



# Jurnal

## Penelitian dan Evaluasi Pendidikan

- Andi Ulfa Tenri Pada, Siti Salina Mustakim, Bambang Subali* Construct Validity of Creative Thinking Skills Instrument for Biology Student Teachers in The Subject of Human Physiology
- Thomas Mbenu Nulangi, Djemari Mardapi* The Accuracy of The Cheating Detection Methods in Large-Scale Tests: Mathematics National Examination
- Kana Hidayati, Budiyono Budiyono, Sugiman Sugiman* Developing Scale to Measure Student's Honesty Characters on Mathematics Learning using Subject Scaling
- Ahmad* Evaluasi Program Mahasiswa Wirausaha dengan Model CIPO di Universitas Brawijaya dan Universitas Negeri Malang
- Raden Roro Yayuk Srirabayu, Indyah Sulistyio Arty* Pengembangan Instrumen *Experiment Performance Assessment* untuk Menilai Keterampilan Proses Sains dan Kerja Sama
- Fajar Nugrobo* Evaluasi Pemanfaatan Laboratorium pada Diklat Pengendali Hama dan Penyakit Ikan di Balai Diklat Aparatur Kementerian Kelautan dan Perikanan
- David, Nasvizar Guspendri* Impact of Admission Type on Students' Achievement in Economy and Business School of Batusangkar State Institute of Islamic Studies
- Ajeng Wahyuni, Badrun Kartowagiran* Developing Assessment Instrument of Qirāatul Kutub at Islamic Boarding School
- Khoirul Basbooir, Supahar Supahar* Validitas dan Reliabilitas Instrumen Asesmen Kinerja Literasi Sains Pelajaran Fisika Berbasis STEM
- Anggarwati Riscaputantri, Sri Wening* Pengembangan Instrumen Penilaian Afektif Siswa Kelas IV Sekolah Dasar di Kabupaten Klaten



  
**Jurnal**  
**Penelitian dan Evaluasi Pendidikan**  
HIMPUNAN EVALUASI PENDIDIKAN INDONESIA (HEPI)  
in cooperation with  
GRADUATE SCHOOL OF UNIVERSITAS NEGERI YOGYAKARTA  
Kampus Karangmalang, Yogyakarta 55281. Phone. 0274 550836 Fax : 0274 520326  
Website: <http://journal.uny.ac.id/index.php/jpep>  
e-mail: [jurnalhepi@uny.ac.id](mailto:jurnalhepi@uny.ac.id)



Jurnal Penelitian dan Evaluasi Pendidikan

Volume 22, No 2, December 2018



# Jurnal

## Penelitian dan Evaluasi Pendidikan

### Publisher

HIMPUNAN EVALUASI PENDIDIKAN INDONESIA  
In Cooperation With  
PROGRAM PASCASARJANA UNIVERSITAS NEGERI YOGYAKARTA  
(MOU Nomor 195 B/J.35.17/LK/04)

---

### Director of Publication

**Djemari Mardapi**, *Universitas Negeri Yogyakarta*

### Editor in Chief

**Samsul Hadi**, *Faculty of Engineering, Universitas Negeri Yogyakarta*

### Associate Editors

**Nur Hidayanto Pancoro Setyo Putro**, *Faculty of Languages & Art, Universitas Negeri Yogyakarta*

### Editors

**Badrun Kartowagiran**, *Faculty of Engineering, Universitas Negeri Yogyakarta*  
**Syukrul Hamdi**, *Faculty of Mathematics and Natural Sciences, Universitas Hamzanwadi - Lombok Timur*  
**Edi Istiyono**, *Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta*  
**Wasis**, *Faculty of Mathematics and Natural Sciences, Universitas Negeri Surabaya*  
**Sudiyatno**, *Faculty of Engineering, Universitas Negeri Yogyakarta*  
**Undang Rosidin**, *Faculty of Mathematics and Natural Sciences, Universitas Lampung*  
**Jailani**, *Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta*  
**Alita Arifiana Anisa**, *Lembaga Pendidikan dan Pengembangan Profesi Indonesia*

---

### Board of Reviewers

**Suratno**  
*Universitas Lambung Mangkurat*  
**Moch Alip**  
*Universita Negeri Yogyakarta*  
**Budiono**  
*Universitas Sebelas Maret Surakarta*  
**Maria Josephine Wantah**  
*Universitas Negeri Manado*  
**Kartono**  
*Universitas Negeri Semarang*  
**Rasmuin**  
*Universitas Dayanu Ikhsanuddin Bau-Bau*  
**Purwo Susongko**  
*Universitas Pancasakti Tegal*  
**Nurul Choyimah**  
*LAIN Tullungagung*

**Sugeng Eko Putro Widoyoko**  
*Universitas Muhammadiyah Purworejo*  
**Primardiana Hermilia Wijayati**  
*Universitas Negeri Malang*  
**Nanik Estidarsani**  
*Universitas Negeri Surabaya*  
**Ekohariadi**  
*Universitas Negeri Surabaya*  
**Mansyur**  
*Universitas Negeri Makassar*  
**Amir Syamsudin**  
*Universitas Negeri Yogyakarta*  
**Zurqoni**  
*STAIN Samarinda*  
**Lilik Sabdaningtyas**  
*Universitas Lampung*

---

Correspondence: Graduate School of Universitas Negeri Yogyakarta  
Kampus Karangmalang, Yogyakarta, 55281, Telp. (0274) 550835, Fax. (0274) 520326

Homepage: <http://journal.uny.ac.id/index.php/jpep>  
e-mail: [jurnalhepi@uny.ac.id](mailto:jurnalhepi@uny.ac.id)

## FOREWORDS

We are very pleased that *Jurnal Penelitian dan Evaluasi Pendidikan* is releasing its issue **Volume 22, No 2, December 2018**. We are also very excited that the journal has been attracting papers from many institutions in Indonesia and many foreign countries. *Jurnal Penelitian dan Evaluasi Pendidikan* was first published in **1998** and since then regularly published online and in print twice a year: June and December.

*Jurnal Penelitian dan Evaluasi Pendidikan* with ISSN 2338-6061 (*online*) has been **re-accredited** by the Ministry of Research, Technology, and Higher Education of Republic of Indonesia under the Decree Number 30/E/KPT/2018 which is valid for 5 (five) years since enacted on 24 October 2018 (Vol. 20, No 2, 2016 until Vol. 24, No 2, 2021). *Jurnal Penelitian dan Evaluasi Pendidikan* successfully achieved accreditation in four periods in a row (in 2007, 2010, 2014, & 2018).

*Jurnal Penelitian dan Evaluasi Pendidikan* is a showcase of original, rigorously conducted educational evaluation, measurement and assessment from primary, secondary, and higher education institutions. Each issue of this journal is not limited to comprehensive syntheses of studies towards developing new understandings of educational evaluation, measurement and assessment only, but also explores scholarly analyses of issues and trends in the field.

Yogyakarta, December 2018

Editor in Chief





## Table of Content

<i>Andi Ulfa Tenri Pada, Siti Salina Mustakim, Bambang Subali</i>	Construct Validity of Creative Thinking Skills Instrument for Biology Student Teachers in The Subject of Human Physiology	119-
<i>Thomas Mbenu Nulangi, Djemari Mardapi</i>	The Accuracy of The Cheating Detection Methods in Large-Scale Tests: Mathematics National Examination	
<i>Kana Hidayati, Budiyono Budiyono, Sugiman Sugiman</i>	Developing Scale to Measure Student's Honesty Characters on Mathematics Learning using Subject Scaling	
<i>Ahmad</i>	Evaluasi Program Mahasiswa Wirausaha dengan Model CIPO di Universitas Brawijaya dan Universitas Negeri Malang	
<i>Raden Roro Yayuk Srirahayu, Indyah Sulistyio Arty</i>	Pengembangan Instrumen <i>Experiment Performance Assessment</i> untuk Menilai Keterampilan Proses Sains dan Kerja Sama	
<i>Fajar Nugrobo</i>	Evaluasi Pemanfaatan Laboratorium pada Diklat Pengendali Hama dan Penyakit Ikan di Balai Diklat Aparatur Kementerian Kelautan dan Perikanan	
<i>David, Nasvizar Guspendri</i>	Impact of Admission Type on Students' Achievement in Economy and Business School of Batusangkar State Institute of Islamic Studies	
<i>Ajeng Wahyuni, Badrun Kartowagiran</i>	Developing Assessment Instrument of Qirāatul Kutub at Islamic Boarding School	
<i>Khoirul Bashooir, Supabar Supabar</i>	Validitas dan Reliabilitas Instrumen Asesmen Kinerja Literasi Sains Pelajaran Fisika Berbasis STEM	
<i>Anggarwati Riscaputantri, Sri Wening</i>	Pengembangan Instrumen Penilaian Afektif Siswa Kelas IV Sekolah Dasar di Kabupaten Klaten	

## CONSTRUCT VALIDITY OF CREATIVE THINKING SKILLS INSTRUMENT FOR BIOLOGY STUDENT TEACHERS IN THE SUBJECT OF HUMAN PHYSIOLOGY

*Andi Ulfa Tenri Pada*<sup>1</sup> \*, *Siti Salina Mustakim*<sup>2</sup>, *Bambang Subali*<sup>3</sup>

<sup>1</sup>Universitas Syiah Kuala, <sup>2</sup>Universiti Putra Malaysia, <sup>3</sup>Universitas Negeri Yogyakarta  
<sup>1</sup>Jl. Teuku Nyak Arief No.441, Kopelma Darussalam, Syiah Kuala, Kota Banda Aceh, 24415  
<sup>2</sup>43400 Seri Kembangan, Selangor, Malaysia

<sup>3</sup>Jl. Colombo No. 1, Depok, Sleman 55281, Yogyakarta, Indonesia

\* Corresponding Author. Email: [andi\\_ulfa@unsyiah.ac.id](mailto:andi_ulfa@unsyiah.ac.id)

### Abstract

This article discusses the construct validity of the creative thinking skill instrument supporting a conation idea aspect for biology student-teachers in the subject of Human Physiology. Two hundred and eighteen students participated in this study. Construct validity was obtained through Confirmatory Factor Analysis technique (CFA). Reliability was estimated by composite reliability. Findings show that the construct validity and the reliability of the instrument are high. A thorough discussion on the findings and future implications are provided towards the end of this paper.

**Keywords:** *construct validity, creative thinking skill, instrument development*

*Permalink/DOI: <http://dx.doi.org/10.21831/pep.v22i2.22369>*

## **Introduction**

Future learning challenges demand the development of creative thinking skills as one of the issues for 21<sup>st</sup> century literacy (Widowati, 2009). Development in information and communication technology requires high-order thinking competencies that include creative thinking skills, to be developed through evaluation processes. It is unfortunately true that, up to now, learners' thinking skills have not been seriously dealt with.

In the study by Rofi'uddin (2000), discontents are found concerning the low levels critical-creative thinking competencies in the graduates of elementary to tertiary levels. This research finding is in line with the results of international creativity survey highlighted in The Global Creativity Index conducted in 2011. The study indicated that Indonesia was at the rank of 81<sup>st</sup> from the 82<sup>nd</sup> countries in the survey, far below neighbouring countries Singapore (Rank 9) and Malaysia (Rank 48) (Florida, Mellander, & Stolarick, 2011). Meanwhile, according to the Human Development Report in 2013, Indonesia's Human Development Index (HDI) was at the rank of 110, while Singapore 11 and Brunei Darussalam 31 respectively. These two neighbouring countries, Singapore and Brunei Darussalam, had a very high HDI. Indonesia was underscored far left behind. Malaysia (HDI 62) was at the high category. In spite of the fact that Indonesia was in the same group as Thailand, at the medium category of HDI, the later was at rank 93, far above Indonesia.

In lieu toward the development of Higher Order Thinking Skills (HOTS), an international survey shows that Indonesian human performances are still minimal. A study by Ramirez & Ganaden's (2008) reveals that poor human performances are caused by incapability of infusing the high-order thinking skills. Learning in higher education seems to be not as effective as to develop creative thinking. Meanwhile, creative thinking is listed as the top skill in the cognitive domain of Bloom's taxonomy as revised by Anderson (Krathwohl, 2002) and the *New Blooms* (Dettmer, 2006). Further-

more, DeHaan's (2009) study shows that creative thinking competencies help learners to find evidence-based analyses, improve HOTS and solve problems. Thus, creative thinking skills need to be further developed and measured.

With respect to the needs of creative thinking skills, it is believed that learners of Biology with HOTS will be able to think creatively and solve problems effectively in completing project tasks provided such as determining objects to observe, gathering information from relevant sources, and using various ideas in and outside the campus. Problem-solving competencies will also be useful for students after they become teachers, who will encounter situations with problems to be solved. This statement goes in line with a study conducted by Husain, Mustapha, Malik, & Mustakim (2014) which highlighted how interaction between environment and student influences student's development in learning. These competencies include developing instructional contents, providing effective classroom management, assessment, and preparing the teaching and learning for 21<sup>st</sup> century.

In the present time, the measurement of creative thinking is conducted in a general manner. One divergent thinking test that is widely known is the Torrance Tests of Creative Thinking (TTCT). Torrance takes creativity as independent cognitive competence and a multidimensional concept. In contrast, Weiping Hu and Philip Adey maintain that creativity is part of intelligence and an independent skill. They developed a scientific creativity test for high-school students (Hu & Adey, 2002). The test, known as A Scientific Creativity Structure Model (SCSM), was built on the analyses of meaning and a scientific creativity aspect based on literary studies. Hu and Adey did an analysis on general themes from science and creativity and built a three-dimensional model from the characteristics of creativity such as features, processes, and products. Fluency, flexibility, and originality were features in SSCM. Scientific products were related to the scientific field such as technical

products, scientific knowledge, scientific phenomena, and scientific problems. They tried to include all their scientific creativity model components in a framework and test. For instance, they included scientific thinking processes and imagination, which often were not asked in a creativity test. Based on these reviews of the measurement of creative thinking, it can be seen that there is a wide opportunity to have creative thinking skills integrated in the school lessons. Studies focusing on the development of creative thinking skill evaluation in school lessons have been conducted (Beghetto, 2013). These studies show that creative performance is moderately a specific domain and can be appraised by a combination of particular power resources (Sternberg, 2006). This is supported by Kaufman and Baer Charyton, Ivcevic, Plucker, & Kaufman (2009) and Barbot, Besançon, & Lubart (2011) that creativity tends to be a specific domain. One may certainly be creative in one domain and a number of persons may be creative in two or more domains. One may most possibly take risks in their own domain for they have a higher level of content.

The choice on the subject of Human Physiology is based on the consideration that, in this subject, students' conative idea responses can be obtained more easily seeing that the analysed cases are contextual. When studying human physiology, students learn about the normal functions of the body organs; therefore, in the instrument, the given stimuli are disrupters to body organs (diseases) or the opposites of the normal functions of the physical organs. The expectation is that, by giving abnormal phenomena as stimuli, learners are able to give various responses in the forms of solutions to problems that are presented in various cases.

Evaluation of learners' creative thinking can be done in an evaluation system of the divergent pattern (Subali, 2011, pp. 130–144). Divergent thinking is a skill in constructing or producing various possible responses, ideas, options, or alternatives against a problem (Isaksen, Dorval, & Treffinger, 1994, p. 18). In other words, divergent think-

ing can be understood as a skill in producing various solutions to a problem using the correct procedure and reasons.

Creating thinking competence is one of the important thinking skills so that it must be mastered by the students. Moreover, the competency-based curriculum in the university brings heavy emphasis on the practice of thinking and reasoning, developing creative activities, problem-solving skills, and communicating ideas.

The evaluation kit to measure creative thinking skills that supports the ideas of teacher-candidate students of biology through a divergent pattern is expected to become one of the ways to improve students' creating thinking skills. The students' logical thinking will be directed to producing arguments based on their concept of understanding in the form of conative ideas. As a result, teachers will be able to see students' divergent production patterns in the form of rational alternatives towards given stimuli so that they can explain concepts that are contradictory to the actions.

The evaluation instrument must be tested for its quality. Quality testing is one way to show that an evaluation instrument has been optimally developed. The primary evidence of the quality of an evaluation instrument is its validity. Messick defines validity as "one that is integrative on how far empirical evidences and theoretical rationales support the feasibilities of interpretations and actions based on the results of the measurement processes" (Reynold, Livingstone, & Wilson, 2009). In line with Messick, the Standards for Educational and Psychological Testing defines validity as "how far evidences and theories support the interpretation of tests scores as a consequence of the using of the test" (American Educational Research; Association American Psychological Association; & National Council on Measurement in Education, 1999).

The validity measure that will be discussed in this article is of the construct type (construct validity), i.e. the construct of the instrument from theories supported by field data (Messick, 1989; Silverlake, 1999). Con-

struct validity is understood as how far the scores of the results of the measurement reflect the theoretical constructs underlying the development of the instrument (Suryabrata, 2000). Hadi (2001) states that construct validity is identical with logical validity or validity by definition. Hadi (2001) also states that, if the theoretical construct of a test is right, then the product of the measurement is valid.

Empirical testing is needed in order to find how far each variable to be measured can be explained by each dimension in the instrument. An instrument is regarded as qualified after it is analysed theoretically and empirically. Theoretical analysis is done by way of instrument review. Empirical analysis is done by way of factor analysis.

Factor analysis techniques come in two types, namely Exploratory Factor Analysis (EFA) and Confirmatory factor analysis (CFA). EFA is used for exploring the theories; CFA for confirming the theories that are obtained in EFA. CFA is a popular statistical technique for providing supports to construct validity in the literature of psychological testing (American Educational Research; Association American Psychological Association; & National Council on Measurement in Education, 1999; Thompson & Daniel, 1996). The present article shows the utilization of EFA and CFA to provide evidence for construct validity in the evaluation instrument development. In particular, the study is aimed at giving evidence to the construct validity of the creative thinking test instrument for teacher-candidate students of biology. A review of CFA, based on theoretical evidence, is an important part of the validation process (DiStefano & Hess, 2005, p. 228).

## **Method**

The study was a confirmatory research which involved 218 students of Biology Programme in the Faculty of Teacher Education. Respondents were selected based on their prior knowledge on Human Physiology learning environment. Implementation of the practical guidelines for the factor analysis

technique followed the ones given by Cattell (1978) dan Guilford (1954) suggesting  $N > 200$  as the number for sample size in the study.

The study used instruments validated by a panel of Subject-Matter Experts (SMEs) in the fields of Educational Assessment, Educational Evaluation, Biology Education, and Human Physiology for their contents (Hammit & Zhang, 2013). The quality of instruments was also observed through their empirical validity and reliability. The instrument validity was appraised by internal consistency of construct indicators to show the degree to which each indicator indicated a general latent construct.

The data analysis began with EFA, which was preceded by Kaiser Meyer Olkin Measure (KMO) testing and Bartlett's Test on the SPSS, and proceeded with CFA using the Lisrel software. According to Hair, Black, Babin, & Anderson (2006), evaluation on the levels of suitability between the data and the model is done through a couple of fits, i.e. the overall model fit and the measurement model fit.

## **Findings and Discussion**

### **Findings**

Data analysis was made first by conducting factor analysis which was preceded by KMO testing and Bartlett's Test in order to determine the sufficiency of the sample. The result showed a KMO MSA value of  $0.821 > 0.05$  with a  $p$ -value  $0.000 < 0.05$ . This finding showed that the data matrix had sufficient correlation measure to be used to conduct a factor analysis. Subsequently, analysis from the communality values showed that not all items had a value higher than 0.30, the acceptable minimum value (Mooi & Sarstedt, 2011). From the EFA, it was found that 10 factors had an eigen value higher than 1, and four had an eigen value higher than 2. In this study, an eigen value higher than 2 is preferred.

Results of the EFA analyses showed that the instrument consisted of 37 test items distributed in four factors. The four factors

were: (1) solution alternative, the skill to produce a number of solutions to respond to a problem formulated in 10 test items; (2) original solution, the skill to produce a number of solutions that are relevant and unique, consisting of eight test items; (3) solution feasibility, the skill to produce a number of solutions effective for the solution of the problem in the given case, consisting of 10 test items; and (4) solution variety, the skill to produce a number of solution categories, consisting of nine test items.

### Overall Model Fit

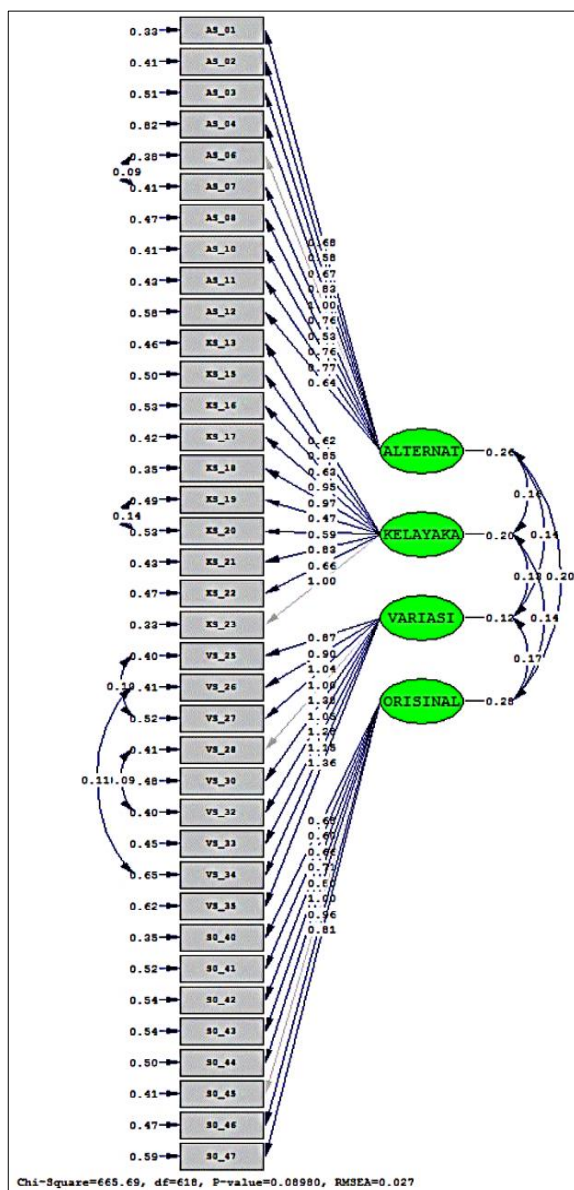


Figure 1. Model of the Creative Thinking Skill Measurement that Supported the Conative Aspect

The first phase of the model-fit testing was directed to evaluating the general degree of the goodness of fit (GOF) between the data and the model. Some comparative-fit indexes emerged that differentiated one model from the base model. Practical guidelines were used to determine the accepted degree of fits since the sampling distribution of the indexes was unknown (Shook, Ketchen, Hult, & Kacmar, 2004). Hence, until a definite measurement was developed should researchers used some steps to provide evidence for their model fits (Breckler, 1990). The use of several indexes would ensure the readers that researchers did not choose only the favourable indexes. The study by Gerbing & Anderson (1992) shows that, among the strong and stable indexes, the normed fit index (NFI), and the comparative index (CFI) are two of these (Hu & Bentler, 1999). Results of the data analysis using the CFA technique can be seen in Figure 1.

In Figure 1, the Chi Square obtained = 665,69,  $df = 618$ ,  $p$ . value = 0.08980 and RMSEA = 0,027. This indicates that the model has a fit. This is in line with the criteria as shown in Table 1.

### Measurement Model Fit

In obtaining the conclusion that the fit between the data and the model was in general good, the next step was the evaluation of the measurement model fit. This was done by looking at the validity and reliability of the measurement model (Hair et al., 2006). These are presented as follows.

Convergent validity is used to test construct validity. The word construct refers to a theoretical view to explain some phenomena (Wiersma, 2000). According to Van Dalen (1973), a construct customarily refers to a complex concept that covers a number of inter-related factors.

Table 1. Results of Overall Model Fit in the Initial Try-Out

GOF Measure	Degree Target of Fit	Estimation Results	Degree of Fit
Normed $\chi^2$	Normed $\chi^2 < 2$	1.07	Good ( <i>good fit</i> )
<i>p</i> -value	<i>p</i> $\geq$ 0.05	0.089	Good ( <i>good fit</i> )
RMSEA	RMSEA $\leq$ 0.05	0.027	Good ( <i>close fit</i> )
ECVI	Smaller values & close to ECVI saturated	M* = 7.67 S* = 12.90 I* = 30.27	Good ( <i>good fit</i> )
AIC	Smaller values & close to AIC saturated	M* = 835.69 S* = 1406.00 I* = 3436.78	Good ( <i>good fit</i> )
NFI	NFI $\geq$ 0.90	0.81	Not so Good ( <i>marginal fit</i> )
GFI	GFI $\geq$ 0.90	0.75	Not so Good ( <i>marginal fit</i> )
CFI	CFI $\geq$ 0.92	1.00	Perfect ( <i>perfect fit</i> )
RMR	SRMR $\leq$ 0.09	0.079	Good ( <i>good fit</i> )

\*M = Model; S\* = Saturated; I\* = Independence

An indicator is said to have good validity towards the construct or latent variable if (a) the *t*-value of its factor loading is higher than the critical value (*t*-value  $\geq$  1.96) (Doll, Xia, & Torkezadeh, 1994; Hair et al., 2006); and (b) the standardized factor loading is  $\geq$  0.30 (Gorsuch, 1983; Mooi & Sarstedt, 2011). The *t*-value and the standardized factor loading of the creative thinking skill measurement model are presented in Table 2.

In Table 2, it can be seen that all the *t*-values of the factor loadings of the variables or items are higher than 2 (*t*-value  $>$  2). The factor loadings of the variables in the model are therefore are significant, or are not zero. Each of the standardized factor loading of each variable is higher than the minimum value (standardized factor loadings  $>$  0.3). Thus, it can be concluded that the validity of all the variables observed towards the latent variable is good.

Evaluation on the measurement model can be done by using the composite reliability measure, or also known as construct reliability (CR) (composite reliability) (Ghadi, Bakar, Alwi, & Talib, 2012; Hair et al., 2006). A construct can have a good reliability measure when its value is the same or higher than 0.70 (Hair et al., 2006; Lance,

Butts, & Michels, 2006). CR can be calculated by the total of the squared factor loading (*Li*) factor loading added with total error variance of a construct (*ei*).

$$CR = \frac{(\sum_{i=1}^n Li)^2}{(\sum_{i=1}^n Li)^2 + (\sum_{i=1}^n ei)}$$

From the results of the computation of the construct reliabilities in Table 2, it can be concluded that the construct reliability of the instrument model is good (CR  $\geq$  0.70). CFA is done to estimate the factor loading of a variable. Factor loading presents the level of the regression path from the latent to its indicator. The CR level is an alternative guide to review convergent validity (Ghadi et al., 2012).

The model being tested consisted of four latent variables and 37 observed variables. The model, supported theoretically and empirically, placed four factors separated but correlated. The factors solution alternative consisted of 10 items, original solution eight, solution feasibility 10, and solution variety nine. CFA results showed a good model fit for the instrument with a sample of 370 students. In addition, the moderate correlation among the four factors (SLF  $\geq$  0.30) showed that these factors considered the constructs



different but having relations (Mezo & Short, 2012).

Table 2. T-Value, Standardized Factor Loadings & Construct Reliability

Item	SLF*	t-Value	CR
<b>Solution Alternative Factor</b>			
AS_01	0.52	4.44	0.74
AS_02	0.42	3.72	
AS_03	0.43	3.81	
AS_04	0.42	3.74	
AS_06	0.64	**	
AS_07	0.52	5.11	
AS_08	0.37	3.29	
AS_10	0.52	4.46	
AS_11	0.52	4.45	
AS_12	0.4	3.52	
<b>Solution Feasibility Factor</b>			
KS_13	0.38	3.31	0.71
KS_15	0.47	4.01	
KS_16	0.36	3.16	
KS_17	0.55	4.53	
KS_18	0.59	4.78	
KS_19	0.29	2.58	
KS_20	0.34	3.01	
KS_21	0.49	4.14	
KS_22	0.39	3.43	
KS_23	0.61	**	
<b>Solution Variety Factor</b>			
VS_25	0.43	3.49	0.73
VS_26	0.43	3.52	
VS_27	0.44	3.57	
VS_28	0.47	**	
VS_30	0.56	4.14	
VS_32	0.5	4.37	
VS_33	0.55	4.08	
VS_34	0.44	3.56	
VS_35	0.51	3.91	
<b>Solution Originality Factor</b>			
S0_40	0.52	4.63	0.73
S0_41	0.41	3.7	
S0_42	0.43	3.91	
S0_43	0.46	4.11	
S0_44	0.51	4.57	
S0_45	0.64	**	
S0_46	0.6	5.17	
S0_47	0.49	4.37	

\*SLF = *Standardized loading factor*

\*\* = *Reference variable, t-value is not estimated. t-value target  $\geq 2$ .*

## Discussion

The research results show that the instrument for testing creative thinking skills

developed for biology teacher-candidate students has a high construct validity. This finding is in agreement with Batey (2012) who stated that creativity has a relation with learning and education, more particularly in the matter of problem solving. This can be combined with the emphasis in the contextual approach that creativity is a social phenomenon underlying the interaction between the individual and the situation (Barbot et al., 2011; DeHaan, 2009). Measurement of learning achievement at the higher-order cognitives, such as creative thinking, demands tasks that require learners to use their knowledge and skills in a new situation (Nitko & Brookhart, 2007). Therefore, learners are demanded not only to be able to understand, but also to be able to analyse, evaluate, and create.

The measurement has an emphasis more on tasks/problems that are oriented to the real world than the school contexts. In line with this, Guilford (1954) stated that stimuli in the forms of conditions/situations may change learners' behaviours. Guilford further stated that creative actions are an example of the outcomes of a learning process that shows changes in behaviours resulting from stimulation and responses. The present study shows that the values of the standardized loading factors are all above 0.30 and the *t*-values of the observed variables are  $\geq 2.00$  with a confidence level of 95%. These show that the validity of all the observed variables against the latent variables is good. In other words, all the items are valid for testing the creative thinking skills in the Human Physiology subject.

Scientific creativity among 130 students was studied in Taiwan by Liang (2002). The primary instruments included Test of Divergent Thinking (TDT) for measuring creativity, Creativity Rating Scale (CRS) and Creative Activities and Accomplishments Check Lists (CAACL) for measuring scientific creativity, Nature of Scientific Knowledge Scale (NSKS) for measuring the nature of science, and Science Attitude Inventory II (SAI II) for measuring attitudes towards science. There were in addition two more

instruments for measuring learners' competences in finding problems and formulating hypotheses. Data analysis techniques included descriptive statistics, Pearson product-moment correlations, and gradual regression. The findings show that, among others, students' scientific creativity is significantly correlated with attitudes towards science, finding problems, formulating hypotheses, nature of science, resistance towards closure, originality, and elaboration. Another finding is related to the emergence of four predictors that are significant for measuring scientific creativity; namely attitudes towards science, finding problems, resistance towards closure, and originality that contribute 48% to the variance of students' scientific creativity. Still another finding is concerned with the big difference between students with high scientific creativity from those with low scientific creativity on the variables family support, career picture, and reading about science. It can be concluded that both cognitive and non-cognitive components are good for predicting scientific creativity.

In the meantime, Barbot et al. (2011) conducted a study to give a real picture and a different way to measure creative potentials that can be used for educational objectives. The study produced a model for evaluating creativity called EPoC (Evaluation of Potential for Creativity) that was multifaceted and domain-specific, and that made it possible for the evaluator to capture the multidimensionality of creative potentials in obtaining potential profiles for creativity. It was a procedure for measuring students' creative potentials by way of multivariate approaches using verbal and graphic tasks for two creative thinking competences: *divergent thinking* (DT) and *integrative thinking* (IT). The creativity phases were designed in two sessions, within each of which each thinking competence was measured. In the DT assignments, each student was asked to produce as many ideas in responding to a unique stimulus. In the IT assignments, students were asked to produce synthetic solution elaborations. The creativity score for each task was based on the total of outputs

(for the DT) and level of originality. The higher score, the higher the potentials of the individual's creativity.

A study in the measurement of creativity in the field of science was done by DeHaan (2009) exploring the relation between creativity and high order cognitive skills (HOCS), reviewing evaluation techniques, and describing learning strategies to improve creative problem solving in the university. DeHaan used Torrence's creative thinking test that dealt with students' divergent thinking competences to measure their scientific creative thinking. In his opinion, creativity is not a thing difficult to measure. Creative processes can be explained by using cognitive competences which are better known such as cognitive flexibility and resistance control that have been widely spread in the society. Creativity is an important element of problem solving and critical thinking. Consequently, creativity application such as creative power and intelligence are components of the HOCS as defined in the taxonomy of educational objectives. It is no wonder that creativity, as other elements in HOCS, can be taught effectively through inquiry-based instruction founded by the constructivist theory.

Finally, reliability of the instrument was estimated by the technique of composite reliability measure or also known as construct reliability (Hair et al., 2006). A high construct reliability shows internal consistency, meaning that all the steps in the measurement consistently represent the same latent constructs. The results of the reliability calculation in Table 2 show that all the reliability values of the constructs are  $\geq 0.7$ ; thus, it can be concluded that the reliability of the measurement is categorically good.

## Conclusion

Based on the research results and discussion, a number of conclusions can be proposed. First, the instrument for measuring the creative thinking skills of biology teacher-candidate students possesses a high degree of construct validity. Second, the instrument is built of four factors namely: (1)

solution alternative, the skill to produce a number of solutions in responding to a problem, consisting of 10 test items; (2) original solution, the skill to produce a number of solutions that are relevant and unique, consisting of eight test items; (3) solution feasibility, the skill to produce a number of solutions that are effective to solve the given problem, consisting of 10 test items; and (4) solution variety, the skill to produce a number of categorical solutions, consisting of nine test items. Third, the instrument consists of 37 items; each having a high loading factor against the latent variable. Fourth, besides having a high construct validity, the instrument is also characterized with a high composite reliability.

Implication from the results of the research study is that the instrument for measuring the creative thinking skills of biology teacher-candidate students is feasible to be used. In relation with the research results, those who need to measure students' creative thinking skills are recommended to use this instrument bearing in mind that the respondents have identical characteristics as the ones in the present study.

## References

- American Educational Research; Association American Psychological Association; & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Barbot, B., Besançon, M., & Lubart, T. I. (2011). Assessing creativity in the classroom. *The Open Education Journal*, 4(1), 58–66.
- Batey, M. (2012). The measurement of creativity: from definitional consensus to the introduction of a new heuristic framework. *Creativity Research Journal*, 24(1), 55–65.
- Beghetto, R. A. (2013). Nurturing creativity in the micro-moments of the classroom. In K. H. Kim, J. C. Kaufman, J. Baer, & B. Sriraman (Eds.), *Creatively Gifted Students are not like Other Gifted Students: Research, Theory, and Practice*. Sense Publishers.
- Breckler, S. J. (1990). Applications of covariance structure modeling in psychology: cause for concern. *Psychological Bulletin*, 107(2), 260–273.
- Cattell, R. B. (1978). *The scientific use of factor analysis*. New York: Plenum.
- Charyton, C., Ivcevic, Z., Plucker, J. A., & Kaufman, J. C. (2009). Creativity assessment for evaluating the effectiveness of educational programs. In C. Schreiner (Ed.), *Handbook of Research on Assessment Technologies, Methods, and Applications in Higher Education* (pp. 78–96). Hershey: IGI Global.
- DeHaan, R. L. (2009). Teaching creativity and inventive problem solving in science. *Life Sciences Education*, 8, 172–181.
- Dettmer, P. (2006). New Blooms in established fields: Four domains of learning and doing. *Roeper Review*, 28(2), 70–78.
- DiStefano, C., & Hess, B. (2005). Using confirmatory factor analysis for construct validation: An empirical review. *Journal of Psychoeducational Assessment*, 23(3), 225–241.
- Doll, W. J., Xia, W., & Torkezadeh, G. (1994). A confirmatory factor analysis of the end-user computing satisfaction instrument. *Management Information Systems Quarterly*, 18(4), 453–461.
- Florida, R., Mellander, C., & Stolarick, K. (2011). *Creativity and prosperity: The global creativity index*. Toronto, Ontario, Kanada: Martin Prosperity Institute.
- Gerbing, D. A., & Anderson, J. C. (1992). Monte Carlo evaluations of goodness of fit indices for structural equation models. *Sociological Methods and Research*, 21(2), 132–160.
- Ghadi, I. N., Bakar, K. A., Alwi, N. H., &

- Talib, O. (2012). Construct validity examination of critical thinking dispositions for undergraduate students in University Putra Malaysia. *Higher Education Studies*, 2(2), 138–145.
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale: Erlbaum.
- Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York: McGraw-Hill.
- Hadi, S. (2001). *Metodologi research (Jilid III)*. Yogyakarta: Andi Offset.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2006). *Multivariate data analysis* (7th ed.). Prentice Hall.
- Hammitt, J. K., & Zhang, Y. (2013). Combining experts' judgments: comparison of algorithmic methods. *Risk Analysis*, 33(1), 109–120.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55.
- Hu, W., & Adey, P. (2002). A scientific creativity test for secondary school students. *International Journal of Science & Education*, 24(4), 389–403.
- Husain, M. Y., Mustapha, R., Malik, S., & Mustakim, S. S. (2014). Validation of instruments learning environment: exploratory factor analysis (EFA). In *10th Conference of AASVET (Asian Academic Society for Vocational Education and Training)*. Tokyo.
- Isaksen, S. G., Dorval, K. B., & Treffinger, D. J. (1994). *Creative approaches to problem solving*. Dubuque: Kendall/Hun.
- Krathwohl, D. R. (2002). *A revision of Bloom's taxonomy: an overview. theory into practice*. Taylor & Francis, Ltd.
- Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria what did they really say? *Organizational Research Methods*, 9(2), 202–220.
- Liang, J.-C. (2002). *Exploring scientific creativity of eleventh grade students in Taiwan*. The University of Texas at Austin. Retrieved from <https://pdfs.semanticscholar.org/470a/a97e1cd2c757315b9f9330abc15a7b4ca41a.pdf>.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13–103). New York: American Council on Education/Collier Macmillan.
- Mezo, P. G., & Short, M. M. (2012). Construct validity and confirmatory factor analysis of the self-control and self-management scale. *Canadian Journal of Behavioural Science*, 44(1), 1–8.
- Mooi, E., & Sarstedt, M. (2011). *A concise guide to market research: the process, data, and methods using IBM SPSS Statistics*. Springer-Verlag.
- Nitko, A. J., & Brookhart, S. M. (2007). *Educational assessment of students*. New Jersey: Pearson Education.
- Ramirez, R. P. B., & Ganaden, M. S. (2008). Creative activities and students' higher order thinking skills. *Education Quarterly*, 66(1), 22–33.
- Reynold, C. R., Livingstone, R. B., & Wilson, V. (2009). *Measurement and assessment in education*. Upper Saddle River: Pearson Education Inc.
- Rofi'uddin, A. (2000). Model pendidikan berpikir kritis-kreatif untuk siswa sekolah dasar. *Majalah Bahasa Dan Seni*, 1(28), 72–94.
- Shook, C. L., Ketchen, D. J., Hult, G. T. M., & Kacmar, K. M. (2004). An assessment of the use of structural equation modeling in strategic management research. *Strategic Management Journal*, 25(4), 397–404.
- Silverlake, A. C. (1999). *Comprehending test manuals: a guide and workbook*. Los Angeles: Pyrczak Publishing.

- Sternberg, R. J. (2006). The nature of creativity. *Creativity Research Journal*, 18(1), 87–98.  
[https://doi.org/10.1207/s15326934crj1801\\_10](https://doi.org/10.1207/s15326934crj1801_10)
- Subali, B. (2011). Pengukuran kreativitas keterampilan proses sains dalam konteks assessment. *Cakrawala Pendidikan*, XXX(1), 130–144.
- Suryabrata, S. (2000). *Pengembangan alat ukur psikologis*. Yogyakarta: Andi.
- Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement*, 56, 197–208.
- Van Dalen, D. B. (1973). *Understanding educational research: an introduction*. McGraw-Hill.
- Widowati, A. (2009). Pengembangan critical thinking melalui penerapan model PBL (problem based learning) dalam pembelajaran sains. In *Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA*. Yogyakarta: Fakultas MIPA, Universitas Negeri Yogyakarta.
- Wiersma, W. (2000). *Research methods in education: an introduction* (7th ed.). Boston: Allyn & Bacon.

## THE ACCURACY OF THE CHEATING DETECTION METHODS IN LARGE-SCALE TESTS: MATHEMATICS NATIONAL EXAMINATION

*Thomas Mbenu Nulangi<sup>1\*</sup>, Djemari Mardapi<sup>1</sup>*

<sup>1</sup>Universitas Negeri Yogyakarta

<sup>1</sup>Jl. Colombo No. 1, Depok, Sleman 55281, Yogyakarta, Indonesia

\* Corresponding Author. Email: [tommymbenu07@gmail.com](mailto:tommymbenu07@gmail.com)

### Abstract

This study aimed to describe (1) the characteristics of items based on the Item Response Theory, (2) the cheating level in the implementation of the national examination based on Angoff's B-Index method, Pair 1 method, Pair 2 method, Modified Error Similarity Analysis (MESA) method, and G2 method, (3) the most accurate method to detect the cheating in the mathematics national examination at the senior secondary school level in the academic year of 2015/2016 in East Nusa Tenggara Province. The result of the item response theory analysis showed that 17 (42.5%) items of the mathematics national examination fit with the 3-PL model, with the maximum information function of 58.0128 at  $\theta = 1.6$ , and the measurement error of 0.1313. The number of pairs detected to be cheating by Angoff's B-Index method was 63 pairs, that by the Pair 1 method was 52 pairs, that by the Pair 2 method was 141 pairs, that by MESA method was 67 pairs, and that by the G2 method was 183 pairs. The methods which could detect most pairs doing cheating were the G2 method, the Pair 2 method, the MESA method, Angoff's B-Index method, and the Pair 1 method successively. The methods which could accurately detect cheating based on the computation of the standard error were Angoff's B-Index method, the G2 method, the MESA method, the Pair 1 method, and the Pair 2 method successively.

**Keywords:** *national examination, item characteristics, cheating detection methods*

*Permalink/DOI: <http://dx.doi.org/10.21831/jpep.v22i2.14930>*

## Introduction

The administration of the national examination as regulated in Government Regulation No 13 Year 2015 (Presiden Republik Indonesia, 2015) about the Educational National Standard aims to reveal how far the students have already mastered the competence. This is in accordance with the competence-based instructional objectives that students should know the development of their learning ability. The competence-based teaching should be implemented to obtain information about students' success and teachers' success in teaching. In the competence-based teaching, it is necessary to establish the graduate competence standard which include attitudes, knowledge, skills so that they can be a reference for schools to implement the teaching activities.

The Educational National Standard Board (*Badan Standar Nasional Pendidikan/BSNP*) as an independent institution is responsible for administering the national examination objectively, fairly, and accountably so that the result of the national examination can provide information about the students' learning success and teachers' success in teaching (Mulyati & Kartowagiran, 2013). Therefore, honesty and achievement become the motto of the administration of the national examination in 2015 (*Badan Standar Nasional Pendidikan*, 2015, p. 4) The 2015 national examination motto should be considered as the common principle which should be owned by all stakeholders in the administration of the national examination so that the national examination administration is of high quality.

Up to the moment, the Indonesian educational system still uses the final achievement as the indicator which the progress and the students' mastery of knowledge. As a result, the community tends to view learning achievement as only seen from the high final scores not in the learning process. This view has caused burden to students to get as high scores as possible (Manoppo & Mardapi, 2014). The burden will make students be oriented to getting high scores, not to mastering knowledge so that students

can do anything including cheating in the examination.

Cheating behaviour is not expected to happen in the administration of the broad scale testing like the national examination. Cheating behaviour is considered as violating the law because this will give benefit to those doing cheating. In line with this, Cizek (2001, p. 7) states that cheating can be defined as behaviour which violates the law for administering a test, every behaviour which gives benefits unfairly in the examination, or the behaviour of the test administrator which can decrease the accuracy of the test scores or the examinees' performance.

Cheating behaviour can be found in all levels and done systematically. As stated by Davis, Drinan, & Gallant (2009, p. 1), students of all educational levels do cheating, from the primary schools to the graduate program, from schools located in villages to those located in cities, from poor schools to rich schools, both public and private schools. Students do cheating because they are afraid that they will get poor scores. They are not honest to themselves. They are not honest their friends and their parents. They sometimes even work hand in hand with teachers and school staff to do cheating.

The problem about cheating is a classical problem faced by all nations. Findings of questionnaires and interviews by Mardapi (2000) when doing an evaluation of the administration of the national examination in 2000 showed that 76% grade I senior secondary school students in Central Java, 73.9% grade I senior secondary school students in South Sulawesi, and 81.8% senior secondary school students in Jambi did cheating in the national examination. The cheating could be done in the form of opening the notes, peeping at others' work, or doing something not allowed according to the regulation. Further findings showed that 20% grade I senior secondary school students in Central Java, 26.1% grade I senior secondary school students in South Sulawesi, and 18.2% grade I senior secondary school students in Jambi said that the proctors assisted the students in doing the exami-



nation. This happened because the proctors were kind so that the standard of the national examination administration had not been implemented in all schools. Another problem was related to the true examination scores (*Nilai Ebtanas Murni/NEM*). There were 12.6% teachers in Central Java, 23.8% teachers in South Sulawesi, and 46.8% teachers in Jambi thought that the true examination scores (NEM) were not true scores in the real sense.

The Ministry of Education and Culture through its website (<http://www.kemdikbud.go.id>) presented news that the national examination integrity index in 2015 was 64.05 nationally. This means that cheating was still done in the administration of the national examination in Indonesia in 2016. In addition, the information provided by the ministry of education and culture in 2015 showed that the national examination integrity index of East Nusa Tenggara province was 62.2%. In 2016 the national examination integrity index of East Nusa Tenggara was 69.3%. Although there was an increase, there were still 30.7% considered as doing cheating in the national examination (Biro Komunikasi dan Layanan Masyarakat Kementerian Pendidikan dan Kebudayaan, 2016).

Cheating behaviour does not only happen in Indonesia but also in other countries such as America. Naghdipour & Emeagwali (2013) conducted a study on 500 consisting of 450 students and 50 lecturers from different faculties in America. The findings showed that there was an indication of academic dishonesty, although students did not make any report on this. In addition, there was a difference between the report as provided by the students and that provided by lecturers in their observation about the number of students doing cheating.

The cheating behaviour will give an impact on the students' future if this is done again and again. The study conducted by Bernardi, Banzhoff, Martino, & Savasta, (2012) showed that the cheating behaviour done again and again from early childhood to adulthood will result in social problems.

The cheating behaviour will influence students' willingness to do the same in their social life in the future. The study conducted by Zastrow (1970) on 45 graduate students showed that 40% of the students did cheating. The reason for doing cheating is the burden from the parents. The parents expected them to get good marks. In addition, this study also found out that there was no difference in the characteristics between students who did cheating and those who did not. This means that the cheating behaviour was done systematically.

There are some ways of doing cheating in the examination. The study conducted by Baird (1980) on 200 students showed that 75% of the students did cheating in the examination. Baird found twelve ways of doing cheating, such as, getting information from other students, copying others' work, copying others' assignment, copying from books, copying others' work, hiding errors from the lecturers, illegal test information, stealing others' test instruments, changing the examination paper, taking a test for others, others taking test for a student, bribery and blackmailing in an examination.

Previously, there were some studies focusing on demographic aspects such as gender Anderman & Murdock (2007, p. 11). The study focused on the difference in gender in doing cheating. The study revealed that male students tended to do more cheating than female students (Calabrese, & Cochran, 1990; Davis, Grover, Becker, & McGregor, 1992; Michaels & Miethe, 1989; Newstead, Franklyn-Stokes, & Armstead, 1996).

However, Whitley, Nelson, & Jones (1999) Nelson and Jones (1999) revealed the effect of the attitude size effect of gender in the medium size and the effect was so small so that they all did cheating. Therefore, both male and female students had the tendency to do cheating.

Doing cheating in the class must be stopped by cutting the missing link. Therefore, teachers should be more proactive in stopping the cheating behaviour in the class. Some simple ways to decrease cheating in the

class such as distributing different test forms, making the seating arrangement randomly, improving the supervision when proctoring an examination, and making spaces between seats (Bernardi, Baca, Landers, & Witek, 2008). Those ways are considered as being effective in reducing the cheating behaviour in the administration of the national examination.

The advancement of technology has contributed greatly to cheating, called e-cheat. A survey conducted by Stogner, Miller, & Marcum (2013) on 534 students in universities showed that 40% of the students did e-cheating the year before. They did cheating using e-cheating because it was free of charge. The findings of the study suggested that teaching restrict cheating behaviour.

Many survey reports showed that copying answers was very common cheating behaviour (Bopp, Gleason, & Misicka, 2001; Brimble & Clarke, 2005; Hughes & McCabe, 2006; Jensen, Arnett, Feldman, & Cauffman, 2002; Lin & Wen, 2007; Rakovski & Levy, 2007; Vandehey, Diekhoff, & LaBeff, 2007). Copying answers could result in uncommon response patterns for the students doing the copying or uncommon answer similarity between two responses, which can be studied through the probabilistic model.

Some statistical procedures have been reported in literature for the last few years related to copying answers. To detect the act of copying answers, some methods are developed, such as Baird Index, Dickenson Index, Anikeef Index (Anderman & Murdock, 2007, pp. 263–264), A, B, and H Index developed by Angoff (1972), Pair 1 method and Pair 2 method developed by Ferry, Tidman, & Wats (1977), G2 method developed by Hanson, Haris, & Brenan (1987), *Error Similarity Analysis Index (ESA)* and *Modified Error Similarity Analysis (MESA)* developed by Bellezza dan Bellezza (Anderman & Murdock, 2007, p. 267) and some other methods to detect the act of cheating.

Based on the above information, what is meant by cheating is every form of dishonest behavior in cheating done by a person to gain success in doing the academic

assignment, especially related to examination. The above facts showed that cheating was an interesting problem to be analyzed further. Therefore, through this study, the researcher examined some methods to detect the possibility of the students doing cheating or sometimes called collusion through observing the response patterns from the examinees.

This study aimed to describe: (1) the characteristics of the mathematics national examination items based on Item Response Theory, (2) the cheating level happening in the administration of the mathematics national examination for senior secondary school students in the academic year of 2015/2016 in East Nusa Tenggara Province based on Angoffs B-Index, the Pair 1 method, the Pair 2 method, Modified Error Similarity Analysis (MESA), and the G2 method, and (3) the most accurate method to detect the act of cheating in the administration of the mathematics national examination for senior secondary school students in the academic year of 2015/2016 in East Nusa Tenggara Province.

## Method

This study belongs to quantitative research using the ex-post facto approach. This study was carried out in East Nusa Tenggara Province. The data analysis was done in the computer laboratory of the Graduate Program, Yogyakarta State University after obtaining the sample response data from *Puspendik, Kemendikbud*, Jakarta. This study was carried out in March 2017.

The target of this study was the senior secondary school mathematics test in the academic year of 2015/2016 Package 3324. The data used in this study were the students' responses in the national examination in East Nusa Tenggara Province, especially on mathematics involving 3,233 students.

