259	0.000	0.00
PI:T (Interaction between Students and Item Nested on the Tester)	3294.00000	66.34000	243	0.27300	0.273	63.64
Total	15491.41667	114.19667	299	4.18127	0.724	100.00

Table 4. Estimation of the Variance Components of Student, Rater, and Item
in the Assessment of Attitude and Sewing Time Allocation

Sources of Variance	SS ₁	SS ₂	df	MS	Variance	% Total Variance
Students (P)	1424.25000	16.57500	9	1.84167	0.144	36.18
Tester (T)	1407.72500	0.05000	2	0.02500	0.000	0.00
I:T	1409.90000	2.17500	9	0.24167	0.000	0.00
PT (Interaction between Students and Tester)	1426.25000	1.95000	18	0.10833	0.000	0.00
PI:T (Students' interaction and Item Nested on the Tester)	1449.00000	20.57500	81	0.25401	0.254	63.82
Total	7117.12500	41.32500	119	2.47068	0.398	100.00

The role of the raters in assessing attitude and sewing time allocation is very important, so if the variance of variance source of rater (T) is very small then the level of consistency inter-raters is high. Therefore, the instrument of assessment of the aspects of attitude and time allocation of sewing women's coat can be used in wider facet.

Table 5. Summary of Coefficient Result of *G Study*

Components/Aspects	Item Number	KKM	Coefficient	Decision
Sewing Preparation	2	0.70	0.77	> KKM
Sewing Process	12	0.70	0.84	> KKM
Sewing Result	10	0.70	0.80	> KKM
Attitude and Sewing Time Allocation	4	0.70	0.76	> KKM
Average			0.79	

2. Analysis Results of *D Study*

Analysis Results of *D Study* covers the aspects of sewing preparation, sewing process, sewing results, attitude, and sewing time allocation.

a. Result of *D Study* for Preparation Assessment

Table 6 shows that the components of assessment of preparation which are used by the rater with one indicator only (01-01 with P = 10, T = 3 and I = 1) have consistency level of 0.70. It means that the raters' level of understanding and agreement toward the use of assessment instrument construct is by 70%, which has met the required criteria of at least 0.70. If the rater in doing the assessment used two indicators at once so the level of agreement and understanding is by 82.6 %. The level of understanding and agreement will achieve the maximum result if all indicators, those are number 1 and 2, are used together so that it can meet the observation level which is accepted at wider facet.

Table 6. Estimation of Generalizability Coefficient and Change Level in Preparation Assessment

<i>D Study</i> Design	Sample Size			Generalizability		The Difference of Genova Coefficient
	P	T	I	Coef.	Phi	
01 – 01	10	3	1	0.70347	0.70347	0.122
01 – 02	10	3	2	0.82593	0.82593	

b. Result of *D Study* for Process Assessment

The analysis results of *D study* for the assessment of sewing process of women's coat are presented in table 7. Information obtained from Table 7 shows that in assessing sewing process, if the rater only used one indicator (03 – 03 with P = 10, T = 3, and I = 12),

the consistency level is 0.54. It shows that understanding and agreement level of the rater toward the use of instrument construct is 54%. If the rater in assessing sewing process used two indicators (03-04 with $P = 10$, $T = 3$, dan $I = 12$), the consistency level is 0.70. It means that the level of understanding and agreement of the raters toward the use of construct in the used assessment instrument is 70% and so on.

Table 7. Estimation of Generalizability Koefficient and Change Level in the Assessment of Sewing Process

D Study Design	Sample Size			Generalizability		The Difference of Genova Coefficient
	P	T	I	Coef.	Phi	
03 – 03	10	3	3	0.54479	0.37325	1.61
03 – 04	10	3	4	0.70533	0.54350	
03 – 05	10	3	5	0.78216	0.64096	0.77
03 – 06	10	3	6	0.82721	0.70408	0.45
03 – 07	10	3	7	0.85682	0.74830	0.29
03 – 08	10	3	8	0.87776	0.78100	0.21
03 – 09	10	3	9	0.89336	0.80616	0.16
03 – 10	10	3	10	0.90543	0.82612	0.12
03 – 11	10	3	11	0.91505	0.84235	0.09
03 – 12	10	3	12	0.92289	0.85579	0.08
03 – 13	10	3	13	0.92940	0.86711	0.06
03 – 14	10	3	14	0,93490	0,87678	0,005