The procedures of the study were as follows: (1) estimating the fitness of the data and the model used using the software Bilog MG version 3.0; (2) analysing the items using the Item Response Theory to find out the

item parameter, the examinee parameter, the test information function (the scale taken was in the range between -4.0 and 4.0 with an interval of 0.25) and the standard error of measurement, (3) detecting the act of cheating using five methods, i.e. the Angoffs B-Index method, the Pair 1 method, the Pair 2 method, the Modified Error Similarity Analysis (MESA), and the G2 method with the help of *Integrity* software, (4) determining the method which could detect most pairs suspected to have done cheating, (5) determining the accuracy of the five methods in detecting the act of cheating through the standard error.

The data used in this study were secondary data in the form of examinees' responses on mathematics national examination for senior high schools in East Nusa Tenggara obtained from the *Puspendik, Kemendikbud*, Jakarta. The number of responses was 3,233 students doing package 3324. The data were raw data ABCDE.

The data collection technique was documentation. The data analysis technique was the quantitative data analysis technique. There were two stages in the data analysis, i.e. the item analysis using the Item Response Theory on the senior secondary school mathematics national examination in the academic year of 2015/2016 and the analysis of cheating using the Angoffs B-Index method, the Pair 1 method, the Pair 2 method, the Modified Error Similarity Analysis (MESA) method, and the G2 method.

The analysis of the test item in this study used the Item Response Theory using the Bilog-MG program. This analysis resulted in item information in accordance with the Item Response Theory used. Then, using the 1 Logistic Parameter, the estimation of the difficulty index was obtained. Using the 2 Logistic Parameter, the difficulty index and discrimination index were obtained. Using the 3 Logistic Parameter, the information on the difficulty index, discrimination index, and the artificial guessing. (DeMars, 2010, p. 34) argued that for 1 Logistic Parameter model, as many as 100 or 200 students could be used. For 2 Logistic Parameter model and

3 Logistic Parameter model, there must be more than 500 students used.

In addition to the item parameter, the goodness of fit statistics was also obtained. The model used for estimating the parameter was the logistic model which could accept many item fits. An item which does not fit is the one with a chi-square value higher than the chi-square critical value. On the other hand, an item which fits is the one with a chi-square value lower than the chi-square critical value or the probability value is more than 0.05.

In addition, there was also another information obtained in this analysis, that is, the test information function and the standard error of measurement (SEM). The value of the test information function was calculated using the excel program. The above criteria were used to determine the good item quality using the Item Response Theory. The criteria were as follows: (1) the discrimination index was in the range between 0 and 2, and (2) the difficulty index was in the range between -2 and +2, and (3) the guessing index was around 0.5.

The analysis of the act of cheating in this study used the help of *Integrity*<sup>TM</sup> software. This software is a safe online application which is designed to analysis multiple choice examination. This software uses five methods, that is, the Angoffs B-Index method, the Pair 1 method, the Pair 2 method, the Modified Error Similarity Analysis (MESA) method, and the G2 method. There were two packages which can be utilized by clients in analysing the act of cheating, i.e. using the free trial package and the purchase license model. This study used the free trial package.

## **Findings and Discussion**

### **The Item Characteristics based on the Item Response Theory**

The test item empirical analysis in this study used the Item Response Theory approach using the Bilog-MG software. Before the item analysis was done, a test on the model fitness was done. It was aimed to find

out the model which fitted with the students' response characteristics.

To know the fitness of the model, a criterion of having the probability value of  $>0.05$  was used. The result of the model fitness test can be seen in Table 1.

Table 1. The Item Fitness to 1-PL, 2-PL, 3-PL models

Parameter Logic	The model fit item	Total
1 PL	12, 29	2
2 PL	12, 23, 24, 31	4
3 PL	5, 12, 23, 25, 29	5

The test of model fitness showed that the senior secondary school mathematics examination in East Nusa Tenggara Province in the academic year of 2015/2016 for package 3324 fitted with the 3 Logistic Parameter model. This is because the 3 Logistic Parameter model could detect most items fitting the model. The number of the items fitting the 3 Logistic Parameter model was 5 items so that the item parameter characteristics analysis used the 3 Logistic Parameter model. The five items were item number 5, 12, 23, 25 and 29.

The next analysis was finding out the item characteristics using the 3 Logistic Parameter model. The good difficulty index criteria were in the range between -2 and 2. When an item has a difficulty index lower than -2, this item is considered as a difficult item, while if the difficulty index is more than 2, the item is considered as easy.

The item characteristics of the mathematics national examination package 3324 viewed from the difficulty level was that there were 33 (82.5%) items belonging to the difficulty at the medium level. The rest or 7 items (17.5%) belonged to the difficult items. The seven items were item no 4, 16, 21, 24, 30, 31, 33, dan 36. However, there were many items belonging to the difficulty of the medium level. Meanwhile, the mean of the difficulty level was 2.565. Therefore, it could be concluded that the mathematics national examination package 3324 for senior sec-

ondary school students in East Nusa Tenggara Province in the academic year of 2015/2016 belonged to the difficult test. This might cause students with low ability to find difficulty to do the test. This means that there were seven items which could only be answered correctly by those with high ability.

Table 2. The Item Difficulty Index of the Mathematics National Examination Package 3324

Difficulty level	Item no	Total
Easy	-	-
Medium	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 18, 19, 20, 22, 25, 26, 27, 28, 29, 32, 34, 35, 37, 38, 39, 40	31
Difficult	4, 16, 21, 23, 24, 30, 31, 33, 36	9

The criteria related to the discrimination index are that the item discrimination index was in the range of 0 and 2. An item having a discrimination index of more than 2 indicated that the item had a good discrimination index. Meanwhile, a lower discrimination index showed contrary information. The result of the discrimination index is presented in Table 3.

Table 3. Mathematics National Examination Package 3324 Discrimination Index

Discrimination index	Item no	Total
Good	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 18, 19, 21, 23, 24, 25, 26, 29, 30, 31, 32, 33, 35, 36, 38	30
Poor	11, 17, 20, 22, 27, 28, 34, 37, 39, 40.	10

Viewed from the characteristics of discrimination index parameter, there were 30 (75%) items belonging to the good category. The rest (10 items or 25%) belonged to

the poor category. The ten items were item no 11, 17, 20, 22, 27, 28, 34, 37, 39, and 40. The mean of the discrimination index was 1.543, which means that the mathematics national examination package 3324 for the senior secondary school students in East Nusa Tenggara Province in the academic year of 2015/2016 had good discrimination index. This means that the test could distinguish students with high and low abilities, while ten items could not distinguish students with high and low abilities.

The criteria for the artificial guessing were in the range between 0 and 0.20. A high artificial guessing index shows that it is highly possible that the examinee guesses the answer correctly while a low artificial guessing index indicates that there is a small possibility that the examinee guesses the answer correctly. The result of the analysis of the artificial guessing can be seen in Table 4.

Table 4. Artificial Guessing of the Mathematics National Examination Package 3324

Artificial guessing	Item no	Total
Good	1, 2, 3, 4, 5, 6, 8, 9, 13, 14, 15, 16, 17, 18, 19, 20, 21, 23, 24, 26, 27, 29, 30, 31, 32, 33, 35, 36, 38, 40	30
Not good	6, 10, 11, 12, 22, 25, 28, 34, 37, 39	10

Viewed from the parameter of the artificial guessing, there were 30 (75%) items belonging to the good category. The rest (10 items or 25%) belonged to the poor category, i.e. item no 6, 10, 11, 12, 22, 25, 28, 34, 37, and 39. The mean of the artificial guessing index was 0.135. Viewed from the mean of the artificial guessing index, the mathematics national examination package 3324 for the senior secondary school students in East Nusa Tenggara Province belonged to the good category. This means that the possibility of the examinees with low ability would be able to answer correctly ten items while the other 30 items which could be

answered correctly without the influence of the guessing factor.

The criteria for the good items for the 3 logistic parameter must meet three requirements, that is, having a good difficulty index with a range between -2 and 2, having discrimination index in the range between 0 and 2, and having good guessing index with a value of less than 0.2. From the result of the analysis, there were 17 items which were categorized as good. The summary of the item analysis result can be seen in Table 5.

Table 5. The result of the Item Characteristics with 3 Logistic Parameter Model

Item	a	b	c	Category
1	1.669	0.32	0.14	Good
2	0.616	0.674	0.001	Good
3	1.17	0.797	0.137	Good
5	1.004	1.436	0.152	Good
7	0.93	-0.012	0.127	Good
8	1.245	1.159	0.193	Good
9	1.588	0.065	0.024	Good
13	1.415	0.482	0.161	Good
14	1.183	0.992	0.166	Good
15	1.751	1.067	0.053	Good
18	0.903	1.026	0.085	Good
19	0.626	1.608	0.165	Good
26	1.044	1.124	0.073	Good
29	1.32	1.811	0.167	Good
32	1.415	0.482	0.133	Good
35	0.981	1.098	0.143	Good
38	0.978	0.646	0.096	Good

The item information function showed the whole power or contribution of the item in measuring the latent trait measured using the test set. The test information function is the sum of the item information functions. The item information function and the test information function are closely related to each other, and, so the test information function will be high when the item information function is also high. A test with high information function will result in the small measurement error so that it provides high

contribution to the information function value in revealing the measurement result.

The item estimation using the 3 logistic parameter model on the mathematics national examination package 3324 resulted in the information function value of 58.0128 with a measurement error of 0.1313. The value was obtained when the students' ability parameter ( $\theta$ ) was 1.6. The measurement error in the mathematics national examination package 3324 for the senior secondary school students in East Nusa Tenggara Province in the academic year of 2015/2016 was 0.1313. The measurement error could happen randomly and systematically. Random measurement error was caused by the physical and mental conditions of the examinees and the very difficult materials. Meanwhile, the systematic error measurement was caused by the measurement instrument, that which is measured and that doing the measurement. The test information function value graph with the parameter of examinees' parameter is presented in Figure 1.

The summary of the analysis result of the mathematics national examination test package 3324 for senior secondary school students in the academic year of 2015/2016 in East Nusa Tenggara Province can be seen in Table 9. The summary of the analysis result can be in the form of test item parameter characteristics, the ability characteristics, the information function, and the measurement error.

Table 6. The Item Characteristics Result using the 3 Logistic Parameter Model

Theory	Characteristics
IRT	The mean of the the item difficulty level 2.565
	The mean of the discrimination index 1.543
	The mean of the artificial guessing index 0.135
	The test maximum information function was 58.0128 at the ability of 1.6
	SEM = 0.1313

### Cheating Analysis

The analysis of the act of cheating in this study used five methods, that is, the Angoffs B-Index method, the Pair 1 method, the Pair 2 method, the Modified Error Similarity Analysis (MESA) method, and the G2 method. The analysis was done in 22 districts or municipalities in East Nusa Tenggara Province. The findings showed that students in 14 districts were indicated to do cheating. The 14 districts were Kupang Municipality, Kupang district, Timur Tengah Selatan district, Manggarai district, Sumba Barat Daya district, Timur Tengah Utara district, Manggarai Barat district, Manggarai Timur district, Malaka district, Belu district, Rote Ndao district, Sabu Raijua district, Lembaya district, and Flores district.

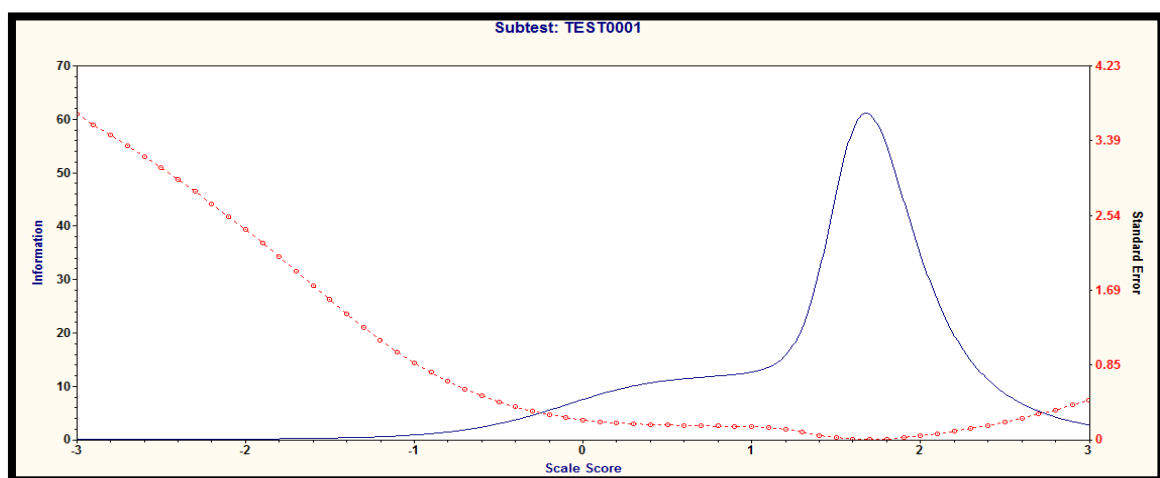


Figure 1. The Information Function Graph, the Ability Scores, and the Standard Error

There was no indication of the act of cheating in 8 districts, that is, Ende district, Sumba Timur district, Sumba Tengah district, Nagekeo district, Sikka district, Sumba Barat district, and Ngada district. The summary of the whole analysis result of the pairs doing cheating in every district/municipality in East Nusa Tenggara Province is presented in Table 7.

Table 7. The Number of Students Doing Cheating in Every District/Municipality in East Nusa Tenggara Province

District/municipality	Number of pairs
Kota Kupang	52
Kabupaten Kupang	63
TTS	67
TTU	4
Belu	8
Malaka	5
Rote Ndao	2
Lembata	1
Sabu Raijua	6
Manggarai	16
Flores Timur	1
Manggarai Barat	7
Manggarai Timur	45
Sumba Barat Daya	17
Jumlah	294

Table 5 presents the frequency of the whole pairs indicated to do cheating.

From the table, it can be seen that the highest number of pairs doing cheating was found in Kupang municipality, Kupang district, and Timur Tengah Selatan district.

In addition, the analysis result of the cheating detection for the whole pairs indicated to have done cheating based on each method, that is, the high category, the medium category, and the low category, is presented in Table 7.

The number of the pairs doing cheating as detected by the G2 method was 183 pairs. This was the highest number of pairs detected. The G2 method detected the cheating based on the number of correct

answers and the number of incorrect answers from the pairs suspected to be involved in the act of cheating. In the next order was the Pair 2 method. This method could detect 141 pairs doing cheating. This method did the detection by looking at the number of the items having the same answers on successive items. In the third order was the MESA method which detected 67 pairs detected as doing cheating. This means that the probability of the pairs to choose the wrong answers in the same test items was as many as 67 pairs.

Table 8. The Pairs Indicated to do Cheating by Methods

Category	The method detecting Cheating				
	B-Index	Pair 1	Pair 2	MESA	G2
High	7	0	57	0	26
Medium	15	6	32	5	76
Total	41	46	52	62	81
Total	63	52	141	67	183

In the fourth order, the method which could detect a great number of pairs was the B-Index method. This method could detect 63 pairs. In detecting pairs doing cheating, this method made a comparison the number of incorrect identical responses to all pairs included in the same interval from the test result. The interval here was based on the multiplication of the number of incorrect responses for 63 pairs of the examinees. Meanwhile, the Pair 1 method was the method which could detect the least number of pairs doing cheating, that is, 52 pairs. This means that the response copying level in the closest pairs was as many as 52 pairs.

The analysis result of the pairs indicated to have done cheating based on the closeness of the seating can be seen from the students' number in one school as presented in the output of the *Integrity*<sup>TM</sup> software. The analysis result of the pairs indicated to have done cheating can be seen in Table 9.



Table 9. The Analysis Result of *Cheating* Based on Seating Closeness

Category	<i>Cheating</i> detection method				
	B-Index	Pair 1	Pair 2	MESA	G2
High	3	1	52	0	29
Medium	4	6	28	5	73
Low	10	41	28	58	74
Total	17	48	108	63	176

Table 7 presented that most pairs were indicated doing cheating based on seating closeness. The G2 method detected the highest number of pairs, followed by the Pair 2 method, the MESA method, the Pair 1 method, and B-Index method successively. The high number of pairs was detected by the G2 method because this method assumed that it is possible for each student to have his or her own response alternative when doing a test.

The analysis result also showed the possibility of cheating happening among pairs from different places or different schools in one district. This indicated that it was possible for some parties to distribute the answer keys in different schools located in districts or municipalities in East Nusa Tenggara Province based on benchmark groups. The result of the analysis is presented in Table 10.

The information obtained from Table 13 was the big number of pairs indicated to have done cheating based on the seating closeness. If this is compared to Table 7, the number of pairs indicated to have done cheating based on seating closeness had the highest frequency compared with the benchmark group. This means that the pairs indicated to have done cheating in the administration of the mathematics national examination were dominated by the behaviour of copying others' work based on seating closeness. Although the number of the cheating behaviour based on benchmark group was small, there was an indication of the distribution of answer key to the examinees.

Table 10. The Analysis Result of the Cheating Behaviour based on the Benchmark Group

Category	<i>Cheating</i> detection method				
	B-Index	Pair 1	Pair 2	MESA	G2
High	4	0	5	0	0
Medium	11	0	4	1	1
Low	31	4	24	4	7
Total	46	4	33	5	8

The result of the analysis also provided information about which method could detect the cheating behaviour accurately viewed from the standard error. The computation of the standard error in this study used the standard deviation from each method. The method with the smallest standard error was the most accurate method. The result of the computation is presented in Table 9.

Table 11. The Accuracy of the Cheating Detecting Method

Method	SD	Sample	SE
<i>B-Index</i> s	0	270888	0
<i>Pair 1</i>	245.261	270888	0.472
<i>Pair 2</i>	557.783	270888	1.072
<i>MESA</i>	33.167	270888	0.064
<i>G2</i>	13.2451	541776	0.018

Table 11 presents the information about the most accurate method in detecting the cheating behaviour. The most accurate method was the B-Index. There was no standard error in this method because there was no standard deviation in this method. The next method was the G2 method because the standard error was only 0.018. The next method was the MESA with the standard error of 0.064. The next method was the Pair 1 method with the standard error of 0.472 and the last is the Pair 2 method with the standard error of 1.072.

As a whole, the cheating detection method analysis is presented in the following figure (Figure 6). In the figure, it can be seen

that the number of pairs detected by the Angoff's B-Index method was 63 pairs, that by the Pair 1 method was 52 pairs, that by the Pair 2 method was 141 pairs, that by the MESA method was 67 pairs, and that by the G2 method was 183 pairs. The pairs here were those sitting close to each other and those pairs who were not at the same place during the administration of the examination but they did cheating were included in the benchmark group to suspect that there was an indication of the distribution of the answer key.

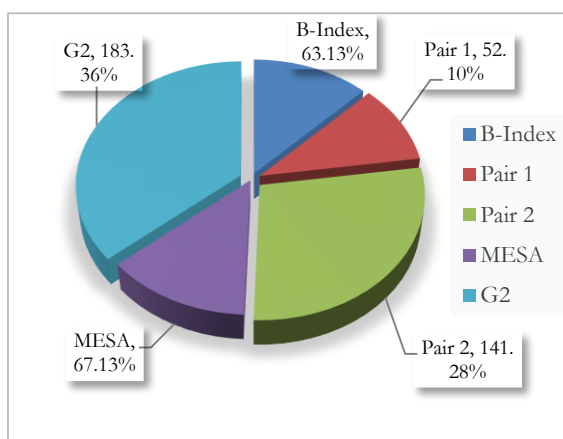


Figure 2. Figure 2. The whole pairs indicated to have done cheating

In the theory of measurement in educational field, the cheating behaviour is not expected to happen because the cheating behaviour will provide inaccurate information about one's ability. In addition, the cheating behaviour is not fair for other students who do not do cheating. As a result, the students who do not do cheating felt that they were treated unfairly. The cheating behaviour also influence the students' lives in the community. This will result in dishonest generation. Ironically, after doing a crosscheck by looking at the response pattern of a pair of students sitting close to each other, that is, student no 170 and student no 172, it was found that the two students had similar answers from item no 1 to the last item. This also indicated that the proctor did not remind the students not to do cheating in this school.

The essence of the problem of cheating behaviour in the administration of the

mathematics national examination for senior secondary school students in the academic year of 2015/2016 in East Nusa Tenggara Province was not only because the scores were used as the graduation criteria, but because of the test form which made it possible for the examinees to do cheating. The test form which is sensitive to the cheating behaviour is the multiple-choice test. Many people say that most examinees seem to copy others' work in the examination because it is easy to do this. The examinees can easily work together with other examinees through using symbols found in the answer sheet. Those symbols facilitate other examinees to do cheating because it does not take a long time to get the answer, let alone with the advancement of technology.

### Conclusion

Based on the result of the analysis and the discussion on the mathematics national examination package 3324 for the senior secondary school students in the academic year of 2015/2016 in East Nusa Tenggara Province, it can be concluded: (1) the analysis result based on the Item Response Theory showed that out of 40 items analysed using the 3 logistic parameter model, there were only 17 (42%) items considered as good with a test information function of 58.0128 at the students' ability ( $\theta$ ) of 1.6 and SEM of 0.1313; (2) based on the analysis, there were 63 pairs detected using the Angoff's B-Index method, 52 pairs detected using the Pair 1 method, 141 pairs detected using the Pair 2 method, 67 pairs detected using the MESA method, 183 pairs detected using the G2 method; (3) based on the analysis result, the method which could detect from the highest number of pairs to the lowest number of pairs doing cheating in the administration of the mathematics national examination in East Nusa Tenggara Province in the academic year of 2015/2016 was the G2 method, the Pair 2 method, the MESA method, the Angoff's B-Index method, and the Pair 1 method successively; (4) based on the computation of the standard error, the most accurate method being able

to detect cheating in the administration of the mathematics national examination in East Nusa Tenggara Province in the academic year of 2015/2016 was the Angoff's B-Index method, the G2 method, the MESA method, the Pair 1 method, and the Pair 2 method successively.

Based on the study, there are some suggestions related to the mathematics national examination test set package 3324 for senior secondary school students in East Nusa Tenggara Province as follows: (1) For the government, it is suggested that the test items developed should be developed so as to meet the good standard so that it can provide accurate information about students' ability. It is suggested that in the future, the test be tried out in remote areas; (2) For schools and teachers, it is suggested that they instill honesty in doing the national examination because cheating behaviour is not an alternative in achieving good scores but through implementing innovative teaching systems, effective teaching methods so that cheating behaviour can be avoided; (3) experts in the educational field are expected to develop accurate detecting method based in good criteria so that the result of the cheating detection can be more accurate.

## References

- Anderman, E., & Murdock, T. (2007). *Psychology of academic cheating*. Cambridge: Academic Press.
- Angoff, H. W. (1972). *The development of statistical indices*. New Jersey, Berkeley: Educational Testing Service Princeton.
- Badan Standar Nasional Pendidikan. (2015). *Buku saku tanya jawab pelaksanaan UN tahun 2016*. Jakarta: Badan Standar Nasional Pendidikan.
- Baird, J. S. (1980). Current trends in college cheating. *Psychology in the Schools*, 17(4), 515–522.  
[https://doi.org/10.1002/1520-6807\(198010\)17:4<515::AID-PITS2310170417>3.0.CO;2-3](https://doi.org/10.1002/1520-6807(198010)17:4<515::AID-PITS2310170417>3.0.CO;2-3)
- Bernardi, R. A., Baca, A. V., Landers, K. S., & Witek, M. B. (2008). Methods of cheating and deterrents to classroom cheating: an international study. *Ethics & Behavior*, 18(4), 373–391.  
<https://doi.org/10.1080/10508420701713030>
- Bernardi, R. A., Banzhoff, C. A., Martino, A. M., & Savasta, K. J. (2012). Challenges to academic integrity: identifying the factors associated with the cheating chain. *Accounting Education*, 21(3), 247–263.  
<https://doi.org/10.1080/09639284.2011.598719>
- Biro Komunikasi dan Layanan Masyarakat Kementerian Pendidikan dan Kebudayaan. (2016). *Indeks integritas ujian nasional (IUN) SMA 2016 meningkat*. Jakarta: Kementerian Pendidikan dan Kebudayaan.
- Bopp, M., Gleason, P., & Misicka, S. (2001). *Reducing incidents of cheating in Adolescence*. Master's thesis. Saint Xavier University.
- Brimble, M., & Clarke, P. S. (2005). Perceptions of the prevalence and seriousness of academic dishonesty in Australian universities. *Australian Educational Researcher*, 32, 19–44.
- Calabrese, R. L., & Cochran, J. T. (1990). The relationship of alienation to cheating among a sample of American adolescents. *Journal of Research & Development in Education*, 23(2), 65–72.
- Cizek, G. J. (2001). An overview of issues concerning cheating on largescale tests. In *The Annual Meeting of the National Council on Measurement in Education*. Seattle, WA.
- Davis, S. F., Drinan, P. F., & Gallant, T. B. (2009). *Cheating in school; what we know and what we can do*. United Kingdom: Wiley-Blackwell.
- Davis, S. F., Grover, C. A., Becker, A. H., & McGregor, L. N. (1992). Academic dishonesty: prevalence, determinants, techniques, and punishments. *Teaching of Psychology*, 19(1), 16–20.  
<https://doi.org/10.1207/s15328023to>

p1901\_3

- DeMars, C. (2010). *Item response theory; understanding statistics measurement*. New York: Oxford University Press.
- Ferry, B. R., Tidman, M. T., & Wats, M. T. (1977). Indices of cheating on multiple-choice tests. *Journal of Educational Statistics*, 2(4), 235–256.
- Hanson, A. B., Haris, J. D., & Brenan, L. R. (1987). *A comparison of several statistical methods for examining allegations of copying*. Iowa: American College Testing.
- Hughes, J. M. C., & McCabe, D. L. (2006). Academic misconduct within higher education in Canada. *Canadian Journal of Higher Education*, 36(2), 1–21.
- Jensen, A. L., Arnett, J. J., Feldman, S. S., & Cauffman, E. (2002). It's wrong, but everybody does it: Academic dishonesty among high school and college students. *Contemporary Educational Psychology*, 27, 209–228.
- Lin, C. S., & Wen, L. M. (2007). Academic dishonesty in higher education: A nationwide study in Taiwan. *Higher Education*, 54, 85–97.
- Manoppo, Y., & Mardapi, D. (2014). Analisis metode cheating pada tes berskala besar. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 18(1). Retrieved from <https://journal.uny.ac.id/index.php/jpep/article/view/2128>
- Mardapi, D. (2000). Evaluasi penyelenggaraan ebtanas. *Jurnal Kependidikan*, XXX(2).
- Michaels, J. W., & Miethe, T. D. (1989). Applying theories of deviance to academic cheating. *Social Science Quarterly*, 70(4).
- Mulyati, B., & Kartowagiran, B. (2013). Analisis hasil ujian nasional mata pelajaran ekonomi SMA di kota Serang. *Jurnal Evaluasi Pendidikan*, 1(1). Retrieved from <http://journal.student.uny.ac.id/ojs/index.php/jep/article/view/45>
- Naghdi-pour, B., & Emeagwali, O. L. (2013). Students' justifications for academic dishonesty: call for action. *Procedia - Social and Behavioral Sciences*, 83, 261–265. <https://doi.org/10.1016/j.sbspro.2013.06.051>
- Newstead, S. E., Franklyn-Stokes, A., & Armstead, P. (1996). Individual differences in student cheating. *Journal of Educational Psychology*, 88(2), 229–241. <https://doi.org/10.1037/0022-0663.88.2.229>
- Presiden Republik Indonesia. Peraturan Pemerintah RI Nomor 13, Tahun 2015, tentang Perubahan Kedua atas Peraturan Pemerintah Nomor 19 Tahun 2005 tentang Standar Nasional Pendidikan (2015).
- Rakovski, C. C., & Levy, E. S. (2007). Academic dishonesty: Perceptions of business students. *College Student Journal*, 41, 466–481.
- Stogner, J. M., Miller, B. L., & Marcum, C. D. (2013). Learning to E-Cheat: a criminological test of internet facilitated academic cheating. *Journal of Criminal Justice Education*, 24(2), 175–199. <https://doi.org/10.1080/10511253.2012.693516>
- Vandehey, M. A., Diekhoff, G. M., & LaBeff, E. E. (2007). College cheating: A twenty-year follow up and the addition of an honor code. *Journal of College Student Development*, 48, 468–480.
- Whitley, B. E., Nelson, A. B., & Jones, C. J. (1999). Gender differences in cheating attitudes and classroom cheating behavior: A meta-analysis. *Sex Roles*, 41(9), 657–680.
- Zastrow, C. (1970). Cheating among college graduate students. *The Journal of Educational Research*, 64(4), 157–160. Retrieved from <http://www.jstor.org/stable/2753609>

## DEVELOPING SCALE TO MEASURE STUDENT'S HONESTY CHARACTERS ON MATHEMATICS LEARNING USING SUBJECT SCALING

*Kana Hidayati*<sup>1</sup> \*, *Budiyono*<sup>2</sup>, *Sugiman*<sup>1</sup>

<sup>1</sup>Universitas Negeri Yogyakarta, <sup>2</sup>Universitas Sebelas Maret

<sup>1</sup>Jl. Colombo No. 1, Depok, Sleman 55281, Yogyakarta, Indonesia

<sup>2</sup>Jl. Ir. Sutami No.36 A, Pucangsawit, Jebres, Kota Surakarta, 57126, Jawa Tengah Indonesia

\* Corresponding Author. Email: [kana@uny.ac.id](mailto:kana@uny.ac.id)

### Abstract

The purpose of this study is to develop a scale to measure honesty characters of students on mathematics learning using subject scaling model. This development research is carried out in several steps namely: preparing initial scales, trials, and interpretation of trials result. The research is conducted on students of SMP N 5 Yogyakarta, SMP Muhammadiyah 2 Yogyakarta, and SMPIT Abu Bakar Yogyakarta. The first trial involves 688 students and the second trial involves 696 students. Content validity is obtained through expert judgment. The construct validity is determined by Lisrel's second order CFA. Reliability coefficients include Omega coefficients ( $\omega$ ) and Alpha coefficients from Cronbach. Based on the results of this research, the scale to measure of student's honesty characters consists of 15 items that theoretically and empirically proved to be of good quality. The scale has good content validity supported by V Aiken index of 0.88-0.92. The scale has good construct validity. Reliability of the scale is also good supported by both the Omega coefficient of 0.86.

**Keywords:** *honesty character, mathematics, subject scaling*

Permalink/DOI: <http://dx.doi.org/10.21831/pep.v22i2.19310>

## **Introduction**

National Education System Indonesia strongly emphasizes education that is able to make Indonesian people not only have intellectual intelligence but also noble character with good spirituality, personality, and self-control. Therefore, character education is highly emphasized in Indonesian education system and continues to undergo reforms following the development of the era.

The implementation of Curriculum 2013 in Indonesia that emphasizes character education and the establishment of the Character Education Strengthening program in schools are several forms of government commitment in order to strengthen the noble character that Indonesian students must possess. It is in line with the statement of Minister of Education and Culture Muhadjir Effendy who mentioned that the portion of science education for junior high school education is 40% and character education by 60% (Maharani, 2016).

According to Lickona (1992), good character education emphasize on three groups of good character namely moral knowing, moral feeling, and moral action. Good character education contains knowledge which further raises the commitment to be a good person and finally really embodied in doing good behavior. According to the taxonomy of (Krathwohl, Bloom, & Masia (1964), this thing corresponds to the affective domain. It becomes the benchmark of attitude competence in the Curriculum 2013 namely receiving, responding, valuing, organization, and characterization by value.

Character education in learning mathematics can be integrated in the form of attitude assessment. That is, through the assessment of attitudes in learning mathematics, character of students can be measured and the results will be very helpful for teachers in order to strengthen good character of the students. Kumaidi (2014) suggest that the Curriculum 2013 requires teachers not only assess the mastery of teaching materials but also include the development of affection such as faith, honesty, discipline, and other good character.

One of the emphasized good character that must be owned by Indonesian students is honesty. Strengthening the honesty character of the students can be integrated into all subjects in school including Mathematics. As one of the material in learning mathematics, statistics is closely related to data in the field. Honesty characters can be developed and strengthened through the learning of statistical materials that are directly applicable in everyday life. Strengthening this honesty character certainly requires accurate information about the actual condition of the honest character of the student. Therefore, the measurement of honest character of students must be done using instruments that are reliable and consistent so that the information obtained from it is truly accurate and can be accounted for.

Measurement of honest character in mathematics learning using the right measurement scale is certainly very helpful for the teacher in assessing the attitude or affective aspects of students. According to Popham (2009, p. 107), affective assessment of students when carried out during school learning activities can contribute to students' behavior after graduation. This means that the measurement and assessment of the honest character of students during school greatly influences student's honest behavior after graduation. Through the measurement and assessment of the honest character done during school, of course, making honest character will be embedded more strongly in students so that it is still carried away even after graduation.

According to the Ministry of Education and Culture (Kementerian Pendidikan dan Kebudayaan Republik Indonesia, 2013), honesty is a behavior based on the effort to make oneself as a person who can always be trusted in words, actions, and work. Asmani (2011) also states that honesty is a behavior that makes oneself as a person who can always be trusted that is manifested in words and actions, both for themselves and for others. In other words, honesty is the character which makes people to act as they are in accordance with the truth, not made up,

not lying, not reduced, not added to, and not hiding information. If someone says something that it is not in accordance with the truth and the actual reality or does not acknowledge something according to what it is, then that person can be said to be dishonest. Honesty is defined as being honorable in principles, intentions, and actions; being upright, fair, sincere, frank, respectable, and truthful; and free from deceit or fraud (Cherrington & Cherrington, 1979, p. 645). State that in line with those opinion, Bisri (2009) also state that honesty means saying or acting according to what is actually in accordance with the reality as it is without being reduced or added. The opposite of honesty is a lie, that is to say or do something that is not in accordance with the actual reality.

The characteristics of honest behavior according to Emosda (2011) are that their behavior is followed by a pure or sincere heart, say something according to reality, and act according to the evidence and truth that exists. The indicators of honesty character in education according to the Ministry of Education and Culture (Kementerian Pendidikan dan Kebudayaan Republik Indonesia, 2015) includes: (a) not cheating in doing tests, (b) not doing plagiarism (copying or taking the work of others without mentioning the source), (c) expressing feelings as they are, (d) submit the findings to those who have the right or authority, (e) make reports based on information or data as they are, (f) acknowledge the shortcomings or errors owned, and (g) convey information according to the facts. Koellhoffer (2009, p. 4) also stated that honesty in academics means being honest in educational regulations. Someone is said to be academically honest if he does not commit acts of plagiarism. That is, not using someone else's work or not copying someone else's work without his permission. According to Koss (2011, p. 39), academic dishonesty is a behavior or action of someone who borrows and copies assignments from other people, including copying answers at the time of the exam. Other academic dishonesty behaviors are writing

test answers on feet or hands, clothes, walls, paper, tables or chairs, and exchanging answers with friends using certain codes.

According to Kibler (Ercegovac & Richardson, 2004, p. 301), academic dishonesty is a form of plagiarism and fraud committed by students in the form of receiving or giving unauthorized assistance in examinations and receiving awards from another person works. Other dishonest behavior in academic activities is that students take advantage of the opportunities available to exchange answers with their friends or cheat, for example when the examiner exits the room. Schmelkin, Gilbert, Spencer, Pincus, & Silva (2008, p. 4) argued that forms of academic dishonesty are plagiarism, imitating each other with exams during tests, stealing tests or questions, and buying answer papers. Dishonest behaviour has severe negative consequences for institutions and students. The academic dishonesty violates scientific principles, as well as study and assessment regulations (Bashir & Bala, 2018, p. 57). Academic dishonesty is clearly an unjustified act in the education system in Indonesia. Based on the various opinions above, it can be concluded that honesty character encompasses several components, namely behaving as they are according to reality, recognizing any positive or negative actions, doing assignments or examinations according to their own abilities, and can be trusted for the ownership of others.

During this time the measurement and assessment of the achievement of competencies for aspects of attitudes related to the level of honest character students in mathematics learning is generally done by observation, interviews, self-assessment, and peer assessment. The form of instruments used mostly in the form of observation sheets, interview guidelines, and questionnaires in the form of response scaling. Character measurement is basically a measurement of an individual's personality so that it requires an instrument that can truly distinguish personality between individuals. Therefore, the use of a scale in the form of subject scaling that is more specific and detailed in



measuring the character of students needs to be done. However, until now there have not been many studies on the development of a scale to measure the honest character of students, especially in mathematics learning that uses the subject scaling model. In fact, the instrument in the form of subject scaling has high practical value and is widely used by designers of psychological scale (Azwar, 2005, p. 32). This is also as stated by Punyanunt-Carter (2017, p. 34) that a scale centered on a subject which is also called an individual difference scale is very suitable to be used to measure a person's personality.

Based on the description above, this study aims to develop a scale to measure the honesty character of students in learning mathematics. This scale uses the subject scaling model and the quality is tested theoretically and empirically. The research problem is formulated as follows: (1) how the content validity of the scale is used to measure the honesty character of students in learning mathematics; (2) how the construct validity of the scale is used to measure the honesty character of students in learning mathematics; (3) how the reliability of the scale is used to measure the honesty character of students in learning mathematics.

## **Method**

This research is a development research conducted by several steps: (1) preparing the initial scale, (2) testing, and (3) interpretation of the results of the trial. Arranging the initial scale is the initial activity of developing a scale which is then discussed in Focus Group Discussion (FGD) and followed up with validation through expert judgment. The final validation results are evidence of the validity of the contents of the scale being developed. The next stage is the trial and followed up by the interpretation of the results of the trial, namely determining the evidence of scale quality related to construct validity and scale reliability and determining whether or not revisions and trials are needed.

The FGD activity involved 4 junior high school math teachers and 3 Mathe-

matics education lecturers. Scale validation with expert judgment involves six experts who examine the scale in terms of its content, namely: (a) suitability of items with indicators, (b) language used, (c) item statements that are unbiased, (d) clarity of answer choices, and (e) accuracy of scoring guidelines. Experts also assess the suitability between items with indicators in the form of a Likert scale with five answer choices. In this case, the content of social desirability is overcome by writing the items following the standard rules in the preparation of the scale (Azwar, 2005, p. 9). In addition to qualitative assessments from experts, evidence of content validity is also supported by the amount of content validity index using V Aiken. The way to determine of the validity of an item is to compare the value of V with the value of  $V_{table}$ . The  $V_{table}$  value is the minimum value of the content validity index based on the number of experts in table V in (Aiken, 1985, p. 134). The number of repeaters in this study were six and the number of ratings was five, so the minimum index of content validity based on V Aiken's table was 0.79.

The trial was conducted on seventh grade junior high school students in the city of Yogyakarta who had implemented the 2013 Curriculum. The selection of junior high schools where the trial was conducted was purposive cluster sampling technique, which was a trial conducted at junior high schools that represented high, medium, and developing quality and represented public and private schools. Determination of the quality of schools in this study is based on the results of the National Examination. The schools selected as trial sites were SMP N 5 Yogyakarta, SMP IT Abu Bakar Yogyakarta, dan SMP Muhammadiyah 2 Yogyakarta.

The results of the trial data in this study are quantitative data. Data is obtained based on the students' responses on the scale to measure the honesty character of students. Data collection techniques is carried out by giving the scale developed for students to be filled out by the participants of the trial.

Data analysis begins with test requirements analysis to find out whether the data meets the requirements to be analyzed by the Confirmatory Factor Analysis (CFA) technique, that is, the data to test the model follows a multivariate normal distribution. The criteria used are if the  $p\text{-value} > 0.05$  then the distribution is normal, and if the  $p\text{-value}$  is  $0.05$  then the distribution is not normal (Yamin & Kurniawan, 2009, p. 29).

After testing the requirements analysis, data analysis is carried out to prove the construct validity of the scale through second order CFA with the help of the Lisrel version 8.51 program. Criteria for an indicator item said to be valid in representing the construct is if the  $t\text{-value}$  is  $> 1.96$  and the value of Standardized Loading Factor (SLF) is at least  $0.3$  (Hair, Black, Babin, & Anderson, 2009; Igbaria, Zinatelli, Cragg, & Cavaye, 1997, p. 290). Referring to these criteria, if there is an invalid indicator item, the item must be eliminated and re-analyzed. Based on the results of re-analysis after invalid indicator items are eliminated, what needs to be considered is the compatibility with the model. If the model is not fit, the model is then respecified by modifying the index as suggested by Lisrel.

The suitability of the model in this study refers to Garson (2009) who stated that the suitability of the model developed through empirical data is at least seen from the fulfillment of two of the three different model suitability measurement categories, namely absolute match test, incremental suitability test, and parsimony suitability test. Suranto, Muhyadi, & Mardapi (2014, p. 102) suggested that the developed model is stated to be compatible with the field data if two criteria have been met from the three criteria that become the absolute compatibility measurement, namely Root Mean Square Error of Approximation (RMSEA)  $0.08$ , Chi-square obtained from the test has a probability greater than  $0.05$  ( $p > 0.05$ ) and Goodness of Fit Index (GFI)  $0.90$ . In addition, it also refers to the opinions of (Hooper, Coughlan, & Mullen, 2008, p. 57) which suggests that the use of Chi square is

sensitive to the sample size used. As according to Wijanto (2008, p. 155) suggests that the consensus of the researchers states that Chi square is not the only measurement of Goodness of Fit (GOF) and there is no particular GOF measurement that exclusively represents the overall suitability of the model. Estimation of reliability is determined by the McDonald's composite reliability coefficient or Omega ( $\omega$ ) coefficient. This refers to Widhiarso (2009, p. 47) which states that reliability which begins with CFA using SEM in multidimensional measurement models can be done with Omega reliability coefficient ( $\omega$ ). Reliability coefficient criteria refers to Hair et al. (2009, p. 688), which mentioned that the estimated reliability  $> 0.7$  is good, while the reliability coefficient  $0.6\text{-}0.7$  is acceptable as long as the construct validity indicator of the model is good.

## Findings and Discussion

Based on the results of the initial scale's preparation which was then discussed in the FGD and followed up by validation of experts, a scale to measure the honesty character of students consisting of 17 points is arranged. The formulation of the scale items is compiled based on the components and indicators of honest construct as presented in Table 1.

Based on the results of validation the initial scale to measure honesty character of the students through mathematics learning, which is consisting of 17 items in the form of subject scaling, has good content validity so that it is feasible to use. The feasibility of the scale being developed is also supported by the magnitude of the V Aiken index for the suitability between items with indicators compared with the V value of  $0.79$ . Based on the calculation results, the calculated V value in the whole item is  $0.88\text{-}0.92$ . This shows that all items on the scale to measure the honest character of students, have good content validity in terms of their suitability with the indicators. That is, in terms of its contents, the scale to measure the honesty character of students developed through this

research can be said to measure what should be measured. Top of Form

Table 1. Components and Indicators for Measuring Honesty Character of Students

Component	Indicator	Item number
A. Acting according to reality.	1. Deliver information as it is without adding or reducing the content.	J1, J2
	2. Say the truth whatever the consequences.	J3
	3. Revealing the real situation or not pretending.	J4, J5
	4. Prepare reports based on the data as they are.	J6
B. Admit every action whether it is positive or negative.	1. Admit the mistake that he/she made.	J7
	2. Appreciate their own skill/ capability	J8
	3. Not cover up other people mistakes.	J9
	4. Appreciate other people skill / capability.	J10
C. Do assignment or exam according to their own abilities.	1. Not cheating.	J11
	2. Not expecting help from friends.	J12
	3. Not taking advantage of the opportunity to cheat.	J13
	4. Not exchanging the result of work with friends.	J14
	5. Not copying the work of friends or other people without mentioning the source.	J15
D. Can be trusted about the ownership of others	1. Give recognition to other people's ownership.	J16
	2. Does not damage or harm other people's ownership.	J17

The scale that was declared feasible by the experts was then tested on students of SMP 5 Yogyakarta, SMPIT Abu Bakar Yogyakarta, dan SMP Muhammadiyah 2 Yogyakarta. Trial I involved 688 students. The results of the initial analysis indicate that the data are not multivariate and normally

distributed so that the second order CFA analysis is carried out using the Weighted Least Square (WLS) estimation model.

Based on the results of the test data analysis I obtained the Chi Square=298.98,  $p=0.00$ , RMSEA=0.048, GFI=0.98, and AGFI=0.97. These results indicate that the suitability of the model can be said to be sufficiently fulfilled. It means that the model is sufficiently compatible with the data. The scale to measure the honesty character of students in the first trial consisted of 17 items in the form of subject scaling. The results of the second order CFA for the value of t-value and Standardized Loading Factor (SLF) for the trial I are presented in Table 2.

Table 2 shows that in terms of the t-value, all items are significant in supporting an honesty construct, but there are four items that have SLF values of less than 0.3 namely points J2, J3, J6, and J10. This means that points J2, J3, J6 and J10 need to be reviewed.

Table 2. Trial I Result of Second Order CFA

Comp	Indicator	Item	Second order CFA		Meaning
			t-value	SLF	
A	A1	J1	**	0,36	Good
		J2	2,61	0,14	Not good (must be revised)
	A3	J3	5,03	0,29	Not good (must be revised)
		J4	6,77	0,58	Good
	A5	J5	5,54	0,30	Good
	A6	J6	0,21	0,01	Not good (must be revised)
B	B1	J7	**	0,75	Good
	B2	J8	16,22	0,69	Good
	B3	J9	13,69	0,65	Good
	B4	J10	5,02	0,24	Not good (must be revised)
C	C1	J11	**	0,69	Good
	C2	J12	13,70	0,57	Good
	C3	J13	18,29	0,81	Good
	C4	J14	17,64	0,80	Good
	C5	J15	14,11	0,59	Good
D	D1	J16	**	0,68	Good
	D2	J17	4,95	0,54	Good

Points J2, J3, J6 and J10 each represent one indicator. Therefore, these four items are not eliminated but are corrected or revised. After the revision, a trial II was conducted involving 696 students.

The results of data analysis in trial II showed that the data analyzed was not normally distributed multivariate so that the second order CFA analysis was carried out using the Weighted Least Square (WLS) estimation model. Based on the results of the initial analysis in trial II it was found that there were two items that had SLF values less than 0.3, namely items J2 and J5. Therefore, further analysis is carried out after eliminating items J2 and J5. Further analysis is also carried out by observing the index modifications suggested by Lisrel. Modification of the model as suggested by Lisrel software can be seen in Figure 1.

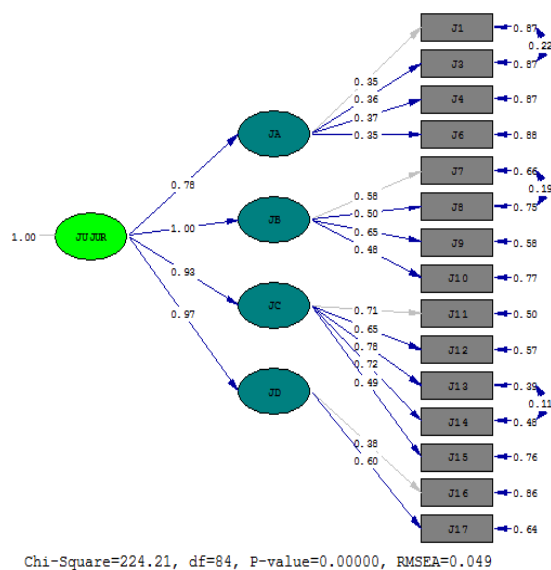


Figure 1. Measurement Model of Scale to Measure Honesty Character

As presented in Figure 1, the modification made by the researcher is to add covariance between two errors in J1 and J2, J7 and J8, J13 and J14. The researchers' considerations that each represents the same latent variable and decreases the chi square value, and it can be supported by strong reason. J1 indicator is to convey information as it is without adding or subtracting and the J2 indicator is to tell the truth whatever the consequences. J7 indicator is to acknowledge

the mistake made and J8 indicator is to give credit to the strengths of self. As for J13 indicator is not taking advantage of the opportunity to cheat and J14 indicator is not exchanging work results with friends. Pay attention to these indicators, the error that occurs in the J1 indicator is clearly very influential on the J2 indicator. Likewise for J7 and J8, J13 and J14. It shows that it is reasonable to add covariance between two errors in J1 and J2, J7 and J8, J13 and J14. Based on the results of the final analysis after modification, Chi Square=224.21,  $p=0.00$ , RMSEA=0.049, GFI=0.98, and AGFI=0.97 were obtained. This shows that the suitability of the model is sufficiently fulfilled, meaning that the model is sufficiently in accordance with the data. The results of the second order CFA for the value of t-value and Standardized Loading Factor (SLF) are presented in Table 3.

Table 3. Trial II Result of Second Order CFA

Comp	Indicator	Item	Second order CFA	
			t-value	SLF
A	A1	J1	**	0,35
	A3	J3	5,60	0,36
	A4	J4	4,64	0,37
	A6	J6	4,11	0,35
B	B1	J7	**	0,58
	B2	J8	10,26	0,50
	B3	J9	12,05	0,65
	B4	J10	8,56	0,48
C	C1	J11	**	0,71
	C2	J12	15,73	0,65
	C3	J13	15,58	0,78
	C4	J14	16,11	0,72
	C5	J15	11,36	0,49
D	D1	J16	**	0,38
	D2	J17	6,42	0,60

Table 3 shows that in terms of t-value, all points are significant in supporting an honest construct. When viewed from the SLF value, it also shows that all items meet the criteria of good construct validity.

The reliability coefficient used is Omega reliability coefficient because honest constructs are multidimensional. Based on the calculation results obtained Omega reliability coefficient of 0.86. These results indi-

cate that the scale for measuring honest characters has good reliability referring to the criteria of Hair et al. (2009, p. 688). This means that in terms of the construct, the scale for measuring honesty character of students can be said to be reliable or consistent.

Based on the results of the trial II, the final results were obtained in the form of a scale to measure the honesty character of students which is theoretically and empirically proved to be good quality. Theoretically, it is proven by the validity of the content of the scale which shows that the content according to the experts is good and the whole item has a good suitability with the indicator. This is also supported by the magnitude of the V Aiken index of 0.88-0.92. Empirically, in terms of its construct validity, all scale items have good support in establishing an honesty construct that is established. As for the reliability review is also included in the good category with the Omega reliability coefficient of 0.85. Thus, the final result of the scale development in this study is to obtain a scale to measure the honest character of students through mathematics learning consisting of 15 items in the form of subject scaling.

Writing instrument items in the form of subject scaling is not easy. However, this form is able to reveal students' affective conditions in more depth. This is as stated by Azwar (2005, pp. 29-32) that the subject scaling aims to put individuals on a continuum so that the relative position of individuals according to an measured attribute can be obtained. In addition, this procedure has high practical value and is widely used by designers of psychological scale. This is due to the fact that in the form of scaling of subjects, questions and choice of answers are presented in the form of statements or narrative.

Here are some examples of items on the scale to measure the honesty character of students through mathematics learning in the form of subject scaling produced in this study namely items J4, J9, J13, and J17. These items are items that have the highest SLF value in each component. The following indicators are represented by each item.

- Item J4 : Revealing the real situation or not pretending (SLF=0,37).
- Item J9 : Not cover up other people mistakes (SLF=0,65).
- Item J13 : Not exchanging the result of work with friends (SLF=0,78)
- Item J17 : Does not damage or harm other people's ownership (SLF=0,60)

The scoring on a scale to measure the honesty character of students is in the form of subject scaling with three choices of answers in this study are: score 2 for the choice of answers that indicate the most honesty character, score 1 for answer choices that indicate less honesty character, and a score of 0 for the answer choices does not indicate honest character. The following is a description of points J4, J9, J13, and J17 along with the scoring guidelines.

#### Item J4:

After explaining the ways to collect data, suddenly the teacher asks your understanding of the material. You have tried to pay attention to the teacher's explanation but have not fully understood the explanation that was just delivered, then you:

- a. Say that you only understand some of it and mention parts that are not understood.
- b. Say that you don't understand at all and beg the teacher to explain more clearly.
- c. Say to the teacher that you already understand and for the material that you don't understand you will just ask your friends .

Scoring guidelines for item J4 are score 2 for choice a, score 0 for choice b, and score 1 for choice c.

#### Item J9:

You see a friend copy your other friends' data when doing individual tasks to collect data on the ideals and hobbies of classmates. One day, suddenly the math teacher called you to the teacher's office

and asked about the actions of your friend who copied the work of another friend, then you:

- a. Say that you don't know what really happened.
- b. Try to cover up your friend's mistakes because you don't want your friend get scolded by the teacher.
- c. Say everything that you know.

Scoring guidelines for item J9 are a score of 1 for choice a, score 0 for choice b, and score 2 for choice c.

Item J13:

When the final exam of mathematics is taking place, suddenly the supervisor comes out of the room and turns out that he hasn't been back in class for a while, then you:

- a. Take advantage of the situation to ask friends just to match the answers to be more confident.
- b. Stay focused on doing the exam eventhough friends start to get noisy because they try to ask each other .
- c. Look around at other friends who ask each other and listen to them.

Scoring guide for item J13 is a score of 0 for choice a, score 2 for choice b, and score 1 for choice c.

Item J17:

Today you borrowed your friend's math notebook because yesterday you didn't attend the class. When you go home from school, you get drenched by the rain and the book you borrow is a little wet and rips a few sheets on the inside. The next day your friend asks for a notebook that you borrow, then what you do is ...

- a. Just be quiet and immediately return it as if nothing happened and will tell only if he found out.

- b. Tell that the book is already damaged when you borrow it.
- c. Tell the damage that occurs and are willing to repair or compensate it.

Scoring guidelines for item J17 are a score of 1 for choice a, score 0 for choice b, and score 2 for choice c.

Through the items with the form of subject scaling, students are expected to provide answers that truly show the true condition of the student's honesty character when facing certain situations as stated in the item. The scale to measure the honesty character of students developed in this study is included in the group of noble characters (good character) at the level of moral feeling (Lickona, 1992). That is, the condition of the character revealed reflects the student's commitment to act honestly. This also corresponds to the taxonomy of the affective domain of Krathwohl, which is until the organization stage. This means that students have formed a value system that is consistent in themselves.

The scale to measure the honesty character of students in mathematics learning produced by this study is very relevant to character education in Indonesia. This scale is certainly very helpful in adding references to teachers and education practitioners. In addition to student affective assessment, this scale can also be used by teachers as a reference for conducting research such as classroom or experimental action research. For other researchers, this scale can be a reference for instruments in the form of psychological scales in the form of subject scaling. This is because this scale is still very rarely developed, especially in the assessment of attitude aspects in mathematics learning.

Scale scoring in this study is still manual, so it requires a relatively long time. Further development can be made in the form of software or presented online so that scoring can be done more easily and quickly. This is like Leachy (2012) who argued that technological development should be used

by teachers in assessment activities because technology can be designed and utilized to provide a rapid response to student work. Further testing activities also need to be carried out in other areas with larger and heterogeneous respondents so that it further broadens generalizations and adds evidence of the quality of the scale produced by this study. Further development of this scale can also be done on the development of a scale to measure other characters which are the priority of character strengthening activities in the education system in Indonesia.

### Conclusion

This study produces a scale to measure the honesty character of students through mathematics learning with a scale component including: acting according to reality, admit every action whether it is positive or negative, do assignment or exam according to their own abilities, and can be trusted about the ownership of others. The resulting scale consists of 15 items in the form of subject scaling which is theoretically and empirically proven to be of good quality. The quality of the resulting scale is as follows: (1) The validity of the content of the scale is good, supported by the suitability of the items with the indicators in all items are valid with V Aiken index of 0.88–0.92, (2) Scale items to measure honesty characters of the student in overall has good construct validity, and (3) the reliability of the scale to measure the honesty character of students is good with the Omega coefficient of 0.86.

### References

- Aiken, L. R. (1985). Three coefficients for analyzing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45(1), 131–142. <https://doi.org/10.1177/0013164485451012>
- Asmani, J. M. (2011). *Buku panduan internalisasi pendidikan karakter di sekolah*. DIVA Press.
- Azwar, S. (2005). *Penyusunan skala psikologi* (1st ed.). Yogyakarta: Pustaka Pelajar.
- Bashir, H., & Bala, R. (2018). Development and validation of academic dishonesty scale (ADS): presenting a multidimensional scale. *International Journal of Instruction*, 11(2), 57–74. <https://doi.org/10.12973/iji.2018.1125a>
- Bisri. (2009). *Akblak*. Jakarta: Direktorat Jenderal Pendidikan Islam Agama RI.
- Cherrington, J. O., & Cherrington, D. J. (1979). Honesty and business ethics in the accounting profession. In *the Annual Meetings of the American Accounting Association in Honolulu*. Hawaii.
- Emosda. (2011). Penanaman nilai-nilai kejujuran dalam menyiapkan karakter bangsa. *Innovatio*, X(1), 151–166.
- Ercegovac, Z., & Richardson, J. V. (2004). Academic dishonesty, plagiarism included, in the digital age: A literature review. *College & Research Libraries*, 301–318.
- Garson, G. D. (2009). Overview structural equation modeling.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate data analysis* (7th ed.). Prentice Hall.
- Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6(1), 53–60.
- Igbaria, M., Zinatelli, N., Cragg, P., & Cavaye, A. L. M. (1997). Personal computing acceptable factors in small firms: A structural equation model. *MIS Quarterly*, 279–299.
- Kementerian Pendidikan dan Kebudayaan Republik Indonesia. (2013). *Modul pelatihan implementasi kurikulum 2013: Materi pelatihan guru implementasi kurikulum 2013 SMP/MTs*. Jakarta: Badan Pengembangan Sumber Daya Manusia Pendidikan dan Kebudayaan

- dan Penjaminan Mutu Pendidikan.
- Kementerian Pendidikan dan Kebudayaan Republik Indonesia. (2015). *Panduan penilaian untuk sekolah menengah pertama*. Jakarta: Direktorat Pembinaan Sekolah Menengah Pertama.
- Koellhoffer, T. (2009). *Character education: being fair and honest*. New York: Infobase Publishing.
- Koss, J. (2011). Academic dishonesty among adolescents. *American Psychological Association*, 11(2), 38–46.
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of educational objectives: Handbook II: affective domain*. New York: David McKay.
- Kumaidi. (2014). Implementasi penilaian autentik dalam pembelajaran di kelas. In *Seminar Nasional Implementasi Penilaian dan Pelaksanaan Kurikulum 2013*. Jakarta: Universitas Negeri Jakarta.
- Leachy, G. (2012). QR code in mathematics classrooms. In *Mathematics Teaching Issue* (Vol. 235, pp. 27–29). Derby: The Association of Teacher of Mathematics.
- Lickona, T. (1992). *Educating for character*. New York: Bantam Books.
- Maharani, E. (2016, June). DIY targetkan 25 persen sekolah terapkan K-13,. *Republika.Co.Id*.
- Popham, W. J. (2009). *Test better, teach better: the instructional role of assessment*. United States of America: Association for Supervision and Curriculum Development.
- Punyanunt-Carter, N. M. (2017). Responseoptions for scales: does it matter what words you use? *Journal of Media Critiques*, 3(9), 33–49. <https://doi.org/10.17349/jmc117103>
- Schmelkin, L. P., Gilbert, K., Spencer, K. J., Pincus, H. S., & Silva, R. (2008). A multidimensional scaling of college students' perceptions of Academic Dishonesty. *The Journal of Higher Education*, 79(5), 587–607. <https://doi.org/10.1353/jhe.0.0021>
- Suranto, Muhyadi, & Mardapi, D. (2014). Pengembangan instrumen evaluasi Uji Kompetensi Keahlian (UKK) administrasi perkantoran di SMK. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 18(1), 98–114. Retrieved from <https://journal.uny.ac.id/index.php/jpep/article/view/2127>
- Widhiarso, W. (2009). Koefisien reliabilitas pada pengukuran kepribadian yang bersifat multidimensi. *Psikobuana*, 1(1), 39–48.
- Wijanto, S. H. (2008). *Structural equation modeling dengan Lisrel 8.8*. Yogyakarta: Graha Ilmu.
- Yamin, S., & Kurniawan, H. (2009). *Structural equation modeling: belajar lebih mudah teknik analisis data kuesioner dengan Lisrel-PLS*. Jakarta: Salemba Infotek.



## EVALUASI PROGRAM MAHASISWA WIRAUSAHA DENGAN MODEL CIPO DI UNIVERSITAS BRAWIJAYA DAN UNIVERSITAS NEGERI MALANG

*Ahmad*

Universitas Gajayana Malang  
Jl. Mertojoyo Blok L, Merjosari Malang, Jawa Timur, 65144 Indonesia  
Email: [ahmad\\_t@unigamalang.ac.id](mailto:ahmad_t@unigamalang.ac.id)

### Abstrak

Penelitian ini bertujuan untuk mengevaluasi implementasi kebijakan Program Mahasiswa Wirausaha pada Universitas Brawijaya dan Universitas Negeri Malang di Jawa Timur. Penelitian ini menggunakan pendekatan evaluatif yang berbasis kasus pada kedua universitas pelaksana PMW dengan metode evaluasi program dengan model CIPO. Hasil penelitian menunjukkan: (1) Konteks evaluasi, suatu perencanaan PMW merujuk pada latar belakang, landasan formal dan operasional pelaksanaan, tujuan, dan manfaat bagi pemangku kepentingan dengan capaian evaluasinya tinggi; (2) Input evaluasi, adanya perbedaan mekanisme pengalokasian dana bantuan antara PMW-UB dan PMW-UM yang menyebabkan capaian evaluasinya moderat, sedangkan pada indikator lain capaian evaluasinya tinggi; (3) Proses evaluasi, terdapat enam belas kriteria evaluasi yang berkategori tinggi, enam kriteria evaluasi yang berkategori moderat dan dua kriteria evaluasi yang berkategori rendah; sedangkan (4) Output evaluasi, indikator mahasiswa yang berwirausaha dan unit usaha yang dikembangkan mahasiswa dengan evaluasi capaiannya rendah.

**Kata kunci:** *program mahasiswa wirausaha, model evaluasi CIPO*

## EVALUATION OF STUDENT ENTREPRENEUR PROGRAM WITH CIPO MODEL IN BRAWIJAYA UNIVERSITY AND THE STATE UNIVERSITY OF MALANG

### Abstract

This research is aimed at evaluating the implementation of the Student Entrepreneurial Program policies at Brawijaya University and the State University of Malang in East Java. This study uses a case-based evaluative approach at both University, Brawijaya University and the State University of Malang with program evaluation methods. The model used in this research was CIPO model. The results showed: Context evaluation, a PMW planning refers to background, formal foundation and operational implementation, purpose, and benefits for the stakeholders to the achievement of a high evaluation; Input evaluation, the difference in the mechanism of allocating funds between PMW-UB and PMW-UM is causing moderate evaluation outcomes, whereas the other indicators of high achievement evaluation; Process evaluation, there are sixteen high-category evaluation criteria, six criteria of moderate categorization and two low evaluation criteria; and outputs evaluation, indicators student entrepreneurship and business unit developed a student with low achievements evaluation.

**Keywords:** *student entrepreneurship program, CIPO evaluation model*

Permalink/DOI: <http://dx.doi.org/10.21831/pep.v22i2.16577>

## Pendahuluan

Jawa Timur yang dinobatkan sebagai provinsi terpadat kedua di Indonesia memiliki populasi penduduk 38.106.600 jiwa dari 245.425.200 jiwa masyarakat Indonesia (Pusat Data Statistik Pendidikan, 2014, p. 13), yang terbagi dalam tiga suku daerah yakni Madura, Using, dan Tengger (MUT) yang memiliki karakter dan budaya yang berbeda-beda. Selain itu tingkat kepadatan penduduk mencapai 15,53% menjadikan Jawa Timur memiliki pengaruh yang besar terhadap perkembangan ekonomi Nasional.

Diasumsikan dengan semakin meningkatnya populasi penduduk suatu daerah akan sebanding dengan peningkatan kesejahteraan rakyatnya (Soenarto, Amin, & Kumaidi, 2017, p. 216). Akan tetapi kondisi terkini di Jawa Timur berbanding terbalik dengan asumsi yang ada, dengan tingkat pengangguran terbuka (TPT) pada Agustus 2014 mencapai 4,19% mendekati TPT rata-rata nasional 6,25% dari 245,425 juta jiwa penduduk Indonesia, (Badan Pusat Statistik Provinsi Jawa Timur, 2014). Angkatan kerja di Jawa Timur usia 15 tahun ke atas per Agustus 2014 berjumlah 20.149.990 jiwa, diantaranya 19.306.510 jiwa bekerja dan 843.490 jiwa masih menganggur. Penduduk yang bekerja tersebut berkurang sebanyak 247 ribu orang dibanding keadaan Agustus 2013 yang jumlahnya sebanyak 19,55 juta orang (Badan Pusat Statistik Provinsi Jawa Timur, 2014, p. 4).

Idealnya bonus demografi tersebut menjadi peluang tersendiri bagi provinsi Jawa Timur menjadi sentral ekonomi nasional yang memiliki sumber daya manusia yang berkualitas, dengan sumber daya manusia yang berkualitas dapat mengelola sumber daya manusia secara optimal (Sutopo, 2012, p. 421). Akan tetapi momok yang bernama pengangguran masih menghantui setiap lini kehidupan. Karena pengangguran merupakan dilema klasik yang terus berevolusi disetiap dinding kehidupan manusia berbangsa dan bernegara. Kondisi demikian terjadi karena tingginya suplai tidak sebanding dengan jumlah permintaan yang tersedia. Pengangguran juga terjadi karena keterbatasan sumber daya ma-

nusia yang potensial, kompetitif, inovatif, kreatif dan mandiri, Suyanto & Ariadi (2015, p. 116). Untuk menggerakkan laju perekonomian nasional, pemerintah mencanangkan gerakan kewirausahaan nasional. Gerakan tersebut dapat berjalan efektif apabila elemen akademisi, bisnis, dan pemerintah berjalan bersama dalam pembangunan (Siradjuddin, 2009, p. 291).

Namun keterpaduan elemen tersebut belum berjalan seperti yang diharapkan, hal ini ditandai dengan masih tingginya angka pengangguran terdidik berpendidikan tinggi di Indonesia mencapai angka 5,96 juta jiwa (Badan Pusat Statistik, 2013), bahkan diproyeksikan akan terus meningkat pada tahun-tahun berikutnya karena masih terdapat 36,81% penduduk usia 20-30 tahun yang bekerja paruh waktu. Kondisi demikian menjadi perhatian Subianto (2013, p. 68) yang menyatakan bahwa jumlah pengangguran dengan tingkat pendidikan sarjana relatif meningkat dari tahun ke tahun, sementara dunia usaha dan dunia kerja setiap tahun juga mengalami kesulitan untuk mendapatkan tenaga kerja baru yang memiliki kualifikasi pendidikan serta keterampilan yang sesuai.

Penyebab utamanya karena pola pikir masyarakat Indonesia masih menganut pola pikir mencari pekerjaan (*job seeker*), minat dan aktivitas kewirausahaan (*entrepreneurial activity*) tenaga berpendidikan tinggi relatif rendah untuk mampu berwirausaha dan bekerja secara mandiri (Suyanto & Ariadi, 2015). Faktor tingkat pendidikan yang tidak relevan dengan kebutuhan lapangan kerja juga menyumbang lahirnya penganggur terdidik di Indonesia. Menurut Tilaar & Nugroho (2012, p. 13) sedikitnya ada tiga hal yang menjadi alasan utama kenapa alumni perguruan tinggi banyak yang tidak bekerja, yaitu hambatan kultural, mutu dan relevansi kurikulum pendidikan, dan pasar kerja.

Menurut McClelland (Ciputra, 2008), suatu negara menjadi makmur apabila didukung oleh minimum jumlah wirausahawannya 2% dari populasi penduduk suatu negara. Namun realitas menunjukkan bahwa jumlah wirausaha di Indonesia baru berada pada angka 1,65%, atau sebanyak 4,20 juta dari

118,17 juta orang penduduk Indonesia yang bekerja, sedangkan Amerika Serikat di tahun 2013 jumlah wirausahanya 11,5%, China 10%, Jepang dengan jumlah wirausaha 10%, Singapura 7,2% dan Malaysia 5% (Ariyanti, 2013).

Istilah wirausaha pertama kali diperkenalkan di dunia pada tahun 1725 oleh ahli ekonomi berkebangsaan Perancis yang bernama (Cantillon, 2010, p. 75), berwirausaha merupakan sebuah upaya untuk bekerja sendiri, membeli dengan harga yang telah ditentukan, kemudian menjualnya di pasar dengan harga yang telah ditentukan. Kemudian tahun 1776 Adam Smith dalam kritiknya yang berjudul *Discussed entrepreneur but associated activities with the capitalist*, pada generasi Jean-Baptiste Say tahun 1803 dalam tulisannya *Defined entrepreneur as improving the state of resources used in production* yang makin mengan-tarkan konsep kewirausahaan makin men-dunia, sehingga di tahun 1871 John Stuart Mill *Associated capitalist and entrepreneur together through risk* dan sampai tahun 1934 Joseph A. Schumpeter melalui tulisannya berjudul *Defined entrepreneurs as sociologically distinct individuals; separated entrepreneurship from the role of the capitalist*.

Dengan demikian, menurut Joseph Schumpeter (Bygrave, 2010, p. 2), wirausaha adalah orang yang mendobrak sistem ekonomi yang ada dengan memperkenalkan barang dan jasa yang baru, dengan menciptakan bentuk organisasi baru atau mengolah bahan baku baru. Orang tersebut melakukan kegiatannya melalui organisasi bisnis yang baru ataupun bisa pula dilakukan dalam organisasi bisnis yang sudah ada.

Menurut Akbar (2007, p. 37) wira-usaha dipergunakan untuk menggambarkan aktivitas manusia yang memaksimalkan sumber daya yang ada untuk menghasilkan produktivitas yang tinggi. Pandangan terhadap wirausaha dapat diartikan sebagai profesi pada dunia usaha di bidang bisnis dan dipan-dang sebagai sikap mental, ciri pribadi, atau nilai-nilai yang dapat dimiliki oleh siapapun.

Dengan demikian wirausaha adalah manusia yang memiliki kepribadian unggul, berani mengambil resiko, kebulatan tekad

secara mandiri, dan tangguh untuk mencipta-kan suatu organisasi baru atau memanfaat-kan organisasi yang sudah ada dan menggu-nakan modal sebagai alat bantu dalam proses penciptaan kreativitas dan inovatif guna mencapai tingkat produktivitas yang tinggi sebagai hasil dari aktivitas wirausaha. Akti-vitas dari wirausaha disebut kewirausahaan.

Terminologi kewirausahaan adalah se-mangat, sikap dan perilaku atau kemampuan seseorang dalam menangani usaha dan atau kegiatan yang mengarah pada upaya mencari, menciptakan, menerapkan cara kerja, tekno-logi, dan produk baru dengan meningkatkan efisiensi dalam rangka memberikan pelayan-an yang lebih baik dan atau memperoleh keuntungan yang lebih besar.

Sebagaimana definisi dari Hisrich, Peters, & Shepherd (2008, p. 10) kewirausa-haan adalah proses penciptaan sesuatu yang baru pada nilai menggunakan waktu dan upaya yang diperlukan, menanggung resiko keuangan, fisik, serta resiko sosial yang mengiringi, menerima imbalan moneter yang dihasilkan serta kepuasan dan kebebasan pri-badi. Schumpeter sebagaimana dikutip Alma (2011, p. 24), mendefinisikan kewirausahaan adalah jiwa seseorang yang mendobrak sis-tem ekonomi yang ada dengan memperke-nalkan barang dan jasa yang baru, dengan menciptakan bentuk organisasi baru. Karena dalam kewirausahaan membutuhkan suatu proses kreasi, dan inovatif, serta beresiko tinggi dalam menghasilkan nilai tambah bagi produk yang bermanfaat bagi masyarakat dan mendatangkan kemakmuran bagi wira-usahawan.

Kewirausahaan itu dapat dipelajari wa-laupun ada juga orang-orang tertentu yang mempunyai bakat dalam hal berwirausaha. Oleh karena itu, untuk menekan angka peng-angguran terdidik di Indonesia, Direktorat Jenderal Pendidikan Tinggi mengeluarkan kebijakan meningkatkan kewirausahaan ma-hasiswa (Direktorat Jenderal Pendidikan Tinggi Kementerian Pendidikan dan Kebu-dayaan, 2009, p. 3) supaya lulusan perguruan tinggi mampu menggerakkan perekonomian serta meningkatkan kesejahteraan masya-rakat.

Hal ini sebagaimana pernyataan Ci-putra (Siradjuddin, 2009, p. 107), pendidikan kewirausahaan adalah senjata penghancur massal mengatasi pengangguran dan kemiskinan sekaligus tangga menuju impian setiap warga negara untuk mandiri secara finansial dan mampu membangun kemakmuran untuk bersama-sama membangun kesejahteraan bangsa.

Kebijakan dan program penguatan kelembagaan perguruan tinggi pun dikembangkan melalui Program Mahasiswa Wirausaha (*Student Entrepreneur Program*) disingkat PMW yang dirintis sejak tahun anggaran 2009. PMW bertujuan untuk membekali mahasiswa dengan pengetahuan, keterampilan dan jiwa wirausaha berbasis ilmu pengetahuan, teknologi dan sains kepada para mahasiswa agar menjadi calon pengusaha yang tangguh dan sukses menghadapi persaingan global. Sasaran lainnya adalah untuk meningkatkan aktivitas kewirausahaan bagi para mahasiswa sebelum lulus agar mampu menjadi wirausaha yang tangguh dimasa yang akan datang.

Untuk menciptakan wirausaha berpendidikan tinggi, sejak tahun 2009-2013 Universitas Brawijaya (UB), dan Universitas Negeri Malang (UM) sebagai representasi dari Dirjen Dikti telah menyelenggarakan PMW secara berkelanjutan dimulai dari proses sosialisasi, rekrutmen, pelatihan, magang, bantuan permodalan, pelaksanaan usaha, mentoring, dan pameran produk PMW. Oleh karena itu, pelaksanaan PMW pada kedua universitas diharapkan dapat menumbuhkembangkan jiwa kewirausahaan dan meningkatkan aktivitas kewirausahaan agar para lulusan perguruan tinggi lebih menjadi pencipta lapangan kerja dari pada pencari kerja.

PMW merupakan kelanjutan dari program-program sebelumnya (PKM, Co-op, KKU, dll) untuk menjembatani para mahasiswa memasuki dunia bisnis riil melalui fasilitas *star-up business*. Disamping itu juga, pelaksanaan PMW pada kedua universitas terfokus dalam mengimplementasikan visi UB menuju *Entrepreneurial University*, dan UM dengan pijakan menuju *the learning university* yang berdasarkan pola pengembangan kemahasiswaan yang bertumpu pada pilar percepatan

adaptasi, perolehan prestasi, dan pendekatan dunia kerja.

Pelaksanaan PMW secara nasional sejak tahun 2009-2014 belum dilakukan evaluasi, karena demikian maka sangat penting dilakukan evaluasi terhadap pelaksanaan program tersebut dengan tujuan untuk mengetahui tingkat keberhasilan program. Dewasa ini banyak model evaluasi program yang dihasilkan, namun menimbulkan permasalahan sendiri bagi para praktisi dan pengguna model evaluasi program tentang model terbaik sesuai dengan tujuan. Setiap model evaluasi dikembangkan berdasarkan asumsi tentang model evaluasi yang seperti apa dan menggunakan pendekatan yang bagaimana untuk mengevaluasi suatu program.

Beragam model evaluasi program tersebut, akan disesuaikan dengan karakteristik program yang akan dievaluasi. Indikator lainnya adalah berangkat dari kerangka teori yang dibangun untuk menunjang komponen program yang akan dievaluasi, walaupun dalam penelitian ini dibatasi pada empat komponen utama yakni komponen konteks, input, proses, dan *output*. Sebagaimana pendapat Scheerens, Glas, & Thomas (2003) sebagai pakar evaluasi *context, input, proses, dan output* (CIPO) memaklumi bahwa suatu model evaluasi program yang dikembangkan tidak semestinya dapat memenuhi semua kebutuhan komponen program, akan tetapi dengan model evaluasi program yang ada disesuaikan dengan komponen apa yang dievaluasi.

Berdasarkan pendapat Scheerens tersebut, dan memperhatikan karakteristik dari kebijakan PMW yang akan dievaluasi, maka dalam penelitian ini dibatasi menggunakan evaluasi model CIPO yang fokus mengkaji komponen konteks, input, proses, dan output yang dikembangkan oleh Scheerens yang diadopsi dari sekolah model.

Penggunaan model evaluasi CIPO merupakan kritikan dari keterbatasan model evaluasi CIPP akan pentingnya komponen output pada setiap tingkatan evaluasi. Karena keterbatasan tersebut, sebagai ciri dari ilmu pengetahuan yang selalu mengalami perkembangan maka dengan sendirinya model evaluasi program yang dapat mengevaluasi

output bermunculan. Oleh sebab itu, untuk mengetahui efektivitas pelaksanaan PMW maka sangat penting untuk dilakukan penelitian dengan judul “Evaluasi Program Mahasiswa Wirausaha Dengan Model CIPO Di Universitas Brawijaya dan Universitas Negeri Malang. Diharapkan dengan penelitian ini dapat memberikan kontribusi nyata dalam perbaikan mutu pembangunan manusia, pendidikan, dan kesejahteraan nasional.

### **Metode Penelitian**

Penelitian telah dilaksanakan di UB dan UM di Jawa Timur. Pemilihan kedua universitas tersebut karena secara bersama-sama telah melaksanakan PMW secara berkelanjutan sejak tahun 2009 sampai tahun 2014 yang dibuktikan dengan laporan pertanggungjawaban program setiap tahun dan kedua universitas tersebut juga sama-sama mendapat predikat 40 besar perguruan tinggi terbaik, dan 25 besar perguruan tinggi negeri terbaik di Indonesia tahun 2014. Subjek dalam penelitian ini berjumlah 39 orang yang berasal dari mahasiswa peserta PMW, perwakilan dari Dikti, perwakilan dari UB, dan perwakilan dari UM. Subjek dalam penelitian ini yakni terdiri dari 15 mahasiswa UB dan 15 mahasiswa UM, perwakilan Dikti diwakili oleh 3 subjek kunci, pelaksana PMW yang terdiri dari 3 subjek kunci dari UB dan 3 subjek kunci dari UM.

Adapun informasi yang dikaji terpeta-kan berdasarkan komponen evaluasi yakni: (1) komponen konteks menggali informasi tentang latar belakang PMW, dasar penyusunan PMW, tujuan penyusunan PMW, dan manfaat PMW, (2) komponen input, untuk mengkaji bentuk pedoman pelaksanaan PMW, struktur organisasi pelaksana PMW, mekanisme pembiayaan/pendanaan, kualitas kurikulum, dan sarana dan prasarana penunjang PMW, (3) komponen proses melalui penelusuran dokumen tentang mekanisme sosialisasi program, sistem seleksi peserta PMW, mekanisme pelatihan dan pembekalan, magang dan pendampingan usaha, mekanisme pencairan modal usaha awal, monitoring dan evaluasi usaha PMW, (4) komponen output untuk mengevaluasi tren perkembangan

jumlah mahasiswa wirausaha, dan jumlah unit bisnis yang berhasil dikembangkan dalam PMW. Informasi tersebut diperoleh melalui penyebaran kuesioner dan wawancara. Kuesioner terbagi dalam dua paket yakni kuesioner untuk pengelola PMW dan kuesioner untuk mahasiswa peserta PMW. Kedua paket kuesioner tersebut di validasi oleh 20 orang panelis terdiri dari unsur akademisi, praktisi pendidikan, dan pelaksana PMW di universitas. Hasil validasi panelis terhadap 46 butir kuesioner untuk pengelola PMW, terdapat 42 butir yang valid, sedangkan 4 butir lainnya tidak valid yakni butir nomor 1, 19, 21, dan 25. Walaupun hasil analisis menunjukkan 42 butir sudah valid, namun peneliti tetap melakukan revisi kuesioner dari segi konstruksi dan bahasa, sedangkan butir yang tidak valid peneliti langsung membuangnya setelah memastikan tidak mempengaruhi aspek, dan indikator evaluasi.

Hasil pengujian validasi butir diperoleh  $CVR_{total} = 0,83$  sedangkan kriteria untuk menentukan valid tidaknya isi butir instrumen dengan jumlah panelis 20 orang yakni dengan syarat  $CVR_{hitung} \geq CVR_{tabel} 0,42$  (Lawshé, 1975, p. 568). Oleh karena  $CVR_{total}$  melebihi kriteria kritis CVR maka butir soal dinyatakan baik untuk digunakan.

Kemudian, hasil validasi panelis terhadap 30 butir kuesioner untuk mahasiswa peserta PMW, diperoleh 27 butir yang valid dan 3 butir tidak valid yakni butir nomor 6, 17, dan 28. Walaupun hasil analisis menunjukkan 27 butir sudah valid, namun peneliti tetap melakukan revisi terhadap kuesioner dari segi konstruksi dan bahasa.

Hasil pengujian validasi butir diperoleh  $CVR_{total} = 0,84$  sedangkan kriteria untuk menentukan valid tidaknya isi butir kuesioner dengan jumlah panelis 20 orang yakni dengan syarat  $CVR_{hitung} \geq CVR_{tabel} 0,42$ . Oleh karena  $CVR_{total}$  melebihi kriteria kritis CVR maka butir soal dinyatakan baik untuk digunakan.

Kriteria keberhasilan PMW dibuat dalam bentuk kategori yakni dengan kategori evaluasi tinggi, moderat, dan rendah terhadap pelaksanaan kebijakan PMW pada kedua universitas.

Tabel 1. Kriteria Evaluasi

Kriteria	Deskripsi
Rendah	Kriteria evaluasi tidak terpenuhi semuanya atau terpenuhi sebagian di bawah 50% dari kriteria yang ditetapkan
Moderat	Kriteria evaluasi terpenuhi sebagian, 50% atau lebih dari kriteria yang ditetapkan
Tinggi	Kriteria evaluasi terpenuhi semua

Penelitian ini menggunakan pendekatan evaluatif yang berbasis kasus pada kedua Universitas pelaksana PMW, dengan metode penelitian evaluasi program.

### Hasil Penelitian dan Pembahasan

Berdasarkan hasil penelitian terhadap pelaksanaan kebijakan PMW pada kedua universitas yang menjadi target tempat penelitian, maka hasil penelitian secara umum dapat didistribusikan seperti pada tabel dua berikut.

Tabel 2. Hasil Evaluasi Program Mahasiswa Wirausaha

Perbandingan Capaian Evaluasi Berdasarkan Aspek Evaluasi Komponen Konteks						
Univ.	Latar Belakang	Dasar PMW	Tujuan	Manfaat		
UB	Tinggi	Tinggi	Tinggi	Tinggi		
UM	Tinggi	Tinggi	Tinggi	Tinggi		
Perbandingan Capaian Evaluasi Berdasarkan Aspek Evaluasi Komponen Masukan						
Univ.	Pedoman PMW	Organisasi PMW	Sumber Dana	Alokasi Dana	Kurikulum	Sapras
UB	Tinggi	Tinggi	Tinggi	Tinggi	Rendah	Tinggi
UM	Tinggi	Tinggi	Tinggi	Sedang	Rendah	Tinggi
Perbandingan Capaian Evaluasi Berdasarkan Aspek Evaluasi Komponen Proses						
Univ.	Sosialisasi PMW	Unsur dalam Sosialisasi	Partisipasi Mahasiswa dalam sosialisasi	Waktu Sosialisasi	Mahasiswa Lolos Seleksi Admin.	Pedoman Seleksi
UB	Tinggi	Tinggi	Sedang	Tinggi	Sedang	Tinggi
UM	Tinggi	Tinggi	Tinggi	Sedang	Rendah	Tinggi
Lanjutan Komponen Proses						
Univ.	Pelibatan unsur PMW dlm seleksi	Jumlah penerima dana	Pelatihan	Waktu Pelatihan	Instruktur Pelatihan	Evaluasi Pelatihan
UB	Tinggi	Rendah	Tinggi	Tinggi	Tinggi	Tinggi
UM	Tinggi	Rendah	Tinggi	Tinggi	Tinggi	Tinggi
Lanjutan Komponen Proses						
Univ.	Magang	Mentoring	Waktu Mentoring	Mekanisme Cair Dana	Proses Cair Dana	Monev
UB	Tinggi	Tinggi	Tinggi	Tinggi	Rendah	Tinggi
UM	Rendah	Tinggi	Tinggi	Tinggi	Sedang	Tinggi
Perbandingan Capaian Evaluasi Berdasarkan Aspek Evaluasi Komponen Keluaran						
Univ.	Jumlah Mahasiswa yang Usaha			Unit Usaha Mahasiswa		
UB	Rendah			Rendah		
UM	Rendah			Rendah		

Berdasarkan hasil penelitian sebagaimana pada Tabel 2, pada komponen evaluasi konteks, secara umum pelaksanaan PMW pada kedua universitas tersebut tergolong tinggi. Hasil komponen evaluasi masukan

menunjukkan pelaksanaan PMW pada kedua universitas tersebut tergolong tinggi, akan tetapi masih terdapat indikator evaluasi yang berstatus moderat dan indikator berstatus rendah. Komponen evaluasi proses menunjukkan bahwa terdapat dua belas indikator berkategori tinggi, lima indikator berkategori moderat, dan satu indikator berkategori rendah. Untuk komponen evaluasi output semua indikator berkategori rendah. Hasil penelitian ini lebih lanjut dapat dijelaskan pada pokok pembahasan berikut.

### Pembahasan

Berdasarkan deskripsi temuan penelitian, maka perlu dilakukan pembahasan pada setiap komponen hasil penelitian. Adapun pembahasan temuan penelitian dapat diuraikan sebagai berikut.

#### Komponen Konteks (*Context*)

##### *Latar Belakang Pelaksanaan*

Hasil penelitian menunjukkan aspek evaluasi tentang latar belakang program dengan capaian kategori tinggi. Hal ini sesuai dengan arah pengembangan PMW pada kedua universitas sesuai dengan karakteristik universitas yang terintegrasi dengan visi, misi, arah pengembangan dan tujuan pada kedua universitas. Semangat *Worldclass Entrepreneur University* dan *The Learning University* pada kedua universitas tersebut berkontribusi perkembangan pendidikan nasional, karena pendidikan dewasa ini bukan saja berorientasi pendidikan untuk pendidikan, namun pendidikan yang memuat/implikasi pengetahuan, skill, dan mental.

Khususnya pendidikan di universitas bukan saja *thinking the job namun creating the job*, karena output dari pendidikan seharusnya mengembangkan ilmu dan teknologi melalui usaha-usaha bisnis salah satunya sebagaimana dalam penerapan PMW dengan tujuan untuk membentuk mahasiswa yang memiliki pola pikir *creating the job*. Untuk menumbuhkembangkan jiwa kewirausahaan dan meningkatkan aktivitas kewirausahaan mahasiswa maka terbentuk program pengembangan budaya kewirausahaan seperti

Program Kreativitas Mahasiswa (PKM), *Co-operative Education* (Co-op), Kuliah Kewirausahaan (KWU), Magang Kewirausahaan (MKU), Kuliah Kerja Usaha (KKU), dan Karya Alternatif Mahasiswa (KAM).

#### *Landasan Penyusunan*

Hasil penelitian pada aspek landasan penyusunan PMW kedua universitas berkategori tinggi. Hasil ini ditunjang baik secara formal dan operasional, landasan pelaksanaan PMW mengacu pada undang-undang, peraturan pemerintah, peraturan menteri, keputusan Direktur Jenderal, regulasi anggaran dan Surat Keputusan Rektor setiap universitas sebagai bentuk komitmen universitas pada setiap kegiatan PMW. Sejumlah regulasi tersebut dapat terbagi dalam dua kelompok utama yakni regulasi formal umum dan regulasi formal operasional. Regulasi formal umum yang terdiri dari undang-undang, peraturan pemerintah, peraturan menteri, keputusan menteri yang bersifat umum sebagai landasan penyelenggaraan program, sedangkan regulasi formal operasional yang merupakan landasan gerak yang mengikat secara langsung terhadap pelaksanaan program di lapangan seperti surat keputusan rektor pada setiap universitas.

#### *Tujuan*

Hasil penelitian pada aspek tujuan PMW kedua universitas tersebut berkategori tinggi. Hasil ini ditandai dengan adanya tujuan PMW yang ditetapkan oleh Dikti sebagai tujuan umum, namun kedua universitas diberikan ruang untuk mengembangkan tujuan khusus sesuai dengan kebutuhan setiap universitas. Berangkat dari tujuan itu pula universitas merancang dan menyusun program, perangkat organisasi pelaksana sehingga sasaran program untuk menciptakan wirausaha yang berpendidikan tinggi, terbentuknya model pembelajaran dan unit kewirausahaan di universitas tercapai. Untuk mencapai target tersebut di atas maka ditetapkan strategi pengembangan kewirausahaan mahasiswa meliputi: (1) pembinaan berkelanjutan wirausaha mahasiswa; (2) pengembangan dan pemantaban kelembagaan program ke-

wirausahaan mahasiswa di universitas; (3) pengembangan kurikulum pendidikan kewirausahaan; (4) pembentukan dan dinamisasi forum nasional kewirausahaan; dan promosi dan desiminasi program kewirausahaan kepada masyarakat luas.

#### *Manfaat*

Hasil penelitian pada aspek manfaat evaluasipun berkategori tinggi. Hal ini ditandai dengan tingginya respon masyarakat terhadap manfaat PMW. selain itu pula keberhasilan pengembangan PMW pada kedua universitas dapat ditinjau dari tiga aspek yakni kebermanfaatannya bagi mahasiswa, pengembangan dan peningkatan produktivitas usaha UKM sebagai bagian dari *stakeholder*, dan tercapainya tujuan universitas.

#### Komponen Input

##### *Pedoman Pelaksanaan*

Hasil penelitian pada aspek evaluasi pedoman pelaksanaan PMW berkategori tinggi. Hal ini dibuktikan dengan kedua universitas telah memiliki buku pedoman PMW yang diterbitkan oleh Dikti. Keberadaan buku pedoman pelaksanaan PMW sebagai rujukan utama dalam penyelenggaraan PMW diperoleh dari hasil sosialisasi PMW yang dilakukan oleh Dikti pada kegiatan-kegiatan kemahasiswaan, walaupun terkadang proses sosialisasi keberadaan buku pedoman mengalami keterlambatan dan terjadinya multi tafsir terhadap pedoman mengakibatkan penyelenggaraan PMW terhambat. Kendala demikian disebabkan karena adanya perubahan strukturisasi, dan regulasi anggaran DIPA pada setiap universitas.

Selain itu juga penyebaran buku pedoman dalam bentuk workshop, dan melalui pemanfaatan media sosial elektronik, sulit bagi universitas tertentu untuk mengimplementasikan dalam bentuk buku petunjuk teknis pelaksanaan program yang disesuaikan dengan kebutuhan setiap universitas dalam hal ini UB bermuara pada bidang sains dan teknologi dan UM bermuara pada bidang pendidikan dan pengajaran yang pada

prinsipnya bertujuan untuk menjaga motivasi dan kreativitas mahasiswa.

### *Struktur Organisasi*

Keberhasilan PMW pada kedua universitas tersebut merupakan cerminan kinerja terbaik dari pelaksana PMW yang memiliki integritas, dan komitmen yang tinggi sehingga hasil evaluasi pada aspek struktur organisasi berkategori tinggi. Hasil penelitian tersebut ditunjang dengan keberadaan panitia khusus pelaksana PMW pada kedua universitas yang berkerja di bawah koordinasi pembantu Rektor III pada setiap universitas. Panitia pelaksana PMW bekerja untuk merencanakan, melaksanakan, dan mengukur efektivitas pelaksanaan program. Berdasarkan aturan yang telah ditetapkan oleh Dikti, untuk menyiapkan/me-rencanakan PMW, tim bekerja selama 1-3 bulan efektif, yang dihitung mulai bulan Januari-Maret, kemudian proses pembekalan dan pelaksanaan program dilakukan selama 7 bulan dihitung dari bulan April-Oktober, sedangkan proses monitoring dan evaluasi program dilakukan pada bulan November-Desember.

### *Mekanisme Pembiayaan*

Berdasarkan hasil penelitian pada aspek sumber dana berkategori tinggi pada kedua universitas sedangkan aspek alokasi dana berkategori moderat. Hal ini ditandai dengan sumber dana yang jelas baik bersumber pihak pemerintah, usaha mandiri universitas, maupun dari mitra kerjasama universitas. Namun hasil penelitian menunjukkan bahwa terdapat perubahan nomenklatur keuangan yang bersumber dari APBN dan PNPB menjadi tantangan tersendiri bagi penyelenggara PMW di kedua universitas.

Dengan adanya perubahan nomenklatur tersebut sangat berpengaruh pada besaran alokasi mata anggaran. Permasalahan yang mendasar lainnya adalah pemahaman konsep HIBAH yang keliru menganggap hibah merupakan pemberian bantuan dana cuma-cuma atau gratis dan tidak sedikit akan muncul embrio-embrio pengusaha berbasis proposal, maksudnya mahasiswa menjalan-

kan usaha jika mendapat dana PMW, setelah tidak mendapat dana bantuan maka usaha yang dikembangkan mahasiswaupun berakhir.

Dengan demikian, konsep hibah PMW kedepan merupakan pelimpahan dana PMW yang berasal dari APBN dan PNPB dari Dikti disalurkan dalam bentuk hibah ke rekening universitas, kemudian universitas memasukannya dalam Daftar Isian Penggunaan Anggaran (DIPA), kemudian dari universitas menyalurkan ke pengelola PMW kemudian pengelola PMW mengkompetisikan dana tersebut dalam bentuk dana bergulir ke peserta PMW.

Skema pembiayaan yang diterapkan pada kedua universitas tersebut berbeda satu dengan yang lain. UB secara konsisten pengalokasian pembiayaan PMW sejak tahun 2011-2014, alokasi pembiayaan sebesar 30% untuk kebutuhan manajemen, dan 70% untuk kebutuhan alokasi bantuan dana mahasiswa. Formulasi pengalokasian pembiayaan PMW-UM, yang sama terjadi pada tahun 2011, pengalokasian sebesar 19,22% untuk manajemen, dan 80,78% untuk bantuan dana peserta PMW, akan tetapi sejak tahun 2012-2014, skema pengalokasian bantuan berbanding terbalik berdasarkan aturan dikti, dimana, alokasi untuk manajemen sebesar 71%, dan untuk bantuan modal peserta PMW sebesar 29%.

Tingginya alokasi anggaran untuk kebutuhan manajemen terjadi karena pelaksana PMW-UM merencanakan RAB untuk belanja honor output kegiatan, belanja barang non operasional lainnya, dan belanja perjalanan lainnya melampaui belanja bantuan modal mahasiswa, sedangkan realisasi dari setiap belanja tersebut hanya pada kisaran 30-40%, sehingga sisa anggaran belanja yang tidak terserap tersebut dikembalikan ke bagian keuangan universitas. Faktor lain yang berpengaruh adalah karena menurunnya minat mahasiswa untuk berwirausaha dalam PMW-UM setiap tahun.

### *Kurikulum*

Penelitian menunjukkan bahwa hasil evaluasi pada aspek kurikulum dengan kategori rendah. Hal ini disebabkan kurikulum



PMW belum terintegrasi dengan kurikulum akademik di universitas. Melalui pelaksanaan PMW idealnya dapat menunjang program kurikuler mahasiswa yang dibuktikan hampir seluruh perguruan tinggi telah mengembangkan mata kuliah kewirausahaan dengan kredit minimal dua Sistem Kredit Semester, dan Dikti telah mengembangkan modul kewirausahaan, buku panduan kewirausahaan, silabus, untuk meningkatkan PMW. Melalui mata kuliah Kewirausahaan dapat memberikan pengetahuan kewirausahaan, pengalihan pengalaman berwirausaha serta mendorong tumbuhnya motivasi berwirausaha, sebagai aktivitas awal mahasiswa yang berminat menjadi wirausahawan baru yang handal.

#### *Sarana dan Prasarana Penunjang*

Hasil evaluasi pada aspek sarana dan prasarana PMW dengan kategori tinggi. Hal ini ditandai dengan kesiapan universitas untuk memfasilitasi mahasiswa dengan mengoptimalkan sarana dan prasarana universitas yang ada. Untuk membekali mahasiswa dalam berwirausaha universitas telah memiliki unit kerja seperti PIBLAM UB, Unit bisnis UB dan lembaga pengembangan pendidikan dan pembelajaran (LP3) UM yang didesain dan dibangun untuk membekali mahasiswa dengan kompetensi tertentu untuk mewadahi minat usaha mahasiswa. Keberadaan lembaga tersebut sebagai upaya untuk mewujudkan pelayanan prima di universitas. Secara praktis, unit kerja tersebut banyak berurusan dengan upaya-upaya untuk menjaga kompetensi mahasiswa agar senantiasa dapat memenuhi tuntutan kebutuhan pasar usaha/pasar kerja. Muara dari pelayanan prima ini adalah kepuasan stakeholders utama universitas yang meliputi mahasiswa, dosen, karyawan, pengguna lulusan, serta pengguna hasil penelitian dan pengabdian masyarakat.

#### *Komponen Pelaksanaan*

Hasil evaluasi pada komponen proses terdiri dari delapan belas kriteria evaluasi yang diukur. Dari delapan belas aspek evaluasi tersebut terdapat dua belas aspek evaluasi dengan kategori tinggi, lima aspek evaluasi dengan kategori moderat, dan satu aspek

yang berkategori rendah. Penjelasan lebih lanjut seperti berikut.

#### *Sosialisasi Program*

Hasil evaluasi pada aspek sosialisasi PMW menunjukkan hasil yang tinggi. Hal ini dikarenakan pelaksanaan sosialisasi keberadaan PMW pada kedua universitas tersebut telah berjalan sesuai dengan mekanisme dan kebutuhan setiap universitas, walaupun ditemukan pendekatan sosialisasi yang diterapkan oleh pelaksana PMW diluar dari ketentuan dikti. Oleh sebab itu, universitas diberi kesempatan untuk menyusun jadwal pelaksanaan sosialisasi PMW kepada mahasiswanya sesuai dengan pedoman yang diterbitkan oleh Dikti, akan tetapi pihak universitas juga melakukan sosialisasi keberadaan PMW pada setiap penerimaan mahasiswa baru serta kegiatan kemahasiswaan lainnya. Penerapan sosialisasi keberadaan PMW pada saat penerimaan mahasiswa baru bertepatan pada bulan September merupakan hal yang tidak seharusnya dilakukan, alasannya karena dapat memperpendek masa pelaksanaan PMW, seharusnya bulan Maret atau April.

#### *Seleksi Peserta Program*

Pada aspek seleksi peserta PMW hasil evaluasi berkategori moderat. Hal ini terjadi karena kedua universitas menerapkan model seleksi yang berbeda. Misalkan saja UB menerapkan sistem kuota dan kompetisi. Sistem penentuan kuota penerima PMW dilakukan oleh UB berdasarkan kesepakatan pelaksana PMW-UB dengan pembantu Dekan setiap fakultas. Kuota standar yang diberikan oleh pelaksana adalah minimal masing-masing 4 (empat) unit usaha untuk setiap fakultas, namun tetap disesuaikan dengan kebutuhan program. Akan tetapi pola yang diterapkan oleh UM cukup berbeda, karena UM menerapkan sistem kompetisi sesuai dengan standar usaha yang telah ditetapkan.

Setiap mahasiswa menempuh proses seleksi yang ketat dengan tahapan telah lulus seleksi administrasi, memiliki perencanaan bisnis sesuai standar, dan lulus wawancara usaha. Berdasarkan hasil seleksi yang telah ditetapkan oleh pelaksana PMW, kebanyakan

an mahasiswa kesulitan dalam pembuatan perencanaan usaha.

#### *Pelatihan dan Pembekalan*

Hasil evaluasi menunjukkan hasil dengan kategori tinggi. Karena kegiatan pendidikan dan pelatihan pada kedua universitas telah melibatkan para dosen, praktisi, pemerhati, dan lembaga mitra PMW lain yang memiliki integritas, komitmen, kompetensi dan pengalaman yang tinggi dalam kewirausahaan. Hal ini sebagaimana hasil penelitian Suwardie (2009, p. 159) menyatakan bahwa tamatan pelatihan memiliki cara pandang baru dalam melihat potensi peluang usaha. Selama kegiatan pelatihan dan pembekalan usaha, mahasiswa juga dilatih dan dibimbing oleh tim pendamping untuk menyusun proses pelaporan keuangan untuk mengetahui seberapa jauh pencapaian hasil dari keseluruhan usaha. Dengan begitu bisa diketahui berapa rata-rata pengeluaran dalam sebulan, dan berapa pendapatan bersih yang diperoleh setiap bulannya. Jadi proses evaluasi usaha secara menyeluruh dapat memberikan gambaran tentang perkembangan usaha dan bahan untuk merencanakan target pertumbuhan usaha mahasiswa selanjutnya.

#### *Magang dan Pendampingan Usaha*

Magang merupakan upaya untuk mempertemukan antara peserta PMW dengan Dosen Pembimbing dan Mitra Magang. Hasil evaluasi program pada aspek magang dan pendampingan usaha berkategori moderat. Hal ini disebabkan karena kegiatan magang kewirausahaan menjadi penting dan wajib diikuti oleh peserta, namun ada pengecualian bagi peserta yang telah mengikuti PKM dan program Co-op. Akan tetapi penafsiran kegiatan magang dari universitas cukup beragam, sehingga data temuan menunjukkan bahwa terdapat universitas menerapkan kegiatan magang usaha, dan terdapat pula universitas yang tidak menerapkan magang bahkan menggantikan dengan kegiatan lain sesuai dengan karakter perguruan tinggi yang lebih memilih hal-hal yang praktis. Untuk program magang ada pengecualian bagi mahasiswa

yang pernah mengikuti program Co-op, KKKU, dan program kewirausahaan lain.

#### *Modal Usaha Awal*

Hasil evaluasi pada aspek modal usaha awal menunjukkan bahwa kedua universitas memiliki metode pencairan yang berbeda sehingga hasil evaluasi berkategori moderat. Hal ini disebabkan bahwa sejak dana PMW masuk dalam DIPA universitas pada awal tahun 2010, proses penyaluran modal usaha awal tidak tepat waktu. Hal ini disebabkan adanya perubahan DIPA anggaran pada kedua universitas. Revisi DIPA juga berdampak pada terlambatnya pelaksanaan program dan anggaran sehingga serapan anggaran universitas pun menjadi rendah. Strategi yang dilakukan oleh dikti adalah memajukan relokasi anggaran DIPA universitas anggaran pada setiap bulan Desember walaupun realisasinya pada bulan Maret, karena pada bulan Januari masih menyelesaikan laporan-laporan pada tahun sebelumnya, dan strategi demikian juga idealnya dilakukan universitas.