The instrument that has met the required criteria is at least 0.70. Thereby, to obtain high level of understanding and agreement (high coefficient of G), it is better to use 12 indicators so that it can meet the level of observation which can be accepted in wider facet.

c. Analysis Result of D Study for Result Assessment

Analysis result of D study for result assessment of sewing clothing is presented in table 8. Table 8 shows that in assessing sewing result, if the rater only used one indicator (15 – 15 with $P = 10$, $T = 3$, and $I = 12$), it obtained consistency coefficient of 0.45. it means that the level of understanding and agreement of the raters toward the use of construct in the used assessment instrument is 45%.

Table 8. Estimation of Generalizability Coefficient and Change Level in the Assessment of Sewing Result

D Study Design	Sample Size			Generalizability		The Difference of Genova Coefficient
	P	T	I	Coef.	Phi	
15 – 15	10	3	15	0.50188	0.45322	0.17
15 – 16	10	3	16	0.66834	0.62375	
15 – 17	10	3	17	0.75141	0.71319	0.08
15 – 18	10	3	18	0.80120	0.76828	0.05
15 – 19	10	3	19	0.83438	0.80562	0.03
15 – 20	10	3	20	0.85806	0.83259	0.03
15 – 21	10	3	21	0.87582	0.85299	0.02
15 – 22	10	3	22	0.88963	0.86896	0.01
15 – 23	10	3	23	0.90068	0.88180	0.01
15 – 24	10	3	24	0.90971	0.89235	0.01

If the rater in making assessment of sewing result used two indicators (15 - 16 with $P = 10$, $T = 3$, and $I = 12$), it was obtained the level of consistency of 0.67. It means that the level of understanding and agreement of the raters toward the use of construct in the used assessment instrument is by 67%. Therefore, to get the level of understanding and agreement which meet the required criteria of minimum 0.70, it is better to use indicator item of 1, 2, and 3. However, if the higher level of consistency which is expected, then it is better to use all indicators to meet the observation level which can be accepted in wider facet.

d. Result of D Study for the Assessment of Attitude and Time Allocation

Table 9 above shows that in making assessment of attitude and time allocation, if the rater only used one indicator (25 – 25 with $P = 10$, $T = 3$, and $I = 12$), it obtained the consistency level of 0.60. G Coefficient of 0.60 shows that the level of understanding and agreement of the raters toward the use of construct in the used assessment instrument is 60%. If the raters in making assessment of attitude and time allocation used two indicators (25 - 26 with $P = 10$, $T = 3$, and $I = 12$), the level of consistency is 0.76.

Table 9. Estimation of Generalizability Coefficient and Change Level on the Assessment of Attitude and Time Allocation

<i>D Study Design</i>	Sample Size			Generalizability		The Difference of Genova Coefficient
	P	T	I	Coef.	Phi	
25 – 25	10	3	25	0.60977	0.60977	0.16
25 – 26	10	3	26	0.75758	0.75758	
25 – 27	10	3	27	0.82418	0.82418	0.06
25 – 28	10	3	20	0.86207	0.86207	0.04

It shows that the level of understanding and agreement of the raters toward the use of construct in the used assessment instrument is by 76%, meaning that it meets the criteria of at least 0.70. Therefore, to get the high level of understanding and agreement of the raters in assessing attitude and time allocation, it is better to use D study design simultaneously so that it can fulfill the observation level which is accepted in wider facet.

3. Analisis of Inter-Rater *Kappa* Coefficient

Competency test in sewing women's coat is assessed by three raters. To find out the level of stability or consistency inter-three raters, it was analyzed by inter-rater coefficient reliability. Analysis technique of inter-rater coefficient reliability used κ (Kappa) analysis technique.

a. Inter-rater Reliability Coefficient of Preparation Assessment

The result of inter-rater coefficient reliability of preparation assessment is shown in table 10. Based on the result of calculation of inter-rater κ (Kappa) coefficient in table 10, there are some differences of index, especially for item 2, but the differences are not significant. The average of κ coefficient in preparation aspect is by 0.77. It is obtained from the average of inter-rater for two items. The reliability index obtained has met the requirement as the instrument to measure preparation aspect in sewing coat in wider facet.