Semenjak diberlakukannya kebijakan anggaran PMW melalui DIPA universitas, berdampak pada pelaporan yang dilakukan oleh universitas tidak tepat waktu. Karena perguruan tinggi memiliki dinamika yang cukup tinggi. Dinamika dimaksudkan berkaitan dengan adanya perubahan strukturisasi dari pejabat lama ke yang baru, sehingga mekanisme kerja yang dipegang oleh pejabat lama sulit diadaptasi oleh pejabat yang baru, selain itu, komitmen dari pelaksana PMW pada setiap universitas yang masih rendah.

Selain itu, perubahan nomenklatur anggaran PMW pun berdampak pada kontrol dan pengawasan dari Dikti menjadi sangat terbatas. Idealnya dana yang dihibahkan harus di kompetisikan, bukan dijadikan grand hibah gratis, apalagi sudah dimasukkan dalam DIPA Universitas karena akan sulit melakukan pengontrolan, evaluasi, dan penyampaian laporan. Berbeda dengan tahun 2009 dengan status dana kompetisi, pengawasan maupun monitoring oleh dikti berjalan intensif sehingga dilihat pesertanya sungguh-sungguh melaksanakan PMW, walaupun ada yang belum berhasil mahasiswa akan diarahkan un-

tuk terus belajar memperbaiki hal-hal yang kurang baik.

### *Monitoring dan Evaluasi*

Hasil evaluasi pada aspek monitoring dan evaluasi pada kedua universitas tersebut berkategori tinggi. Hal ini ditandai dengan kedua universitas mengacu pada pedoman monitoring dan evaluasi yang telah ditetapkan oleh Dikti. Akan tetapi pada pelaksanaannya monitoring dan evaluasi memiliki kendala tersendiri bagi dikti, karena universitas memiliki respon yang berbeda-beda untuk menunjang setiap program laporan DIKTI. Kontrol yang dilakukan oleh dikti menganalisis hasil audit dari Inspektorat Jendral Dikti, rujukan inspektorat dalam melakukan monitoring pelaksanaan program adalah dari buku panduan PMW yang telah dikembangkan. Karena tugas monitoring yang dilakukan oleh inspektorat adalah berkaitan dengan laporan keuangan bukan pada laporan pelaksanaan program. Jika dalam pelaksanaan PMW keluar dari arahan dan petunjuk buku pedoman, maka menjadi temuan yang kemudian berdampak pada pengalokasian anggaran tahun berikutnya. Kalaupun sanksi yang diberikan dari pengurangan alokasi DIPA, universitas tetap mencari sumber pendanaan lain melalui kerjasama dengan mitra lembaga keuangan lainnya.

### *Komponen Keluaran*

Evaluasi pada komponen output terbagi dalam dua aspek evaluasi yakni: jumlah mahasiswa peserta PMW dengan kategori berhasil dalam berwirausaha; dan jumlah unit usaha baru mahasiswa. Adapun hasilnya sebagai berikut.

### *Jumlah Mahasiswa yang Wirausaha*

Hasil evaluasi pada aspek jumlah mahasiswa yang menjadi wirausaha dengan capaian evaluasi berkategori rendah. Hal ini disebabkan dalam kurung waktu 2011-2014, terdapat 820 mahasiswa yang berusaha. Hal ini menunjukkan bahwa motivasi mahasiswa untuk berwirausaha cukup tinggi. Akan tetapi dari hasil penelitian PMW dengan banyaknya jumlah wirausaha tersebut bukan berarti

setiap tahun mengalami peningkatan, malah cenderung menurun sebesar 35,5%. Hal ini disebabkan karena pola pikir mahasiswa dengan berwirausaha sambil kuliah akan menghambat proses akademik sehingga minat mahasiswa berkurang.

Sebagaimana hasil penelitian Suyanto & Ariadi (2015, p. 120) menunjukkan bahwa hanya 32% lulusan perguruan tinggi yang berhasil membangun usaha mandiri dengan rentang waktu rintisan membangun usaha antara 3-6 tahun. Kesulitan yang dihadapi dalam membangun usaha diantaranya modal yang terbatas, pemilihan jenis usaha dan terbatas koneksitas usaha.

Dari aspek PMW perubahan nomenklatur anggaran PMW juga memberikan kontribusi terhadap menurunnya jumlah mahasiswa yang terlibat dalam wirausaha, karena alokasi anggaran yang terbatas, disamping itu rendahnya daya serap anggaran untuk bantuan mahasiswa akibat dari kesalahan perencanaan anggaran seperti yang terjadi pada pelaksanaan PMW-UM pada tahun 2013 dan 2014 sehingga dana dikembalikan ke bagian keuangan universitas.

### *Unit Usaha Mahasiswa*

Hasil evaluasi pada aspek jumlah unit usaha mahasiswa dengan kategori rendah. Hal ini disebabkan terjadinya penurunan minat dan motivasi mahasiswa untuk bergabung dan menyelenggarakan PMW. Hasil penelitian ini pula sesuai dengan penelitian yang dilakukan oleh Budiyanto, Suprpto, & Poerwoningsih (2017, p. 393) menyimpulkan bahwa terdapat 40% mahasiswa yang berhasil menjadi wirausaha baru dengan keunggulan ipteks pada bidang produk kerajinan, makanan, dan jasa persewaan.

Dari perpektif jenis usaha, rata-rata mahasiswa mengembangkan usaha tidak sesuai dengan disiplin ilmu yang dimiliki, karena pengembangan usaha berdasarkan potensi dan kondisi pasar, misalnya karena Kota Malang Raya merupakan kota pariwisata, mahasiswa cenderung memilih usaha di bidang boga/makanan dan minuman.

Pada umumnya kondisi ini terjadi pada mahasiswa yang tidak mempunyai kemam-

puan teknis yang cukup misalkan mahasiswa program studi teknik yang bergerak dibidang budidaya ikan, budidaya lele, atau membuat restoran warung makan dan lain-lain sebenarnya bukan itu prioritas dalam PMW, akan tetapi yang diinginkan dari mahasiswa teknik itu dari sisi keilmuan mengembangkan usaha PMW sesuai dan relevan dengan kompetensi dan disiplin ilmu yang dimiliki.

## **Simpulan**

Berdasarkan hasil evaluasi dari tiga puluh kriteria keberhasilan program sebagaimana pada Tabel 2 penelitian ini, ditemukan bahwa terdapat dua puluh kriteria evaluasi yang berkategori tinggi, enam kriteria evaluasi yang berkategori moderat, dan empat kriteria evaluasi yang berkategori rendah. Ini berarti bahwa secara umum pelaksanaan PMW pada kedua universitas yang menjadi fokus penelitian ini telah berjalan berdasarkan kriteria standar objektif yang ditetapkan.

Berdasarkan hasil penelitian yang diuraikan pada pokok kesimpulan tersebut, rekomendasi yang dapat diberikan pada penelitian ini yaitu sebagai berikut.

Pertama, rekomendasi yang dapat diberikan kepada Direktorat Pembelajaran dan Kemahasiswaan Pendidikan Tinggi sebagai pengelola PMW Nasional diharapkan untuk: (1) mengevaluasi pelaksanaan PMW secara nasional, (2) mengupayakan agar nomenklatur dana PMW dikembalikan menjadi dana bergulir yang dikompetisikan ke setiap universitas, (3) meningkatkan alokasi dana PMW setiap tahun, (4) proses monitoring dan evaluasi pada PMW diharapkan untuk diberlakukan sama seperti tridarma lainnya yang dilakukan secara intensif dan sistematis; (5) mendorong gerakan kewirausahaan nasional melalui pemberdayaan universitas dengan program-program kewirausahaan; dan (6) memberikan sanksi dan penghargaan bagi universitas yang melanggar dan berhasil dalam pelaksanaan PMW.

Kedua, rekomendasi yang dapat diberikan kepada pelaksana dan civitas akademika di universitas yaitu supaya dapat: (1) mengevaluasi pelaksanaan PMW di tingkat universitas; (2) mempercepat proses pencairan DI-

PA universitas; (3) melaksanakan dan menyalurkan dana bantuan PMW tepat waktu sesuai dengan standar dan mekanisme yang telah ditetapkan secara nasional; (4) merencanakan dan mengalokasikan anggaran PMW kepada mahasiswa sesuai ketentuan, (5) berkoordinasi dengan Direktorat Pembelajaran dan Kemahasiswaan Pendidikan Dikti berkaitan dengan perubahan nomenklatur dana PMW yang semula DIPA universitas menjadi Hibah bergulir yang dikompetisikan ke mahasiswa; (6) membentuk tim pelaksanaan PMW yang permanen yang terintegrasi dengan struktur universitas; (7) membentuk unit-unit kewirausahaan mahasiswa pada setiap fakultas; (8) meningkatkan kerjasama dengan lembaga mitra universitas dengan memperdayakan dana bantuan sosial/dana CSR untuk kegiatan kewirausahaan mahasiswa; dan (9) menetapkan dan menerapkan mata kuliah kewirausahaan sebagai mata kuliah wajib di universitas.

Ketiga, rekomendasi yang dapat diberikan kepada mahasiswa adalah supaya: (1) meningkatkan minat dan partisipasinya dalam kegiatan kewirausahaan di universitas; (2) mengembangkan PMW secara berkelompok dan terintegrasi dari berbagai disiplin ilmu; (3) mengembangkan usaha yang berbasis teknoentrepreneur sesuai dengan disiplin ilmu; (4) memberikan pelaporan perkembangan secara berkala kepada pelaksana PMW di tingkat universitas; dan (5) meningkatkan kapasitas dan kapabilitas individu dan kelompok dalam PMW.

Keempat, rekomendasi yang dapat diberikan kepada lembaga mitra PMW yaitu agar dapat: (1) meningkatkan kerjasama dengan universitas melalui pemberian dana bantuan sosial/CSR lembaga mitra; (2) memperdayakan mahasiswa sukses berusaha sebagai mitra lembaga; (3) memberikan bantuan dana kredit ke mahasiswa yang sukses dalam berusaha; (4) meningkatkan peran dalam pembinaan dan pendampingan usaha mahasiswa; dan (5) bekerja sama dengan mahasiswa untuk meningkatkan profit usaha mahasiswa.

## Daftar Pustaka

- Akbar, S. (2007). *Pembelajaran nilai kewirausahaan dalam perspektif pendidikan umum: prinsip-prinsip dan vektor-vektor percepatan proses internalisasi nilai kewirausahaan*. Malang: UM Press.
- Alma, B. (2011). *Kewirausahaan untuk mahasiswa dan umum*. Bandung: Alfabeta.
- Ariyanti, F. (2013). RI mimpi punya wirausaha sebesar Amerika Serikat. Retrieved December 8, 2013, from <https://www.liputan6.com/bisnis/read/631916/ri-mimpi-punya-wirausaha-sebesar-amerika-serikat>
- Badan Pusat Statistik. (2013). *Perkembangan beberapa indikator utama sosial-ekonomi Indonesia Agustus 2013*. Jakarta: Badan Pusat Statistik.
- Badan Pusat Statistik Provinsi Jawa Timur. (2014). *Tingkat pengangguran terbuka (TPT) di Jawa Timur pada Agustus 2014*. Jawa Timur: Badan Pusat Statistik Provinsi Jawa Timur.
- Budiyanto, H., Suprpto, A., & Poerwoningsih, D. (2017). Program pengembangan kewirausahaan dalam bentuk inkubator bisnis di perguruan tinggi bagi mahasiswa pemilik usaha pemula. In *Seminar Nasional Sistem Informasi* (pp. 385–394). Malang: Fakultas Teknologi Informasi Universitas Merdeka Malang. Retrieved from <https://www.researchgate.net/publication/319999057>.
- Bygrave, W. D. (2010). The entrepreneurial proces. In W. D. Bygrave & A. Zacharakis (Eds.), *The Portable MBA in Entrepreneurship*. New Jersey: John Wiley & Sons, Inc.
- Cantillon, R. (2010). *Essay sur la nature du commerce un general (An essay on economic theory)*. (C. Saucier, Trans.). Auburn: Mises Intitute.
- Ciputra. (2008). *Ciputra quantum leap: Entrepreneurship mengubah masa depan bangsa dan masa depan anda*. Jakarta: PT. Elex Media Komputindo.
- Direktorat Jenderal Pendidikan Tinggi Kementerian Pendidikan dan Kebudayaan. (2009). *Rencana Strategis 2010-2014*. Jakarta: Direktorat Jenderal Pendidikan Tinggi Kementerian Pendidikan dan Kebudayaan.
- Hisrich, R. D., Peters, M. P., & Shepherd, D. A. (2008). *Kewirausahaan*. (C. Sungkono & D. Angelica, Trans.). Jakarta: Salemba Empat.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>
- Pusat Data Statistik Pendidikan. (2014). *Proyeksi data tingkat perguruan tinggi tahun 2013/2014-2020/2021*. Jakarta: Pusat Data Statistik Pendidikan.
- Scheerens, J., Glas, C., & Thomas, S. M. (2003). *Educational Evaluation, Assessment, and Monitoring: A Systemic Approach*. Lisse: Swets & Zeitlinger B. V.
- Siradjuddin, E. (2009). *Memerangi sindrom negara gagal transformasi Indonesia 2020 mencapai negara entrepreneur maju*. Jakarta: Kata Hasta Pustaka.
- Soenarto, S., Amin, M. M., & Kumaidi, K. (2017). Evaluasi implementasi kebijakan Sekolah Menengah Kejuruan program 4 tahun dalam meningkatkan employability lulusan. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 21(2), 215. <https://doi.org/10.21831/pep.v21i2.17076>
- Subianto, P. (2013). *Membangun kembali Indonesia Raya: strategi besar transformasi bangsa*. Jakarta: Institut Garuda Nusantara.
- Sutopo, A. (2012). Evaluasi efektivitas unit produksi dalam mempersiapkan kompetensi kerja siswa SMK. *Jurnal*

- Penelitian Dan Evaluasi Pendidikan*,  
16(2). Retrieved from  
<https://journal.uny.ac.id/index.php/jp>  
[ep/article/view/1124](https://journal.uny.ac.id/index.php/jp/ep/article/view/1124)
- Suardie, S. (2009). Model evaluasi kinerja tamatan pelatihan kewirausahaan Balai Diklat Pertanian DIY. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 13(2). Retrieved from  
<https://journal.uny.ac.id/index.php/jp>  
[ep/article/view/1407](https://journal.uny.ac.id/index.php/jp/ep/article/view/1407)
- Suyanto, B., & Ariadi, S. (2015). Upaya pengembangan usaha mandiri di kalangan pengangguran terdidik di Jawa Timur. *Jurnal Masyarakat, Kebudayaan, Dan Politik*, 28(3), 115–124.
- Tilaar, H. A. R., & Nugroho, R. (2012). *Kebijakan pendidikan*. Yogyakarta: Pustaka Pelajar.

**PENGEMBANGAN INSTRUMEN  
*EXPERIMENT PERFORMANCE ASSESSMENT*  
UNTUK MENILAI KETERAMPILAN PROSES SAINS DAN KERJA SAMA**

Raden Roro Yayuk Srirahayu<sup>1</sup> \*, Indyah Sulistyio Arty<sup>1</sup>

<sup>1</sup>Universitas Negeri Yogyakarta

<sup>1</sup>Jl. Colombo No. 1, Depok, Sleman 55281, Yogyakarta, Indonesia

\* Corresponding Author. Email: [ysrirahayu1@gmail.com](mailto:ysrirahayu1@gmail.com)

**Abstrak**

Penelitian ini bertujuan untuk: menghasilkan *instrumen experiment performance assessment* dengan model pembelajaran *guided inquiry*, membuktikan validitas dan estimasi reliabilitas produk instrumen, dan mengetahui keefektifan produk instrumen untuk meningkatkan keterampilan proses sains dan kerja sama. Metode yang digunakan adalah R & D (*Research and Development*) yang diadopsi dari Dick et al yang terbagi menjadi tahap pendahuluan, tahap perencanaan, tahap pengembangan, tahap uji coba terbatas dan tahap uji coba luas. Berdasarkan hasil uji coba diperoleh reliabilitas hasil rating instrumen penilaian keterampilan proses sains sebesar 0,707 - 0,989, dan reliabilitas skor komposit instrumen penilaian kerja sama sebesar 0,923 - 0,968. Hasil *t-test* menunjukkan terdapat peningkatan keterampilan proses sains dan kerja sama. Hasil penelitian ini adalah: (1) produk instrumen *experiment performance assessment* untuk pembelajaran IPA SMP materi sifat-sifat cahaya, pembentukan bayangan, dan prinsip kerja alat optik; (2) produk instrumen valid dan reliabel untuk menilai keterampilan proses sains dan kerja sama; (3) *instrumen experiment performance assessment* efektif untuk menilai keterampilan proses sains dan kerja sama.

**Kata kunci:** *experiment performance assessment, guided inquiry, keterampilan proses sains, kerja sama*

**DEVELOPMENT OF EXPERIMENT PERFORMANCE ASSESSMENT TO  
ASSESS SCIENCE PROCESS SKILLS AND COOPERATION**

**Abstract**

This study aims to: produce experiment performance assessment instrument with guided inquiry learning model, proof the validity and estimation of reliability of instrument products, and know the effectiveness of instruments to improve science process skills and cooperation. The method used is R & D (Research and Development) adopted from Dick et al., which is divided into preliminary stage, planning stage, development stage, limited and extensive testing stage. Based on limited and extensive testing trials, the reliability of the rating results of the science process skill instrument are 0.707 - 0.989, while the reliability of the composite scores of the cooperation assessment instrument obtained are 0.923 - 0.968. The t-test results show that there is an increase of science process skills and cooperation of learners. The results of this research are: (1) an experiment performance assessment product for science lesson of Junior High School material of light properties, shadow formation, and working principle of optical instrument; (2) the instrument product has valid and reliable categories for assessing science process skills and cooperation; (3) an effective experiment performance assessment instrument to assess science process skills and cooperation.

**Keywords:** *experiment performance assessment, guided inquiry, science process skills, cooperation*

Permalink/DOI: <http://dx.doi.org/10.21831/pep.v22i2.20270>

## Pendahuluan

IPA berkembang dari interpretasi data yang dikumpulkan melalui pengamatan, percobaan dan penyelidikan, sehingga IPA memiliki dimensi produk ilmiah (*scientific product*), proses ilmiah (*scientific process*), sikap ilmiah (*scientific attitude*), dan aplikasi dalam kehidupan sehari-hari (Bundu, 2006, pp. 4–13; National Science Teachers Association., 2009, pp. 61–65). Sesuai lampiran Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 22 Tahun 2016 tentang Standar Proses Pendidikan dan Menengah (Menteri Pendidikan dan Kebudayaan Republik Indonesia, 2016, pp. 3&11) maka pembelajaran IPA hendaknya menggunakan pendekatan saintifik dan berbasis penelitian (*discovey/ inquiry learning*).

IPA dicirikan oleh proses penyelidikan (National Science Teachers Association., 2009, p. 65). Hal ini diperkuat pendapat Trowbridge & Bybee (1990, pp. 231–240) bahwa kegiatan eksperimen merupakan kegiatan yang berperan dalam mengembangkan keterampilan proses. Peraturan Menteri Pendidikan dan Kebudayaan Nomor 21 Tahun 2016 tentang Standar Isi Pendidikan dan Menengah bagian lampiran (Menteri Pendidikan dan Kebudayaan, 2016, pp. 138–139) secara tegas menyatakan kompetensi muatan IPA pada peserta didik SMP adalah: (1) memiliki sikap ilmiah: rasa ingin tahu, logis, kritis, analitis, jujur, dan tanggung jawab melalui IPA; (2) mengajukan pertanyaan tentang fenomena IPA, melaksanakan percobaan, mencatat dan menyajikan hasil penyelidikan dalam bentuk tabel dan grafik, menyimpulkan, serta melaporkan hasil penyelidikan secara lisan maupun tertulis untuk menjawab pertanyaan tersebut; (3) memahami konsep dan prinsip IPA serta saling keterkaitannya dan diterapkan dalam menyelesaikan masalah.

Hasil penelitian Rahman, Wahyuni, & Rifqiwati (2017, pp. 1–6) di SMP Satu Atap menyatakan bahwa keterampilan proses sains dan sikap ilmiah masih rendah. Penyebabnya yaitu tidak tepatnya metode atau model pembelajaran. Penelitian Sukarno, Permasari, & Hamidah, (2013, pp. 79–83) juga mem-

berikan hasil bahwa kemampuan rata-rata keterampilan proses sains peserta didik SMP di Jambi masih relatif rendah. Demikian pula hasil observasi dan tes tertulis pada penelitian Rahmasiwi, Santosari, & Sari (2015, pp. 428–433) menunjukkan bahwa keterampilan proses sains siswa relatif rendah karena model pembelajaran yang diterapkan belum melatih keterampilan proses sains.

Pembelajaran IPA juga harus mengembangkan sikap ilmiah. Menurut National Curriculum Council (Bundu, 2006, p. 39) bahwa salah satu sikap ilmiah yang sangat penting dimiliki pada semua tingkatan pendidikan sains adalah bekerja sama dengan orang lain.

Penelitian Anjani, Suciati, & Maridi (2017, pp. 94–98) menyatakan bahwa keterampilan kerja sama dalam kelompok siswa masih tergolong kurang. Demikian juga hasil observasi Susilowati (2016, pp. 45–55) menunjukkan bahwa kerja sama antarsiswa dalam menyelesaikan tugas di kelompok masih rendah dan tanggung jawab masih kurang. National Science Teachers Association (2009, p. 173) menjelaskan bahwa dengan menggabungkan dan menekankan keterampilan bekerja sama bersamaan dengan pembelajaran IPA maka peserta didik akan mengembangkan keterampilan yang penting bagi mereka saat memasuki dunia kerja dan keterampilan itu penting bagi semua individu sepanjang hayat.

*Performance assessment* merupakan salah satu bentuk *authentic assessment* yang sesuai untuk diterapkan dalam proses penilaian pembelajaran IPA. Hal ini juga ditegaskan dalam Permendikbud Nomor 22 Tahun 2016 tentang Standar Proses Pendidikan Dasar dan Menengah bab V (Menteri Pendidikan dan Kebudayaan Republik Indonesia, 2016) yang menyatakan bahwa penilaian proses pembelajaran menggunakan pendekatan penilaian otentik (*authentic assesment*) yang menilai kesiapan peserta didik, proses, dan hasil belajar secara utuh. Okukawa (2007, p. 21) menyatakan bahwa *performance assessment* merupakan serangkaian strategi untuk merepresentasikan hasil dari suatu pengetahuan, keterampilan, dan kerja yang dilakukan oleh



peserta didik melalui tugas-tugas yang bermakna dan menarik.

Studi pendahuluan dengan subjek penelitian sejumlah 67 guru IPA SMP bertaraf Sekolah Standar Nasional (SSN) dan rintisan SSN di Kabupaten Gunungkidul, Bantul, Kulonprogo, dan Kota Yogyakarta menunjukkan hasil bahwa guru sudah melakukan berbagai metode pembelajaran, namun guru kesulitan melakukan penilaian kinerja eksperimen (*experiment performance assessment*) dan penilaian kerja sama selama proses pembelajaran. Alasannya adalah: guru belum memahami penerapannya, penyusunan instrumen dan pelaksanaannya memakan waktu yang lama, sementara waktu efektif kegiatan belajar mengajar cukup singkat, serta belum berkorelasi dengan nilai Ujian Nasional. Kendala dalam pelaksanaannya yaitu guru harus jeli mengamati setiap indikator penilaian yang ditampilkan peserta didik, harus menghafal seluruh peserta didik, harus menyiapkan alat dan bahan eksperimen, kemampuan peserta didik kurang memadai, serta rawan subjektivitas.

Hasil monitoring dan evaluasi pelaksanaan Kurikulum 2013 tingkat SMP (Kementerian Pendidikan dan Kebudayaan, 2017, p. 1) menunjukkan bahwa salah satu kesulitan pendidik mengimplementasikan Kurikulum 2013 adalah melaksanakan penilaian. Sekitar 60% responden menyatakan belum dapat merancang, melaksanakan, mengolah, melaporkan, dan memanfaatkan hasil penilaian dengan baik. Mereka kesulitan merumuskan indikator, menyusun butir-butir instrumen, melaksanakan penilaian sikap menggunakan berbagai macam teknik, dan melaksanakan penilaian keterampilan. Hal ini disebabkan karena mereka belum sepenuhnya memahami cara menyusun instrumen dan rubrik penilaian keterampilan.

Hal tersebut juga dinyatakan oleh Fadiawati & Astuti (2007, p. 24) bahwa pelaksanaan *performance assessment* belum banyak dikembangkan karena kemungkinan para pendidik belum mampu merencanakan dan mengembangkan *performance assessment*. Majid (2014, p. 59) menjelaskan bahwa sebagian guru tidak mau menggunakan penilaian

kinerja karena membuang waktu dan energi serta terlalu mahal.

Berdasarkan latar belakang tersebut, terdapat beberapa masalah yang ditemukan di lapangan. Pelaksanaan Kurikulum 2013 membawa implikasi pada perubahan pembelajaran menggunakan pendekatan saintifik berbasis penelitian dan penilaian autentik, sementara guru dan peserta didik belum siap. Penilaian proses pembelajaran IPA belum dilaksanakan secara optimal dan masih menekankan pada aspek kognitif, selain itu *experiment performance assessment* belum berkorelasi dengan penilaian Ujian Nasional, sementara guru dan peserta didik masih kesulitan untuk melakukan penilaian autentik. Padahal penilaian keterampilan proses sains dan penilaian kerja sama yang dilaksanakan melalui eksperimen tidak dapat dilakukan dengan tes tertulis. Guru kesulitan dalam mengembangkan instrumen *experiment performance assessment* terutama dalam hal menyusun indikator dan rubrik penilaian. Pelaksanaan *experiment performance assessment* memerlukan kejelian dan keahlian guru, sementara guru belum terbiasa melakukannya.

Berdasarkan uraian yang telah disampaikan tersebut, penelitian bertujuan menghasilkan instrumen *experiment performance assessment* dengan model pembelajaran *guided inquiry*, membuktikan validitas dan estimasi reliabilitas produk instrumen untuk menilai keterampilan proses sains dan kerja sama peserta didik SMP pada materi sifat-sifat cahaya, pembentukan bayangan dan prinsip kerja alat optik secara *expert judgement* dan secara empiris, dan mengetahui keefektifan produk instrumen untuk meningkatkan keterampilan proses sains dan kerja sama.

## Metode Penelitian

Penelitian ini adalah penelitian pengembangan instrumen *experiment performance assessment* untuk menilai keterampilan proses sains dan kerja sama dalam pembelajaran IPA di SMP dengan model pembelajaran *guided inquiry*. Instrumen yang dikembangkan meliputi Silabus, Rencana Pelaksanaan Pembelajaran (RPP), *handout*, Lembar Kerja Peserta Didik (LKPD), Instrumen Penilaian

Keterampilan Proses Sains, Instrumen Penilaian Kerja sama dan Soal Tes Hasil Belajar Kognitif.

Model penelitian pengembangan mengacu pada model yang dikemukakan oleh Dick, Carey, & Carey (2005, pp. 1–8) terdiri dari 10 tahap. Pada penelitian ini, keseluruhan tahapan pengembangan yang dilakukan, namun kesepuluh tahapan pengembangan tersebut dikelompokkan, diadopsi menjadi lima tahap, yakni: tahap (1) *identify instructional goals*, (2) *conduct instructional analysis*, (3) *analyze learners and contexts* dikelompokkan menjadi Tahap Pendahuluan, tahap (4) *write performance objectives* merupakan Tahap Perencanaan, tahap (5) *develop assessment instruments* dan (6) *revise instruction* dikelompokkan menjadi Tahap Pengembangan Produk, dan tahap (7) *develop instructional strategy*, (8) *develop and select instructional materials*, (9) *design and conduct formative evaluation of instruction*, dan (10) *design and conduct summative evaluation* dilaksanakan pada Tahap Uji Coba Terbatas dan Tahap Uji Coba Luas.

Pada tahap pendahuluan dilakukan identifikasi tujuan pengembangan, kajian pustaka dan studi lapangan. Pada tahap perencanaan dilakukan penetapan kriteria penilaian dan desain produk. Pada tahap pengembangan produk dilakukan penyusunan draf produk instrumen *experiment performance assessment* yang kemudian divalidasi secara *expert judgement* oleh sepuluh validator yang terdiri dari tiga rekan sejawat, dua dosen dan 5 guru IPA SMP. Draft produk instrumen juga divalidasi secara empiris kepada peserta didik untuk mengetahui tingkat keterbacaan produk. Keterlaksanaan proses pembelajaran model *guided inquiry* diobservasi oleh guru dan observer mahasiswa yang sudah dilatih. Masukan dari validator, observer, dan peserta didik digunakan sebagai revisi draft produk instrumen.

Kegiatan uji coba skala terbatas dalam penelitian ini dilakukan di SMP Negeri 2 Playen. Uji coba skala luas dilakukan di SMP Negeri 2 Karangmojo, SMP Negeri 1 Playen, SMP Negeri 1 Ponjong dan SMP Negeri 1 Semanu. Subjek uji coba terbatas melibatkan

32 peserta didik, sedangkan untuk uji coba skala luas melibatkan 124 peserta didik.

Teknik pengumpulan data menggunakan teknik tes dan non-tes. Teknik tes dilakukan untuk menilai validitas empiris produk instrumen *experiment performance assessment* untuk menilai keterampilan proses sains dan kerja sama, dan hasil belajar kognitif. Teknik pengumpulan data non tes digunakan untuk menilai validitas konstruk produk instrumen.

Instrumen yang digunakan untuk mengumpulkan data penelitian terdiri dari: (1) Lembar Validasi Produk Instrumen *Experiment Performance Assessment*, untuk mendapatkan data kuantitatif berupa skor kelayakan produk serta data kualitatif berupa saran dari validator; (2) Angket Respon Peserta Didik, untuk mengetahui penilaian peserta didik terhadap *handout* dan LKPD, serta respon terhadap proses pembelajaran; (3) Instrumen Observasi Pelaksanaan Proses Pembelajaran, untuk mengetahui keterlaksanaan sintaks model pembelajaran *guided inquiry*; (4) Soal keterampilan proses sains yang didesain dalam bentuk LKPD sebagai panduan melakukan kinerja; (5) Lembar observasi dan rubrik penilai keterampilan proses sains; (6) Lembar observasi dan rubrik penilai kerja sama; (7) Soal Tes Hasil Belajar Kognitif

Teknik analisis data dilakukan dengan cara sebagai berikut.

Penilaian produk silabus, RPP, *handout*, LKPD oleh validator dan penilaian produk *handout* dan LKPD berdasar angket respon peserta didik dianalisis secara deskriptif menggunakan indeks skala baku skala empat mengikuti panduan yang diberikan oleh Mardapi (2004, p. 123).

Data kuantitatif yang berupa skor penilaian validator terhadap produk berupa Instrumen Penilaian Keterampilan Proses Sains, Instrumen Penilaian Kerja sama, dan Soal Tes Hasil Belajar Kognitif dianalisis menggunakan indeks validitas Aiken (Aiken, 1980, pp. 955–957; Azwar, 2012, pp. 112–113).

Data kuantitatif dari validasi empiris soal hasil belajar kognitif setelah diujicobakan kepada peserta didik dianalisis dengan model *Rasch* dengan bantuan *software Winstep*

(Sumintono & Widhiarso, 2013, pp. 109–111) untuk mengetahui validitas dan reliabilitas soal. Reliabilitas soal dapat dicermati dari nilai *person reliability*, *item reliability* dan *alpha cronbach*.

Untuk mengetahui reliabilitas instrumen penilaian keterampilan proses sains, hasil penilaian dari tiga observer dianalisis total skornya dengan metode rating (Azwar, 2012, pp. 88–90). Untuk mengetahui reliabilitas instrumen penilaian kerja sama, maka hasil penilaian observer, penilaian diri sendiri dan penilaian antarteman dianalisis menggunakan reliabilitas skor komposit yang disarankan oleh Mosier (Azwar, 2012, pp. 84–86). Untuk mengetahui keefektifan instrumen *experiment performance assessment* dalam menilai keterampilan proses sains dan kerja sama maka dilakukan analisis *t-test* antara eksperimen pertama dengan eksperimen kelima. Selanjutnya untuk mengetahui peningkatan hasil belajar kognitif dilakukan analisis dengan teknik *gain score* (Hake, 1999, pp. 1–4) (Hake, 1999, pp.1-4) terhadap hasil *pretest-posttest*.

Data kuantitatif respon peserta didik terhadap proses pembelajaran dan skor hasil observasi pelaksanaan proses pembelajaran dianalisis secara deskriptif menggunakan indeks skala baku (Mardapi, 2004, p. 123).

## Hasil Penelitian dan Pembahasan

### Hasil Pengembangan Produk Awal

Produk awal instrumen *experiment performance assessment* merupakan produk yang telah melewati beberapa tahap pengembangan.

### Hasil Tahap Pendahuluan

Hasil dari tahap pendahuluan yaitu dirumuskannya tujuan penelitian, yakni menghasilkan produk *instrumen experiment performance assessment* dengan model pembelajaran *guided inquiry* untuk peserta didik SMP pada mata pelajaran IPA materi sifat-sifat cahaya, pembentukan bayangan, dan prinsip kerja alat optik. Survey lapangan telah dilakukan di 16 SMP di Kota Yogyakarta, Kabupaten Bantul, Kabupaten Gunungkidul dan Kabu-

paten Kulonprogo, dan melibatkan 67 guru IPA sebagai responden, dan disimpulkan bahwa guru kesulitan melakukan *experiment performance assessment*.

### Hasil Tahap Perencanaan

Hasil pada tahap perencanaan adalah perumusan kriteria *experiment performance assessment* berdasarkan kajian Kompetensi Inti dan Kompetensi Dasar menjadi indikator penilaian keterampilan proses sains dan kerja sama. Keterampilan proses sains yang dinilai terdiri dari keterampilan: menyusun hipotesis, mengamati, melakukan eksperimen, klasifikasi, pengukuran, komunikasi, dan inferensi (Rezba, McDonnough, Matkins, & Sprague, 2007, p. 5). Kemampuan kerja sama yang dinilai terdiri dari empat tingkatan, yaitu tahap membentuk, tahap memfungsikan, tahap merumuskan, dan tahap mengembangkan (Goodwin, 1999, p. 30).

### Hasil Tahap Pengembangan

Hasil tahap pengembangan berupa draf produk instrumen *experiment performance assessment* terdiri: (1) silabus; (2) RPP; (3) *handout*; (4) LKPD; (5) instrumen penilaian keterampilan proses sains; (6) instrumen penilaian kerja sama; dan (7) soal tes hasil belajar kognitif yang terdiri dari soal pilihan ganda dan soal uraian. Hasil validasi produk oleh sepuluh validator telah dianalisis secara deskriptif kualitatif.

Berdasarkan hasil analisis secara deskriptif menggunakan indeks skala baku skala 4, keseluruhan aspek dan indikator penilaian produk silabus dan RPP mendapat penilaian sangat baik. Skor rerata silabus aspek kelayakan isi: 3,95, aspek kebahasaan: 3,98 dan aspek penyajian 4,00. Skor rerata RPP aspek kelayakan isi: 3,93, aspek kebahasaan: 3,97 dan aspek penyajian 4,00.

Produk *handout* mendapat penilaian sangat baik dari validator. Skor rerata aspek kelayakan isi: 3,96, aspek kebahasaan: 3,91, aspek penyajian 3,94 dan aspek kegrafisan 3,93. Validator memberikan masukan agar dalam *handout* diberi contoh soal dan penge-tikan persamaan dengan menggunakan *equation*. Produk Lembar Kerja Peserta Didik

(LKPD) dinilai sangat baik dari validator. Skor rerata aspek kelayakan isi: 3,94, aspek kebahasaan: 3,94, aspek penyajian 3,98 dan aspek kegrafisan 3,97. Gambar pelangi direvisi menjadi gambar orang bercermin, dan pengetikan persamaan dengan menggunakan *equation*.

Instrumen penilaian keterampilan proses sains berupa LKPD sebagai pedoman dalam melaksanakan kinerja, lembar observasi dan pedoman penskoran untuk menilai kinerja peserta didik. LKPD didesain dengan model pembelajaran *guided inquiry*. Instrumen penilaian disusun dalam bentuk LKPD dan tidak ditampilkan dalam bentuk soal agar *setting* pembelajaran bersifat alami dan peserta didik dapat memaksimalkan kinerjanya (Jacobsen, Eggen, & Kauchak, 2009, p. 302). Hal ini juga sesuai dengan prinsip-prinsip *performance assessment* yaitu proses penilaian merupakan bagian yang tak terpisahkan dari proses pembelajaran. Berdasarkan hasil analisis indeks validitas Aiken, keseluruhan butir instrumen penilaian keterampilan proses sains dinilai sangat berguna.

Instrumen penilaian kerja sama berupa lembar observasi dan sekaligus pedoman penskoran dan lembar penilaian diri dan antarteman untuk menilai kerja sama peserta didik dalam melaksanakan eksperimen dan mendiskusikan hasil eksperimen. Penilaian kompetensi sikap dapat dilakukan melalui observasi, penilaian diri, penilaian teman sejawat (*peer evaluation*) dan jurnal (Majid, 2014, p. 77). Berdasarkan hasil analisis indeks validitas Aiken, keseluruhan aspek penilaian produk instrumen lembar observasi penilaian kerja sama dinilai sangat berguna oleh validator.

Data kuantitatif yang berupa skor penilaian validator terhadap produk berupa Soal Tes Hasil Belajar Kognitif, dianalisis menggunakan indeks validitas Aiken, dan hasilnya keseluruhan butir soal pilihan ganda dan soal uraian dinilai sangat berguna.

Produk yang telah divalidasi oleh ahli kemudian digunakan di satu kelas di SMP Negeri 1 Playen. Hasil Validasi Empiris *Handout* dan LKPD dianalisis secara deskriptif menggunakan indeks skala baku, dan

hasilnya keseluruhan aspek produk *handout* dan LKPD direspon sangat baik oleh peserta didik.

Peserta didik memberikan respon bahwa pembelajaran menarik, mendorong aktif, memotivasi, mendukung peningkatan keimanan, mengembangkan sikap ilmiah, meningkatkan keterampilan proses sains, membelajarkan kerja sama, dan menyenangkan.

Hasil Validasi Empiris Instrumen Penilaian Keterampilan Proses Sains dan Instrumen Kerja sama diketahui bahwa keseluruhan indikator-indikator penilaian keterampilan proses sains dan kerja sama dapat diamati dengan jelas.

Soal hasil belajar kognitif, diujikan kepada 50 peserta didik dan dianalisis menggunakan bantuan *software winstep*. Secara umum keseluruhan soal memperoleh nilai *person reliability* 0,86 (bagus) dan *item reliability* 0,93 (bagus sekali), serta *alpha cronbach* 0,87 (bagus sekali). Namun ada soal pilihan ganda yang *misfit*, sehingga akhirnya dipilih 40 butir soal pilihan ganda. Hasil analisis 40 butir soal pilihan ganda mendapat nilai *person reliability* 0,90 (bagus) dan *item reliability* 0,93 (bagus sekali), serta *alpha cronbach* 0,92 (bagus sekali). Hasil analisis tingkat kesulitan dan ukuran keacakan menunjukkan ada 5 butir soal yang kurang bagus tetapi tidak menurunkan kualitas. Hasil kesesuaian data dengan model menunjukkan semua soal mempunyai perkiraan logis. Untuk soal uraian, tidak ada yang *misfit* sehingga semua dapat digunakan.

Berdasarkan hasil analisis indeks skala baku, dapat diketahui pelaksanaan proses pembelajaran adalah sangat baik. Artinya proses pembelajaran sudah sesuai dengan RPP dan sesuai sintaks *guided inquiry*.

#### Hasil Uji Coba Produk

Tahap uji coba terdiri dari uji coba terbatas dan uji coba luas.

#### Hasil Uji Coba Terbatas

Uji coba terbatas dilakukan di SMP Negeri 2 Playen. Berdasarkan hasil analisis menggunakan indeks skala baku, keseluruhan aspek produk *handout* direspon sangat baik oleh subjek uji coba terbatas. Skor rerata

aspek kelayakan isi: 3,75, aspek kebahasaan: 3,82 dan aspek penyajian 3,80. Produk LKPD juga direspon sangat baik. Skor rerata aspek kelayakan isi: 3,78, aspek kebahasaan dan kegrafisan: 3,83 dan aspek penyajian 3,81. Observer memberikan masukan supaya disajikan gambar susunan peralatan yang lengkap untuk setiap letak benda pada eksperimen pembentukan bayangan pada cermin cekung dan lensa cembung sehingga mempermudah pelaksanaan eksperimen.

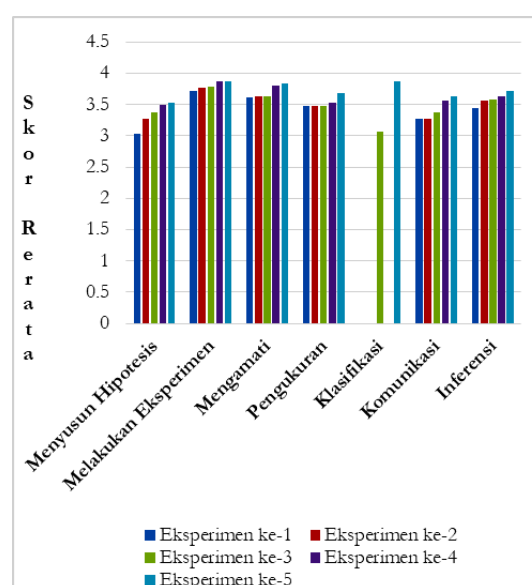
Respon peserta didik terhadap pembelajaran adalah sangat baik. Berdasarkan pengamatan selama pembelajaran dengan model *guided inquiry*, peserta didik harus dipandu oleh guru dalam melakukan eksperimen, dan dibimbing untuk menemukan dan menyimpulkan konsep berdasarkan hasil percobaan.

Tabel 1 dan grafik pada Gambar 1 menyajikan hasil uji coba terbatas untuk penilaian keterampilan proses sains pada Uji Coba Terbatas. Berdasarkan Tabel 1 dan grafik pada Gambar 1 dapat diketahui bahwa secara umum keterampilan proses sains peserta didik meningkat 9%. Hasil *t-test* hasil *t* hitung = 8,423 lebih besar dari *t* tabel = 1,696. Ini berarti bahwa penggunaan instrumen *experiment performance assessment* dapat menilai keterampilan proses sains peserta didik.

Hasil perhitungan reliabilitas penilaian keterampilan proses sains disajikan pada Tabel 2. Kategori koefisien reliabilitas adalah sebagai berikut:  $0,90 < r_{11} \leq 1,00$  reliabilitas sangat tinggi,  $0,70 < r_{11} \leq 0,90$  reliabilitas tinggi,  $0,40 < r_{11} \leq 0,70$  reliabilitas sedang,  $0,20 < r_{11} \leq 0,40$  reliabilitas rendah, dan kurang dari 0,20 reliabilitas sangat rendah (Guilford, 1956, p. 145). Berdasarkan Tabel 2, dapat diketahui bahwa koefisien reliabilitas penilaian keterampilan proses sains adalah tinggi dan sangat tinggi, berarti instrumen penilaian keterampilan proses sains dapat digunakan dengan baik di SMP Negeri 2 Playen pada uji coba terbatas.

Tabel 1. Hasil Penilaian Keterampilan Proses Sains pada Uji Coba Terbatas

Aspek Keterampilan Proses Sains	Skor Eksperimen ke-				
	1	2	3	4	5
Menyusun hipotesis	3,03	3,28	3,38	3,50	3,53
Melakukan eksperimen	3,71	3,77	3,79	3,88	3,88
Mengamati	3,61	3,63	3,63	3,81	3,84
Pengukuran	3,47	3,47	3,47	3,53	3,69
Klasifikasi			3,06		3,88
Komunikasi	3,27	3,27	3,37	3,56	3,64
Inferensi	3,44	3,56	3,58	3,64	3,72
Rerata	3,42	3,50	3,47	3,65	3,74



Gambar 1. Hasil Penilaian Keterampilan Proses Sains pada Uji Coba Terbatas

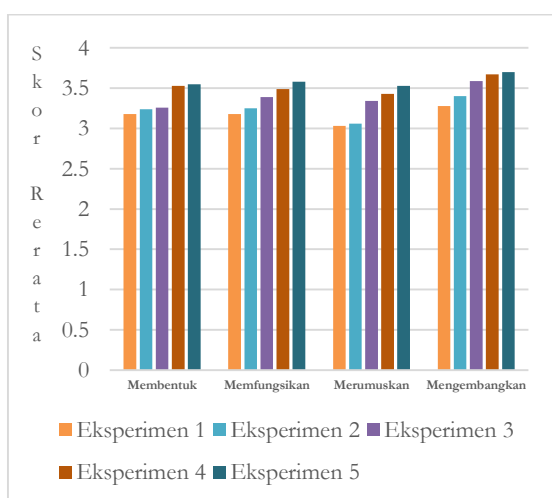
Tabel 2. Hasil Koefisien Reliabilitas Rating Penilaian Keterampilan Proses Sains pada Uji Coba Terbatas

Eksperimen ke	Koefisien Reliabilitas	
	Rata-rata rating 3 rater	Rata-rata skor orang rater
1	0,989	0,967
2	0,954	0,874
3	0,984	0,956
4	0,959	0,887
5	0,962	0,894

Kerja sama dinilai oleh tiga penilai, yakni observer, diri sendiri dan teman. Panduan Penilaian oleh Pendidik dan Satuan Pendidikan untuk SMP (Kementerian Pendidikan dan Kebudayaan, 2017, pp. 21–37) menjelaskan bahwa penilaian diri dan penilaian antarteman dapat dilakukan dalam rangka pembinaan dan pembentukan karakter peserta didik, yang hasilnya dapat dijadikan sebagai salah satu data konfirmasi penilaian sikap oleh pendidik. Selain itu penilaian diri peserta didik juga dapat digunakan untuk menumbuhkan nilai-nilai kejujuran dan meningkatkan kemampuan refleksi atau mawas diri. Penilaian antarteman juga dapat digunakan untuk menumbuhkan nilai kejujuran, tenggang rasa, dan saling menghargai. Hasil rerata penilaian kerja sama peserta didik SMP Negeri 2 Playen pada uji coba terbatas dapat dilihat pada tabel 3 dan Gambar 2.

Tabel 3. Hasil Penilaian Kerja sama Peserta Didik SMP Negeri 2 Playen pada Uji Coba Terbatas

Aspek Kerja sama	Skor Eksperimen ke-				
	1	2	3	4	5
Membentuk	3,18	3,24	3,26	3,53	3,55
Memfungsikan	3,18	3,25	3,39	3,49	3,58
Merumuskan	3,03	3,06	3,34	3,43	3,53
Mengembangkan	3,28	3,40	3,59	3,67	3,70
Rerata	3,17	3,23	3,40	3,53	3,59



Gambar 2. Hasil Penilaian Kerja sama Peserta Didik SMP Negeri 2 Playen pada Uji Coba Terbatas

Tabel 3 dan Gambar 2 memperlihatkan bahwa secara umum kerja sama peserta didik meningkat 13%. Hasil *t-test* terhadap kerja sama peserta didik SMP Negeri 2 Playen diperoleh hasil  $t$  hitung = 16,574 lebih besar dari  $t$  tabel = 1,696. Ini berarti bahwa terjadi peningkatan kemampuan kerja sama pada peserta didik.

Hasil perhitungan koefisien reliabilitas skor komposit penilaian kerja sama disajikan pada Tabel 4. Koefisien reliabilitas skor komposit penilaian kerja sama menunjukkan hasil sangat tinggi (Guilford, 1956, p. 145). Berdasarkan hasil perhitungan koefisien reliabilitas skor komposit dapat disimpulkan bahwa instrumen penilaian kerja sama dapat digunakan dengan baik di SMP Negeri 2 Playen pada uji coba terbatas.

Tabel 4. Koefisien Reliabilitas Skor Komposit Penilaian Kerja sama pada Uji Coba Terbatas

Eksperimen ke	Koefisien Reliabilitas Skor Komposit	Hasil
1	0,945	Sangat Tinggi
2	0,935	Sangat Tinggi
3	0,941	Sangat Tinggi
4	0,944	Sangat Tinggi
5	0,959	Sangat Tinggi

Rerata nilai *pretest* adalah 26,22 dan rerata nilai *posttest* adalah 82,66. Berdasarkan perhitungan diperoleh *gain score* sebesar 0,76 yang artinya terdapat peningkatan hasil belajar kognitif peserta didik subjek uji coba terbatas dalam kategori tinggi.

Pelaksanaan proses pembelajaran diobservasi dan dinilai menggunakan Instrumen supervisi akademik berdasarkan Kurikulum 2013. Kategori nilai proses pembelajaran berdasarkan instrumen supervisi akademik adalah sebagai berikut: 86-100 proses sangat baik, 70-85 proses baik, 55-69 proses cukup, dan kurang dari 55 nilai kurang. Hasil observasi proses pembelajaran memperoleh nilai 89-95 dengan kategori sangat baik. Artinya proses sesuai dengan RPP dan sesuai sintaks *guided inquiry*.



Hasil Uji Coba Luas

Uji coba luas dilakukan di SMP Negeri 2 Karangmojo, SMP Negeri 1 Playen, SMP Negeri 1 Ponjong dan SMP Negeri 1 Semanu. Eksperimen yang dilakukan sebanyak 5 eksperimen, yakni: pemantulan cahaya, pembentukan bayangan pada cermin datar, pembentukan bayangan pada cermin cekung, pembiasan cahaya dan pembentukan bayangan pada lensa cembung. Selama kegiatan eksperimen dilakukan penilaian terhadap keterampilan proses sains dan penilaian kerja sama. Pada awal kegiatan dilakukan *pretest* dan pada akhir pembelajaran dilakukan *posttest*.

Berdasarkan hasil analisis indeks skala baku, keseluruhan aspek produk *handout* direspon sangat baik oleh subjek uji coba luas. Skor rerata aspek kelayakan isi: 3.53, 3.43, 3.38, dan 3.34. Skor rerata aspek kebahasaan: 3.56, 3.67, 3.41, dan 3.33. Skor rerata aspek penyajian: 3.69, 3.66, 3.44, dan 3.37. Produk LKPD juga direspon sangat baik oleh peserta didik. Skor rerata aspek kelayakan isi: 3.47, 3.50, 3.45, dan 3.31. Skor rerata aspek kebahasaan dan kegrafisan: 3.56, 3.62, 3.48, dan 3.29. Skor rerata aspek penyajian LKPD: 3.66, 3.53, 3.49, dan 3.32. Peserta didik dapat menggunakan *handout* dan LKPD dengan baik dan merasa terbantu.

Berdasarkan perhitungan indeks skala baku, diketahui respon peserta didik terhadap pembelajaran adalah sangat baik. Peserta didik memberikan respon bahwa pembelajaran menstimulasi rasa ingin tahu, menarik, mendorong aktif, memotivasi, mendukung peningkatan keimanan, mengembangkan sikap ilmiah, meningkatkan keterampilan proses sains, membelajarkan kerja sama, dan menyenangkan. Berdasarkan pengamatan selama pelaksanaan pembelajaran dengan model *guided inquiry*, peserta didik antusias dan berpartisipasi aktif dalam melakukan eksperimen. Namun selama pembelajaran, peserta didik harus dibimbing supaya hasil pembelajaran sesuai dengan harapan.

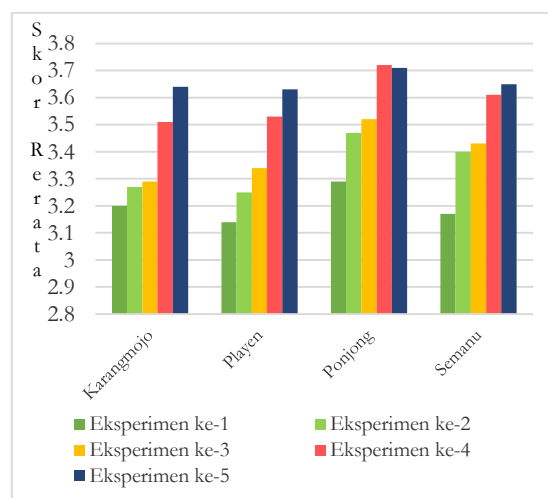
Penilaian keterampilan proses sains peserta didik SMP Negeri 2 Karangmojo, SMP Negeri 1 Playen, SMP Negeri 1

Ponjong dan SMP Negeri 1 Semanu pada uji coba luas disajikan Tabel 5 dan grafik pada Gambar 3.

Hasil *t-test* terhadap keterampilan proses sains peserta didik di SMP Negeri 2 Karangmojo diperoleh hasil  $t$  hitung = 15,688, di SMP Negeri 1 Ponjong diperoleh hasil  $t$  hitung = 6,052, dan di SMP Negeri 1 Semanu diperoleh hasil  $t$  hitung = 12,276. Keseluruhan hasil  $t$  hitung lebih besar dari  $t$  tabel = 1,696. Demikian juga hasil *t-test* terhadap keterampilan proses sains peserta didik di SMP Negeri 1 Playen diperoleh hasil  $t$  hitung = 8,193 lebih besar dari  $t$  tabel = 1,701. Ini berarti bahwa penggunaan instrumen *experiment performance assessment* dapat meningkatkan keterampilan proses sains pada peserta didik uji luas.

Tabel 5. Hasil Penilaian Keterampilan Proses pada Uji Coba Luas

SMP	Skor Keterampilan Proses Sains Eksperimen ke-				
	1	2	3	4	5
SMP N 2 Karangmojo	3,20	3,27	3,29	3,51	3,64
SMP N 1 Playen	3,14	3,25	3,34	3,53	3,63
SMP N 1 Ponjong	3,29	3,47	3,52	3,72	3,71
SMP N 1 Semanu	3,17	3,40	3,43	3,61	3,65



Gambar 3. Grafik Ringkasan Hasil Penilaian Keterampilan Proses Sains Peserta Didik pada Uji Coba Luas

Tabel 5 dan Gambar 3 memperlihatkan bahwa keterampilan proses sains peserta didik SMP Negeri 2 Karangmojo, SMP

Negeri 1 Playen, SMP Negeri 1 Ponjong dan SMP Negeri 1 Semanu, secara berturut-turut meningkat sebesar 14%, 16%, 13%, dan 15%. Senada dengan hasil temuan Suryandari (2013, p. 30), selama penelitian, dapat diamati bahwa peserta didik merasa dinilai dengan rubrik yang kriteria penilaiannya jelas dan detail, sehingga peserta didik lebih sungguh-sungguh dalam melakukan eksperimen, serta menunjukkan kinerja terbaiknya.

Pembelajaran kooperatif melalui kegiatan eksperimen secara berkelompok yang terdiri dari 4 peserta didik juga dapat menjadi sebab peningkatan keterampilan proses sains. Bertucci, Conte, Johnson, & Johnson (2010, pp. 270–271) mengungkapkan bahwa peserta didik dalam kelompok yang berisi 2 atau 4 orang menunjukkan perolehan nilai yang lebih tinggi dibandingkan yang hanya mengerjakan secara individu atau dalam kelompok besar.

Raj & Devi (2014, p. 2442) menyarankan guru sains sebaiknya mempersempit celah antara sains dan aplikasinya di kehidupan sehari-hari melalui kegiatan eksperimen, karena eksperimen mampu mengembangkan kemampuan peserta didik dan keterampilan prosedural yang digunakan mereka dalam karirnya kelak. Karamustafaoğlu (2011, p. 33) juga menegaskan bahwa keterampilan proses sains merupakan sesuatu yang krusial bagi pembelajaran yang bermakna karena belajar berlangsung selama hidup manusia, dan setiap manusia perlu menemukan, menginterpretasikan dan menilai bukti-bukti berdasarkan keadaan yang berbeda-beda.

Hasil perhitungan reliabilitas rating instrumen penilaian keterampilan proses sains adalah 0,707-0,980 dengan kategori tinggi dan sangat tinggi (Guilford, 1956, p. 145). Berarti instrumen dapat digunakan dengan baik di SMP Negeri 2 Karangmojo, SMP Negeri 1 Playen, SMP Negeri 1 Ponjong dan SMP Negeri 1 Semanu.

Penilaian kerja sama diperoleh dari skor komposit hasil penilaian observer, diri sendiri dan antarteman. Penilaian diri merupakan teknik penilaian di mana peserta didik diminta untuk menilai dirinya sendiri berkaitan dengan status, proses dan tingkat

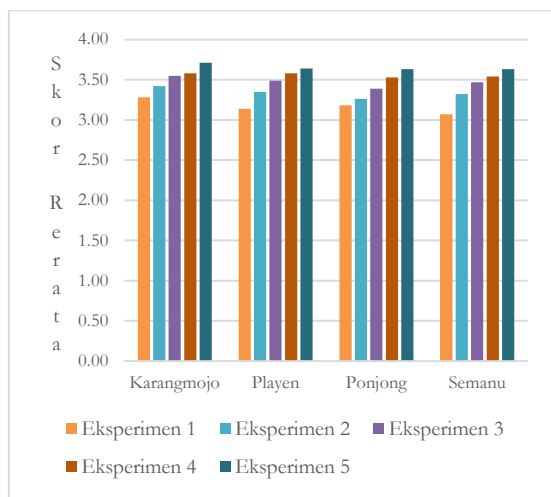
pencapaian kompetensi yang dipelajarinya (Majid, 2014, pp. 65–66). Teknik penilaian diri dapat digunakan untuk mengukur kompetensi kognitif, afektif dan psikomotor dan bermanfaat untuk menumbuhkan rasa percaya diri, membuat peserta didik menyadari kekuatan dan kelemahan dirinya, melatih berperilaku jujur, dan menumbuhkan semangat untuk maju secara personal.

Penilaian antarteman merupakan teknik penilaian dengan cara meminta peserta didik untuk saling menilai temannya terkait dengan pencapaian kompetensi, sikap, dan perilaku. Karami & Rezaei, 2015 (p. 96) melakukan studi literatur dan memperoleh informasi bahwa penilaian antarteman memiliki pengaruh positif terhadap kerja kelompok, kemampuan sosial dan motivasi. Studi literatur tersebut juga menyarankan sebaiknya lebih melibatkan peserta didik dalam pelaksanaan penilaian antarteman, memberikan dukungan dan pelatihan supaya peserta didik dapat menilai temannya dengan baik, dan membuat kriteria penilaian yang jelas.

Panduan Penilaian oleh Pendidik dan Satuan Pendidikan untuk SMP (Kementerian Pendidikan dan Kebudayaan, 2017, pp. 21–38) menegaskan bahwa penilaian pencapaian hasil belajar seharusnya lebih mengutamakan pendekatan penilaian sebagai pembelajaran (*assessment as learning*) yaitu dengan melibatkan peserta didik secara aktif dalam kegiatan penilaian seperti penilaian diri dan penilaian antarteman. Peserta didik diberi pengalaman untuk belajar menjadi penilai bagi dirinya sendiri dan juga dapat dilibatkan dalam merumuskan prosedur penilaian, kriteria, maupun rubrik penilaian sehingga mereka mengetahui dengan pasti apa yang harus dilakukan agar memperoleh capaian belajar yang maksimal. *Assessment as learning* dilakukan selama proses pembelajaran berlangsung dan juga digunakan sebagai dasar untuk melakukan perbaikan proses belajar mengajar.

Hasil penilaian kerja sama peserta didik SMP Negeri 2 Karangmojo, SMP Negeri 1 Playen, SMP Negeri 1 Ponjong dan SMP Negeri 1 Semanu disajikan pada Tabel 6 dan grafik pada Gambar 4





Gambar 4. Grafik Ringkasan Hasil Penilaian Kerja sama Peserta Didik pada Uji Coba Luas

Tabel 6. Hasil Penilaian Kerja sama pada Uji Coba Luas

SMP	Skor Kerja sama Eksperimen ke				
	1	2	3	4	5
SMP N 2 Karangmojo	3,28	3,42	3,55	3,58	3,71
SMP N 1 Playen	3,14	3,35	3,49	3,58	3,64
SMP N 1 Ponjong	3,18	3,26	3,39	3,53	3,63
SMP N 1 Semanu	3,07	3,32	3,47	3,54	3,63

Hasil *t-test* terhadap kemampuan kerja sama peserta didik di SMP Negeri 2 Karangmojo diperoleh hasil *t* hitung = 20,480, di SMP Negeri 1 Ponjong diperoleh hasil *t* hitung = 28,606, dan di SMP Negeri 1 Semanu diperoleh hasil *t* hitung = 18,698. Keseluruhan hasil *t* hitung lebih besar dari *t* tabel = 1,696. Demikian juga hasil *t-test* terhadap kemampuan kerja sama peserta didik di SMP Negeri 1 Playen diperoleh hasil *t* hitung = 22,898 lebih besar dari *t* tabel = 1,701. Secara umum diperoleh hasil bahwa kerja sama peserta didik SMP Negeri 2 Karangmojo meningkat 13%, di SMP Negeri 1 Playen meningkat 16%, di SMP Negeri 1 Ponjong meningkat 14%, dan di SMP Negeri 1 Semanu meningkat 18%.

Membelajarkan kerja sama melalui pembelajaran kooperatif merupakan salah satu strategi pengajaran yang efektif untuk meningkatkan prestasi dan sosialisasi dan berkontribusi pada peningkatan sikap terha-

dap pembelajaran dan bekerja sama dengan orang lain, melatih demokrasi, mengembangkan rasa hormat, kebaikan hati, persahabatan, menghargai perbedaan dan memelihara hubungan sosial yang positif (Gillies & Ashman, 2003, p. 8; Kagan & Kagan, 2009, p. 2.16). Oleh sebab itu penting untuk mendesain pembelajaran yang mampu menumbuhkan sikap kerja sama di antara peserta didik.

Reliabilitas instrumen penilaian kerja sama pada uji coba luas dianalisis menggunakan reliabilitas skor komposit yang disarankan oleh Mosier (Azwar, 2012, pp. 84–86). Skor dari observer diberi bobot 3, skor diri sendiri diberi bobot 1 dan skor teman diberi bobot 2. Hasil perhitungan koefisien reliabilitas skor komposit penilaian kerja sama disajikan pada Tabel 7.

Tabel 7. Koefisien Reliabilitas Skor Komposit Penilaian Kerja sama pada Uji Coba Luas

SMP	Koefisien Reliabilitas Skor Komposit Eksperimen ke				
	1	2	3	4	5
SMP N 2 Karangmojo	0,936	0,933	0,950	0,953	0,923
SMP N 1 Playen	0,928	0,948	0,968	0,932	0,932
SMP N 1 Ponjong	0,949	0,950	0,941	0,950	0,963
SMP N 1 Semanu	0,953	0,956	0,952	0,964	0,955

Perhitungan koefisien reliabilitas skor komposit menunjukkan hasil bahwa instrumen penilaian kerja sama dapat digunakan dengan baik di SMP Negeri 2 Karangmojo, SMP Negeri 1 Playen, SMP Negeri 1 Ponjong dan SMP Negeri 1 Semanu.

Hasil nilai *pretest*, *posttest* dan perolehan *gain score* di empat sekolah disajikan pada Tabel 8.

Tabel 8. Hasil Belajar Kognitif Peserta Didik pada Uji Coba Luas

Perolehan Nilai	Peserta didik			
	SMP 2 Karangmojo	SMP N 1 Playen	SMP N 1 Ponjong	SMP N 1 Semanu
<i>Pretest</i>	22,94	31,75	37,94	31,50
<i>Posttest</i>	73,31	82,36	77,78	74,53
<i>Gain Score</i>	0,64	0,76	0,64	0,63
Kategori Peningkatan	Sedang	Tinggi	Sedang	Sedang

Kategori keberhasilan peningkatan hasil belajar dilihat dari nilai *gain* yang diperoleh, yaitu bila nilai *gain score* < 0,3 berarti peningkatan hasil belajar berada pada kategori rendah;  $0,3 \leq \textit{gain score} < 0,7$  kategori sedang; dan *gain score*  $\geq 0,7$  berarti peningkatan hasil belajar berada pada kategori tinggi (Hake, 1999, pp. 1–4). Berdasarkan hasil tersebut di atas maka guru tidak perlu khawatir menerapkan *experiment performance assessment* karena menunjukkan peningkatan hasil belajar kategori sedang dan tinggi. (Zeidan & Jayosi (2014, p. 20) mengungkapkan bahwa keterampilan proses sains memiliki korelasi dengan sikap positif peserta didik terhadap sains dan membuat mereka tertarik untuk fokus pada keterampilan proses sains. Ketika peserta didik memahami keterampilan proses sains, bagi mereka sains menjadi lebih menarik.

Hasil observasi pelaksanaan proses pembelajaran pada uji coba luas adalah sangat baik. Proses pembelajaran dengan *guided inquiry* sejalan dengan rekomendasi *National Science Teachers Association* (Kuhlthau, Maniotes, & Caspari, 2007, p. 103) yang memandang *inquiry* sebagai “dasar” bagi pendidikan IPA, menekankan bahwa para guru fokus untuk melakukan *inquiry* dan mengembangkan pemahaman. Hasil penelitian Ergül et al. (2011, pp. 48–68) menunjukkan bahwa penerapan pembelajaran *inquiry*, secara signifikan meningkatkan keterampilan proses sains dan sikap peserta didik. Pembelajaran *guided inquiry* melibatkan peserta didik secara langsung dalam mencari dan menemukan sendiri, melatih memecahkan permasalahan, dan juga mengembangkan sikap kerja sama, toleransi, teliti, percaya diri dan bertanggung jawab terhadap hasil temuannya.

## Simpulan

Berdasarkan hasil penelitian mengenai pengembangan instrumen *experiment performance assessment* dapat disimpulkan sebagai berikut: (1) produk penelitian ini adalah instrumen *experiment performance assessment* untuk pembelajaran IPA SMP materi sifat-sifat cahaya, pembentukan bayangan, dan prinsip kerja alat optik yang mengacu pada

model pengembangan Dick et al.; (2) produk instrumen *experiment performance assessment* memiliki kategori valid dan reliabel untuk menilai keterampilan proses sains dan kerja sama; (3) produk instrumen efektif untuk menilai keterampilan proses sains dan kerja sama peserta didik. Peserta didik memberikan respon senang dengan pembelajaran *guided inquiry* dan *experiment performance assessment* berdasarkan analisis dari data isian angket respon peserta didik.

Beberapa saran untuk pemanfaatan produk instrumen *experiment performance assessment* adalah sebagai berikut: (1) produk instrumen *experiment performance assessment* yang dihasilkan dinilai valid dan reliabel sehingga layak digunakan untuk penilaian di tingkat SMP kelas VIII; (2) produk instrumen menekankan pada pengembangan keterampilan kinerja peserta didik sehingga dapat digunakan sebagai referensi dan masukan bagi guru untuk pembelajaran dan penilaian dalam pelaksanaan Kurikulum 2013; (3) produk instrumen diperuntukkan bagi peserta didik kelas VIII SMP dan terbatas pada materi materi sifat-sifat cahaya, pembentukan bayangan, dan prinsip kerja alat optik sehingga disarankan kepada peneliti lain agar dapat mengembangkan instrumen untuk materi IPA yang lain.

## Daftar Pustaka

- Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, 40(4), 955–959.  
<https://doi.org/10.1177/001316448004000419>
- Anjani, D., Suciati, & Maridi. (2017). Profil keterampilan kerjasama dalam kelompok siswa kelas XI SMA Negeri 8 Surakarta pada materi sistem peredaran darah. In *Prosiding Seminar Nasional Pendidikan Sains II; Inovasi dan Pengembangan Kualitas Pembelajaran Sains Berbasis Pendidikan Karakter dan Teknologi di Era MEA*. Salatiga: Program Studi Pendidikan Biologi

- Fakultas Biologi Universitas Kristen Satya Wacana Salatiga.
- Azwar, S. (2012). *Reliabilitas dan validitas* (4th ed.). Yogyakarta: Pustaka Pelajar.
- Bertucci, A., Conte, S., Johnson, D. W., & Johnson, R. T. (2010). The impact of size of cooperative group on achievement, social support, and self-esteem. *The Journal of General Psychology*, 137(3), 256–272. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/20718226>
- Bundu, P. (2006). *Penilaian keterampilan proses dan sikap ilmiah dalam pembelajaran sains SD*. Jakarta: Departemen Pendidikan Nasional.
- Dick, W., Carey, L., & Carey, J. O. (2005). *The systematic design of instruction* (6th ed.). Boston: Pearson/Allyn and Bacon.
- Ergül, R., Şimşekli, Y., Çalış, S., Özdilek, Z., Göçmençelebi, Ş., & Şanlı, M. (2011). The effects of inquiry-based science teaching on elementary school students' science process skills and science attitudes. *Bulgarian Journal of Science and Education Policy (BJSEP)*, 5(1).
- Fadiawati, N., & Astuti, S. P. (2007). *Performance assessment: produk dan keterampilan. assessment dalam pembelajaran sains*. Bandung: Sekolah Pascasarjana UPI.
- Gillies, R. M., & Ashman, A. F. (2003). *Cooperative Learning*. New York: RoutledgeFalmer.
- Goodwin, M. W. (1999). Cooperative learning and social skills: what skills to teach and how to teach them. *Intervention in School and Clinic*, 35(1), 29–33.
- Guilford, J. P. (1956). *Fundamental statistic in psychology and education* (3rd ed.). New York: McGraw-Hill Book Company, Inc.
- Hake, R. R. (1999). Analyzing change/gain scores. Retrieved from <http://www.physics.indiana.edu/~sdi/AnalyzingChange-Gain.pdf>
- Jacobsen, D. A., Eggen, P., & Kauchak, D. (2009). *Methods for teaching: Metode-metode pengajaran meningkatkan belajar siswa TK-SMA*. (A. Fawaid & K. Anam, Trans.). Yogyakarta: Pustaka Pelajar.
- Kagan, S., & Kagan, M. (2009). *Kagan cooperative learning*. California: Kagan Publishing.
- Karami, A., & Rezaei, A. (2015). An overview of peer-assessment: the benefits and importance. *Journal for the Study of English Linguistics*, 3(1), 93. <https://doi.org/10.5296/jsel.v3i1.7889>
- Karamustafaoglu, S. (2011). Improving the science process skills ability of science student teachers using I diagrams. *Eurasian Journal of Physics and Chemistry Education*, 3(1), 26–38.
- Kementerian Pendidikan dan Kebudayaan. (2017). *Panduan penilaian oleh pendidik dan satuan pendidikan untuk sekolah menengah pertama*. Jakarta: Kementerian Pendidikan dan Kebudayaan Direktorat Jenderal Pendidikan Dasar dan Menengah Direktorat Pembinaan Sekolah Menengah Pertama.
- Kuhlthau, C. C., Maniotes, L. K., & Caspari, A. K. (2007). *Guided Inquiry: Learning in the 21st century school*. London: Libraries Unlimited, Inc.
- Majid, A. (2014). *Penilaian autentik: Proses dan hasil belajar*. Bandung: Remaja Rosdakarya.
- Mardapi, D. (2004). *Penyusunan tes hasil belajar*. Yogyakarta: Program Pascasarjana Universitas Negeri Yogyakarta.
- Menteri Pendidikan dan Kebudayaan. Peraturan Menteri Pendidikan dan Kebudayaan Nomor 21 Tahun 2016 tentang Standar Isi Pendidikan Dasar dan Menengah (2016).

- Menteri Pendidikan dan Kebudayaan Republik Indonesia. Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 22 Tahun 2016 tentang Standar Proses Pendidikan Dasar dan Menengah (2016).
- National Science Teachers Association. (2009). *The biology teacher's handbook* (4th ed.). Virginia: National Science Teachers Association.
- Okukawa, H. (2007). Vocational certificate curriculum in Thailand: The ultimate authentic assessment is gone. *International Forum of Teaching and Studies*, 3(1), 13–22.
- Rahman, A., Wahyuni, I., & Rifqiawati, I. (2017). Profil keterampilan proses sains dan sikap ilmiah siswa di SMP Satu Atap Pulau Tunda. *School Education Journal PGSD FIP Unimed*, 7(1), 1–7.  
<https://doi.org/10.24114/sejpgsd.v7i1.6827>
- Rahmasiwi, A., Santosari, S., & Sari, D. P. (2015). Peningkatan keterampilan proses sains siswa dalam pembelajaran biologi melalui penerapan model pembelajaran inkuiri di kelas XI MIA 9 (ICT) SMA Negeri 1 Karanganyar tahun pelajaran 2014/2015. In *Prosiding Seminar Nasional XII Pendidikan Biologi FKIP UNS*.
- Raj, R. G., & Devi, S. N. (2014). Science process skills and achievement in science among high school students. *Scholarly Research Journal for Interdisciplinary Studies*, II(XV), 2435–2443.
- Rezba, R. J., McDonough, J. T., Matkins, J. J., & Sprague, C. (2007). *Learning and assessing science process skills*. Iowa: Kendall/Hunt Publishing Company.
- Sukarno, Permanasari, A., & Hamidah, I. (2013). The profile of science process skill (SPS) student at secondary high school (case study in Jambi). *International Journal of Scientific Engineering and Research (IJSER)*, 1(1).
- Sumintono, B., & Widhiarso, W. (2013). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial*. Cimahi: Tim Komunikata Publishing House.
- Suryandari, E. T. (2013). Performance assessment sebagai instrumen penilaian untuk meningkatkan keterampilan proses pada praktikum kimia dasar di tadaris kimia. *Jurnal Phenomenon*, 3(2), 19–34.
- Susilowati, E. (2016). Penggunaan pembelajaran kooperatif tipe TGT pada materi struktur tumbuhan untuk peningkatan hasil belajar dan keaktifan siswa kelas VIII-F SMP Negeri 32 Semarang. *Jurnal Scientia Indonesia*, 1(1). Retrieved from <https://journal.unnes.ac.id/nju/index.php/jsi/article/view/7941>
- Trowbridge, L. W., & Bybee, R. W. (1990). *Becoming a secondary school science teacher*. Columbus: Merrill Publishing Company.
- Zeidan, A. H., & Jayosi, M. R. (2014). Science process skills and attitudes toward science among Palestinian secondary school students. *World Journal of Education*, 5(1).  
<https://doi.org/10.5430/wje.v5n1p13>

## EVALUASI PEMANFAATAN LABORATORIUM PADA DIKLAT PENGENDALI HAMA DAN PENYAKIT IKAN DI BALAI DIKLAT APARATUR KEMENTERIAN KELAUTAN DAN PERIKANAN

*Fajar Nugrobo*

Balai Diklat Aparatur, Kementerian Kelautan dan Perikanan  
Jl. Raya 2 Sukamandi Ciasem, Subang, 41256, Jawa Barat, Indonesia  
Email: [fajarfathi2010@gmail.com](mailto:fajarfathi2010@gmail.com)

### Abstrak

Keberadaan laboratorium menjadi hal yang penting dalam pelaksanaan Diklat Pengendali Hama dan Penyakit Ikan (PHPI) oleh karena itu Balai Diklat Aparatur (BDA) Kementerian Kelautan dan Perikanan (KKP) bekerja sama dengan Stasiun Karantina Ikan Kelas II Cirebon pada tahun 2014 membangun laboratorium uji karantina ikan di lingkungan kampus BDA KKP. Kajian ini bertujuan untuk mengevaluasi pemanfaatan laboratorium Diklat PHPI dengan mengukur perubahan kepuasan peserta terhadap ketersediaan media pembelajaran dan kelengkapan fasilitas belajar mengajar, nilai hasil belajar, peningkatan pengetahuan dan keterampilan peserta sebelum dan setelah laboratorium digunakan. Data diperoleh melalui pengisian angket/kuesioner evaluasi pelaksanaan diklat kepada peserta. Validitas dan reabilitas instrumen telah diuji sehingga instrumen dapat digunakan dengan valid dan reliabel. Uji t dilakukan untuk mengetahui perbedaan nilai pengukuran sebelum dan sesudah laboratorium beroperasi. Hasil kajian menunjukkan bahwa semua indikator meningkat setelah operasionalisasi laboratorium di BDA. Secara statistik kepuasan peserta terhadap kelengkapan fasilitas pembelajaran dan keterampilan peserta meningkat signifikan. Adapun peningkatan pengetahuan peserta, peningkatan kepuasan peserta terhadap kelengkapan media pembelajaran dan peningkatan nilai hasil belajar peserta, secara statistik tidak meningkat signifikan.

**Kata kunci:** *evaluasi, diklat, laboratorium, pemanfaatan, PHPI*

## THE EVALUATION OF LABORATORY UTILIZATION ON FISH QUARANTINE OFFICERS TRAINING IN MARINE AND FISHERIES OFFICERS TRAINING CENTER, MINISTRY OF MARINE AFFAIRS AND FISHERIES

### Abstract

The laboratory become inseparable in Fish Quarantine Officer Training. Therefore, Marine and Fisheries Officers Training Center (MFO TC) in collaboration with Cirebon Fish Quarantine Office build the fish quarantine laboratory in MFO TC campus in 2014. This study was conducted to evaluate the lab utilization by measure the change of training participans satisfaction in learning media and facilities, participans final grade, and cognitive/knowledge and psychomotor skill enhancement before and after laboratory operationalization. Data were collected by distributed the evaluation questionnaire. The instruments were used in this study valid and reliable according to the result of the instrument's validity and reability test. T-test two samples for means was used to analyse the differences of measurement before and after laboratory operationalization. Statistically, psychomotor skill enhancement and training participans satisfaction in facilities were significant increased after laboratory operationalization. Knowledge enhancement, training participans satisfaction in learning media and final grade were insignificant increased.

**Keywords:** *fish quarantine officer, laboratory, evaluating, utilization, training*

Permalink/DOI: <http://dx.doi.org/10.21831/pep.v22i2.18458>

## **Pendahuluan**

Peningkatan kualitas sumber daya manusia masih menjadi prioritas pembangunan nasional (Presiden Republik Indonesia, 2015, p. 5). Pendidikan dan pelatihan (diklat) berperan penting dalam menghasilkan sumber daya manusia yang berkualitas, memiliki kompetensi keahlian dan berdaya saing, serta mempunyai karakter dan budi pekerti unggul. Sasaran dari kegiatan tersebut salah satunya adalah perbaikan mutu proses belajar mengajar di kelas dengan berorientasi pada setiap aspek perkembangan peserta diklat. Karena secara naluriah peserta diklat menginginkan pengalaman belajar yang konkret, menyenangkan, dan mencakup semua aspek kompetensi yang diperlukan. Pengalaman belajar dapat diperoleh melalui metode pembelajaran praktik di kelas maupun di luar kelas. Pembelajaran praktik merupakan suatu sistem pengajaran dengan memberikan materi pengajaran menggunakan peralatan seperti diperagakan dengan harapan peserta menjadi jelas menerima materi sekaligus memperoleh kesempatan untuk melaksanakan suatu pekerjaan sehingga mereka memiliki pengetahuan dan keterampilan serta bersikap sesuai dengan keterampilan tersebut (Fathurrahman & Sutikno, 2007, p. 64). Ketersediaan sarana prasarana praktik yang berkualitas dan sesuai dengan standar menentukan keberhasilan tujuan pembelajaran dalam sebuah diklat.

Diklat perkarantina ikan baik fungsional maupun teknis sebagai salah satu kegiatan peningkatan kompetensi bagi petugas karantina ikan, sangat terkait erat dengan sarana dan prasarana yang dapat menunjang optimalisasi pencapaian tujuan pembelajaran. Tujuan pembelajaran dalam diklat ini dicapai melalui penyampaian teori dan praktik. Sesuai dengan kurikulum diklat perkarantina ikan, bobot penguasaan teori sebanyak 40 persen dan praktik 60 persen (Pusat Pelatihan Kelautan dan Perikanan, 2013, p. 8). Salah satu sarana yang diperlukan dalam pembelajaran diklat perkarantina ikan adalah ketersediaan fasilitas laboratorium dengan segala perlengkapannya sebagai penunjang kegiatan praktik bagi peserta diklat.

Dilihat dari fungsi idealnya, laboratorium menjadi tempat bagi pendidik/widyaiswara untuk mendalami konsep, mengembangkan metode pembelajaran, memperkaya pengetahuan dan keterampilan. Selain itu sebagai tempat bagi peserta diklat untuk belajar memahami materi melalui optimalisasi keterampilan proses serta mengembangkan sikap ilmiah. Fungsi laboratorium dalam posisi tersebut adalah sebagai tempat untuk memahami, membuktikan berbagai konsep dan melakukan penelitian ilmiah. Laboratorium sangat diperlukan dalam pembentukan sikap ilmiah peserta diklat. Dengan adanya laboratorium yang sesuai dengan standar, diharapkan peserta diklat mampu mengaplikasikan semua teori yang telah diajarkan dalam kelas.

Balai Diklat Aparatur (BDA) Kementerian Kelautan dan Perikanan (KKP) sebagai salah satu penyelenggara diklat bagi aparatur KKP bekerjasama dengan Stasiun Karantina Ikan Kelas II Cirebon membangun laboratorium uji karantina ikan sebagai sarana pendukung kegiatan kediklatan perkarantina ikan yang dilaksanakan di BDA KKP. Sejak diresmikan pada bulan Februari 2014, laboratorium uji karantina ikan BDA KKP telah digunakan sebagai tempat praktik pada Diklat Jabatan Fungsional Pengendali Hama dan Penyakit Ikan (PHPI) Tingkat Ahli dan Terampil dan Diklat Penjurangan Jabatan Fungsional PHPI Tingkat Ahli dan Terampil. Sebelumnya, kegiatan praktik pada diklat PHPI sebagian besar dilaksanakan di dalam ruang kelas sementara untuk beberapa mata diklat tertentu, praktik dilakukan di Balai Uji Standar Karantina Ikan (BUSKI) yang berlokasi di Jakarta. Secara khusus laboratorium uji karantina ikan didirikan sebagai syarat akreditasi penyelenggaraan Diklat Dasar Jabatan Fungsional PHPI tingkat Ahli maupun Terampil. Akreditasi penyelenggaraan diklat menjadi suatu kebijakan penting dalam rangka menentukan kelayakan dalam penyelenggaraan program diklat, sehingga kemandirian pelaksanaan diklat lebih menjamin peningkatan kualitas diklat (Lembaga Administrasi Negara, 2017, p. 1). Akreditasi lembaga kediklatan, memberikan jaminan kepada ma-

syarakat bahwa lembaga tersebut mampu dan kompeten melaksanakan kegiatan diklat secara profesional.

Keberadaan laboratorium dalam penyelenggaraan Diklat PHPI mempunyai peran strategis dalam meningkatkan pengetahuan, keterampilan dan sikap peserta diklat. Sehingga dalam rangka mengukur kemajuan pemanfaatan laboratorium uji karantina ikan BDA KKP pada kegiatan kediklatan PHPI maka evaluasi perlu dilakukan. Evaluasi program diklat merupakan proses untuk menjamin akuntabilitas dan peningkatan berkelanjutan (Edward, Scott, & Raju, 2007, p. 4). Evaluasi kualitas pelaksanaan diklat dapat dilihat dari keseluruhan tahapan mulai dari persiapan, pelaksanaan serta dampak diklat terhadap peningkatan produktivitas kinerja layanan organisasi (Edward et al., 2007, p. 6). (Kirkpatrick & Kirkpatrick, 2006, p. 21) mengemukakan model evaluasi melalui empat level yaitu level reaksi (*reaction*), pembelajaran (*learning*), perilaku (*behavior*), dan hasil (*result*). Pengukuran level reaksi dilakukan untuk mengetahui tingkat kepuasan peserta terhadap pelaksanaan program diklat; Pengukuran level pembelajaran dilakukan untuk mengetahui pencapaian pengetahuan, keterampilan dan sikap peserta setelah mengikuti diklat; Pada level perilaku, pengukuran dilakukan untuk mengetahui dampak diklat terhadap perubahan perilaku setelah peserta bekerja kembali ke instansi masing-masing; Sedangkan pengukuran level hasil dilakukan untuk mengetahui seberapa besar pengaruh program diklat terhadap kinerja instansi peserta (Kirkpatrick & Kirkpatrick, 2006, pp. 21–26).

Level evaluasi yang berkaitan dengan ketersediaan laboratorium sebagai pendukung kegiatan diklat dan dapat diukur saat diklat berlangsung adalah level reaksi dan pembelajaran. Oleh karena itu kajian ini dibatasi pada bahasan pemanfaatan laboratorium terhadap level reaksi dan pembelajaran peserta Diklat PHPI. Hal ini dilakukan dengan mengukur perubahan level reaksi dan pembelajaran peserta Diklat PHPI sebelum dan sesudah laboratorium digunakan untuk kegiatan diklat. Level reaksi meliputi komponen kepuasan peserta dengan kriteria ke-

puasan peserta terhadap ketersediaan media pembelajaran dan kepuasan terhadap kelengkapan fasilitas pembelajaran. Level pembelajaran meliputi komponen daya serap peserta terhadap materi dengan kriteria nilai hasil belajar, serta peningkatan pengetahuan dan keterampilan peserta.

Berdasarkan latar belakang tersebut maka kajian ini bertujuan untuk mengevaluasi pemanfaatan laboratorium uji karantina ikan pada kegiatan diklat PHPI dengan (1) mengukur perubahan kepuasan peserta terhadap ketersediaan media pembelajaran dan kelengkapan fasilitas pembelajaran sebelum dan sesudah laboratorium digunakan; dan (2) Mengukur perubahan nilai hasil belajar, peningkatan pengetahuan dan keterampilan peserta sebelum dan setelah laboratorium digunakan.

## Metode Penelitian

Kajian ini dilaksanakan di Balai Diklat Aparatur KKP Sukamandi, Jawa Barat pada Januari 2018. Kajian ini merupakan penelitian evaluatif yaitu penelitian yang mengetahui akhir dari sebuah kebijakan dalam rangka menentukan rekomendasi sebagai pertimbangan penentuan kebijakan selanjutnya (Arikunto & Jabar, 2014, p. 59).

Subjek penelitian ini adalah populasi dari peserta Diklat Pengendali Hama dan Penyakit Ikan (PHPI) yang dilaksanakan sebelum dan sesudah laboratorium uji karantina ikan beroperasi. Secara keseluruhan terdapat 8 (delapan) angkatan yang terdiri dari 4 (empat) angkatan yang dilaksanakan sebelum operasional laboratorium dan 4 (empat) angkatan diklat yang dilaksanakan setelah operasional laboratorium (Tabel 2).

Model yang dipakai dalam penelitian ini adalah model evaluasi Kirkpatrick. Model ini menggunakan 4 (empat) level evaluasi yaitu level reaksi (*reaction*), pembelajaran (*learning*), perilaku (*behavior*), dan hasil (*result*) (Kirkpatrick & Kirkpatrick, 2006, p. 21).

Pengumpulan data hanya pada 2 (dua) level evaluasi yaitu level reaksi dan level pembelajaran karena level tersebut yang berkaitan dengan ketersediaan laboratorium dalam mendukung kegiatan diklat. Instrumen eva-



luasi pada penelitian ini merupakan bagian dari pelaksanaan evaluasi penyelenggaraan diklat yang telah dilakukan secara kontinyu pada setiap diklat yang dilaksanakan BDA KKP. Instrumen telah dilakukan uji validitas isi yang melibatkan beberapa pakar dalam bidang evaluasi dan beberapa widyaiswara. Reabilitas instrumen atau keajegan instrumen diestimasi melalui penggunaan instrumen pada obyek yang sama dalam beberapa kali waktu yang berbeda (uji retest). Hasil uji menunjukkan bahwa tidak ada perbedaan signifikan pada penggunaan instrumen dalam waktu yang berbeda sehingga dapat dikatakan bahwa instrumen reliabel untuk digunakan sebagai alat ukur.

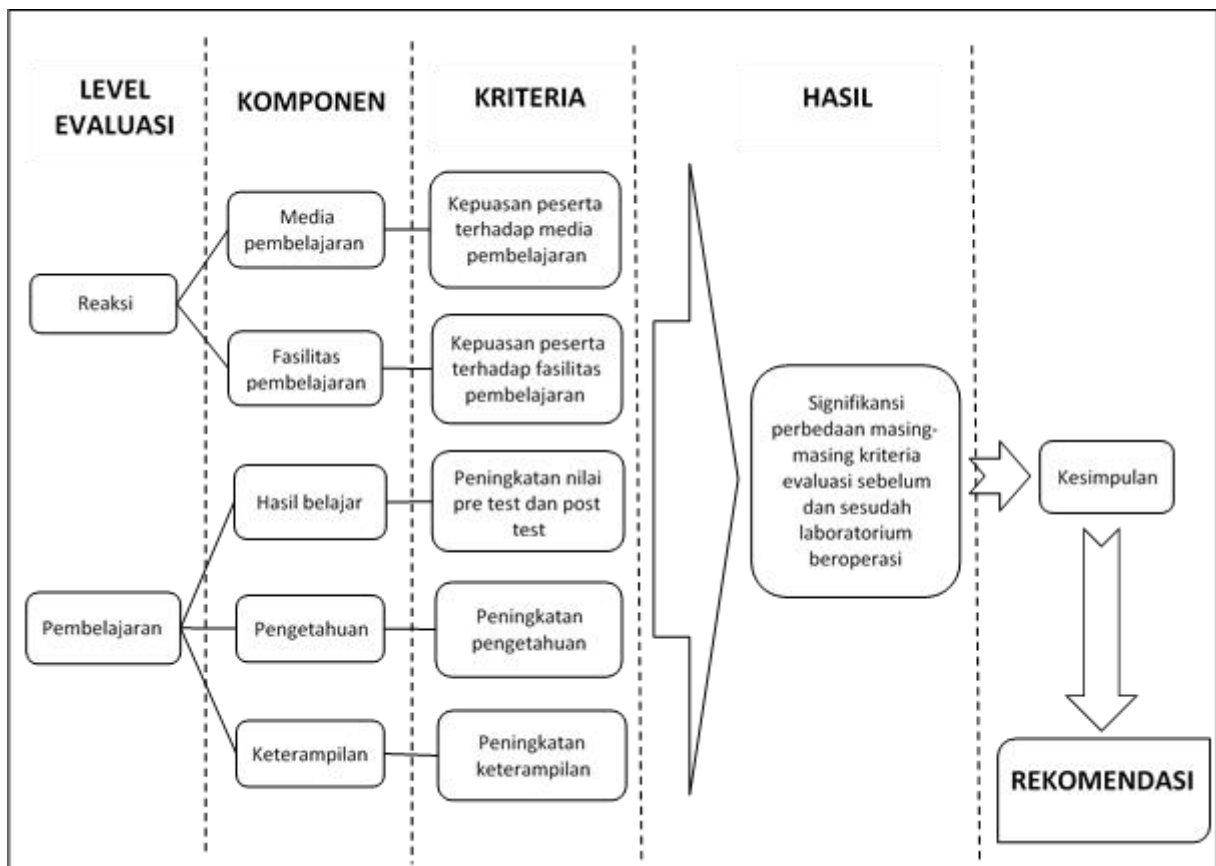
Komponen yang diukur pada instrumen level reaksi meliputi: (1) media pembelajaran dengan kriteria kepuasan peserta terhadap media pembelajaran; (2) fasilitas pembelajaran dengan kriteria kepuasan peserta terhadap fasilitas pembelajaran.

Data diperoleh dari pengisian angket/kuesioner evaluasi peserta (responden) setelah mengikuti program Diklat PHPI. Data yang dikumpulkan berupa nilai pada kriteria kepuasan peserta terhadap media pembelajaran dan kepuasan peserta terhadap fasilitas pembelajaran dengan kategori sebagai sebagaimana tersaji pada Tabel 1.

Tabel 1. Nilai dan Kategori pada Kriteria Evaluasi Pembelajaran

Nilai	Kategori
1 – 2	Tidak baik
3 – 4	Kurang baik
5 – 6	Cukup baik
7 – 8	Baik
9 – 10	Sangat baik

Sumber: (Pusat Pelatihan Kelautan dan Perikanan, 2013).



Gambar 1. Konstruk Umum Desain Penelitian



Komponen yang diukur pada instrumen level pembelajaran adalah pengetahuan dan keterampilan yang meliputi: (1) hasil belajar peserta dengan kriteria peningkatan nilai pre-test dan post test peserta diklat; (2) pengetahuan dengan kriteria peningkatan pengetahuan peserta setelah mengikuti diklat; (3) keterampilan dengan kriteria peningkatan keterampilan peserta setelah mengikuti diklat.

Data peningkatan nilai pre-test dan post-test peserta diklat diperoleh dari nilai hasil peserta setelah mengerjakan test. Sedangkan untuk peningkatan pengetahuan dan keterampilan, data diperoleh dengan pengisian nilai pada angket/kuesioner evaluasi peserta setelah mengikuti program Diklat PHPI. Adapun kategorinya sebagaimana tersaji pada Tabel 1.

Teknik analisis data yang digunakan pada penelitian ini adalah kuantitatif deskriptif yaitu dengan menggambarkan fenomena yang ada berdasarkan hasil uji kuantitatif data

yang diperoleh (Syamsuddin & Damaianti, 2011, p. 14).

Secara teknis setelah responden mengisi angket/kuesioner evaluasi, nilai/skor dijumlahkan kemudian direratakan. Khusus untuk kriteria peningkatan nilai hasil belajar peserta, data diperoleh dari selisih nilai pre-test dan post-test. Kemudian dilakukan uji t untuk mengetahui signifikansi perbedaan nilai setiap kriteria sebelum dan sesudah operasional laboratorium.

Kesimpulan ditarik berdasarkan hasil uji kuantitatif signifikansi perbedaan setiap kriteria evaluasi sebelum dan setelah laboratorium beroperasi. Konstruksi umum desain penelitian ini tersaji pada Gambar 1.

Data sekunder berupa kelengkapan fasilitas pembelajaran diperoleh dari daftar inventaris ruangan di BDA KKP. Informasi tentang berjalannya diklat serta evaluasi penyelenggaraan diklat, diperoleh dari laporan akhir panitia pelaksanaan Diklat PHPI.

Tabel 2. Kegiatan Diklat PHPI di BDA KKP sebelum dan sesudah Operasional Laboratorium.

No	Nama diklat	Tanggal pelaksanaan	Jumlah peserta (orang)	Sebelum/ sesudah operasional laboratorium
1	Diklat Dasar PHPI Tingkat Ahli Angkatan I Tahun 2013	26 Febr. – 11 Mar. 2013	39	sebelum
2	Diklat Dasar PHPI Tingkat Ahli Angkatan II Tahun 2013	14 - 26 Mar. 2013	40	sebelum
3	Diklat Dasar PHPI Tingkat Terampil Angkatan I Tahun 2013	26 Febr. - 9 Mar. 2013	39	sebelum
4	Diklat Dasar PHPI Tingkat Terampil Angkatan II Tahun 2013	14 - 25 Mar. 2013	40	sebelum
5	Diklat Penjurusan PHPI Tingkat Keahlian Muda Angkatan I Tahun 2014	9 – 17 Sept. 2014	40	sesudah
6	Diklat Penjurusan PHPI Tingkat Keahlian Muda Angkatan II Tahun 2014	16 – 24 Sept.2014	40	sesudah
7	Diklat Penjurusan PHPI Tingkat Keterampilan Mahir Angkatan I Tahun 2014	7 – 13 Okt. 2014	42	sesudah
8	Diklat Penjurusan PHPI Tingkat Keterampilan Mahir Angkatan II Tahun 2014	14 – 20 Okt. 2014	37	sesudah

## Hasil Penelitian dan Pembahasan

### Gambaran umum

Diklat bagi Jabatan Fungsional PHPI yang dilaksanakan di BDA KKP meliputi Diklat Dasar PHPI dan Diklat Penjenjangan PHPI. Diklat dasar maupun diklat penjenjangan terdiri dari diklat tingkat ahli dan terampil sesuai dengan jabatan peserta sebagaimana diatur dalam Peraturan Menteri Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Nomor 22 Tahun 2010 tentang Jabatan Fungsional Pengendali Hama dan Penyakit Ikan dan Angka Kreditnya. Diklat Jabatan Fungsional PHPI bertujuan untuk meningkatkan kompetensi PHPI agar mampu mengetahui karakteristik tugas, berwawasan luas, profesional dan bertanggung jawab sehingga dapat mendukung pelaksanaan tugas dalam mewujudkan visi dan misi organisasi di unit kerja masing-masing (Menteri Negara Pendayagunaan Aparatur Negara dan Reformasi Birokrasi, 2010, p. 25). Selain itu penyelenggaraan Diklat Jabatan Fungsional PHPI memfasilitasi peserta untuk mendapatkan sertifikat yang digunakan sebagai syarat kenaikan jenjang karier bagi Pejabat Fungsional PHPI.

Kurikulum diklat terdiri dari mata diklat yang dikelompokkan menjadi kelompok kompetensi umum, kompetensi inti dan kompetensi penunjang. Kompetensi umum adalah kompetensi yang berisi wawasan aspek pengetahuan dalam hal kebijakan, moti-

vasi dan perilaku. Kompetensi inti adalah kompetensi teknis yang berkaitan langsung dengan tugas jabatan PHPI. Sedangkan kompetensi penunjang adalah kompetensi yang bertujuan menambah dan melengkapi wawasan terkait dengan tugas jabatan PHPI. Setiap kelompok kompetensi mempunyai jenis mata diklat dan jumlah jam pelajaran yang berbeda, namun secara umum bobot praktik pada Diklat PHPI lebih banyak dibandingkan bobot teori. Secara detail jumlah jam pelajaran pada Diklat Dasar dan Penjenjangan PHPI disajikan pada Tabel 3.

Pejabat Fungsional PHPI mempunyai tugas, tanggung jawab, wewenang dan hak secara penuh untuk melakukan pengendalian hama dan penyakit ikan serta lingkungan. Pengendalian hama dan penyakit ikan diartikan sebagai upaya untuk mencegah penyebaran, mempersempit/memperkecil area atau daerah tertular, area atau daerah wabah, dan menekan kerugian yang ditimbulkan akibat serangan hama penyakit ikan. Melalui upaya ini, hama penyakit ikan yang timbul di kawasan budidaya dapat ditanggulangi, begitu pula dengan hama dan penyakit ikan yang terbawa saat terjadi lalu lintas media/komoditas. Melihat tugas pokok dan fungsi (tupoksi) PHPI yang berkaitan langsung dengan pemeriksaan hama dan penyakit ikan, maka ketersediaan sarana penunjang kegiatan praktik menjadi bagian yang tidak terpisahkan dari kegiatan diklat sebagai upaya peningkatan profesionalisme PHPI.

Tabel 3. Jumlah Jam Pelajaran pada Diklat PHPI

No	Diklat	Jumlah Jam Pelajaran pada Kelompok Kompetensi						Jumlah
		Umum		Inti		Penunjang		
		T	P	T	P	T	P	
1	Diklat Dasar PHPI Tingkat Ahli	8	0	32	61	4	4	109
2	Diklat Dasar PHPI Tingkat Terampil	6	0	26	53	4	4	93
3	Diklat Penjenjangan PHPI Tingkat Keahlian Muda	4	0	16	48	4	8	80
4	Diklat Penjenjangan PHPI Tingkat Keterampilan Mahir	4	0	16	36	4	4	64

Keterangan:

T = teori

P = praktik

Wallace (1991, p. 52) menyampaikan bahwa terdapat dua sumber pengetahuan yaitu pengetahuan yang diperoleh melalui belajar (*received knowledge*) dan pengetahuan yang diperoleh melalui pengalaman (*experiential knowledge*). Kedua sumber pengetahuan tersebut merupakan unsur kunci bagi pengembangan profesionalisme. Setiap peserta pelatihan membawa pengetahuan dan pengalaman ketika memasuki diklat baru. Efektifnya pelatihan tergantung pada bagaimana peserta pelatihan melakukan refleksi mengaitkan antara pengetahuan dan pengalaman serta praktik untuk memperbaiki pembelajarannya lebih lanjut (Wallace, 1991, p. 55). Kemampuan melakukan refleksi dari praktik yang didasarkan pada pengalaman dan pengetahuan menentukan pencapaian kompetensi profesional. Pendekatan utama dalam pembelajaran diklat khususnya bagi Jabfung PHPI adalah pendekatan praktik sehingga peserta mampu menemukan sendiri, melalui langkah-langkah kerja ilmiah, seperti mengamati, mengumpulkan data, mengukur, memaknai data dan menarik kesimpulan. Pertimbangan tersebut kemudian dituangkan dalam kurikulum sehingga jumlah jam pelajaran kompetensi inti mempunyai porsi pembelajaran praktik paling banyak dibandingkan dengan kelompok kompetensi yang lain. Selain itu kompetensi inti juga sangat berkaitan dengan hal-hal teknis yang menyangkut tupoksi Pejabat Fungsional PHPI.

Pembelajaran orang dewasa akan lebih efektif jika peserta lebih banyak terlibat langsung dibandingkan hanya pasif menerima dari pengajar (Kolb, 1984, p. 300). Praktik di laboratorium merupakan upaya untuk memberi kesempatan kepada peserta mendapatkan pengalaman langsung. Ide dasar belajar berdasarkan pengalaman mendorong peserta pelatihan untuk merefleksikan atau melihat kembali pengalaman-pengalaman yang mereka pernah alami. Peranan laboratorium pada kegiatan kediklatan merupakan bagian dari proses belajar-mengajar berupa kegiatan praktik yang obyeknya sesuai dengan kurikulum yang tersedia. Kegiatan praktik pada Diklat Dasar PHPI meliputi praktik persiapan pelaksanaan kegiatan pengendalian

Hama dan Penyakit Ikan (HPI)/Hama dan Penyakit Ikan Karantina (HPIK), identifikasi HPI/HPIK golongan parasit, cendawan dan bakteri. Sedangkan untuk praktik pada Diklat Penjenjangan Jabatan Fungsional PHPI meliputi praktik penyiapan tempat, alat dan bahan laboratorium serta praktik pelaksanaan pengendalian HPI.

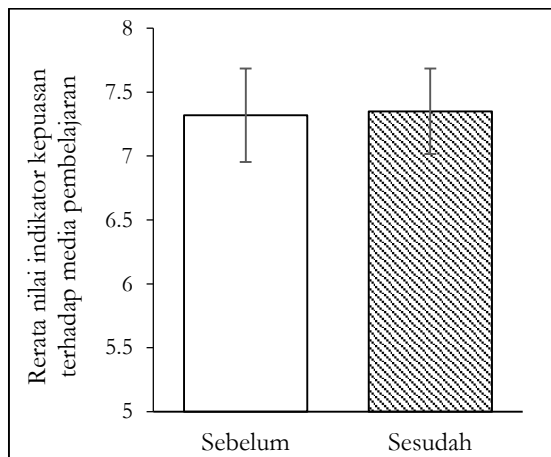
#### Pemanfaatan Laboratorium terhadap Peningkatan Kepuasan Peserta pada Media Pembelajaran

Berdasarkan pengukuran evaluasi kepuasan peserta terhadap kelengkapan media pembelajaran, diperoleh hasil bahwa terjadi kenaikan kepuasan setelah laboratorium uji karantina ikan digunakan sebagai sarana praktik kegiatan diklat (Gambar 2). Namun kenaikan tersebut secara statistik tidak signifikan (Tabel 4). Hal ini berarti keberadaan laboratorium tidak begitu bermanfaat terhadap peningkatan kepuasan peserta terhadap kelengkapan media pembelajaran.