Table 10. Results of κ (Kappa) Coefficient in the Aspect of Sewing Preparation Inter-Three Raters

Item Number	Rater			Average
	V vs K	V vs R	K vs R	
1.	0.80	0.80	0.80	0.80
2.	0.80	0.80	0.62	0.74
The Average of κ Coefficient				0.77

b. Inter-rater Reliability Coefficient in Process Assessment

The result of inter-rater reliability coefficient in the assessment of sewing process is shown in table 11. Based on the result of estimation, κ reliability coefficient inter-three raters in the aspect of sewing process is by 0.76. It means that the instrument of assessment in the aspect of sewing process can be used in wider facet.

Table 11. Result of κ (Kappa) Coefficient in the Aspect of Sewing Process Inter-Three Raters

Item number	Rater			Average
	V vs K	V vs R	K vs R	
3.	0.60	0.80	0.78	0.73
4.	0.84	0.80	0.84	0.83
5.	0.69	0.70	0.68	0.69
6.	0.85	1.00	0.85	0.90
7.	0.80	0.70	0.80	0.77
8.	0.68	0.59	0.85	0.71
9.	0.83	0.83	0.64	0.75
10.	0.80	0.60	0.80	0.73
11.	0.84	0.69	0.83	0.79
12.	0.82	0.82	0.63	0.76
13.	0.60	0.80	0.80	0.73
14.	0.80	0.60	0.80	0.73
The average of κ coefficient				0.76

Based on the table above, there are some differences of inter-rater reliability index, however, the differences are not significant.

c. *Inter-rater* Coefficient Reliability of Result Assessment

Inter rater penilaian coefficient reliability of sewing result is shown in table 12. Based on the estimation of κ reliability coefficient inter-three raters for clothing sewing

result aspect which is shown in table 11 is 0.78. It means that the instrument of assessment of clothing sewing result aspect can be used in wider facet.

Table 2. Result of κ (Kappa) Coefficient in the Aspect of Sewing Result Inter-Three Raters

Item Number	Rater			Average
	V vs K	V vs R	K vs R	
15.	0.80	0.63	0.80	0.74
16.	0.83	0.80	0.60	0.76
17.	0.84	0.68	0.84	0.79
18.	0.80	0.80	1.00	0.87
19.	0.83	0.83	0.83	0.83
20.	0.80	0.80	0.62	0.74
21.	0.60	0.60	0.78	0.66
22.	0.80	0.80	0.80	0.80
23.	0.80	0.80	0.80	0.80
24.	0.80	0.80	0.80	0.80
The Average of κ Coefficient				0.78

d. Inter-rater Reliability Coefficient in the Assessment of Attitude and Time Allocation

Based on the result shown in table 13, the reliability coefficient of κ inter-three raters for the aspect of working attitude and time allocation is 0.78. It means that the instrument of assessment of attitude and time allocation in sewing clothing can be used in wider facet. Based on the result of estimation shown in table 13, κ reliability coefficient inter-three raters for the aspect of working attitude and time allocation is by 0.78. The reliability index is 0.78 meaning that the assessment instrument in the aspect of attitude and time allocation in sewing clothing can be used in wider facet.

Table 13. Result of κ (Kappa) Coefficient inter-Three Raters in the Aspects of Attitude and Time Allocation in Sewing

Item Number	Rater			Average
	V vs K	V vs R	K vs R	
25.	0.60	0.78	0.80	0.73
26.	0.68	0.83	0.83	0.78
27.	0.82	0.83	0.83	0.83
28.	0.83	0.83	0.65	0.77
The Average of κ Coefficient				0.78

Table. 14. Summary of Reliability Coefficient of *Genova* and *Kappa*

No	Sewing Aspects	Genova Coefficient	Kappa Coefficient
1.	Sewing Preparation	0.77	0.77
2.	Sewing Process	0.84	0.76
3.	Sewing Result	0.80	0.78
4.	Attitude and Time Allocation	0.76	0.78
Average		0.79	0.77

Based on the estimation using *Genova* and *Kappa* program in table 14, it results the reliability coefficient of > 0.70 . It means that the instrument of assessment of competency test of sewing women's coat has met the standard of determined reliability coefficient, which is 0.70. Therefore, the instrument of competency test of sewing women's coat can be used in wider facet.