Hal ini mungkin dikarenakan ketersediaan alat uji pemeriksaan sebagai sarana pembelajaran belum lengkap. Laboratorium hanya menyiapkan media praktik untuk pemeriksaan parasit, sementara untuk pemeriksaan bakteri, jamur/cendawan dan virus belum tersedia. Pembelajaran praktik untuk materi pemeriksaan bakteri dan cendawan terbatas pada penyiapan media tanam untuk kultur. Sedangkan identifikasi cendawan dan bakteri disampaikan secara teori melalui buku identifikasi. Media pembelajaran didefinisikan sebagai sarana yang dipakai untuk menyampaikan informasi dalam perspektif belajar mengajar sehingga tercapai pembelajaran yang efektif (Naz & Akbar, 2008, p. 35). Media pembelajaran bermanfaat untuk mendukung pembelajaran dan membantu memahami sebuah pengetahuan (Kozma, 1991, p. 205). Media pembelajaran yang dirancang dengan benar, akan memberikan implikasi pada proses pembelajaran berupa efektifitas waktu, peningkatan minat peserta, klarifikasi gagasan, memperkuat konsep, mempertahankan memori, serta membantu membuktikan sebuah teori (Mohan, McGregor, Saunders, & Arceche, 1997).

Tabel 4. Hasil Uji T dan Rerata Nilai Kepuasan Peserta terhadap Ketersediaan Media Pembelajaran sebelum dan sesudah Laboratorium Beroperasi

Diklat	Rerata nilai indikator ( $\pm$ SD)	df	P	t hitung	Keterangan
Sebelum	7,323 $\pm$ 0,366	4	0,464	0,095	Tidak berbeda signifikan ( $\alpha= 0,05$ )
Sesudah	7,357 $\pm$ 0,335				



Gambar 2. Grafik Rerata Nilai Kepuasan Peserta terhadap Media Pembelajaran sebelum dan sesudah Laboratorium Beroperasi

Faktor yang menentukan efektifitas media pembelajaran adalah bagaimana media mampu membantu memfasilitasi peserta untuk terlibat aktif dalam proses pembelajaran (Killen, 2003, p. 19). Keterlibatan peserta dalam proses pembelajaran akan terlihat jelas saat peserta sedang mengerjakan tugas, fokus pada isu penting dan secara sadar berusaha untuk serius dalam belajar (Killen, 2003, p. 19). Dampak penggunaan media pembelajaran dapat meningkat dengan menerapkan beberapa prinsip yaitu ketepatan pemilihan media yang relevan, keterkaitan media dengan materi sebelumnya dan hubungannya dengan tujuan yang akan dicapai pada pembelajaran sekarang, kesempatan untuk mendapatkan umpan balik dari peserta, dan evaluasi penggunaan media (Heinich, Molenda, Russell, & Smaldino, 1996, p. 318). Agar pengalaman belajar peserta lebih konkret dan realistis,

maka pengajar harus menggunakan dan menyiapkan media yang spesifik. Media pembelajaran membantu pengajar untuk mengklarifikasi, menetapkan dan mengkorelasikan konsep peserta didik. Media pembelajaran juga membantu mengajarkan konsep abstrak dengan cara yang nyata (Naz & Akbar, 2008, p. 38). Penggunaan berbagai media pengajaran telah berhasil mengubah paradigma pembelajaran tradisional, dimana pengajar melakukan sebagian besar komunikasi searah dan para peserta adalah pendengar pasif, ke dalam pusat pembelajaran partisipatif yang memfasilitasi pembelajaran agar lebih produktif (Thomas, 2008, p. 106).

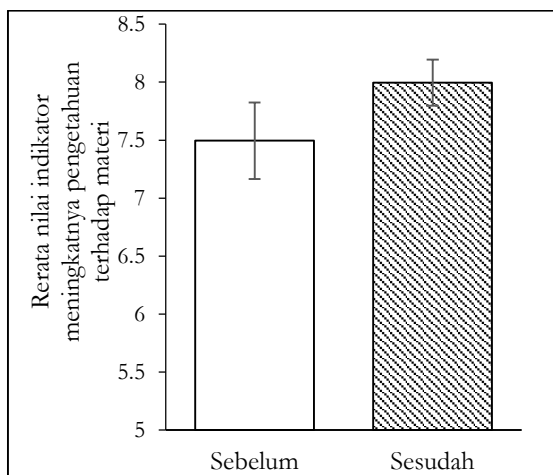
Melihat hasil kajian dan beberapa konsep tersebut diatas maka kelengkapan media pembelajaran dalam laboratorium menjadi hal yang cukup penting untuk mencapai tujuan Diklat PHPI. Ketersediaan media pembelajaran yang lengkap juga berpotensi memberikan kepuasan bagi peserta diklat. Ketercapaian kepuasan mendorong peserta untuk menyerap materi dengan maksimal.

#### Pemanfaatan Laboratorium terhadap Peningkatan Pengetahuan Peserta Diklat

Berdasarkan hasil kajian, terjadi peningkatan pengetahuan peserta terhadap materi pembelajaran setelah laboratorium uji dioperasikan untuk kegiatan diklat (Gambar 3). Namun rerata peningkatan tersebut secara statistik tidak signifikan (Tabel 5). Sehingga dapat diartikan bahwa keberadaan laboratorium secara umum belum begitu bermanfaat terhadap peningkatan pengetahuan peserta diklat.

Tabel 5. Hasil Uji t dan Rerata Nilai Indikator Meningkatnya Pengetahuan terhadap Materi Pembelajaran sebelum dan sesudah Laboratorium Beroperasi

Diklat	Rerata nilai indikator ( $\pm$ SD)	df	P	t hitung	Keterangan
Sebelum	7,583 $\pm$ 0,330	4	0,096	0,168	Tidak berbeda signifikan ( $\alpha = 0,05$ )
Sesudah	7,950 $\pm$ 0,199				



Gambar 3. Grafik Rerata Peningkatan Pengetahuan sebelum dan sesudah Laboratorium Beroperasi

Peningkatan pengetahuan secara efektif diperoleh dari kegiatan di kelas seperti diskusi, penyampaian teori, presentasi, pembelajaran berbasis masalah (*problem based learning*), seminar dan tanya jawab dengan para ahli (Kasilingam, Ramalingam, & Chinnavan, 2014, p. 28). Sedangkan praktik dalam laboratorium efektif dalam meningkatkan keterampilan/psikomotorik. Sehingga keberadaan laboratorium sebagai sarana praktik kurang begitu berdampak terhadap peningkatan pengetahuan/kognitif peserta.

Ranah pengetahuan/kognitif merupakan salah satu dari tiga ranah kompetensi yang harus dimiliki peserta dalam menjalankan tugas profesionalismenya. Ranah yang lain adalah keterampilan/psikomotorik dan sikap kerja/afektif (Bloom, 1956, p. 7). Pengetahuan/kognitif didefinisikan sebagai pengenalan dan pengembangan dari kemampuan keterampilan intelektual (Bloom, 1956, p. 7). Ranah pengetahuan/kognitif berfokus pada kemampuan intelektual peserta. Hasil atau produk yang terkait dengan pengetahuan/kognitif adalah (Bloom, 1956, p. 18) (a)

pengetahuan atau kemampuan untuk mengenali atau mengingat informasi; (b) pemahaman atau kemampuan untuk menunjukkan pemahaman dengan menggambarkan; (c) penerapan informasi pembelajaran untuk memecahkan masalah atau menjawab pertanyaan; (d) menganalisis atau membedah suatu masalah menjadi bagian penyusunnya; (e) mengevaluasi atau menilai sebuah gagasan dengan menggunakan kriteria eksplisit; dan (f) menciptakan atau menata ulang pengetahuan menjadi sebuah pola baru. Melihat teori diatas, pengetahuan/kognitif masih dalam bentuk yang abstrak artinya belum teraktualisasikan dalam perbuatan fisik. Pengetahuan kemudian diimplementasikan dalam bentuk keterampilan/psikomotorik.

#### Pemanfaatan Laboratorium terhadap Peningkatan Keterampilan Peserta Diklat

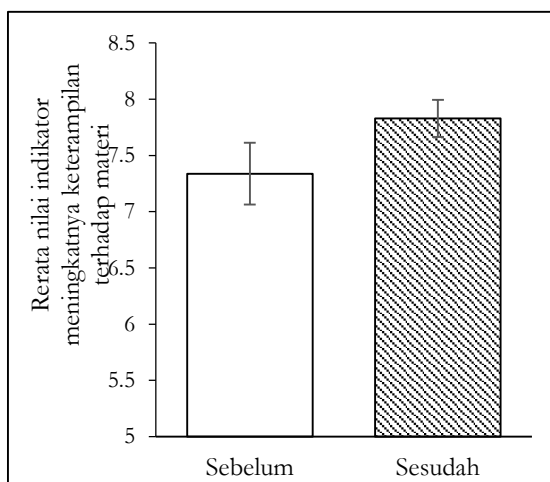
Berdasarkan hasil kajian, terjadi kenaikan yang signifikan rerata nilai indikator peningkatan keterampilan peserta diklat setelah laboratorium beroperasi (Gambar 4). Meskipun kenaikan hanya 0.357 namun kenaikan tersebut secara statistik berbeda nyata sebelum dan setelah operasionalisasi laboratorium (Tabel 6). Dengan demikian dapat diartikan bahwa keberadaan laboratorium bermanfaat dalam meningkatkan keterampilan/psikomotorik bagi peserta diklat.

Keterampilan praktis peserta didik di laboratorium dikaitkan dengan ranah psikomotorik. Keterampilan/psikomotorik berkaitan dengan kemampuan fisik seperti kecepatan, ketangkasan, akurasi, penggunaan instrumen, gerakan ekspresif dan penggunaan anggota tubuh (Simpson, 1966, p. 11). Menurut Zaghoul (2001, p. 3), pikiran dan tubuh manusia terhubung bersama saat melakukan aktivitas fisik. Produk yang berkaitan dengan keterampilan/psikomotorik adalah:

Tabel 6. Hasil Uji t dan Rerata Nilai Indikator Meningkatnya Keterampilan terhadap Materi Pembelajaran sebelum dan sesudah Laboratorium Beroperasi

Diklat	Rerata nilai indikator ( $\pm$ SD)	df	P	t hitung	Keterangan
Sebelum	7,450 $\pm$ 0,275	4	0,043	2,258	Berbeda signifikan ( $\alpha = 0,05$ )
Sesudah	7,807 $\pm$ 0,165				

(a) persepsi, yaitu mampu mendeteksi isyarat untuk bertindak; (b) respon, yaitu mampu melakukan tindakan nyata; (c) mekanisme atau kemampuan untuk melakukan tugas tanpa pengawasan; (d) respon yang kompleks, atau kemampuan untuk menyusun pola aksi yang kompleks; (e) adaptasi, atau kemampuan untuk mengubah tindakan untuk merespons situasi baru; dan (f) oriniasi, atau kemampuan untuk mengembangkan tindakan baru yang original (Simpson, 1966, pp. 25–30).



Gambar 4. Grafik Rerata Peningkatan Keterampilan sebelum dan sesudah Laboratorium Beroperasi

Keterampilan (psikomotorik) dapat diperoleh melalui kegiatan praktik laboratorium, kursus kejuruan, pelatihan menggunakan peralatan tertentu (Kasilingam et al., 2014, p. 29). Kemampuan keterampilan/psikomotorik secara efektif dapat diobservasi melalui demonstrasi atau simulasi. Simulasi membantu peserta untuk mempelajari langkah-langkah dari suatu pekerjaan. Dengan demikian peningkatan keterampilan peserta diklat mampu ditingkatkan melalui

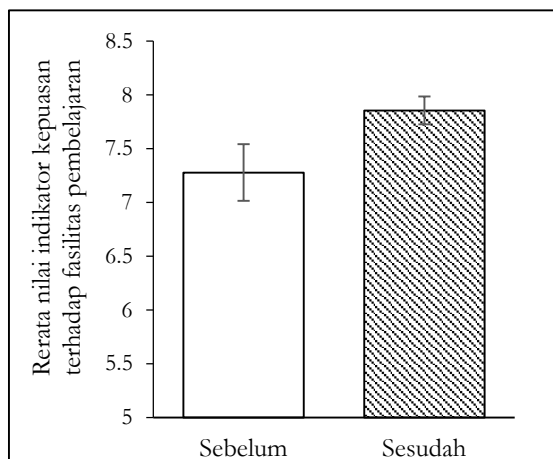
kegiatan praktik berupa demonstrasi dan simulasi di laboratorium.

Metode paling efektif untuk mengajar peserta didik melakukan sesuatu adalah dengan meminta mereka melakukan tugasnya (Schank, Berman, & Macpherson, 1999, p. 235). Melalui praktik di laboratorium, peserta didik belajar dengan mempraktikkan keterampilan yang tidak bisa dipelajari secara teoritis. Saat melakukan praktik di laboratorium, peserta memiliki kesempatan untuk mengembangkan dan melatih keterampilan praktis dan keterampilan langsung mereka (Hunter, Mccosh, & Wilkins, 2003, p. 73; Krivickas & Krivickas, 2007, p. 196). Selain itu, praktik di laboratorium juga mampu mengembangkan kemampuan lain seperti keterampilan penulisan laporan, keterampilan kerja tim dan komunikasi saat melakukan eksperimen di laboratorium (Krivickas & Krivickas, 2007, p. 191). Kegiatan praktik di laboratorium mampu menambah keterampilan praktis bagi para peserta dan memaparkannya kedalam bidang pekerjaannya yang relevan (Salim, Puteh, & Daud, 2012, p. 546). Tupoksi petugas PHPI tidak bisa dilepaskan dari kegiatan di laboratorium. Pemeriksaan penyakit ikan dilakukan oleh petugas PHPI dalam skala lapangan maupun laboratorium. Oleh karena itu kegiatan praktik diperlukan bagi peserta diklat PHPI agar terbiasa dengan alat yang digunakan dalam pelaksanaan tugasnya di laboratorium.

#### Pemanfaatan Laboratorium terhadap Peningkatan Kepuasan pada Fasilitas Pembelajaran

Berdasarkan pengukuran evaluasi kepuasan peserta terhadap kelengkapan fasilitas pembelajaran, diperoleh hasil bahwa terjadi kenaikan kepuasan yang signifikan

setelah laboratorium digunakan sebagai sarana praktik kegiatan diklat (Gambar 5). Uji t menunjukkan perbedaan nyata nilai indikator kepuasan peserta sebelum dan setelah laboratorium beroperasi (Tabel 7). Hal ini berarti keberadaan laboratorium bermanfaat dalam menaikkan kepuasan peserta terhadap fasilitas pembelajaran.



Gambar 5. Grafik Rerata Nilai Kepuasan Peserta terhadap Fasilitas Pembelajaran sebelum dan sesudah Laboratorium Beroperasi

Kepuasan peserta PHPI terhadap fasilitas pembelajaran setelah adanya laboratorium cukup beralasan karena pada diklat sebelumnya (sebelum laboratorium beroperasi) mereka melaksanakan kegiatan praktik hanya di kelas. Sedangkan untuk praktik pelaksanaan pengendalian hama dan penyakit ikan berupa pengujian pestisida dan kontaminan, serologi serta molekuler dilaksanakan di Balai Uji Standar Karantina Ikan (BUSKI)

Tabel 7. Hasil Uji t dan Rerata Nilai Kepuasan Peserta terhadap Fasilitas Pembelajaran sebelum dan sesudah Laboratorium Beroperasi

Diklat	Rerata nilai indikator ( $\pm$ SD)	df	P	t hitung	Keterangan
Sebelum	7,278 $\pm$ 0,264	4	0,024	2,798	Berbeda signifikan ( $\alpha = 0,05$ )
Sesudah	7,855 $\pm$ 0,130				

Tabel 8. Hasil Uji t dan Rerata Persentase Peningkatan Nilai Peserta sebelum dan sesudah Laboratorium Beroperasi

Diklat	Rerata % peningkatan nilai peserta ( $\pm$ SD)	df	P	t hitung	Keterangan
Sebelum	41,465 $\pm$ 10,795	4	0,172	1,073	Tidak berbeda signifikan ( $\alpha = 0,05$ )
Sesudah	52,510 $\pm$ 24,379				

yang berlokasi di Jakarta. Dengan dibangunnya laboratorium, beberapa aktifitas praktik mampu difasilitasi di BDA KKP sehingga waktu dan sumber daya yang ada bisa digunakan secara efektif.

Fasilitas pembelajaran merupakan hal penting untuk menyediakan lingkungan bagi percepatan peningkatan pengetahuan peserta (Fleming & Storr, 1999, p. 235). Selain itu fasilitas pembelajaran menjadi penting karena memberikan kondisi dalam meningkatkan pengalaman belajar.

Studi oleh Fleming & Storr (1999, p. 234) mengungkapkan bahwa kualitas fasilitas pembelajaran memberikan pengaruh besar terhadap pengalaman belajar peserta. Laboratorium uji karantina menjadi fasilitas untuk mengeksplorasi keterampilan belajar peserta diklat PHPI sehingga mampu menambah pengalaman belajar mereka.

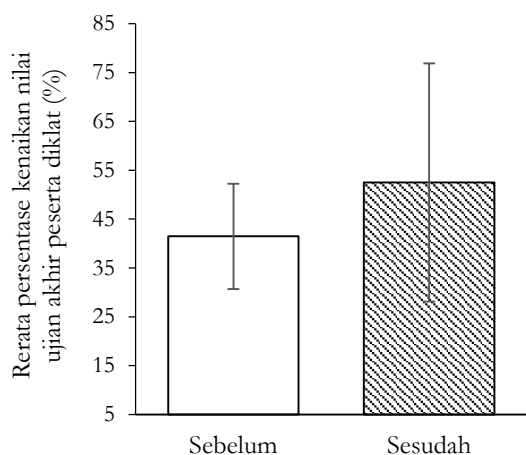
#### Pemanfaatan Laboratorium terhadap Peningkatan Nilai Hasil Belajar Peserta Diklat

Terjadi peningkatan rerata nilai akhir peserta diklat sebesar 11.045 persen setelah laboratorium beroperasi (Gambar 6). Namun kenaikan tersebut secara statistik tidak signifikan (Tabel 8). Artinya bahwa keberadaan laboratorium tidak begitu berpengaruh terhadap kenaikan nilai hasil belajar peserta diklat. Hal ini mungkin disebabkan karena sebagian besar soal ujian akhir yang diberikan hanya mencakup pengukuran kemampuan pengetahuan/kognitif peserta saja. (Tabel 9).

Tabel 9. Contoh Soal Ujian Akhir dari Beberapa Mata Diklat pada Diklat PHPI di BDA KKP

No	Diklat	Mata Diklat	Contoh Soal
1	Dasar PHPI Ahli	<p>Persiapan Pelaksanaan Kegiatan Pengendalian Hama dan Penyakit Ikan</p> <p>Identifikasi Jenis, Jumlah, dan Klasifikasi Media Pembawa HPI/HPIK</p> <p>Sistem Manajemen Mutu Laboratorium Pengujian HPI/HPIK (ISO/IEC 17025)</p> <p>Pemeriksaan Klinis Media Pembawa HPI/HPIK</p> <p>Pengambilan Sampel Media Pembawa HPI/HPIK</p> <p>Identifikasi HPI/HPIK Golongan Parasit</p> <p>Identifikasi HPI/HPIK Golongan Cendawan</p> <p>Identifikasi HPI/HPIK Golongan Bakteri</p>	<p>Ikan yang dixxx tidak dapat digunakan untuk pemeriksaan ...</p> <p>Apabila ikan xxx xxx mengalami xxx, maka sistem xxx akan terganggu dan biasanya xxx xxx ikan akan mengalami ...</p> <p>Sebuah xxx yang menjalankan prosedur xxx dan pengendalian xxx dengan menjaga xxx dari xxx xxx maupun tersebarinya xxx disebut ...</p> <p>Apabila ikan mengalami gejala klinis xxx, maka kemungkinan ikan mengalami ...</p> <p>Media pembawa/ikan yang dijadikan sampel untuk xxx diantaranya, kecuali ...</p> <p>Dibawah ini jenis parasit xxx yang tidak bisa diwarnai adalah ...</p> <p>Penyakit jamur yang hanya menginfeksi xxx adalah ...</p> <p>Apabila ikan mengalami xxx, maka kemungkinan ikan tersebut mengalami, kecuali ...</p>
2	Dasar PHPI Terampil	<p>Persiapan pelaksanaan kegiatan pengendalian HPI/HPIK</p> <p>Persiapan pelaksanaan kegiatan pengendalian HPI/HPIK tingkat laboratorium</p> <p>Pemeliharaan media pembawa HPI/HPIK</p> <p>Pengawasan lalulintas media pembawa HPI/HPIK</p> <p>Pemeriksaan Dokumen Karantina dan Persyaratan Lainnya</p> <p>Perlakuan media pembawa HPI/HPIK</p>	<p>Yang dimaksud Persiapan xxx adalah ...</p> <p>Yang termasuk <i>good laboratory</i> faktor xxx adalah ...</p> <p>Komponen dasar pemeliharaan xxx adalah ...</p> <p>Berikut ini merupakan xxx yang sering dilalulintaskan dengan status xxx xxx adalah ...</p> <p>Dalam penerbitan sertifikat xxx masih sering terjadi adanya perbedaan dalam hal jumlah dan jenis sebagaimana tertulis dalam sertifikat xxx bahkan untuk jenis-jenis yang dilindungi, untuk mengantisipasi hal tersebut dilakukan pemeriksaan ulang sebelum keberangkatan. Sesuai ketentuan xxx maka prosentase pemeriksaan adalah ...</p> <p>Dibawah ini adalah kondisi sampel yang tidak dapat digunakan untuk keperluan xxx kecuali ...</p>
3	Penjenjangan PHPI Tingkat Keahlian Muda	<p>Merencanakan pengendalian HPI/HPIK</p> <p>Menyiapkan tempat, alat, dan bahan laboratorium</p> <p>Mengembangkan sistem manajemen laboratorium uji HPI/HPIK</p> <p>Melaksanakan pengendalian hama dan penyakit ikan</p> <p>Memahami peraturan perundangan tentang pengendalian hama dan penyakit ikan</p>	<p>Apa tujuan pemantauan daerah xxx ?</p> <p>Alat xxx yang digunakan untuk xxx harus dikalibrasi besaran ...</p> <p>Salah satu cara untuk xxx alat laboratorium, dilakukan ...</p> <p>Jaminan mutu hasil xxx dilakukan dengan ...</p> <p>Ketentuan tentang xxx diatur dalam ...</p>
4	Penjenjangan PHPI Tingkat Keterampilan Mahir	<p>Menyiapkan data bahan penyusunan rencana kerja</p> <p>Menyiapkan tempat, alat, dan bahan laboratorium</p> <p>Melaksanakan pengendalian hama dan penyakit ikan</p> <p>Memahami peraturan perundangan tentang pengendalian hama dan penyakit ikan</p>	<p>Di bawah ini manfaat dari data bulanan xxx adalah ...</p> <p>Sterilisasi ruang xxx dilakukan pada ...</p> <p>Kegiatan pelabelan koleksi xxx meliputi data berikut ini, kecuali ...</p> <p>Sesuai dengan UU Nomor xxx jabatan xxx terdiri dari ...</p>





Gambar 6. Grafik Rerata Persentase Peningkatan Nilai Ujian Akhir sebelum dan sesudah Laboratorium Beroperasi

Sesuai hasil kajian, peningkatan pengetahuan/kognitif peserta tidak signifikan dipengaruhi oleh keberadaan laboratorium. Beberapa studi diantaranya yang dilakukan oleh Hiltz & Wellman (1997, p. 47) menyatakan bahwa nilai ujian akhir adalah hal yang umum digunakan untuk mengukur pencapaian kognitif peserta didik. Akan tetapi penggunaan nilai untuk pengukuran pencapaian pembelajaran tidak selalu memberikan hasil terbaik. Nilai ujian cenderung memiliki rentang pengukuran yang sangat terbatas dan hanya mencerminkan pencapaian pada ranah (kognitif, afektif dan psikomotorik) yang seragam. Selain itu, nilai memiliki sedikit hubungan dengan apa yang dipelajari peserta. Sebagai contoh, peserta mungkin sudah mengetahui materi sebelum mereka mengikuti diklat sehingga nilai peserta mungkin lebih terkait dengan partisipasi di kelas, penyerahan tugas tepat waktu, atau kehadiran dalam kelas. Selanjutnya, nilai mungkin bukan ukuran pembelajaran yang andal, terutama untuk tes kinerja, karena pengajar yang berbeda atau bahkan pengajar yang sama pada waktu yang berbeda cenderung tidak memberikan nilai secara konsisten. Oleh karena itu, dengan menggunakan nilai sebagai satu-satunya ukuran pembelajaran dapat menimbulkan masalah, terutama saat mengukur hasil pembelajaran di berbagai diklat keterampilan. Pengembangan instrumen komprehensif yang mampu mengukur ketiga ranah pembelajaran akan lebih bermanfaat karena pembel-

ajaran akan selalu melibatkan komponen kognitif, afektif, dan psikomotorik.

Peningkatan peran laboratorium untuk mengukur pencapaian pembelajaran peserta dapat dilakukan dengan ujian praktik yang dilaksanakan dengan memanfaatkan media pembelajaran di laboratorium. Pengukuran pencapaian pembelajaran peserta Diklat PHPI secara umum hanya dilakukan dengan tes tertulis pada akhir diklat. Oleh karena itu instrumen pengukuran seperti ujian praktek di laboratorium perlu ditambahkan agar pembelajaran peserta dapat terukur secara komprehensif.

### Simpulan

Berdasarkan hasil penelitian dan pembahasan dapat disimpulkan bahwa pemanfaatan laboratorium uji karantina ikan di BDA KKP, secara rinci adalah: (1) ada perbedaan kepuasan peserta dan keterampilan peserta diklat dan kepuasan peserta terhadap kelengkapan fasilitas pembelajaran; (2) tidak ada perbedaan hasil belajar dan pengetahuan peserta terhadap materi pembelajaran antara sebelum dan sesudah laboratorium beroperasi.

Penambahan media pembelajaran di laboratorium perlu dilakukan dalam rangka peningkatan kepuasan peserta terhadap kelengkapan media pembelajaran. Selain itu perlu adanya instrumen pengukuran pencapaian peserta yang komprehensif sehingga ketiga ranah pembelajaran (kognitif, psikomotorik dan afektif) dapat dievaluasi dengan baik.

### Daftar Pustaka

- Arikunto, S., & Jabar, C. S. Abdul. (2014). *Evaluasi Program Pendidikan* (2nd ed.). Jakarta: Bumi Aksara.
- Bloom, B. S. (1956). *Taxonomy of educational objective cognitive domain*. New York: Longmans, Green and Co.
- Edward, J. E., Scott, J. C., & Raju, N. S. (2007). *Evaluating human resources programs: A 6-phase approach for optimizing performance*. San Fransisco: Pfeiffer.
- Fathurrahman, P., & Sutikno, S. (2007).

- Strategi belajar mengajar*. Bandung: PT Refika Aditama.
- Fleming, D., & Storr, J. (1999). The impact of lecture theatre design on learning experience. *Facilities*, 17(7/8), 231–236. <https://doi.org/10.1108/02632779910270186>
- Heinich, R., Molenda, M., Russell, J., & Smaldino, S. (1996). *Instructional media and technologies for learning*. Englewood Cliffs New Jersey: Prentice-Hall, Inc, Asimon & Scuster Company.
- Hiltz, S. R., & Wellman, B. (1997). Asynchronous learning networks as a virtual classroom. *Communications of the ACM*, 40(9), 44–49. <https://doi.org/10.1145/260750.260764>
- Hunter, C., Mccosh, R., & Wilkins, H. (2003). Integrating learning and assessment in laboratory work. *Chem. Educ. Res. Pract.*, 4(1), 67–75. <https://doi.org/10.1039/B2RP90038F>
- Kasilingam, G., Ramalingam, M., & Chinnavan, E. (2014). Assessment of learning domains to improve student's learning in higher education. *Journal of Young Pharmacists*, 6(1), 27–33. <https://doi.org/10.5530/jyp.2014.1.5>
- Killen, R. (2003). *Effective teaching strategies: lessons from research & practice*. Melbourne: Thomson Social Science Press.
- Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating training programs: the four levels* (3rd ed.). San Francisco: Berret-Koehler Publisher, Inc.
- Kolb, D. A. (1984). *Experiential learning: experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice Hall.
- Kozma, R. B. (1991). Learning with Media. *Review of Educational Research*, 61(2), 179–211. <https://doi.org/10.3102/00346543061002179>
- Krivickas, R. V., & Krivickas, J. (2007). Laboratory instruction in engineering education. *Global Journal of Engineering Education*, 11(2), 191–196.
- Lembaga Administrasi Negara. (2017). *Petunjuk teknis akreditasi diklat*. Jakarta: Pusat Pengembangan Program dan Pembinaan Diklat, Lembaga Administrasi Negara.
- Menteri Negara Pendayagunaan Aparatur Negara dan Reformasi Birokrasi. Peraturan Menteri Negara Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Nomor 22 Tahun 2010 tentang Jabatan Fungsional Pengendali Hama dan Penyakit Ikan dan Angka Kreditnya (2010).
- Mohan, T., McGregor, H., Saunders, S., & Archee, R. (1997). *Communicating!: theory & practice* (4th ed.). Sydney: Harcourt Brace.
- Naz, A. A., & Akbar, R. A. (2008). Use of media for effective instruction its importance: some consideration. *Journal of Elementary Education*, 18(1–2), 35–40. Retrieved from [http://pu.edu.pk/images/journal/JEE/PDF-Files/JEE-18\(1-2\) No\\_3.pdf](http://pu.edu.pk/images/journal/JEE/PDF-Files/JEE-18(1-2) No_3.pdf)
- Presiden Republik Indonesia. Peraturan Presiden Republik Indonesia Nomor 2 Tahun 2015 tentang Rencana Pembangunan Jangka Menengah Nasional Tahun 2015 - 2019 (2015).
- Pusat Pelatihan Kelautan dan Perikanan. (2013). *Pedoman penyelenggaraan diklat PHPI*. Jakarta: Pusat Pelatihan Kelautan dan Perikanan, Kementerian Kelautan dan Perikanan.
- Salim, K. R., Puteh, M., & Daud, S. M. (2012). Assessing students' practical skills in basic electronic laboratory based on psychomotor domain model. In *International Conference on Teaching and Learning in Higher Education (ICTLHE)* (pp. 546–555).

- Schank, R. C., Berman, T. R., & Macpherson, K. A. (1999). Learning by doing. In C. M. Reigeluth (Ed.), *Instructional-Design Theories and Model: A New Paradigm of Instructional Theory 2* (pp. 161–175). New Jersey: Lawrence Erlbaum Associates.
- Simpson, E. J. (1966). The classification of educational objectives: the psychomotor domain. *Illinois Journal of Home Economics*, 10(4).
- Syamsuddin, A. R., & Damaianti, V. S. (2011). *Metode penelitian pendidikan bahasa*. Bandung: Remaja Rosdakarya.
- Thomas, M. (2008). *Effective teaching: a measure of excellence*. New Delhi: S.Chand & Company.
- Wallace, M. (1991). *Training foreign language teachers*. Cambridge: Cambridge University Press.
- Zaghloul, A.-R. M. (2001). Assessment of lab work: A three-domain model; cognitive, affective and psychomotor. In *ASEE Annual Conference & Exposition*. Albuquerque.

---

**IMPACT OF ADMISSION TYPE ON STUDENTS' ACHIEVEMENT IN  
ECONOMY AND BUSINESS SCHOOL OF  
BATUSANGKAR STATE INSTITUTE OF ISLAMIC STUDIES**

*David<sup>1</sup>\*, Nasviżar Guspendri<sup>1</sup>*

<sup>1</sup>IAIN Batusangkar

<sup>1</sup>Jl. Jend. Sudirman No.137, Limo Kaum, Lima Kaum, Tanah Datar, Sumatera Barat 27217

\* Corresponding Author. Email: davidazis@gmail.com

**Abstract**

The study is aimed at finding the differences in students' academic achievement based on the admission system type to IAIN Batusangkar and the size of the contribution of admission systems to students' achievement. The study uses the quantitative research approach. The comparative multivariate technique is used for the data analysis. Subjects include 159 student candidates for the local-exam stream, 161 for the national-exam stream, and 123 for the achievement-based stream. Findings show that there is a significant difference in academic achievement between students admitted through the achievement stream and students admitted through the local and national-exam streams. There is no significant difference between the local-exam system and the national-exam system. The contribution of admission systems over academic achievement is 5.4%; specified: 5.2% for Semester I, 2.3% for Semester II, 3.8% for Semester III, 4.1% for Semester IV, and 4.4% for Semester V.

**Keywords:** *university admission system, entrance examination, academic achievement*

*Permalink/DOI: <http://dx.doi.org/10.21831/jpep.v22i2.22367>*

## **Introduction**

The student selection system in the university is aimed at grouping candidates who are qualified to be accepted and who are not. Accepted candidates are predicted to be able to finish the education program without much difficulty. Declined candidates are those who are predicted to have a lot of difficulty in finishing the education program. There are at least four things to consider in the student entrance system; namely: (1) prediction effectiveness, (2) economic efficiency, (3) teaching-learning incentive, and (4) equity (Suryabrata, 2004). Accuracy in prediction is the ability of the screening instrument to be able to predict that the accepted students will be successful in their classes. Economic efficiency is integrated in the prediction effectiveness in knowing whether, by increasing fund, time, and effort, it will be worthwhile in terms of the prediction (social benefit and cost). In the equality aspect, equal opportunity is given to those who register.

Further, Suryabrata (2004) stated that selection results will group candidates into four categories: (1) those are predicted to be successful and are successful; (2) those who are predicted to fall short will fall short; (3) those are predicted to be successful but fail; and (4) those who are predicted to fail but actually make it. If 1 and 2 happen, it will not be a problem; however, if it is 3 and 4, this will raise problems. In the case of 3 and 4, the prediction is said to be positively deficient or negative prediction. In order to minimize negative prediction, Suryabrata suggested six alternatives: (1) success criteria, (2) treatment strategy, (3) candidate origin, (4) kind and number of predictors, (5) technique for combining the predictors, and (6) technique for determining passing grades. A selection system functions effectively when the predictors function well. This will happen if the instruments are valid. Further up, Mardapi (1993) stated university entrance selection system needs instruments to measure candidates' qualifications, one of which is a test that is valid and reliable.

Instrument validity, as explained by Allen & Yen (1979), is the accuracy of the instrument in its measurement function. Meanwhile, Azwar (2007) stated that validity can be explained by how far an X observed score can detect its pure T score. A score may not be identical with its real score; they can be identical when the test is valid, complete without errors. In other words, validity is related to how far an instrument measures what it is supposed to measure. In addition, an instrument must be able to give a picture of how accurate the data are. By accuracy is meant that an instrument must be able to depict the smallest differences among the subjects (Mardapi, 2008). Whether or not an instrument is valid substantially depends on the measurement objective. Validity is not applicable for all; an instrument may be valid for one objective but not for another. As it has been stated by Cronbach (1984), an instrument that is valid for one decision making may not be as valid for another and for another group.

In terms of its estimation, for both the characteristics and functions of a test, Azwar (2007b) categorized validity into three types: (1) content validity, (2) construct validity, and (3) criterion-related validity. Content validity answers the question how far the test items cover all the domains of contents, objects, or situations being measured. Analysis of content validity does not involve statistics, but expert judgement. Content validity is categorized into two types: face validity and logical validity. Face validity places emphases on appearance. Face validity is achieved when each test item measures a relevant aspect and the underlying principle is common sense. Construct validity shows how far an instrument measures the theoretical constructs of what is supposed to be measured. Construct validity looks at how far the scores resulting from the measurement reflect the theoretical constructs that underline the development of the instrument. To measure construct validity, complex statistical analyses are needed such as factor analysis and multi-trait multi-method analysis. Criterion-related validity is based on criteria; i.e.

it is detected by the relation between the scores of the test and the scores of the criteria. Suryabrata (2004) stated criterion-related validity is divided into two types: *concurrent validity* and *predictive validity*. Concurrent validity is obtained by using the test scores as a predictor and the scores of another test as a criterion in the same time. Correlation between these two score types gives concurrent validity.

Besides being valid, an instrument must be reliable. Reliability refers to how far the instrument and results of the measurement are trusty. Measurement results are trusty if, in a number of measurement on the same subjects, the results are relatively the same (Suryabrata, 2005). Reliability, according to Mehrens (1980), is the degree of consistency between two measurements on the same subjects. Furthermore, Allen & Yen (1979) refer to reliability as a condition wherein the observed scores have a high correlation with the actual scores. Reliability has two concepts: internal consistency and stability reliability. Internal consistency involves one-time testing, known as single-trial administration, while stability reliability refers to two score groups, results of two test administrations, using the same, or identical, test on the same subjects but at different times (Mardapi, 2008). Further on, Surapranata (2006) stated that there are at least four concepts of reliability: parallel or equivalent, test-retest, split-half, and internal consistency.

Different screening objectives use different instruments. The university selection instrument is aimed at predicting that accepted candidates will not have difficulties in finishing their study. Student candidates registering into IAIN Batusangkar in the academic year 2013 were selected by way of three streams (1) local, (2) national, and (3) academic achievement. Local admission is done by way of written tests and interviews conducted solely by IAIN Batusangkar. The preparation of the test instruments, administering of the tests, and issuance of results were done by the university. For the national stream, the test instruments were prepared

by the central committee; the administering of the test by the local committee. Meanwhile, for the academic achievement stream, selection was done since students were still in the high school by using scores in the report cards. In the present time, these three admission systems are known as (1) Academic Achievement Selection of Islamic State University (SPAN-PTKIN), (2) Entrance Examination of Islamic State University (UM-PTKIN), and (3) Local Screening. However, since the used data are those of the 2013 students, the old names are used. All the three admission systems have the same objective in admitting candidates who are predicted to be successful and rejecting the ones who are predicted to fail. It is a fact, however, that there is no information, up to the present time, which among the three modes of selection is the better one seen from the academic achievement of the students during their study in the university. Thus, the present study is conducted under the title of the impact of the selection types on the academic achievement of students of the Islamic Economy and Business, IAIN Batusangkar.

## Method

In view of the nature of the data and the way the data were obtained, it can be said that the study is *ex-post facto* research. No manipulation or treatment was given to the research variables. *Ex post facto* research, as stated by Kartowagiran (1998), is one that tries to reveal the impact of a variable with no manipulation that may obstruct the genuine cause-effect relation of the variables.

The study took place in IAIN Batusangkar, and the gathered data were students' academic achievement (GPA) of the 2013 academic year, semesters I, II, III, IV, and V. The study was conducted in four months. The study used multivariate analysis of variance (MANOVA) as the data analysis technique, with generalized randomized block design (GRB-p) suggested by Kirk (1995). The design of the comparison of academic achievements can be seen in Table 1.

Table 1. Comparison Design of Students' Academic Achievement among Local, National and Achievement

Admission Stream	Grade Point Average					
	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>
Local (1)	X <sub>111</sub>	X <sub>112</sub>	X <sub>113</sub>	X <sub>114</sub>	X <sub>115</sub>	X <sub>116</sub>
	X <sub>211</sub>	X <sub>212</sub>	X <sub>213</sub>	X <sub>214</sub>	X <sub>215</sub>	X <sub>216</sub>
	Etc.	Etc.	Etc.	Etc.	Etc.	Etc.
National (2)	X <sub>121</sub>	X <sub>122</sub>	X <sub>123</sub>	X <sub>124</sub>	X <sub>125</sub>	X <sub>126</sub>
	X <sub>221</sub>	X <sub>222</sub>	X <sub>223</sub>	X <sub>224</sub>	X <sub>225</sub>	X <sub>226</sub>
	Etc.	Etc.	Etc.	Etc.	Etc.	Etc.
Achievement(3)	X <sub>131</sub>	X <sub>132</sub>	X <sub>133</sub>	X <sub>134</sub>	X <sub>135</sub>	X <sub>136</sub>
	X <sub>231</sub>	X <sub>232</sub>	X <sub>233</sub>	X <sub>234</sub>	X <sub>135</sub>	X <sub>236</sub>
	Etc.	Etc.	Etc.	Etc.	Etc.	Etc.

Notes

- X<sub>1</sub> = IP Semester 1
  - X<sub>2</sub> = IP Semester 2
  - X<sub>3</sub> = IP Semester 3
  - X<sub>4</sub> = IP Semester 4
  - X<sub>5</sub> = IP Semester 5
  - X<sub>6</sub> = Grade Point Average
- Admission Stream  
1 = Local  
2 = National  
3 = Achievement

Findings and Discussion

Findings

The general findings can be divided into two: descriptive and inferential. These can be presented as follows.

Descriptive Data

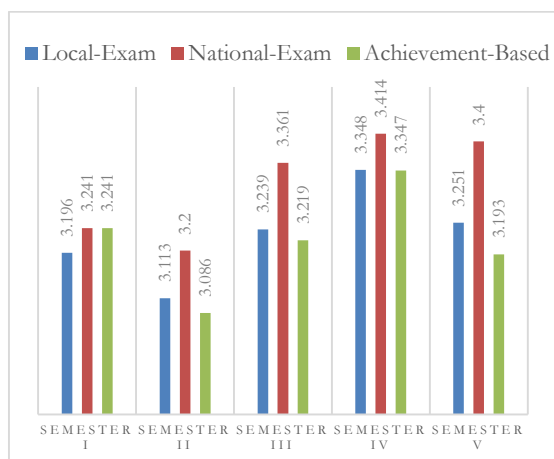


Figure 1. GPA Averages of Students of the Syariah Accounting Department, Academic Year 2013

In Figure 1, it can be seen that, in each semester, the highest GPA average is owned by students who are admitted through the

national stream. The second highest GPA average belongs to the local-stream students. Students who are admitted by way of their school academic achievements have the lowest GPA average. For the local admission system, the lowest GPA average is during Semester I; while in Semesters II to V they are superior to those of students of the third admission type. The data for students of the Syariah Management Department can be seen in Figure 2.

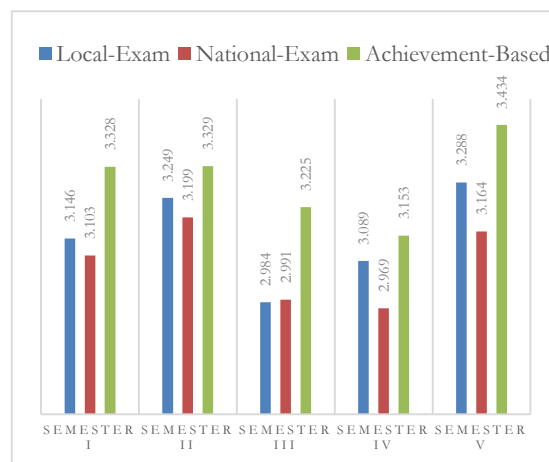


Figure 2. GPA Averages of Students of the Syariah Management Department, Academic Year 2013

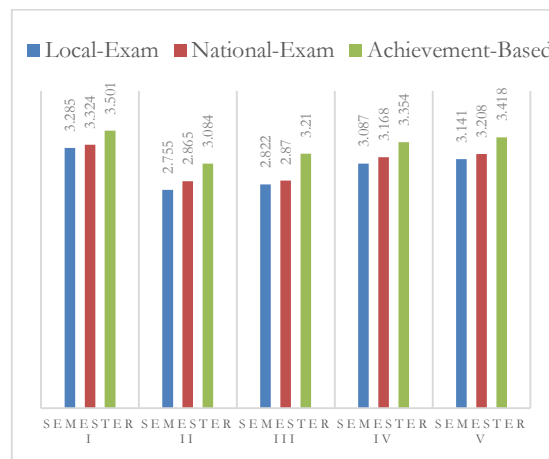


Figure 3. GPA Averages of Students of the Syariah Banking Department, Academic Year 2013

Different from the Syariah Accounting Department, where the highest GPAs are obtained by students of the national admission, in the Syariah Management Department, students of the school achievement admis-

sion type are superior to other students from Semesters I to V. After this, students of the local admission system follow. Next is the nature of academic achievements of students of the Syariah Banking Department in Figure 3.

For the *Syariah* Banking Department, the highest GPA average belongs to the achievement-stream students. This is the same as what happens in the *Syariah* Management Department. Figure 4 presents data of the Informatics Management Department.

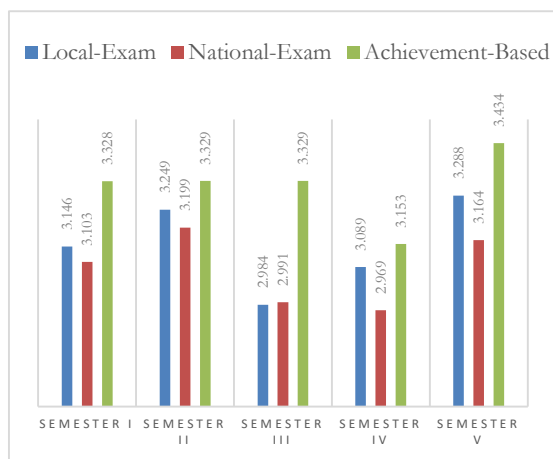


Figure 4. GPA Averages of Students of the *Informatics Management* Department, Academic Year 2013

In Figure 4, students who are admitted through the school academic achievements have higher achievements than those from both the local and national admission systems throughout the five semesters. Seeing results from the four departments, it turns out that students of the national recruitment system do better in their academic achievements are only in those of the *Syariah* Accounting Department. For the Departments of *Syariah* Management, *Syariah* Banking, and Informatics Management, students who have higher achievements are those who are admitted through the school academic achievement stream. So, based on the data description, it can be seen that there is an impact of the admission types on the students' academic achievements. However, to know whether or not the difference is significant, an analysis is conducted that compares the differences among departments and among semesters.

### *Multivariate Analysis of Variance (MANOVA)*

An assumption of the =MANOVA is that each dependent variable should have the same variance for all the groups. To test this assumption, the Levene's test is used. Results of the Levene's test can be seen in Table 2.

Table 2. Levene's Test of Equality of Error Variances

	F	df1	df2	sig
Semester I	1.883	2	440	.153
Semester II	.533	2	440	.587
Semester III	.052	2	440	.949
Semester IV	.544	2	440	.581
Semester V	1.041	2	440	3.54
Grade Point Average	.105	2	440	.900

The null hypothesis test: Error variance of the dependent variable is equal across groups

#### *a. Design: Intercept + JM*

In Table 2, it can be seen that all dependent variables have a significance level higher than 0.05. This means that each dependent variable has the same variance fulfilling the requirement of the MANOVA. Multivariate tests are used to find the differences among students' academic achievements across semesters I, II, III, IV, and V and among the academic achievements of students who are admitted by the three modes of local, national, and achievement.

Results of the analyses can be seen in Table 3.

The multivariate test to know whether each factor has an impact on the dependent variable groups can be done by any of the four multivariate tests for significance; they are Pillai's Trace, Wilks' Lambda, Hotelling's Trace, and Roy's Largest Root. Based on Table 3, it can be seen that the Wilks' Lambda  $F = 3.562$  and  $sig. = 0.000$ , so that the  $H_0$  is rejected meaning that there is a difference in academic achievements among students of semesters I, II, III, IV, and V and in the GPAs among students who are admitted through the modes local, national, and achievement. To know the differences in the students' achievements in each semester among the three recruitment systems, data analysis is presented in Table 4.



Table 3. Multivariate Tests

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.993	10091.262 <sup>b</sup>	6.000	435.000	.000
	Wilks' Lambda	.007	10091.262 <sup>b</sup>	6.000	435.000	.000
	Hotelling's Trace	139.190	10091.262 <sup>b</sup>	6.000	435.000	.000
	Roy's Largest Root	139.190	10091.262 <sup>b</sup>	6.000	435.000	.000
JM	Pillai's Trace	.093	3.555	12.000	872.000	.000
	Wilks' Lambda	.909	3.562 <sup>b</sup>	12.000	870.000	.000
	Hotelling's Trace	.099	3.568	12.000	868.000	.000
	Roy's Largest Root	.070	5.091 <sup>c</sup>	6.000	436.000	.000

a. Design: Intercept + JM

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level

Table 4. Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	Semester I	2.253 <sup>a</sup>	2	1.126	12.009	.000
	Semester II	1.512 <sup>b</sup>	2	.756	5.073	.007
	Semester III	3.340 <sup>c</sup>	2	1.670	8.798	.000
	Semester IV	2.298 <sup>d</sup>	2	1.149	9.294	.000
	Semester V	2.316 <sup>e</sup>	2	1.158	10.012	.000
	Grade Point Average	2.451 <sup>f</sup>	2	1.226	12.485	.000
Intercept	Semester I	4699.200	1	4699.200	50099.666	.000
	Semester II	4218.383	1	4218.383	28296.248	.000
	Semester III	4228.578	1	4228.578	22276.030	.000
	Semester IV	4595.648	1	4595.648	37180.221	.000
	Semester V	4605.016	1	4605.016	39815.163	.000
	Grade Point Average	4469.043	1	4469.043	45524.328	.000
JM	Semester I	2.253	2	1.126	12.009	.000
	Semester II	1.512	2	.756	5.073	.007
	Semester III	3.340	2	1.670	8.798	.000
	Semester IV	2.298	2	1.149	9.294	.000
	Semester V	2.316	2	1.158	10.012	.000
	Grade Point Average	2.451	2	1.226	12.485	.000
Error	Semester I	41.271	440	.094		
	Semester II	65.595	440	.149		
	Semester III	83.524	440	.190		
	Semester IV	54.386	440	.124		
	Semester V	50.890	440	.116		
	Grade Point Average	43.194	440	.098		
Total	Semester I	4789.900	443			
	Semester II	4332.029	443			
	Semester III	4351.165	443			
	Semester IV	4697.794	443			
	Semester V	4706.235	443			
	Grade Point Average	4558.086	443			
Corrected Total	Semester I	43.524	442			
	Semester II	67.107	442			
	Semester III	86.864	442			
	Semester IV	56.684	442			
	Semester V	53.206	442			
	Grade Point Average	45.645	442			

a. R Squared = .052 (Adjusted R Squared = .047)

b. R Squared = .023 (Adjusted R Squared = .018)

c. R Squared = .038 (Adjusted R Squared = .034)

d. R Squared = .041 (Adjusted R Squared = .036)

e. R Squared = .044 (Adjusted R Squared = .039)

f. R Squared = .054 (Adjusted R Squared = .049)

Table 5. Summary of the Results of Analysis of Differences in Students' Achievements

No	Achievement	F	Sig.	Decision	Contribution of Recruitment Type on Achievement
1	Semester I	12.009	.000	H <sub>0</sub> rejected	5.2%
2	Semester II	5.073	.007	H <sub>0</sub> rejected	2.3%
3	Semester III	8.798	.000	H <sub>0</sub> rejected	3.8%
4	Semester IV	9.294	.000	H <sub>0</sub> rejected	4.1%
5	Semester V	10.012	.000	H <sub>0</sub> rejected	4.4%
6	GPA	12.485	.000	H <sub>0</sub> rejected	5.4%

Table 6. Multiple Comparisons

Dependent Variable	(I) Stream	(J) Stream	Mean Difference		Sig.
			(I-J)	Std. Error	
Semester I	Local	National	-.0138	.03424	.922
		Achievement	-.1656*	.03678	.000
	National	Local	.0138	.03424	.922
		Achievement	-.1519*	.03668	.000
	Achievement	Local	.1656*	.03678	.000
		National	.1519*	.03668	.000
Semester II	Local	National	-.0425	.04317	.617
		Achievement	-.1455*	.04636	.008
	National	Local	.0425	.04317	.617
		Achievement	-.1030	.04624	.085
	Achievement	Local	.1455*	.04636	.008
		National	.1030	.04624	.085
Semester III	Local	National	.0040	.04871	.997
		Achievement	-.1918*	.05232	.001
	National	Local	-.0040	.04871	.997
		Achievement	-.1958*	.05218	.001
	Achievement	Local	.1918*	.05232	.001
		National	.1958*	.05218	.001
Semester IV	Local	National	.0165	.03931	.915
		Achievement	-.1517*	.04222	.002
	National	Local	-.0165	.03931	.915
		Achievement	-.1683*	.04210	.000
	Achievement	Local	.1517*	.04222	.002
		National	.1683*	.04210	.000
Semester V	Local	National	-.0680	.03802	.203
		Achievement	-.1822*	.04084	.000
	National	Local	.0680	.03802	.203
		Achievement	-.1142*	.04073	.020
	Achievement	Local	.1822*	.04084	.000
		National	.1142*	.04073	.020
Grade Point Average	Local	National	-.0242	.03503	.787
		Achievement	-.1767*	.03762	.000
	National	Local	.0242	.03503	.787
		Achievement	-.1525*	.03752	.000
	Achievement	Local	.1767*	.03762	.000
		National	.1525*	.03752	.000

Based on Table 4, differences in students' academic achievements are summarized in Table 5. From Table 5, it is known that for every semester, including for grade point averages, the  $H_0$  is rejected. This means that, for each semester, there is a difference in academic achievements among students who are admitted by the local, national, and achievement systems. It can therefore be concluded that there is an impact of recruitment systems on students' achievements. In terms of the contribution of recruitment systems on semesters, the highest is on GPA, then Semester I, then Semester V last.

To further know in more details the differences among the recruitment systems, the Scheffe test is done. The results can be seen in Table 6.

From the results of the multiple comparison calculations, the *mean difference* between the local and national systems is -0.0138 with a significance 0.922, between the local and achievement -0.1656 and 0.000, and between the national and achievement systems 0.1519 and 0.000. It can be stated that there is a difference in achievements between students of the achievement recruitment and local recruitment and between students of the achievement recruitment and national recruitment. There is no significant difference in achievement between students of the local and national recruitment systems for Semester I. For more detailed pictures, more information on the results of the Scheffe test is presented in Table 7.

Table 7. GPA Average for Semester I

Recruitment Stream	N	Subset	
		1	2
Local	159	3.2223	
National	161	3.2360	
Achievement	123		3.3879
Sig.		.929	1.000

In terms of the homogenous subset, it can be seen that, for students' GPAs for Semester I, the local and national streams can be one group and the achievement stream stands alone as a different group.

This means that there is no difference in students' achievement for Semester I between the local and national groups. Seen from the averages, it turns out that higher achievements are obtained by students of the achievement recruitment system.

In Table 6, for Semester II, a mean difference of -0.0435 is found between the local and national streams with a significance of 0.617; local and achievement streams -0.1455 with a significance of 0.008; and national and achievement -0.1030 with a significance of 0.085. It turns out that differences only occur between the local and achievement groups; no difference between the local and national groups and between the national and achievement groups. This can be seen in Table 8.

Table 8. GPA Average for Semester II

Recruitment Stream	N	Subset	
		1	2
Local	159	3.0470	
National	161	3.0894	3.0894
Achievement	123		3.1924
Sig.		.645	.076

In terms of the homogenous subset, it can be seen that students' achievements for Semester II are categorizable into two groups, however achievements of students of the national group overlap with those of the two other groups. This means that the achievements of the national stream could be one group with those of the local stream and with those of the achievement stream. Seen from the score averages, higher achievements are with students of the achievement system.

As can be seen in Table 6, for Semester III, a mean difference is obtained between the local and national streams (0.0040) with a significance = 0.997; between the local and achievement streams (-0.1918) with a significance = 0.001; and between the national and achievement streams (0.1958) with a significance = 0.001. It can be concluded that there are differences in students' achievements between the local and achievement groups and between the national and

achievement groups. There is no difference for the local and national groups. This is presented in a summary form as Table 9.

Table 9. GPA Average for III

Scheffe <sup>a,b,c</sup>			
Recruitment Stream	N	Subset	
		1	2
Local	161	3.0468	
National	159	3.0508	
Achievement	123		3.2426
Sig.		.997	1.000

From the homogeneous subsets, it can be seen that the GPAs of students recruited through the national system are in one group with those of the achievement system. From the averages, it turns out that achievements in Semester III are higher with students who are admitted through the school achievement system.

For Semester IV, as can be seen in Table 6, the local and national groups have a mean difference of 0.0165 with a significance of 0.915; the local and achievement groups -0.1517 with 0,002; and the national and achievement groups -0.1683 with 0.000. There is a difference in the academic achievements of the students who are accepted through the local and achievement systems, and through the national and achievement systems for Semester IV. These differences are further shown in Table 10.

Table 10. GPA Average for Semester IV

Scheffe <sup>a,b,c</sup>			
Recruitment Stream	N	Subset	
		1	2
Local	159	3.1656	
National	161	3.2336	
Achievement	123		3.3478
Sig.		.235	1.000

From the homogeneous subsets, it can be seen the academic achievements of Semester V students from the local and national streams become one group and those of the achievement stream stands alone as a different group. Based on their average score, the students of the achievement re-

cruitment system tend to have higher GPAs than those of the other two streams.

Finally, for Semester V, as can be seen in Table 6, there is a mean difference between the students admitted by the local and national streams by -0.0242 with a significance of 0.787, between the local and achievement streams by -0.1767 with a significance of 0.000, and between the national and achievement streams by -0.1525 with a significance of 0.000. Based on these data, it can be seen that there is a difference between the local and achievement groups and the national and achievement groups. However there is no difference between the local and national groups. For a more detailed picture, this is represented in Table 11.

Table 11. GPA Average for Semester V

Scheffe <sup>a,b,c</sup>			
Recruitment Stream	N	Subset	
		1	2
Local	159	3.1337	
National	161	3.1579	
Achievement	123		3.3104
Sig.		.805	1.000

Seen from the homogeneous subsets, it can be seen that the GPAs of the students of the local and national streams become one group, while those of the achievement stream form a different group. Based on the GPA average scores, the academic achievement of the achievement group is higher than those of the other two groups.

## Discussion

The academic achievement of students admitted by the national recruitment system is higher in the Accounting Department than that in the local and achievement systems. This high achievement occurs in all semesters except for Semester I where there are equal levels. On the contrary, there is a different phenomenon for the other three departments; *Syariah* Management, *Syariah* Banking, and Informatics Management. For these three departments, high academic achievement is achieved by students of the

school achievement system. Up from Semester I through Semester V, students from the achievement stream consistently show higher academic performances. In this recruitment system, the primary consideration is laid on the achievements of the candidates through the scores in the high-school report cards from Grade X to Grade XII. Candidates are nominated by the school, taken from the best ten of the class. Meanwhile, candidates who are admitted through the national or local entrance examinations rely solely on the results of the examination.

From the multiple comparisons of students' GPAs, it can be seen that there are significant differences between the achievement stream and the national and local streams. Meanwhile, there is no significant difference in the GPAs between the national and local streams. The absence of differences in this matter may be caused by the fact that the national and local systems are the same in that they rely solely on the results of the entrance tests. It is different from the school achievement system in which criteria for accepting the candidates are based on the academic achievements of the candidates in their three years of high-school education. It is true that selection is influenced by several factors such as, among others, procedures, motivation, testing contexts, candidates' experiences, scoring methods, and exam supervisors' characteristics (Gregory, 2013). Included in these are validity and reliability of the test items. It has been further proposed by Ghani (2008) that types and autonomy levels have impacts on learning outcomes.

Contribution of the admission system toward GPA is not that high, only 5.4%; to be more specific: 5.2% for Semester I, 2.3% for Semester II, 3.8% for Semester III, 4.1% for Semester IV, and 4.4% for Semester V. Stronger influences come from some other different factors. In general, factors affecting achievement can be internal or external. Internal factors include health, psyche, intelligence, attention, interest, talents, maturity, and readiness. Fatigue can also be one. External factors include family, school, and society (Slameto, 2010). A study by Waluyo

(2006) shows that learning styles and approaches have impacts on learning outcomes. Finally, teachers' evaluation techniques and instruments can also influence learning achievements; considering that the instruments used by the teacher may not have been valid and reliable.

## **Conclusion**

Based on the results of the multivariate statistical analyses, it is found that there is a significant difference in the students' academic achievements between those admitted by the school achievement recruitment stream and those admitted by the national or local entrance examinations. In another fact, there is no significant difference in the academic achievement between students who are accepted through the national examination stream and those who are accepted through the local examination stream. The GPAs of students of the achievement system are higher than GPAs of students of the national and local systems through Semesters I to V. Contribution of the admittance type towards students' GPAs is 5.4%; to be more specific: 5.2% for Semester I, 2.3% for Semester II, 3.8% for Semester III, 4.1% for Semester IV, and 4.4% for Semester V .

In view of the research findings, it is suggested that university admission systems have more say for the achievement-based recruitment method added with considerations of the candidates' performances during their three-year educational processes in the high school. Since the contribution of admission systems towards academic achievement is not so high, it is also suggested that sufficient attention be given to students' instructional processes after being admitted to study in the university.

## **Daftar Pustaka**

- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey, CA: Brooks/Cole Publishing Company.

- Azwar, S. (2007a). *Dasar-dasar psikometri*. Yogyakarta: Pustaka Pelajar.
- Azwar, S. (2007b). *Reliabilitas dan validitas*. Yogyakarta: Pustaka Pelajar.
- Cronbach, L. J. (1984). *Essentials of psychological testing* (5th ed.). New York: Harper Dayan.
- Ghani, A. R. A. (2008). Pengaruh tes formatif dan kemandirian belajar terhadap hasil belajar Ekonomi siswa SMA. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 12(2). Retrieved from <https://journal.uny.ac.id/index.php/jpep/article/view/1425>
- Gregory, R. J. (2013). *Tes psikologi: sejarah, prinsip, dan aplikasi*. Jakarta: Erlangga.
- Kartowagiran, B. (1998). Pengenalan metode penelitian eksperimen dan ex post facto. In *Penataran Berbagai Pendekatan dan Metode Penelitian* (p. 7). Yogyakarta: Lemlit IKIP Yogyakarta.
- Kirk, R. E. (1995). *Experimental design: procedures for the behavioral sciences*. New York: International Thomson Publishing Inc.
- Mardapi, D. (1993). *Daya prediksi tes masuk IKIP Jakarta terhadap prestasi dan lama studi mahasiswa Pascasarjana KPK IKIP Yogyakarta*. Yogyakarta.
- Mardapi, D. (2008). *Teknik penyusunan instrumen tes dan nontes*. Yogyakarta: Mitra Cendikia Press.
- Mehrens, W. A. (1980). *Measurement and evaluation in education and psychology*. Forth Worth: Holt, Rinerhart, and Winston, Inc.
- Slameto. (2010). *Belajar dan faktor-faktor yang mempengaruhi*. Jakarta: Rineka Cipta.
- Surapranata, S. (2006). *Analisis, validitas, reliabilitas dan interpretasi hasil tes, implementasi kurikulum 2004*. Bandung: PT. Remaja Rosdakarya.
- Suryabrata, S. (2004). Pengembangan sistem seleksi calon mahasiswa perguruan tinggi yang akurat dan berkeadilan. In *Seminar Nasional Rekrutasi Sistem Penilaian dalam Rangka Meningkatkan Kualitas Pendidikan* (p. 3). Yogyakarta: Himpunan Evaluasi Pendidikan Indonesia (HEPI).
- Suryabrata, S. (2005). *Pengembangan alat ukur psikologis*. Yogyakarta: Penerbit Andi.
- Waluyo, W. (2006). Pengaruh pendekatan konstruktivistik dan gaya belajar siswa terhadap prestasi belajar Kimia siswa kelas I semester 2 SMA Negeri 1 Ngaglik Sleman tahun pelajaran 2003/2004. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 8(1). Retrieved from <https://journal.uny.ac.id/index.php/jpep/article/view/2016>

## DEVELOPING ASSESSMENT INSTRUMENT OF QIRĀATUL KUTUB AT ISLAMIC BOARDING SCHOOL

*Ajeng Wahyuni<sup>1</sup> \*, Badrun Kartowagiran<sup>1</sup>*

<sup>1</sup>Universitas Negeri Yogyakarta

<sup>1</sup>Jl. Colombo No. 1, Depok, Sleman 55281, Yogyakarta, Indonesia

\* Corresponding Author. Email: [ajengwahyuni77@gmail.com](mailto:ajengwahyuni77@gmail.com)

### Abstract

The purpose of this study was to develop and describe the quality and characteristics of an assessment instrument of by name of *Qirāatul Kutub*. This development research is based on Borg & Gall model. The steps of the development were (1) planning, (2) primary developing product, (3) preliminary field testing, (4) main field testing, and (5) final product revision. The subjects for preliminary field testing were 28 testees for performance test instrument and 96 testees for grammar mastery test derived from students of Darul Huda Islamic Boarding School, whereas the subjects for main field testing were 80 students for performance test instrument and 425 students for grammar mastery test. The research instruments were assessment instrument of *Qirāatul Kutub* which consisted of performance assessment instrument and grammar mastery essay test, observation and interview guide. The research data were analyzed based on Rasch Model. The results showed that: (1) the results of the development were 12 items of the performance test and 17 items of the grammar mastery test which were fit with the proposed model, (2) the item difficulty of the performance test ranged from -1.32 to 0.99, while for the grammar mastery test from -1.13 to 1.01; (3) the *person reliability* for the *performance test* was 0.86, *item reliability* was 0.97, and *Cronbach's alpha* value was 0.97. The grammar mastery test has person reliability 0.97 and item reliability 0.99 with Cronbach's alpha 0.90; (4) based on the test information function (TIF) and Standard Error of Measurement (SEM), the instruments were found to be good; the performance test and the grammar mastery test were suitable for learners with ability masteries between -4 and 4.

**Keywords:** *qirāatul kutub, tarkib, assessment instrument, Islamic boarding school*

Permalink/DOI: <http://dx.doi.org/10.21831/pep.v22i2.16592>

## Introduction

The *Kitab Kuning* (lit.: Yellow Book) is part of a long history of the oldest education institution in Indonesia, the *Pesantren* (Departemen Agama Republik Indonesia, 2003). The *Kitab Kuning*, along with other classical books in Arabic, that has been learned through generations is one of the characteristics of the traditional Islamic boarding school *Salafiyah* (Arifin, 2012, p. 42). The Yellow Book is used as a learning source for various materials such as law, faith, worship, etc. In addition to these purposes, the learning of the Yellow Book requires the mastery of Arabic even in places where Moslems are a minority (Wekke, 2015, p.314). The proficiency in reading the Yellow Book becomes an obligatory skill for students in the traditional *pesantren*. This skill of reading the Yellow Book is called *Qirāatul Kutub*, literary meaning reading the book.

In Arabic learning, *Qirāatul Kutub* is included in the area of language reading skills (*qiraah* = reading). Reading (Nurgiyantoro, 2001, p. 24) is an effort to understand what is told through writing and, thus, knowledge about the writing system, the alphabet, and spelling is important. Meanwhile, Iskandarwassid & Sunendar (2015, p. 246) state that reading is done to get meaning and understanding what is being written. In general, *Qirāatul Kutub* is an activity conducted to understand the meaning and content of the Yellow Book.

As in other reading activities, *Qirāatul Kutub* needs the accompaniment of other skills such as grammar, semantics, and comprehension. Besides, one other important skill of *Qirāatul Kutub* is fluency. These four components then become the criteria for evaluating *Qirāatul Kutub*.

Text reading fluency refers to the quality of reading that is fast and accurate with a natural intonation (Veenendaal, Groen, & Verhoeven, 2015, p. 213). Arabic grammar mainly consists of word forms (Al-Hasyimi, 1971, p. 1). Comprehension is the main component and purpose of reading (Wheeler, Cartwright, & Swords, 2012, p. 416).

In the tradition of the *pesantren*, *Qirāatul Kutub* uses the *Sorogan* method. It is an individual learning method where each learner is faced to the teacher. The learning process runs naturally, proceeding as it does, as it has traditionally and culturally run for hundreds of years. Evaluation is left for the teacher to make to decide whether a student is given a pass or is obliged to re-read the book. This type of evaluation is obviously highly subjective. Learners' achievement is known only to the teacher and decided on only by the teacher.

This subjective evaluation tends to raise questions. According to Nitko & Brookhart (2007, p. 8) evaluation is a process of decision making about the quality and achievement of the learner. To know whether or not an instructional process is successful much depends on the information from an evaluation process. Such information may reveal the quality of the learning product, achievements, and learning difficulties of the students. It may even reveal whether the teacher is successful or not.

The present study is aimed at developing an evaluation instrument for *Qirāatul Kutub* consisting of indicators, assessment criteria, performance test, *tarkib* mastery test, and evaluation report. The presence of a *Qirāatul Kutub* evaluation instrument in the *pesantren* is expected to offer the possibility that evaluation may truly measure success in relation to the objectives and components in the learning. Eventually, actions can be taken as to what aspects can be improved.

## Method

The study was developmental research of the Borg & Gall (1989) model with four developmental phases. The first phase was planning which consisted of research and information collecting to include literary reviews related to the problems under study, needs analysis, and preparation for formulating the research framework. In addition, this phase also involved the development of the research procedure, formulation of the objectives to be achieved at each phase, and research design and steps.



The second phase was developing the preliminary form of the product. This phase was intended to develop the initial form of the product. This included preparing for supporting components, guides, and feasibility appraisals of the supporting elements.

The next phase was preliminary field testing, which consisted of an initial field testing of a limited scale. The results of this preliminary try-out was used to make improvements of the instrument items. The main field testing was used to reveal the final characteristics of the instrument. The products of the two try-outs were a performance test instrument and a *Tarkib* achievement test instrument of the *Safinah An-Najah* book, both being valid and reliable.

The fourth phase was *operational product revision*, which was revision/perfection of the results of the expanded try-out.

Lastly, the final phase was final product revision. This consisted of the final revision of the developed product.

The product developed in the study consists of several parts. First, the indicators and evaluation criteria are to be used as a guideline by the teacher in giving evaluation to the students' learning. Second, evaluation sheets of *Qiraatul Kutub* achievement contain essay test items to measure the student's achievement in *tarkib*. And last, evaluation report sheets are used to show the results of the evaluation of the student's performances.

The research subjects were male and female teachers and the students of the Yellow Book learning using the *sorogan* method of the "Darul Huda" *Pesantren* Boarding School in Mayak, Ponorogo. The population consisted of 1,098 students. Respondents for the performance test instrument were five female teachers who were credible in the field, who were members of the examination board for class promotion in *Sorogan*; one as the limited-scale try-out respondent and four as the extended field try-out respondents. The limited try-out involved 28 students for the performance test and 96 students for the *Tarkib* mastery test. The students belonging to the four female teachers were 80 in num-

ber. Assignment of the Kretjie & Morgan Table (Wagiran, 2015) for determining the number of the subjects for the essay-type test resulted in 225 students of the marked *Safinah* and 200 of the unmarked *Safinah*. Considering the sample numbers above, sampling was done by simple random sampling within which all students in the group had the equal chance to be selected as a research subject.

In relation to data analyses, qualitative data analyses were used for data derived from the interviews concerning criteria, indicators, and variables. Quantitative data analyses were used for the results of the product try-outs. These analyses were also used to find out the improvements on the product and the validity and reliability measures of the instrument. These analyses also showed the characteristics of the instrument items as well as the achievement of the students.

The first analysis was testing the content validity by expert agreement on the Aiken indexes. According to Retnawati (2016, p. 19), an index is regarded as having a low validity when it is lower than or equal to 0.4, moderate when it is between 0.4 and 0.8, and high when it is higher than 0.8. The second described the quality of the evaluation instrument based on the *Cronbach alpha* value and the Rating Scale Quality Instrument Criteria by looking at the Standard Error of Measurement. Concerning the criteria for the Cronbach alpha value reliability, the instrument will be more reliable as its reliability value is close to 1. Finally, the analyses of test items and evaluation items were done by the Rasch Model since *Partial Credit Model* or PCM was used (Mardapi & Kartowagiran, 2011, p. 329) by the aid of the *Winstep 4.3*. Software. The analyses concentrated on looking at the item fits toward the model (*Goodness of Fit*), level of difficulty, and dispersion of the students' performances.

Levels of difficulty were indicated by item measure values in the logit scale ranging from  $-\infty$  to  $+\infty$ . The higher the logit values, the higher the levels of difficulty (Sumintono & Widhiarso, 2015, p. 70).

The next analysis was conducted on the Test Information Function (TIF) and Standard Error of Measurement (SEM). TIF was used to describe the strengths of a test in conveying the testee's abilities (Retnawati, 2016). It had a reverse correlation with SEM in that the higher the TIF, the lower the SEM. TIF also gave information as how far the test best explained test information. The ability range where the TIF curve crossed the SEM curve was the spot where a test gave the best test information.

The testee's abilities in the analysis output results used the *Winstep 3.73* program in the form of the logit scale. The ability range in the logit scale was from  $-\infty$  to  $+\infty$ . The values of the abilities in the logit scale were then converted into the 0-to-100 range to make it easy for the teacher or reader to read them. The converted scores were finally grouped into numbers of categories of ability levels.

## Findings and Discussion

The product of the development of the study was a *Qiraatul Kutub* evaluation instrument consisting of 17 essay-type test items of *tarkib* mastery of the *Safinah An-Najah* book. Validation was given by experts in the fields of *Qiraatul Kutub* and assessment, including qualitative item analyses covering contents, construction, and language. The content validity was calculated using the Aiken formula. This phase involved two university lecturers.

Using the Aiken formula for content validity, all the 17 test items for *tarkib* mastery were viable to be used. The Aiken index scores ranged from 0.6 to 1. According to (Retnawati, 2016), an Aiken index score was accepted when it was higher than 0.4.

The results of the field try-out were seen from the test item characteristics and the goodness of fit of the model. This also involved the use of TIF and SEM of the aid of the *Winstep 3.73* software. Before the analyses were conducted, three pre-requisite assumptions should be fulfilled of the instrument, namely unidimension, local independence, and parameter invariance. The as-

sumption for unidimension was seen from the Principal Component Analysis (PCA) of the eigenvalue score. When the eigenvalue was higher than 2.0 and the percentages of the variance were higher than the variance percentages of the item, then other factors or dimensions were present and the test was multidimensional. To satisfy the pre-requisite of unidimension, the eigenvalue score should be lower than 2.0 (Linacre, 2011).

From the results of the analyses, it was found that the eigenvalue scores for the first to the last contrasts were lower than 2.0, with a percentage of 5.0%. The conclusion was that there were no other dominant dimensions in the *Qiraatul Kutub* evaluation instrument. The absence of other dominant dimension showed that the instrument measured only one dimension, and thus the unidimension pre-requisite was fulfilled.

The local independence assumption was satisfied when the residual correlation among the items was not higher than 0.3 (Sumintono & Widhiarso, 2015). The results of the analyses showed that none of the items had a residual correlation above 0.3, and thus the pre-requisite for local independence was fulfilled.

The parameter invariance assumption was obtained by dividing the respondents into odd and even groups and comparing their parameters, in the case of the study, levels of difficulty. The results of the analyses showed that the instrument had fulfilled the assumption for parameter invariance.

## Reliability of the Instrument

The reliability of the *Qiraatul Kutub* instrument can be seen in Table 1. The Rasch Model reliability can be seen from the values of the *Cronbach Alpha*, item reliability, and person reliability. It can also be seen, from the separation for item and person, that the higher the separation value, the better the test. Person separation is used to classify testees; if the score is low, it is possible that the instrument is not too sensitive in differentiating between high-scoring and low-scoring testees (Pada, Kartowagiran, & Subali, 2016, p. 9). The reliability criteria are

based on the rating scale instrument quality criteria of (Fisher, 2007).

Table 1. Reliability Measure of the *Qirāatul Kutub* Evaluation Instrument

No	Reliability	Performance Test	Notes	Tarkib Test	Notes
1	Person Reliability	0,86	Good	0,91	Very good
	Person Separation	2,45	Moderate	3,17	Good
2	Item Reliability	0,97	Very good	0,99	Very good
	Item Separation	6,03	Very good	12,98	Very good
3	Test Reliability (Cronbach Alpha)	0,85	Good	0,90	Good

Goodness of Fit

The goodness of fit on the Rasch model was seen from the values of the outfit mean square MNSQ, Z-Standard (ZTSD), and point measure correlation (PR Corr) by the following criteria. The accepted MNSQ value was  $0.5 < \text{MNSQ} < 1.5$ , or between 0.5 and 1.5. The accepted ZTSD value was  $-2.0 < \text{ZTSD} < +2.0$ , or between -2.0 and +2.0. For the Pt measure correlation, the accepted value was  $0.4 < \text{Pt Measure Corr} < 0.85$ , or between 0.4 and 0.85. According to Sumintono & Widhiarso (2015, p. 73) an item is considered to fulfill the fit if it has the three criteria. However, an item can still be retain if it has one or two of the three. Results of analyses of the data from the field try-outs are explained through Tables 2 and 3.

Table 2. Goodness of Fit of Field Try-outs Performance Test

Item	Outfit		Pt-Measure Corr.	Notes
	MNSQ	ZTSD		
1	0,99	00	0,71	Fit
2	0,62	-2,9	0,83	Fit
3	0,76	-1,7	0,63	Fit
4	0,78	-1,5	0,58	Fit
5	0,77	-1,6	0,72	Fit
6	0,96	-0,2	0,70	Fit
7	0,90	-0,6	0,64	Fit
8	1,02	0,2	0,63	Fit
9	1,35	2,2	0,61	Fit
10	0,82	-1,1	0,69	Fit
11	1,37	2,1	0,22	Fit
12	1,44	2,4	0,20	Fit

In Table 2, it can be seen that 12 items of the performance test have a fit although two items, 11 and 12, do not satisfy the criteria for ZTSD and Pt. Corr. The two items still fulfill the criteria for MNSQ; therefore, they may still be retained (Sumintono & Widhiarso, 2015, p. 73).

Table 3. Goodness of Fit of *Tarkib* Mastery Test

Item	Outfit		PT-Measure Corr.	Notes
	MNSQ	ZSTD		
b1	1,24	3,0	0,63	Fit
b2	0,80	-3,1	0,68	Fit
b3	0,96	-0,4	0,55	Fit
b4	0,91	-1,0	0,62	Fit
b5	0,96	-0,5	0,57	Fit
b6	1,01	0,2	0,62	Fit
b7	1,12	1,7	0,65	Fit
b8	0,97	-0,4	0,61	Fit
b9	1,17	2,2	0,53	Fit
b10	1,04	0,4	0,62	Fit
b11	1,00	0,0	0,62	Fit
b12	1,12	1,6	0,58	Fit
b13	0,92	-1,1	0,67	Fit
b14	0,99	-0,1	0,60	Fit
b15	0,95	-0,5	0,64	Fit
b16	1,07	-1,0	0,63	Fit
b17	0,86	-1,4	0,68	Fit

In Table 3, it can be seen that all the essay-type items of the *tarkib* of the *Qirāatul Kutub* have a fit on the Rasch model, or that the items can be subjected to analyses of this model. As in the above case, although items b1, b2, and b9 do not fulfill the criteria for ZTSD, they satisfy the other two criteria and, hence, can still be retained. Thus, all 17 items have the fit.

Item Characteristics

The item characteristics of the *Qirāatul Kutub* instrument by the Rasch Model was represented by item difficulty levels. Assignment of the categories was done by seeing the Standard Deviation scores. The results are explained through Tables 4 and 5.

In Table 4, it can be seen that items 6 and 7 have the highest difficulty level with a logit value of 1.25. Meanwhile, items 11 and 12 have the lowest difficulty levels with a logit value of -1.76 and -1.95 respectively.

Table 4. Item Difficulty of Performance Test of Field Try-outs

Item	Total Score	Measure (Item Difficulty)	Notes
1	230	-0,80	Moderate
2	209	-0,07	Moderate
3	217	-0,40	Moderate
4	216	-0,37	Moderate
5	186	0,48	Moderate
6	159	1,25	Difficult
7	159	1,25	Difficult
8	165	1,07	Moderate
9	196	0,20	Moderate
10	161	1,19	Difficult
11	258	-1,76	Easy
1212	263	-1,95	Easy

Table 5. Item Difficulty of *Tarkib* Mastery Test

Item	Total Score	Measure (Item Difficulty)	Notes
B1	646	0.19	Moderate
B2	879	-0.32	Moderate
B3	1231	-1.13	Easy
B4	473	0.63	Moderate
B5	1210	-1.07	Easy
B6	792	-0.13	Moderate
b7	862	-0.28	Moderate
b8	727	0.01	Moderate
B9	1005	-0.59	Moderate
B10	383	0.90	Difficult
b11	349	1.01	Difficult
B12	976	-0.53	Moderate
B13	711	0.04	Moderate
B14	1004	-0.59	Moderate
B15	452	0.69	Difficult
b16	639	0.21	Moderate
b17	361	0.97	Difficult

In Table 5, it can be seen that, for the *tarkib* mastery test of *qiraatul kutub* of the *Safinah An-Najah* book, the item difficulty levels range from -1 to 1. Item 11 has the highest difficulty level with 1.01 logit value while item 3 has the lowest difficulty level with a logit value of -1.13.

#### Information Function (TIF) and Standard Error of Measurement (SEM)

Results of the computation of the data analyses, gave the following curves for information function (TIF) and standard error of

measurement (SEM) for the performance test.

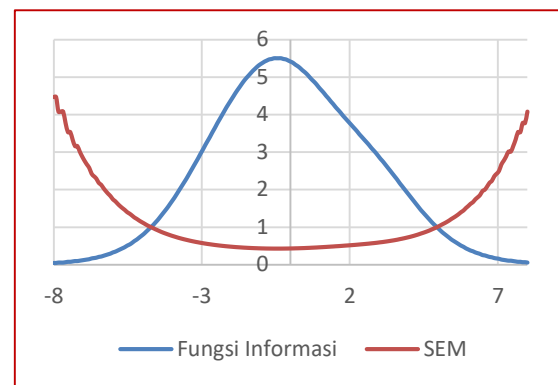


Figure 1. Information Function (TIF) and Standard Error of Measurement (SEM) (Performance Test)

The curves in Figure 1 shows the TIF and SEM of the performance test, wherein the highest is at 5.24 for TIF and 0.24 for SEM. The curves show that the test is suitable for testees with abilities between -4 and 4. The test is said to have good information function as the TIF curve has a reverse form from that of the SEM, in which the TIF highest value crosses the SEM lowest.

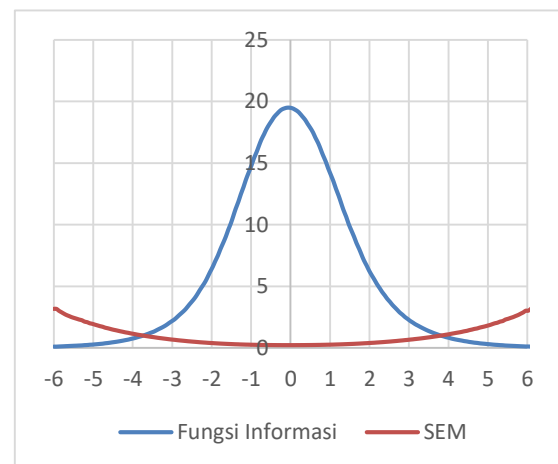


Figure 2. Information Function (TIF) and Standard Error of Measurement (SEM) (*Tarkib* Test)

For the *tarkib* test (Figure 2), the highest TIF is at 19.49 with its SEM 0.11. This means that the test is suitable for testees with moderate abilities between -3.6 and 3.8. Similarly, the test has good information function as the highest value of the TIF is also the lowest of the SEM.

Ability Profile of Safinah An-Najah Book in *Qirāatul Kutub*

Based on the information from the performance test and *tarkib* mastery test, a general description of the performances of the learners of the Darul Huda boarding school in Mayak, Ponorogo can be explained using the information found in Table 6.

Table 6. Summary of Learners' Mastery (Logit Scale)

No.	Measure (Person)	Performance Tarkib Test	
		Test	Test
1	Mean	-.02	-.23
2	S. D.	1.21	.90
3	Max.	2.31	3.12
4	Min.	-2.81	-2.66

Table 6 shows that the means of the learners' mastery are -0.02 for the performance test and -0.23 for the *tarkib* test. The highest mastery for the *performance test* is 2.31 and for the *tarkib* test 3.12. The lowest logit value for the *performance test* is at -2.81 and for the *tarkib* test -2.66. Dispersion of the learners' mastery of the performance test can be seen in Figure 3.

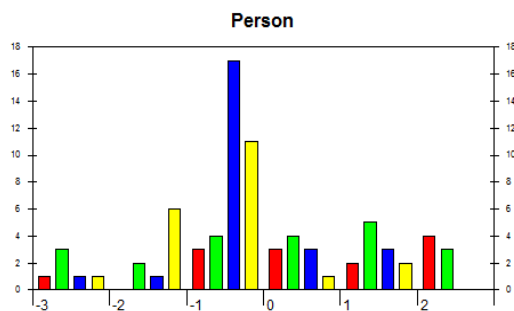


Figure 3. Dispersion of Learners' Mastery (Performance Test)

In Figure 3, it can be seen that the learners' mastery in *Qirāatul Kutub*, based on the Performance Test, stretch from -3 to 3. Most of the abilities lie between logits 0 and -1.

In Figure 4, it can be seen that the majority of the learners have *tarkib* mastery ranging from -1 to 1, or at the moderate level. Some learners show a mastery of above 2 and the remaining between -1 and -3. The

following presents the diagrams for learners' mastery for the performance test in Figure 5 and for the *tarkib* test in Figure 6.

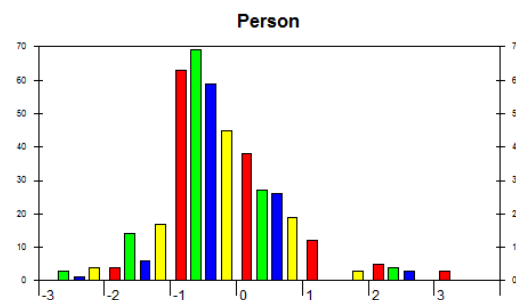


Figure 4. Dispersion of Learners' Mastery (Tarkib Test)

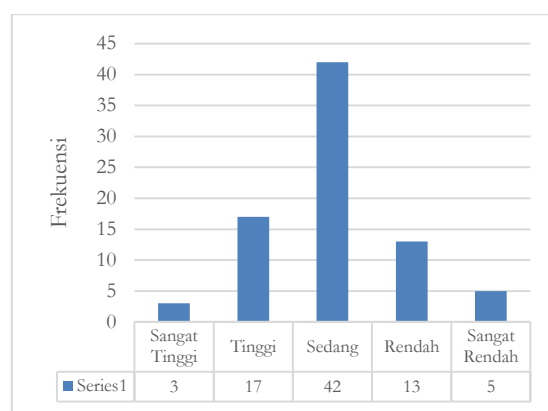


Figure 5. Diagram of Learners' Mastery (Performance Test)

Figure 5 shows that the majority of the learners (42 in number) have mastery of the moderate level. Three learners have mastery of the very high level, and 17 of the high. Then, 5 learners are noted to have a very low mastery level, and 13 others have a low mastery level.

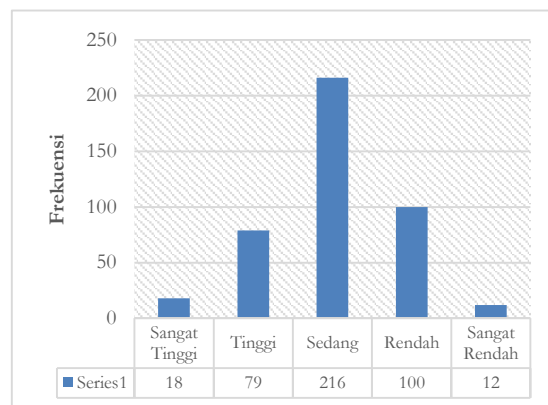


Figure 6. Diagram of Learners' Mastery (Tarkib Mastery Test)

Looking at Figure 6, it can be seen that a large number of learners (256) show a mastery level of the moderate category. The very low category is occupied by 12 learners while the low category 100 learners. On the top levels, 18 learners note the very high category and 79 note the high category. The following presents summary of the learners' mastery in converted scores for both of the tests.

#### Final Product Revision

Revision of the preliminary draft was given by the validators in the use of vocabulary words from foreign languages (transliteration), such as *mabni*, *tasrif istilahi*, *mubtada' khabar*, and others; and the spelling and writing system of such words. In interrogative sentences, a question must be ended with a question mark. In imperatives, the instruction "sebutkan" (=mention) must be replaced with "tuliskan" (=write down). Another revision was on the answer-key items in the scoring guides. In terms of clarity, items 1 and 2 should be merged in order to avoid ambiguities on the part of the respondents.

Table 7. Revision on the Items of the Essay-type Test

Item before Revision	Item after Revision
شهادة انلا الهالا اللهو أنمحمدر سوالله وإقام الصلاة وإيتاء الزكاة صوم رمضان وحج البيت من استطاع عليه سبيلا	شهادة انلا الهالا اللهو أنمحمدر سو لالله وإقام الصلاة وإيتاء الزكاة وصوم رمضان وحج البيت من استطاع عليه سبيلا
Item 1: Mention the words that are <i>mabni</i> in the paragraph above!	Item 1 : Write down the words that are <i>mabni</i> in the paragraph above, and give your reasons!
Item 2 : Based on item number 1, explain the state of the <i>mabni</i> words you have found in the paragraph above!	

After being revised, and the revision agreed by the experts, the *Qirāatul Kutub* evaluation instrument was then tried out in

the limited forum and then the extended field forum. Following the results of the try-outs, it was found out which items were to be revised and which to be deleted. Based on the results of the data analyses, it was decided that item numbers 11 and 12 could be revised or removed.

Table 8. Revision on Performance Test

Item no	Item description
11	Politeness in behaviour ( <i>Adab</i> )
12	Politeness in dressing

Field wise, in which *Qirāatul Kutub* learning was done after the sunset (*maghrib*) prayers, most of the girls were still wearing the *mukena* (attire covering whole body); meanwhile, the learners were sitting leg-crossed in front of the and the teacher was also sitting leg-crossed. In this position, it was hard to conduct evaluation on politeness in both behavior and dressing. As a result, these two items had the lowest level of difficulty and were far different from the other items. In the Goodness of Fit analysis, these two items satisfied only one of the three criteria. This could be understood that these two items were not applicable for a *pesantren* learning condition. However, for possibly a different condition of seating, the items can still be used.

The final product of the *Qirāatul Kutub* evaluation instrument, there are 12 items for the performance test and 17 for the essay-type.

#### Final Product

The final product of this developmental research is an evaluation instrument for measuring *tarkib* mastery and a test instrument for measuring performance in *qirāatul kutub*. The two instruments have undergone two try-outs, first in a limited scope and second in the extended field scope. Beforehand, the drafts of the instruments have also undergone validation processes with experts. The research findings have given indication that the two instruments were valid and reliable and that they have contained items with good quality character-

istics. The following presents the validity and reliability of the instruments and the good quality of the item characteristics.

### Validity

The instruments are valid since they have fulfilled the criteria for content validity as they have been examined by two experts from the Yogyakarta State University (YSU). In addition, the instruments have been empirically shown to be capable of measuring what they are supposed to measure.

### Reliability

The reliability co-efficients of the Cronbach's alpha have been found to have achieved the requirements, 0.85 for the performance test and 0.90 for the tarkib mastery test. The highest information function is at TIF 5.24 and SEM 0.24 for the performance test, suitable for testees with an ability range from -4 to 4. Meanwhile, for the *tarkib* mastery test, the highest value is at TIF 19.49 and SEM 0.11, suitable for testees with an ability range from -4 to 4.

### Item Characteristics

Based on the results of the analyses of item characteristics, the two instruments of *Qirāatul Kutub* have a difficulty of the moderate level with a mean of difficulty levels of 0.00. Summary of the levels of the difficulty measures of the two instruments can be seen in Table 9.

Table 9. Summary of Levels of Difficulty

	Performance Test	Test <i>Tarkib</i>
Mean	.00	.00
S. D.	.84	.65
Max.	.99	1.01
Min.	-1.32	-1.13

Table 9 shows the following. For the performance test, the highest difficulty level is 0.99 and the lowest -1.32. Meanwhile, for the *tarkib* test, the highest is 1.01 and the lowest -1.13. The distribution of the difficulty levels of the two instruments ranges from -1 to 1. This shows that the two instruments

have a difficulty level of the moderate category.

### Conclusion

From the results of the research and development study, the following items of conclusion can be drawn. A number of items can be drawn as conclusion of the study. First, the learning process of *Qirāatul Kutub* runs after the sunset (*maghrib*) prayers, each group consisting of 10 to 20 learners taught by one female teacher. The learning steps run as follows: (1) the teacher opens the session by taking attendance and reciting prayers; (2) one by one, the learners come up to the front of the teacher to submit her reading; (3) the teacher listens to the reading, and asks questions on the *tarkib*, *murad*, or understanding of what has been read; (4) while one learner sits in front of the teacher, the next learners prepare or recite what she is going to submit to the teacher.

The *Qirāatul Kutub* program in the Darul Huda boarding school in Mayak, Ponorogo uses two evaluation formats: a) classroom evaluation conducted by the class teacher, running intuitively without many formal rules and regulations; and b) end-of-year evaluation for class promotion, from which learners' mastery of the *Qirāatul Kutub* is known.

Second, by the Rasch model, quality of the instruments, performance test and mastery test, of the *Safinah An-Najah* is high. The two instruments have passed the validity tests and have empirically been stated as valid. The reliability level of the performance test is marked by a high category of 0.85 and the mastery test 0.90 on the Cronbach's *alpha* co-efficient.. Item and person reliability measures are also found as high. The highest information function TIF is at the 5.24 point with a standard error of measurement SEM 0.24 for the performance test, suitable for learners with ability levels between -4 and 4. Meanwhile, for the mastery test, the highest TIF is at 19.49 with a SEM of 0.11, suitable for learners with ability levels between -4 and 4.

The highest level of difficulty for the performance test is 0.99 and the lowest -1.32. For the *tarkib mastery test*, the highest difficulty level is at the logit value 1.01 and the lowest level -1.13.

Third, the average measure for the learners' mastery is -0.02 for the performance test and -0.23 for the *tarkib* test. The highest learners' performance is at the logit 2.31 and 3.12 for the *tarkib* test. The lowest logit is at the point -2.81 for the performance test and -2.66 for the *tarkib* test. Learners' mastery of the *Qirāatul Kutub* lies between -3 and 3 on the performance instrument. The majority of the learners show mastery between logits 0 and -1. For the *tarkib* instrument, learners' mastery on the *tarkib* ranges between -1 and 1. A number of learners show a mastery above 2 and the remaining between -1 and -3.

Results of this study is expected to inspire various parties to initiate efforts to develop an instrument applicable for *Qirāatul Kutub*. For the time being, it is recommendable for teachers and other practitioners of *Qirāatul Kutub* to use the product of this study, that has gone through careful empirical examinations, for the assessment purposes of *Qirāatul Kutub* and *tarkib* mastery.

## References

- Al-Hasyimi, A. (1971). *Qowaidh al-asaasiyah li al-lughab al-'arabiyah*. Beirut: Dar Al-Kotob Al-Ilmiyah.
- Arifin, Z. (2012). Perkembangan pesantren di Indonesia. *Pendidikan Agama Islam*, IX(1), 40–53.
- Borg, W. R., & Gall, M. D. (1989). *Educational research: an introduction* (4th ed.). New York: Longman.
- Departemen Agama Republik Indonesia. (2003). *Pondok pesantren dan madrasah diniyah: pertumbuhan dan perkembangannya*. Jakarta: Direktorat Jenderal Kelembagaan Agama Islam.
- Fisher, W. P. (2007). Rating scale instrument quality criteria. Retrieved September 20, 2017, from <https://www.rasch.org/rmt/rmt211m.htm>
- Iskandarwassid, I., & Sunendar, D. (2015). *Strategi pembelajaran bahasa*. Bandung: Rosda Karya.
- Linacre, J. M. (2011). *User's guide to Winsteps Ministeps Rasch-Model computer programs*. Retrieved from <https://www.winsteps.com/winman/copyright.htm>
- Mardapi, D., & Kartowagiran, B. (2011). Pengembangan instrumen pengukur hasil belajar nirbias dan terskala baku. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 15(2), 326–341. Retrieved from <https://journal.uny.ac.id/index.php/jpep/article/view/1100>
- Nitko, A. J., & Brookhart, S. M. (2007). *Educational assesment of students*. New Jersey: Pearson Education.
- Nurgiyantoro, B. (2001). *Penilaian dalam pengajaran bahasa dan sastra Indonesia*. Yogyakarta: BPFE.
- Pada, A. U. T., Kartowagiran, B., & Subali, B. (2016). Separation index and fit items of creative thinking skills assessment. *Research and Evaluation in Education*, 2(1), 1. <https://doi.org/10.21831/reid.v2i1.8260>
- Retnawati, H. (2016). *Validitas reliabilitas dan karakteristik butir*. Yogyakarta: Parama Publishing.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan RASCH pada assesment pendidikan*. Cimahi: Trim Komunikata.
- Veenendaal, N. J., Groen, M. A., & Verhoeven, L. (2015). What oral text reading fluency can reveal about reading comprehension. *Journal of Research in Reading*, 38(3), 213–225. <https://doi.org/10.1111/1467-9817.12024>
- Wagiran. (2015). *Metodologi penelitian pendidikan*. Yogyakarta: Budi Utama.



Wekke, I. S. (2015). Antara tradisionalisme dan kemodernan: Pembelajaran bahasa Arab madrasah minoritas muslim Papua Barat. *Tsaqafah*, 11(2), 313–332.

Wheeler, R., Cartwright, K. B., & Swords, R. (2012). Factoring AAVE into

reading assessment and instruction. *Reading Teacher*, 65(6), 416–425.  
<https://doi.org/10.1002/TRTR.01063>

## VALIDITAS DAN RELIABILITAS INSTRUMEN ASESMEN KINERJA LITERASI SAINS PELAJARAN FISIKA BERBASIS STEM

*Khoirul Bashooir<sup>1\*</sup>, Supabar<sup>2</sup>*

<sup>1</sup>SMPIT Ihsanul Fikri, <sup>2</sup>Universitas Negeri Yogyakarta

<sup>1</sup>Jl. Pabelan 1, Pabelan Satu, Pabelan, Mungkid, Magelang, Jawa Tengah 56511, Indonesia

<sup>2</sup>Jl. Colombo No. 1, Depok, Sleman 55281, Yogyakarta, Indonesia

\* Corresponding Author. Email: [khoirul.bashooir@gmail.com](mailto:khoirul.bashooir@gmail.com)

### Abstrak

Penelitian ini merupakan bagian dari penelitian pengembangan asesmen kinerja literasi sains berbasis *STEM* pada pembelajaran fisika. Tujuan dari penelitian ini adalah untuk mengungkapkan validitas isi, validitas empiris, dan reliabilitas instrumen asesmen kinerja literasi sains berbasis *STEM* yang sebelumnya telah disusun. Instrumen yang dikembangkan berupa lembar pengamatan dan tes pilihan ganda. Analisis validitas isi dari lembar pengamatan menggunakan Koefisien V oleh Aiken sedangkan validitas isi instrumen tes dianalisis dengan menggunakan *CVI (Content Validity Index)* oleh Lawshe. Validitas empiris reliabilitas instrumen tes diestimasi dengan *IRT (Item Response Theory)*. Reliabilitas lembar pengamatan ditentukan dengan *ICC (Item Correlation Coefficient)*. Hasil dari penelitian ini menunjukkan bahwa (1) Lembar pengamatan berupa rubrik penskoran dan penilaian diri terbukti valid dengan koefisien V Aiken 0,75 dan reliabel dengan koefisien Reliabilitas Alfa > 0,8 dan ICC yang *Excellent*. (2) Instrumen tes terbukti reliabel untuk digunakan pada peserta didik dengan kategori sedang sampai dengan tinggi (-0,7 sampai dengan 6,7) dengan *CVI=1* dan *INFIT MNSQ* sesuai model Rasch. Berdasarkan hasil penelitian tersebut maka asesmen kinerja Literasi Sains berbasis *STEM* layak digunakan.

**Kata kunci:** *validitas isi, validitas empiris, asesmen kinerja, literasi sains, STEM*

## VALIDITY AND RELIABILITY INSTRUMENT OF SCIENTIFIC LITERACY PERFORMANCE ASSESSMENT IN PHYSICS TEACHING BASED ON STEM

### Abstract

This research is part of the development of scientific literacy performance assessment based on STEM in teaching physics. The aim of this research is to reveal the validity (content and also empiric) and reliability of scientific literacy performance assessment instrument based on STEM. The kind of instruments were developed are observational sheet and multiple choice test. The content validity of observational sheet was revealed by used the Aiken's V Coefficient. The content validity of multiple choice tests was revealed by used Content Validity Index (CVI) which proposed by Lawshe. The empirical validity and reliability of multiple choice tests was revealed by used Item Response Theory Analysis. The reliability of observational sheet was revealed by used ICC (Item Correlation Coefficient) Analysis. The results of this study are the validity from the contents and empirical trials from the developed instruments. The observation sheet from scoring rubric and self-assessment has been valid with Aiken's V value that exceeds the standard of 0,75. The reliability of the scoring rubric has Alfa Reliability > 0.8 and Excellent of ICC. Validity values from The written test is shown with CVI of 1 and the MNSQ INFIT value which match to the Rasch model. Based on the TIC and SEM graphs, the written test is stated to be reliable for use in students with moderate to high categories (-0.7 to 6.7). STEM-based Science Literacy performance assessment with caloric material is appropriate to use.

**Keywords:** *content validity, empirical validity, performance assessment, scientific literacy, STEM*

Permalink/DOI: <http://dx.doi.org/10.21831/jpep.v22i2.19590>

## Pendahuluan

Mencerdaskan kehidupan bangsa adalah salah satu tujuan dari Pemerintah Negara Indonesia. Tujuan tersebut termaktub dalam pembukaan Undang-Undang Dasar 1945 dan menjadi salah satu tujuan dari pendidikan nasional negara Indonesia yang diatur dalam undang-undang Republik Indonesia Nomor 20 Tahun 2003 (Presiden Republik Indonesia, 2003). Oleh karena itu, pendidikan yang diterapkan di Indonesia haruslah bertujuan untuk kecerdasan warga negara Indonesia dalam hal ini adalah peserta didik agar mampu bersaing dengan negara-negara yang lain.

Pendidikan yang diselenggarakan melalui proses pembelajaran di sekolah menengah diharapkan dapat menghasilkan lulusan yang mampu mengaplikasikan ilmu pengetahuan yang dimiliki untuk pemecahan masalah dalam kehidupan sehari-hari (Badan Standar Nasional Pendidikan, 2006, p. 443). Berdasarkan acuan tersebut, kegiatan pembelajaran dapat didesain dengan memberikan tugas-tugas sebagai bentuk latihan untuk menyelesaikan permasalahan. Hal ini selaras dengan konsep literasi sains yang dikemukakan oleh *PISA (Programme for International Student Assessment)*. Konteks literasi sains dalam penelitian ini dikhususkan hanya pada pembelajaran fisika.