CONCLUSION

1. The form of instrument for the assessment of competency test on sewing woman's coat in Vocational School in Yogyakarta Special Region includes: (a) performance worksheet on sewing woman's coat completed by the coat's design; (b) assessment sheet of sewing women's coat; (c) rubric or assessment guideline
2. The content validity of the assessment instrument of sewing woman's coat based on expert judgment in clothing field was categorized as good.
3. The reliability coefficient of the test on sewing woman's coat used *Genova* program is 0.79, while *Kappa*'s reliability coefficient is 0.77. The reliability index has met the determined standard that is > 0.70 , so the test can be used to measure the students' competence in competency test on sewing woman's coat.

SUGGESTION

1. In conducting assessment through competency test on sewing woman's coat, the teacher should arrange the assessment instrument as the guideline of the assessment.
2. The school should design, develop, and apply the assessment instrument so that in assessing the student's real capability it can avoid subjectivity factor, based on SKKNI which is determined in Minister's Regulation for the field of sewing clothing.
3. In conducting the assessment on sewing clothing, the teacher should not do him/herself in order to get the high consistency level.

4. Related to the demand of test reliability that is the level of consistency and the level of understanding among raters in conducting assessment, the instrument of assessment should be analyzed to find out its consistency and understanding. The program used for the analysis of performance test's reliability is done by Genova program and/or Kappa.

References:

- Berk, R.A. (1986). *Performance assessment*. Baltimore: The John Hopkins University Press
- Brennan, R.L. (2006). *Educational measurement*. Westport: Praeger
- Bresciani, M.J, et al. (2009). Examining design and inter-rater reliability of a rubric measuring research quality across multiple disciplines. *Practical Assessment, Research & Evaluation*, Vol. 14, No 12
- Djemari Mardapi. (2004). Pengembangan sistem penilaian berbasis kompetensi. *Proceeding: Rekayasa sistem penilaian dalam rangka meningkatkan kualitas pendidikan*. Yogyakarta: HEPI
- Depdiknas. (2007). *Panduan kebijakan pemanfaatan hasil ujian nasional untuk perbaikan mutu pendidikan*. Jakarta: Badan Penelitian dan Pengembangan Pendidikan
- Ernawati, Izwerni, & Weni, N. (2008). *Tata busana untuk SMK jilid 3*. Jakarta: Direktorat Pembinaan Sekolah Menengah Kejuruan
- Johnson, R.L., Penny, J.A., & Gordon, B. (2009). *Assessing performance: designing, scoring, and validating performance task*. London: The Guilford Pres
- Lichtenberg, J.W, at al. (2007). Challenges to the assessment of competence and competencies. *Article*. Professional Psychology: Research and Practice, Vol. 38, No. 5, pp 474–478
- Moore, D.R, Cheng, M.I, & Dainty, A.R.J. (2002). Competence, competency, and competencies : performance assessment in organization. *Work Study*, 51 (6), 314-316
- Orlich, D. C., et al. (2007). *Teaching strategies: A guide to effective instruction*. New York: Houghton Mifflin Company
- Smith, M.V. (2007). Inter-rater reliability of flight school instructors: a foundational study. Diambil pada tanggal 20 September 2010 dari: <http://www.public.asu.edu/~mvsmith/IRRinAviation.pdf>.
- Tri Hartiti Retnowati. (2009). Pengembangan Instrumen Penilaian Seni Lukis Anak di Sekolah Dasar. *Disertasi*. Tidak dipublikasikan. Yogyakarta: Program Pascasarjana Universitas Negeri Yogyakarta.
- Yorkovich, S. A, Waddell, G.S, & Gerwig, R.K. (2008). Competency-based assessment systems: Encouragement toward a more holistic approach. Diambil pada tanggal 5 Januari 2010 dari: http://spiritoforganization.com/documents/Waddell_CompentencyBasedAssessment.pdf

- Yosephine Flori Setiarini. (2009). Pembelajaran praktikum pada program keahlian busana di SMK. *Cakrawala Pendidikan*, Februari 2009, Th. XXVIII, No. 1
- Yoyoh Jubaedah. (2009). Model penilaian keahlian tata busana berbasis kompetensi nasional di Sekolah Menengah Kejuruan. *Disertasi*. Tidak dipublikasikan Bandung: Universitas Pendidikan Indonesia