Sains sebagai ilmu pengetahuan memiliki empat dimensi (Chiappetta & Koballa, 2010, p. 102) yaitu (1) *science as a way of thinking*, (2) *science as a way of investigating*, (3) *science as a body of knowledge*, dan (4) *science and its interactions with technology and society*. Keempat dimensi tersebut diperlukan dalam kegiatan pembelajaran sains.

Pembelajaran sains yang dilaksanakan harus dapat menjawab tantangan pada abad ke-21 yaitu untuk membentuk manusia yang dapat bertahan dengan perkembangan teknologi. Wagner (2008, pp. 14–38) merumuskan beberapa keterampilan yang dibutuhkan oleh manusia agar dapat bertahan hidup pada abad ke-21. Beberapa keterampilan tersebut adalah: (1) keterampilan berpikir kritis dan menyelesaikan masalah; (2) berkolaborasi dalam komunitas; (3) ketangkasan dan kemam-

puan beradaptasi; (4) berinisiasi dan jiwa bisnis; (5) kemampuan komunikasi secara lisan dan tulis; (6) mengakses dan menganalisis informasi; dan (7) imajinasi dan keingintahuan. Selain itu, literasi sains juga merupakan keterampilan yang penting pada abad ke-21 (P21, 2009, p. 2). Hal tersebut sejalan dengan hasil penelitian Yore & Treagust (2006, p. 304) yang menunjukkan bahwa literasi sains dibutuhkan oleh individu pada abad 21 untuk dapat hidup dan memecahkan permasalahannya terkait dengan fenomena alam.

*PISA* merupakan program asesmen yang diadakan oleh *OECD (Organization for Economic Co-operation and Development)* untuk menilai kemampuan membaca, literasi matematika dan literasi sains peserta yang mengikuti tes tersebut. Pada tahun 2012 (OECD, 2014, p. 19) tes yang dilakukan oleh *PISA* diikuti oleh 510.000 peserta didik dari 65 negara. Indonesia berada pada ranking 64 dari 65 negara. Data tersebut menunjukkan bahwa kemampuan membaca, literasi matematika dan literasi sains peserta didik di Indonesia masih sangat rendah bila dibandingkan dengan 64 negara lain yang mengikuti program tersebut. Peringkat yang diperoleh Indonesia dalam *PISA* mengindikasikan bahwa peserta didik Indonesia memiliki literasi sains yang rendah. Selain itu, peringkat tersebut juga mengindikasikan bahwa Indonesia belum dapat bersaing dengan negara-negara lain.

Ilmu fisika merupakan salah satu konteks dalam sains yang diukur dalam *PISA*. Fisika juga menjadi salah satu ilmu pengetahuan yang mendasari perkembangan teknologi. Dalam fisika terdapat integrasi antara konsep fisika, konsep matematika, cara merangkai peralatan sehingga menghasilkan teknologi yang inovatif.

Berdasarkan observasi yang dilakukan di empat sekolah di kabupaten Magelang yang memiliki peringkat tinggi, sedang dan rendah menurut hasil ujian nasional diketahui bahwa (1) kegiatan pembelajaran masih berpusat kepada guru (*Teacher Oriented*), (2) metode pembelajaran yang digunakan adalah ceramah, dan (3) asesmen yang digunakan

oleh guru untuk mengukur kemampuan peserta didik adalah asesmen tes tertulis. Lebih lanjut diketahui pula bahwa asesmen tes tulis yang digunakan tidak disusun berdasarkan konsep pemecahan masalah dalam kehidupan sehari-hari. Hal tersebut belum mencerminkan tujuan pendidikan yang ditentukan oleh BSNP yang menuntut peserta didik untuk dapat mengaplikasikan pengetahuan yang dimilikinya untuk memecahkan masalah dalam kehidupan sehari-hari.

*STEM (Science Technology Engineering dan Mathematic)* adalah salah satu alternatif solusi bagi pembelajaran abad 21. Pendekatan *STEM* merupakan pembelajaran yang mengintegrasikan *Science Technology Engineering* dan *Mathematic* dalam pembelajaran *Engineering* (Department of Education, 2009, p. 19; Breiner, Harkness, Johnson, & Koehler, 2012, p. 3; Gonzalez & Kuenzi, 2012, p. 1; Reeve, 2013, p. 3; Hernandez et al., 2014, p. 108). *STEM* memenuhi keempat dimensi dalam pembelajaran Sains (Chiappetta & Koballa, 2010, p. 102; Departemen Pendidikan dan Kebudayaan, 2001, p. 1473). *Science* sebagai kerangka utama (*Science as a way of thinking* dan *science as a body of knowledge*). *Science* sebagai *way of investigating*, *Science* dikombinasikan dengan *Engineering* dan *Mathematics*. *Engineering* atau keterampilan teknik digunakan dalam merancang dan melaksanakan percobaan. *Mathematics* digunakan dalam menganalisis data percobaan. *Technology* digunakan untuk membantu dalam menganalisis data serta mengomunikasikan hasil percobaan.

Penggunaan pendekatan *STEM* dimaksudkan agar peserta didik dapat memiliki kemampuan dan pemahaman dalam keempat aspek *STEM* yang saling terkait pada satu pokok bahasan. Peserta didik juga diharapkan dapat mengaplikasikan hal-hal yang dipelajarinya dalam dunia nyata atau lingkungan kerja.

Pengembangan keterampilan literasi sains peserta didik dapat dilakukan melalui pembelajaran *STEM* (Ismail, Permanasari, & Setiawan, 2016, p. 10). Pengembangan literasi sains dan *STEM* peserta didik dapat melalui pendidikan formal dan informal. Pendi-

dikan formal yang dimaksud adalah pendidikan yang dilaksanakan oleh sekolah-sekolah pada umumnya. Pendidikan informal adalah pendidikan yang didapatkan oleh individu karena interaksinya dalam masyarakat di sekitarnya. Penelitian ini membatasi literasi sains dan *STEM* ada pendidikan formal yang dilaksanakan di sekolah, khususnya sekolah menengah pertama.

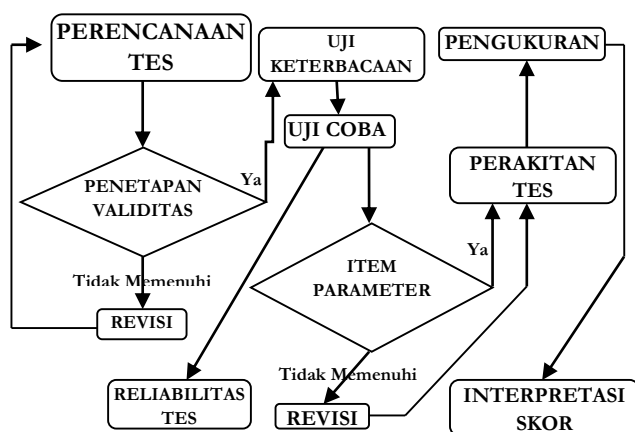
Selain dengan memperbaiki proses pembelajaran yang dilaksanakan, peningkatan kualitas pembelajaran dapat dilakukan melalui perbaikan pada asesmen yang digunakan untuk mengukur hasil belajar peserta didik. Perbaikan dalam aspek penilaian dapat dilakukan dengan memberikan panduan dan pelatihan dalam melakukan penilaian (Kartowagiran & Jaedun, 2016, p. 140). Asesmen yang digunakan untuk memperoleh, menyintesis dan menginterpretasikan informasi yang diperoleh dari peserta didik dengan tujuan untuk menarik keputusan tentang peserta didik di dalam kelas dikenal dengan istilah *classroom assessment*. *Classroom assessment* yang digunakan untuk menilai kinerja peserta didik selama pembelajaran dikenal sebagai *assessment for learning*. Dalam pembelajaran, *assessment for learning* menyatu dengan rencana pelaksanaan pembelajaran yang disusun atau digunakan oleh guru.

Salah satu bentuk *assessment for learning* adalah asesmen kinerja. Asesmen kinerja digunakan untuk mengases kemampuan peserta didik dalam menerjemahkan pemahamannya menjadi suatu kerja nyata, sehingga asesmen kinerja dapat menunjukkan kemampuan peserta didik yang sesungguhnya.

Instrumen yang digunakan dalam asesmen memiliki beberapa syarat agar layak digunakan yaitu valid dan reliabel (Mardapi, 2012, pp. 37–96; Sumintono & Widhiarso, 2015, pp. 7–12). Oleh karena itu, instrumen kinerja berbasis *STEM* yang telah dikembangkan sebelumnya harus memenuhi kedua syarat tersebut. Oleh karena itu, uji validitas dan reliabilitas instrumen asesmen kinerja literasi sains perlu dilakukan.

## Metode Penelitian

Penelitian ini merupakan bagian dari penelitian pengembangan asesmen kinerja literasi sains berbasis *STEM* pada pembelajaran fisika. Instrumen yang telah memenuhi standar kevalidan dan standar kereliatan dapat digunakan untuk tahap pengukuran. Tahapan penelitian secara keseluruhan yang disajikan pada Gambar 1 mengacu pada pembuatan instrumen tes yang dibuat oleh Supahar & Prasetyo (2015, p. 99).



Gambar 1. Skema Prosedur Penelitian Pengembangan Asesmen Kinerja Berbasis *STEM*

Instrumen asesmen dikembangkan berdasarkan aspek dan indikator literasi sains. Penelitian tersebut memperoleh 11 aspek literasi sains (Bashooir & Supahar, 2016, p. 93), yaitu: (1) merumuskan masalah, (2) menyusun variabel fisika, (3) menyusun kajian teori, (4) mendesain eksperimen, (5) mengumpulkan data, (6) menganalisis data, (7) menyusun tabel, (8) menyusun grafik, (9) menyimpulkan, (10) menyusun solusi dan (11) mengomunikasikan solusi. Dalam penelitian ini, instrumen dikembangkan dengan menggunakan 9 aspek yang mengacu pada penelitian tersebut. Ke-9 aspek tersebut adalah 1, 2, 3, 4, 6, 7, 8, 9 dan 10.

Data yang diperlukan untuk penentuan validitas dan reliabilitas instrumen diperoleh melalui beberapa cara. Validitas isi instrumen diperoleh dengan memberikan angket kepada pakar, yaitu pakar asesmen, pakar pembelajaran, pakar ilmu fisika dan pakar pembelajaran fisika. Selanjutnya, uji keter-

bacaan instrumen juga dilakukan dengan memberikan instrumen kepada 6 peserta didik. Hasil dari uji keterbacaan digunakan untuk memperbaiki instrumen.

Validitas empiris dan reliabilitas diperoleh instrumen yang dikembangkan diperoleh melalui uji coba instrumen. Instrumen tes tertulis bentuk pilihan ganda diujicobakan kepada 350 peserta didik di empat sekolah di Kabupaten Magelang. Keempat sekolah tersebut mewakili sekolah dengan peringkat tinggi, sedang dan rendah ditinjau dari hasil ujian nasional di Kabupaten Magelang. Keempat sekolah tersebut adalah SMAIT Ihsanul Fikri, SMA N 1 Muntilan, SMA N 1 Salaman dan SMA Muhammadiyah 1 Muntilan. Selanjutnya, Lembar pengamatan diujicobakan kepada 57 peserta didik yang berasal dari SMAIT Ihsanul Fikri.

### Validitas Isi

Validitas isi suatu butir dapat dibuktikan dengan menggunakan *CVR* dan *CVI* atau koefisien *V* Aiken. *CVR* dan *CVI* diusulkan oleh Lawshe pada tahun 1975 dengan menggunakan 3 skala rating (Lawshe, 1975, p. 567). Pada tahun 1985, Aiken mengusulkan konsep validitas isi dengan lebih rinci. Kerincian tersebut terlihat dari standar kevalidan yang dipengaruhi oleh jumlah rater dan skala rating yang digunakan (Aiken, 1985, p. 133). Penggunaan cara pembuktian harus memperhitungkan asumsi atau syarat yang harus dipenuhi, sedangkan, Pembuktian validitas dengan menggunakan konsep dari Lawshe (1975, p. 567) hanya menggunakan tiga skala rating, yaitu (1) esensial, (2) berguna tetapi tidak esensial, dan (3) tidak perlu. Pembuktian validitas isi oleh Lawshe (Lawshe, 1975, p. 567; Azwar, 2012, p. 114) dapat ditentukan dengan formula:

$$CVR = \frac{2ne}{n} - 1 \quad (1)$$

Keterangan:

*ne* = banyaknya *SME* (*Subject Matter Experts*) yang menilai suatu butir 'esensial'

*n* = banyaknya *SME* yang melakukan penilaian

Nilai *CVR* memiliki rentang antara -1 sampai dengan 1. Jika setengah dari *SME* menyatakan esensial maka nilai *CVR* akan bernilai 0. *CVR* akan bernilai 1 jika seluruh *SME* menyatakan esensial untuk suatu butir. Acuan standar kevalidan item menurut Lawshe (1975, p. 568) ditunjukkan pada Tabel 1.

Tabel 1. Standar Minimum *CVR* berdasarkan *SME*

Jumlah <i>SME</i>	Nilai Minimal <i>CVR</i>
5	0,99
6	0,99
7	0,99
8	0,75
9	0,78
10	0,62

Butir yang memperoleh nilai *CVR* negatif merupakan butir yang tidak valid dan harus dieliminasi. Nilai validitas tes secara keseluruhan dapat ditentukan dengan menggunakan *CVI* (*Content Validity Index*). Penentuan *CVI* (Lawshe, 1975, p. 568; Azwar, 2012, p. 115) dilakukan dengan menggunakan formula:

$$CVI = \frac{\sum CVR}{k} \quad (2)$$

Keterangan:

*CVR* = *Content Validity Ratio* dari tiap item  
*k* = banyaknya butir soal

Pada pembuktian validitas isi (Azwar, 2012, p. 113), peneliti dapat menentukan banyaknya kategori rating yang dikehendaki. Banyaknya kategori rating mempengaruhi standar validitas isi yang ditentukan oleh Aiken. Banyaknya kategori rating terkecil yang dirumuskan oleh Aiken adalah 2 dan terbanyak adalah 7 (Aiken, 1985, p. 134). Penelitian ini menggunakan 5 kategori rating dan 7 Rater. Berdasarkan standar yang telah ditetapkan Aiken (Aiken, 1985, p. 134), standar minimal *V* Aiken untuk penelitian ini adalah 0,75 dengan probabilitas 0,041.

Koefisien validitas Aiken didapatkan perhitungan dengan skor mentah dari ahli yang berjumlah *n*. Nilai koefisien *V* Aiken

memiliki rentang antara -1 sampai dengan 1 (Supahar, 2015, p. 66). Validitas isi dengan menggunakan koefisien *V* Aiken (Azwar, 2012, p. 113; Aiken, 1985, p. 133) diperoleh dengan menerapkan formula:

$$V = \frac{\sum(r_i - l_o)}{[n(c-1)]} \quad (3)$$

Keterangan:

*r* = angka yang diberikan oleh penilai  
*l<sub>o</sub>* = angka penilaian validitas terendah  
*c* = angka penilaian validitas yang tertinggi  
*n* = banyaknya ahli & praktisi yang melakukan penilaian  
*i* = bilangan bulan dari 1,2,3 sampai ke *n*  
*n* = banyaknya penilai

Analisis validitas isi dengan menggunakan koefisien Aiken *V* digunakan untuk menguji validitas instrumen lembar pengamatan, sengan validitas isi yang diperkenalkan oleh Lawshe digunakan untuk menguji validitas isi instrumen tes.

#### Validitas Empiris

Validitas empiris diperoleh dari analisis respon terhadap tes yang diberikan kepada responden. Respon didapatkan dari uji coba tes kepada para responden atau testi. Validitas empiris (Sumintono & Widhiarso, 2015, pp. 12-47; Retnawati, 2016, pp. 113–146) dapat ditentukan dengan menggunakan *Classical Test Theory (CTT)* atau *Item Response Theory (IRT)*.

*Rasch model* merupakan bagian dari *IRT* yang dapat dilakukan dengan bantuan program *QUEST*. Butir soal dinyatakan valid apabila nilai *INFITMNSQ* berada pada rentang 0,77 sampai dengan 1,30 (Subali & Suyata, 2012, p. 61). Butir-butir yang berada dalam rentang tersebut dapat digunakan untuk keperluan pengukuran keterampilan literasi sains peserta didik.

#### Koefisien Alfa

Salah satu cara untuk menentukan tingkat reliabilitas instrumen adalah menggunakan koefisien Alfa. Azwar (2015, p. 67) menjelaskan bahwa koefisien alfa dapat digunakan selama masing-masing belahan sama

panjang atau berisi item yang sama banyaknya. Koefisien alfa dapat ditentukan dengan menggunakan persamaan (Azwar, 2015, p. 45):

$$\rho_{XX'} \geq \alpha = \frac{\left[\frac{k}{k-1}\right](\sigma_x^2 - \sum \sigma_{Y_i}^2)}{\sigma_x^2} \quad (4)$$

Keterangan:

$\rho_{XX'}$  = koefisien reliabilitas

$\alpha$  = koefisien alfa

$k$  = banyaknya butir dalam tes

$\sigma_x^2$  = varian skor tes

$\sigma_{Y_i}^2$  = varian skor belahan tes dengan  $i = 1, 2, 3, \text{ dst.}$  (belahan tes)

Dalam penelitian ini estimasi reliabilitas instrumen dengan menggunakan koefisien alfa digunakan untuk mengestimasi reliabilitas instrumen tes. Reliabilitas alfa memiliki rentang nilai antara 0 sampai dengan 1. Berdasarkan (Streiner, 2003, p. 103) sebuah instrumen pada tahap penelitian pendahuluan dinyatakan reliabel apabila nilai reliabilitas alfa 0,7; pada penelitian dasar sebesar 0,8 dan pada penelitian dalam bidang kedokteran dengan tujuan klinis sebesar 0,95.

#### Interclass Correlation Coefficient (ICC)

ICC (*Interclass Correlation Coefficient*) merupakan reliabilitas yang digunakan untuk hasil rating dari pengamatan beberapa rater (Shrout & Fleiss, 1979). Hal yang sama juga disampaikan oleh Mardapi (2012). ICC dapat diestimasi dengan menggunakan persamaan (Mardapi, 2012, p. 89).

$$\rho = \frac{MSrs - MSe}{MSr + (k-1)MSe} \quad (5)$$

Keterangan:

$MSrs$  = rerata kuadrat antar penilai

$MSe$  = varians skor kesalahan

$k$  = jumlah penilai

Analisis ICC pada penelitian ini menggunakan bantuan SPSS 16. Dalam penelitian ini, reliabilitas ICC digunakan untuk mengestimasi reliabilitas instrumen lembar pengamatan.

#### Reliabilitas Skala Logit: Item Response Theory (IRT)

Sebagai pendukung, pengujian reliabilitas tiap butir soal dapat ditentukan dengan menggunakan ICC dan reliabilitas perangkat tes dapat diperoleh dari TIC. Grafik ICC dan TIC didapatkan dengan pengolahan data melalui program BILLOG MG (Mathilda du Toit, 2003, pp. 505-508). Melalui grafik ICC dengan 3 parameter dapat diketahui tingkat kehandalan soal. Grafik ICC dan TIC memiliki rentang skala logit antara -3 sampai dengan 3. Sumintono & Widhiarso (2015, p. 39) menjelaskan bahwa kala logit merupakan skala yang menggambarkan abilitas atau kemampuan peserta didik. Skala minus menunjukkan abilitas yang rendah dan semakin positif menunjukkan abilitas yang tinggi.

#### Hasil Penelitian dan Pembahasan

##### Lembar Pengamatan: Rubrik Penskoran dan Penilaian Diri

Skor yang diperoleh untuk perangkat asesmen dalam bentuk rubrik penskoran dianalisis dengan menggunakan analisis validitas V Aiken. Hasil analisis dapat dikategorikan valid apabila memenuhi batas koefisien V Aiken. Syarat batas koefisien V Aiken untuk 5 skala rating dan 7 rater adalah 0,75 dengan probabilitas 0,41. Tiap aspek penilaian dalam rubrik penskoran aspek psikomotorik, LKPD (lembar Kerja Peserta Didik), laporan dan penilaian diri mendapatkan skor Aiken di atas 0,75. Oleh karena itu, ketiga rubrik penskoran tersebut dinyatakan valid sehingga layak digunakan dalam eksperimen. Skor rata-rata V Aiken disajikan dalam Tabel 2.

Tabel 2. Skor Rata-rata V Aiken untuk Rubrik Penskoran dan Penilaian Diri

Perangkat Asesmen	Rata-rata V Aiken
Penilaian Diri	0,90
Rubrik Penskoran Aspek Psikomotorik	0,88
Rubrik Penskoran LKPD	0,88
Rubrik Penskoran Laporan	0,89

### Perangkat Asesmen: Tes Tertulis

Hasil dalam validasi isi untuk perangkat asesmen tes tertulis dianalisis dengan menggunakan Validitas isi Lawshe dimana standar kevalidan *CVR* tergantung pada jumlah *SME*. Nilai *CVR* harus memenuhi 0,99 agar butir dapat dinyatakan valid. Hal tersebut berlaku untuk validasi isi dengan menggunakan 7 *SME* (Lawshe, 1975, p. 568). Nilai *CVR* yang diperoleh dari tiap butir adalah 1 dan secara lengkap disajikan dalam lampiran. Nilai *CVI* yang diperoleh dari rata-rata *CVR* sebesar 1. Berdasarkan nilai *CVR* yang melebihi 0,99 maka seluruh butir dinyatakan valid (Lawshe, 1975, p. 568) dan layak digunakan untuk penelitian lebih lanjut.

### Uji Coba Tes Tulis

Butir soal yang telah terbukti validitasnya kemudian disusun ke dalam 2 paket soal (Paket A dan Paket B). Setiap paket terdiri dari 34 butir soal dengan 7 butir yang digunakan sebagai angkor, sehingga terdapat 7 butir soal yang termasuk dalam Paket A dan Paket B. Uji coba keterbacaan soal dilakukan dengan melibatkan 6 peserta didik. Hasil uji coba keterbacaan butir soal digunakan sebagai bahan untuk memperbaiki instrumen tes tulis sebelum diujicobakan pada responden dengan jumlah lebih besar.

Uji coba ke dua dilakukan terhadap 350 peserta didik yang berada di Kabupaten Magelang. Respon peserta didik dianalisis dengan menggunakan *IRT* model *Rasch* berbantuan program *QUEST* dan *BILOG MG*. Berikut ini adalah hasil analisis yang diperoleh:

### Kecocokan Butir Instrumen Tes Tulis dengan Model *Rasch*

Butir dinyatakan valid apabila memiliki nilai *INFIT MNSQ* dalam rentang 0,77 sampai dengan 1,30 (Subali & Suyata, 2012, p. 61; Supahar, 2014, p. 141). Analisis yang dilakukan dengan menggunakan bantuan program *QUEST* menunjukkan bahwa ke 61 butir memiliki *INFIT MNSQ* yang berada dalam rentang 0,77 sampai dengan 1,30. Oleh karena itu, seluruh butir dinyatakan valid. Ilustrasi secara lengkap kevalidan butir soal disajikan dalam Gambar 2.

### Hasil Estimasi

Hasil estimasi 61 butir tes yang dikerjakan oleh 350 responden dengan tingkat peluang 0,5 dan menggunakan penskoran dikotomus berdasarkan Model *Rasch* dinyatakan dalam Tabel 3. Dalam Tabel 3 dapat diketahui estimasi untuk butir dan estimasi untuk testi.

### Tingkat Kesulitan

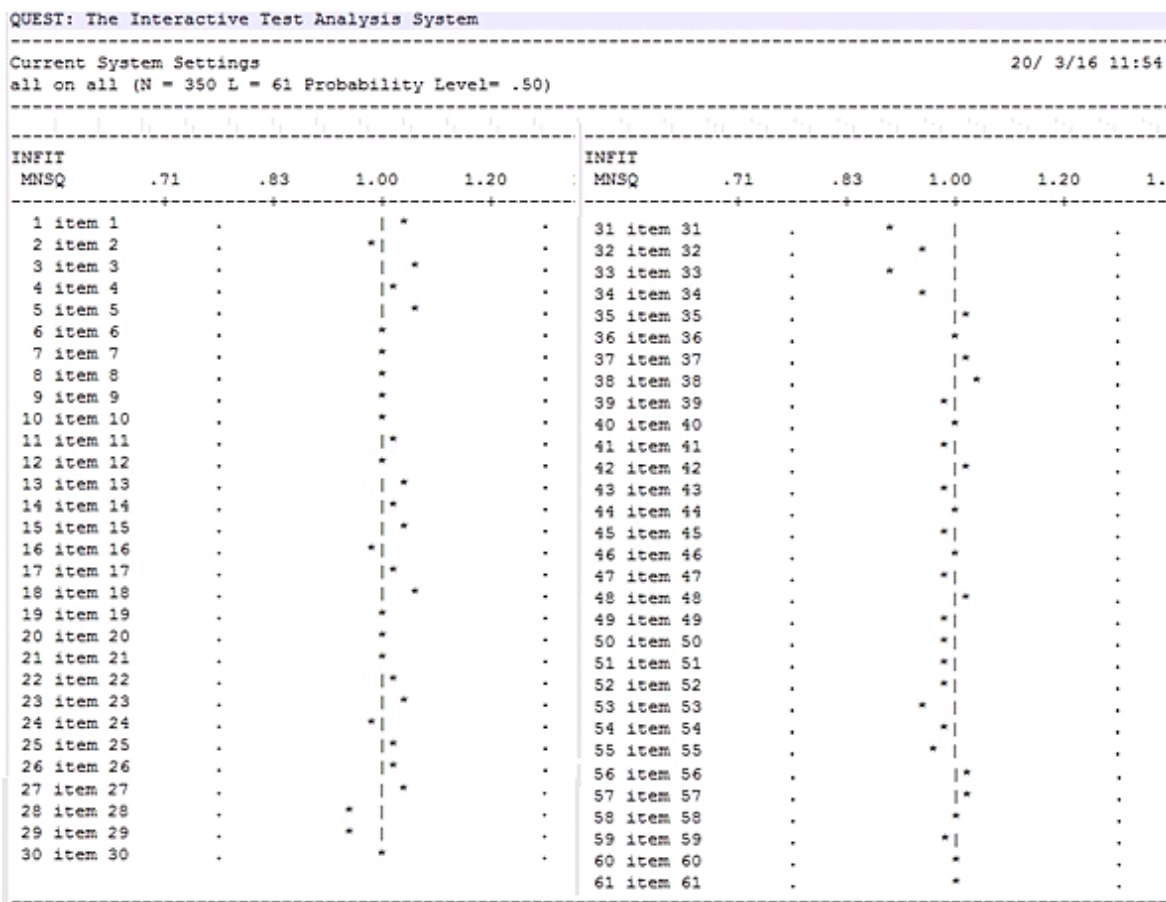
Tingkat kesulitan tiap aspek dalam penelitian ini diperoleh dari rerata tingkat kesulitan tiap butir dalam satu aspek. Semakin tinggi nilai tingkat kesulitan maka butir soal dinilai semakin sulit (Subali & Suyata, 2012, p.56). Hal yang sebaliknya, jika nilai tingkat kesulitan semakin rendah maka butir soal dinilai semakin mudah.

Indikator dengan tingkat kesulitan terendah adalah indikator 2 pada aspek menyusun tabel. Indikator tersulit adalah indikator 3 pada aspek menyusun kesimpulan. Ilustrasi perbandingan tingkat kesulitan antar aspek disajikan dalam Gambar 3.

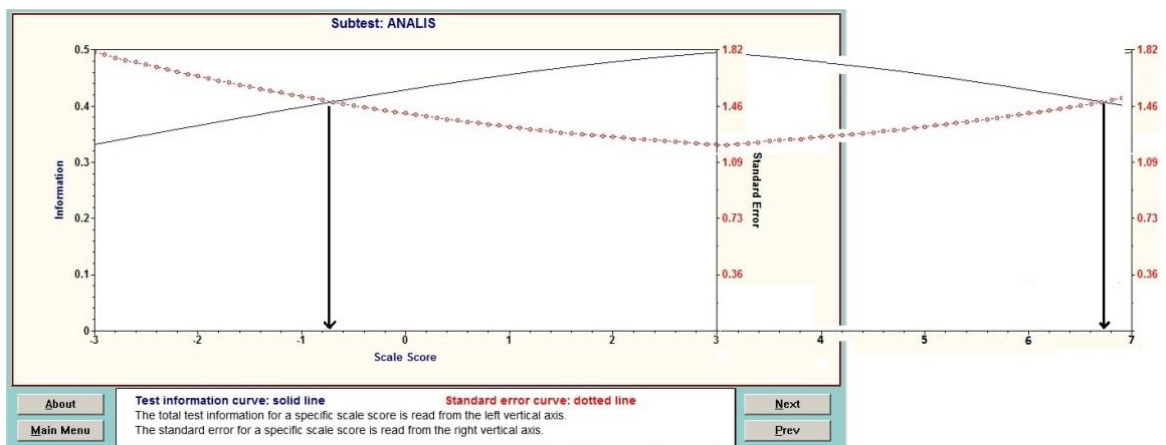
Tabel 3. Hasil Estimasi Butir dan Estimasi Testi Keterampilan Literasi Sains Peserta Didik SMA pada Kegiatan Uji Coba

No.	Uraian	Estimasi untuk butir	Estimasi untuk testi
1.	Nilai rata-rata dan simpangan baku	101,74 ± 64,74	17,73 ± 3,88
2.	Nilai rata-rata dan simpangan baku yang sudah disesuaikan	0,00 ± 1,13	-1,11 ± 0,41
3.	Reliabilitas	0,98	
4.	Nilai rata-rata dan simpangan baku <i>INFIT MNSQ</i>	1,00 ± 0,03	1,00 ± 0,10
5.	Nilai rata-rata dan simpangan baku <i>OUTFIT MNSQ</i>	1,01 ± 0,10	1,01 ± 0,32

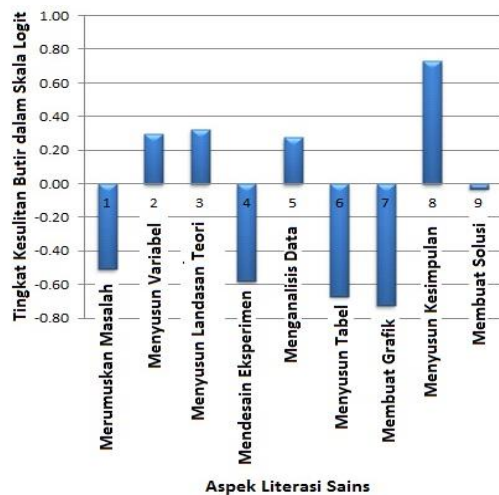




Gambar 2. Distribusi Nilai *INFI* *MNSQ* tiap butir Soal dalam *Fit Model*



Gambar 3. *Total Information Curve (TIC)* untuk Instrumen Tes Tulis



Gambar 4. Tingkat Kesulitan Butir Tiap Aspek Keterampilan Literasi Sains

Gambar 4 menunjukkan perbandingan tingkat kesulitan butir soal dari tiap indikator dan tiap aspek keterampilan literasi sains. Sumintono & Widhiarso (2015, p. 70) mengategorikan tingkat kesulitan skala logit dalam 4 kategori, yaitu sangat sukar, sukar, mudah dan sangat mudah. Kategori sangat sukar memiliki kriteria skala logit lebih dari 1. Kategori sukar berada dalam kriteria 0 sampai 1. Kategori mudah dan kurang dari -1 merupakan kategori sangat mudah. Secara keseluruhan, tingkat kesulitan tiap aspek berada dalam rentang -1 sampai dengan 1. Hal tersebut berarti bahwa tingkat kesulitan rata-rata dari butir pada tiap aspeknya memiliki kategori mudah dan sukar.

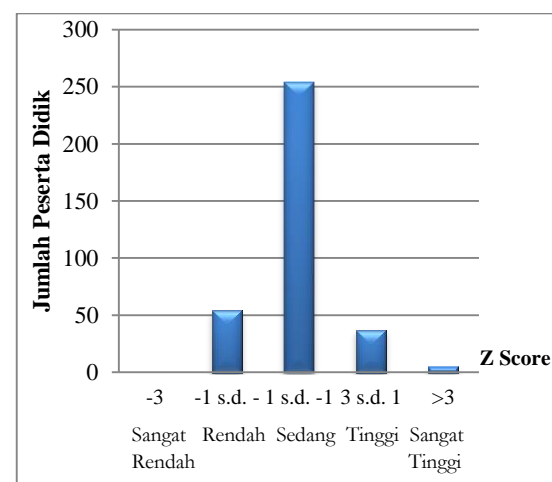
#### Reliabilitas Tes

Reliabilitas perangkat tes tulis dapat diperoleh dengan menggunakan bantuan program *BILOG MG*. Reliabilitas tersebut dapat dilihat pada grafik *Total Information Curve (TIC)* dan *Standard Error of Measurement (SEM)*. Instrumen tes dinyatakan reliabel untuk digunakan kepada peserta didik dengan kategori sedang (-0,7) sampai dengan tinggi (6,7). Hal tersebut memiliki arti bahwa instrumen tes akan tepat diujikan kepada responden dengan kemampuan sedang (-0,7) sampai dengan tinggi (6,7) (Sumintono & Widhiarso, 2015, pp. 7-12; Supahar &

Prasetyo, 2015, p. 104). Hasil analisis reliabilitas tes dapat dilihat pada Gambar 2.

#### Kemampuan Peserta Didik

Kemampuan peserta didik dapat diketahui dari *output* program *BILOG MG* yaitu *file* dengan format *PH3* dan *SCO*. Kemampuan peserta didik dalam literasi sains disajikan dalam kolom *ability* dengan skala logit. Nilai *ability* merupakan  $\theta$  (*tetha*). Nilai  $\theta$  dalam uji coba memiliki rentang antara -2,03 sampai dengan 1,19. Nilai tersebut jika dikonversikan dalam  $z$  score adalah -3,82 sampai dengan 3,92. Keterampilan literasi sains diinterpretasikan sebagai dalam 5 skala yaitu (1) sangat tinggi, (2) tinggi, (3) sedang, (3) rendah, dan (3) sangat rendah. Ilustrasi tingkat keterampilan literasi sains peserta didik disajikan dalam Gambar 5.



Gambar 5. Tingkat Keterampilan Literasi Sains Peserta Didik Uji Coba

#### Uji Coba Rubrik Penskoran

Pada saat mengujicobakan rubrik penskoran, peserta didik diberikan tugas-tugas dalam bentuk LKPD. Pengamat atau rater menggunakan rubrik penskoran untuk menilai ketuntasan kinerja yang dilakukan peserta didik. Skor yang diperoleh peserta didik dianalisis untuk ditentukan reliabilitas Alfa dan *ICC*. Analisis dilakukan dengan menggunakan bantuan program *SPSS 16*. Hasil analisis dirangkum dalam Tabel 4.

Interpretasi berdasarkan analisis skor peserta didik dari pengamatan adalah sebagai berikut (1) Reliabilitas Alfa yang diperoleh mendapatkan nilai di atas 0,8 (Streiner, 2003, p.103) yang merupakan batas minimal pada penelitian dasar. (2) Skor ICC menunjukkan tingkat reliabilitas sangat baik (*Excellent*) yang berada dalam rentang 0,75 sampai dengan 1,00 (Cicchetti, 1994, p. 286).

Tabel 4. Hasil Analisis Reliabilitas Alfa dan *Interclass Coefficient Correlation (ICC)*

No	Statistik	Reliabilitas Alfa	ICC
1.	Rubrik Penskoran Aspek Psikomotorik	0,983	0,936
2.	Rubrik Penskoran LKPD	0,993	0,971
3.	Rubrik Penskoran Laporan	0,999	0,995

### Simpulan

Lembar pengamatan berupa rubrik penskoran dan penilaian diri dinyatakan layak berdasarkan nilai validitas isi V. Rubrik penskoran juga dinyatakan reliabel dengan nilai Reliabilitas Alfa melebihi 0,8 dan ICC dengan kriteria *Excellent*. Perangkat tes tulis terbukti valid dilihat dari sudut pandang *content* atau isi dengan nilai *CVI* sebesar 1 dan valid secara empirik karena sesuai dengan nilai *INFIT MNSQ* model *Rasch*. Berdasarkan Grafik *TIC* dan *SEM*, perangkat tes tulis dinyatakan reliabel untuk digunakan pada peserta didik dengan kategori sedang sampai dengan tinggi (-0,7 sampai dengan 6,7). Oleh karena itu, Asesmen kinerja Literasi Sains berbasis *STEM* layak digunakan karena telah memenuhi syarat dalam aspek validitas isi, validitas empiris dan reliabilitas.

Saran bagi peneliti yang hendak melakukan penelitian yang sejenis adalah sebagai berikut (a) Semakin banyak pakar yang menelaah instrument maka akan semakin baik kualitas instrumen dalam aspek validitas isi, (b) penting untuk mempersiapkan lebih dari satu untuk tiap indikator agar dapat menghindari ketidakterwakilan suatu indikator jika butir tertentugugur dalam prses validasi (c) penting untuk memilih res-

ponden yang mewakili peserta didik dengan kemampuan rendah, sedang dan tinggi dalam uji coba empiris sebagai upaya untuk mendapatkan validitas yang baik.

### Daftar Pustaka

- Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, 45(1), 131–142. <https://doi.org/10.1177/0013164485451012>
- Azwar, S. (2012). *Reliabilitas dan validitas* (4th ed.). Yogyakarta: Pustaka Pelajar.
- Azwar, S. (2015). *Metode penelitian*. Yogyakarta: Pustaka Pelajar.
- Badan Standar Nasional Pendidikan. (2006). *Panduan penyusunan kurikulum tingkat satuan pendidikan jenjang pendidikan dasar dan menengah*. Jakarta: BSNP. Retrieved from [http://bsnp-indonesia.org/id/wp-content/uploads/kompetensi/Panduan\\_Umum\\_KTSP.pdf](http://bsnp-indonesia.org/id/wp-content/uploads/kompetensi/Panduan_Umum_KTSP.pdf)
- Bashooir, K., & Supahar, S. (2016). Analisis aspek kinerja literasi sains pada materi kalor Fisika. *UPEJ Unnes Physics Education Journal*, 5(1). Retrieved from <https://journal.unnes.ac.id/sju/index.php/upej/article/view/12711>
- Breiner, J. M., Harkness, S. S., Johnson, C. C., & Koehler, C. M. (2012). What is STEM? A discussion about conceptions of STEM in education and partnerships. *School Science and Mathematics*, 112(1), 3–11. <https://doi.org/10.1111/j.1949-8594.2011.00109.x>
- Chiappetta, E. L., & Koballa, T. R. (2010). *Science instruction in the middle and secondary schools: developing fundamental knowledge and skills* (7th ed.). USA: Pearson Education, Inc.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological*

- Assessment*, 6(4), 284–290.  
<https://doi.org/10.1037/1040-3590.6.4.284>
- Departemen Pendidikan dan Kebudayaan. (2001). *Kamus besar bahasa Indonesia* (3rd ed.). Jakarta: Balai Pustaka.
- Department of Education. (2009). *Report of the STEM review*. Retrieved from [https://www.education-ni.gov.uk/sites/default/files/publications/de/Report of the STEM Review 2009\\_1.PDF](https://www.education-ni.gov.uk/sites/default/files/publications/de/Report%20of%20the%20STEM%20Review%202009_1.PDF)
- Gonzalez, H. B., & Kuenzi, J. J. (2012). Science, technology, engineering, and mathematics (STEM) education: a primer. Retrieved from <https://fas.org/sgp/crs/misc/R42642.pdf>
- Hernandez, P. R., Bodin, R., Elliott, J. W., Ibrahim, B., Rambo-Hernandez, K. E., Chen, T. W., & de Miranda, M. A. (2014). Connecting the STEM dots: measuring the effect of an integrated engineering design intervention. *International Journal of Technology and Design Education*, 24(1), 107–120. <https://doi.org/10.1007/s10798-013-9241-0>
- Ismail, I., Permanasari, A., & Setiawan, W. (2016). Efektivitas virtual lab berbasis STEM dalam meningkatkan literasi sains siswa dengan perbedaan gender. *Jurnal Inovasi Pendidikan IPA*, 2(2), 190. <https://doi.org/10.21831/jipi.v2i2.8570>
- Kartowagiran, B., & Jaedun, A. (2016). Model asesmen autentik untuk menilai hasil belajar siswa sekolah menengah pertama (SMP): implementasi asesmen autentik di SMP. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 20(2), 131. <https://doi.org/10.21831/pep.v20i2.10063>
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>
- Mardapi, D. (2012). *Pengukuran, penilaian dan evaluasi pendidikan*. Yogyakarta: Nuha Medika.
- OECD. (2014). *PISA 2012 results: what students know and can do student performance in mathematics, reading and science volume I*. Paris: OECD Publishing.
- P21. (2009). 21st century skills map. Retrieved from [http://www.p21.org/storage/documents/21st\\_century\\_skills\\_english\\_map.pdf](http://www.p21.org/storage/documents/21st_century_skills_english_map.pdf)
- Presiden Republik Indonesia. Undang-Undang Republik Indonesia nomor 20 tahun 2003 tentang Sistem Pendidikan Nasional (2003). Indonesia.
- Reeve, E. M. (2013). Implementing science, technology, mathematics, and engineering (STEM) education in Thailand and in ASEAN. Retrieved from [http://dpst-apply.ipst.ac.th/specialproject/images/IPST\\_Global/document/Implementing STEM in ASEAN - IPST May 7 2013 - Final.pdf](http://dpst-apply.ipst.ac.th/specialproject/images/IPST_Global/document/Implementing%20STEM%20in%20ASEAN%20-%20IPST%20May%207%202013%20-%20Final.pdf)
- Retnawati, H. (2016). *Validitas reliabilitas dan karakteristik butir*. Yogyakarta: Parama Publishing.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420–428. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/18839484>
- Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80(1), 99–103. [https://doi.org/10.1207/S15327752JPA8001\\_18](https://doi.org/10.1207/S15327752JPA8001_18)
- Subali, B., & Suyata, P. (2012). *Pengembangan item tes konvergen dan divergen: penyelidikan*

- validitasnya Secara empiris*. Yogyakarta: Diandra Pustaka Indonesia.
- Sumintono, B., & Widhiarso, W. (2015). *Aplikasi pemodelan RASCH pada assessment pendidikan*. Cimahi: Tim Komunikata Publishing House.
- Supahar, & Prasetyo, Z. K. (2015). Pengembangan instrumen penilaian kinerja kemampuan inkuiri peserta didik pada mata pelajaran fisika SMA. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 19(1), 96–108. Retrieved from <https://journal.uny.ac.id/index.php/jpep/article/view/4560>
- Supahar, S. (2014). The estimation of inquiry performance test items of high school physics subject with quest program. In *International Conference on Research, Implementation And Education of Mathematics And Sciences*. Yogyakarta: Yogyakarta State University.
- Supahar, S. (2015). Applying content validity ratios (CVR) to the quantitative content validity of physics learning achievement tests. In *International Conference on Research, Implementation And Education of Mathematics And Sciences*. Yogyakarta: Yogyakarta State University.
- Wagner, T. (2008). *The global achievement gap*. New York: Basic Book.
- Yore, L. D., & Treagust, D. F. (2006). Current realities and future possibilities: language and science literacy—empowering research and informing instruction. *International Journal of Science Education*, 28(2–3), 291–314. <https://doi.org/10.1080/09500690500336973>

## PENGEMBANGAN INSTRUMEN PENILAIAN AFEKTIF SISWA KELAS IV SEKOLAH DASAR DI KABUPATEN KLATEN

Anggarwati Riscaputantri<sup>1</sup> \*, Sri Wening<sup>1</sup>

<sup>1</sup>Universitas Negeri Yogyakarta

Jl. Colombo No. 1, Depok, Sleman 55281, Yogyakarta, Indonesia

\* Corresponding Author. Email: [anggarwati.riscaputantri2015@student.uny.ac.id](mailto:anggarwati.riscaputantri2015@student.uny.ac.id)

### Abstrak

Penelitian ini bertujuan untuk menghasilkan instrumen yang memenuhi syarat untuk mengukur afektif siswa kelas IV, dan mengetahui gambaran afektif siswa kelas IV tahun ajaran 2016/2017 di Kabupaten Klaten. Pendekatan yang digunakan dalam penelitian ini adalah *Research and Development* dengan Model Thiagarajan. Instrumen penilaian afektif telah dinyatakan valid dan reliabel. Validitas isi diuji oleh ahli dan diperoleh indeks Aiken sebesar 0,85. Validitas konstruk dilakukan dengan analisis faktor eksploratori dan konfirmatori. Hasil analisis faktor eksploratori menunjukkan nilai KMO sebesar 0,726 sedangkan Bartlett's Test sebesar Sig. 0,000. Hasil analisis faktor konfirmatori menunjukkan nilai Chi-Kuadrat sebesar 0,039; RMSEA= 0,012; GFI=0,93; SRMR=0,033; AGFI=0,90; CFI=1,00; RFI=1,00. Hasil reliabilitas menunjukkan bahwa koefisien Cronbach's Alpha sebesar 0,653 sedangkan pada reliabilitas konstruk menunjukkan bahwa pada sikap disiplin koefisien CR=0,99 dan VE=0,99; sikap jujur CR=0,99 dan VE=0,99; sikap peduli CR=0,99 dan VE=0,99; sikap percaya diri CR=0,94 dan VE=0,95; sikap santun CR=0,99 dan VE=0,81; sikap tanggung jawab CR=0,99 dan VE=0,79. Gambaran afektif siswa adalah (a) berdasarkan tingkatan afektif Krathwohl, Bloom & Masia pada tingkatan karakterisasi sebesar 42,9%, mengorganisasi sebesar 50,9%, menilai 6,2% sedangkan tahapan merepon dan menerima sebesar 0%, (b) berdasarkan acuan Kurikulum 2013 pada kategori baik sekali sebesar 62%, kategori baik 38% dan kategori cukup 1%, kategori butuh bimbingan sebesar 0%.

**Kata kunci:** *penilaian afektif, sikap, tingkatan afektif Krathwohl, Bloom & Masia, analisis faktor eksploratori, analisis faktor konfirmatori*

## DEVELOPING AFFECTIVE ASSESMENT INSTRUMEN OF FOURTH GRADE STUDENT IN KLATEN REGENCY

### Abstract

This study aimed to: produce qualified instrument to measure an affective of grade IV students and find out the affective picture of fourth grade students in academic year 2016/2017 Klaten. This study used Research and Development approach by using Thiagarajan Model. Affective assessment instruments were stated as valid and reliable. Content validity was judged by experts. The Aiken index of instruments is 0.85. Construct validity was measured by using both exploratory and confirmatory factor analysis. The results of exploratory factor analysis showed that the KMO value is 0.726 while the Bartlett's Test is equal to Sig. 0,000. Confirmatory factor analysis showed that Chi-Square value is 0.039; RMSEA = 0.012; GFI = 0.93; SRMR = 0.033; AGFI = 0.90; CFI = 1.00; RFI = 1.00. Cronbach's Alpha coefficient is 0.653 while the construct reliability showed the coefficient of discipline CR = 0.99 and VE = 0.99; honest attitude CR = 0.99 and VE = 0.99; attitude caring about CR = 0.99 and VE = 0.99; self-confidence CR = 0.94 and VE = 0.95; polite attitude CR = 0.99 and VE = 0.81; the attitude of responsibility CR = 0.99 and VE = 0.79. The description of students' affective was (a) based on the affective level of Krathwohl, Bloom & Masia, level of characterization 42.9%, organizing 50.9%, assessing 6.2% while phoning and receiving 0%, (b) based on the 2013 Curriculum reference, it could be stated as excellent category 62%, favorable category 38% and fair category 1%, the category needs guidance at 0%.

**Keywords:** *affective assessment, attitude, Krathwohl affective level, Bloom & Masia, exploratory factor analysis, confirmatory factor analysis*

Permalink/DOI: <http://dx.doi.org/10.21831/pep.v22i2.16885>

## **Pendahuluan**

Sekolah dasar merupakan salah satu jenjang dalam pendidikan yang harus ditempuh oleh siswa. Berbagai visi dan misi disusun untuk mencapai tujuan dalam mendidik siswa. Salah satu visi yang saat ini diterapkan dan menjadi populer adalah membentuk karakter siswa. Hal tersebut bukan tanpa dasar namun merupakan wujud dari pelaksanaan Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 23 Tahun 2015 tentang Budi Pekerti (Menteri Pendidikan dan Kebudayaan Republik Indonesia, 2015) yang menyatakan bahwa pendidikan karakter seharusnya menjadi gerakan bersama melibatkan pemerintah, pemerintah daerah, masyarakat, dan atau orang tua. Berlandaskan hal tersebut, sekolah merupakan bagian dari pemerintah yang berupaya melaksanakan peraturan sebagai bentuk peran aktif membangun sumber daya manusia yang lebih baik melalui pendidikan.

Pendidikan karakter menurut Samani & Hariyanto (2012, p. 45) menjelaskan bahwa pendidikan karakter merupakan proses pemberian tuntunan kepada siswa untuk menjadi manusia seutuhnya yang berkarakter dalam dimensi hati, pikir, raga serta karsa. Hal ini merujuk pada kesimpulan bahwa siswa yang hidup di lingkungan sosial dapat mempengaruhi pembentukan karakternya. Sekolah sebagai salah satu lingkungan yang membentuk karakter siswa diharapkan mampu memberikan perlakuan yang baik sehingga terbentuk karakter yang baik pula. Siswa pada jenjang sekolah dasar termasuk dalam tahapan operasional konkret. Hal ini menunjukkan bahwa siswa perlu dilibatkan secara langsung dalam kegiatan-kegiatan di sekolah maupun proses pembelajaran terkait penanaman karakter secara maksimal untuk membentuk karakter siswa. Selain itu, pendidikan yang berbasis karakter perlu memperhatikan proses pembelajaran terkait dengan ketercapaian pendidikan karakter tersebut.

Pembelajaran yang memuat nilai-nilai karakter akan menghasilkan siswa yang memiliki karakter yang diharapkan sesuai dengan tujuan pembelajaran. Hal tersebut didukung penelitian yang dilakukan oleh

Wening (2012, p. 1) menjelaskan dalam hasil penelitiannya bahwa pembentukan karakter siswa yang diberi intervensi pembelajaran nilai menggunakan buku cerita memberikan efek yang bermakna pada aspek pembentukan karakter siswa. Hal senada juga disampaikan oleh Adibatin (2016, p. 1) dalam penelitiannya menjelaskan bahwa hasil pembelajaran menggunakan alat peraga inovatif ini menunjukkan bahwa melalui strategi pembelajaran PAKEM dengan bermain cincin akik di jempol tangan dapat membangun karakter siswa sampai 11,33%. Berdasarkan penelitian-penelitian yang telah dilakukan maka dapat disimpulkan bahwa pembelajaran disekolah memiliki peran dalam membentuk karakter siswa. Hal tersebut sesuai dengan tujuan sekolah yang disampaikan Mardapi (2012, p. 2) bahwa sekolah sebagai lembaga pendidikan adalah mengembangkan potensi peserta didik secara optimal menjadi kemampuan untuk hidup dimasyarakat.

Pengembangan potensi peserta didik dapat dilakukan melalui proses belajar mengajar disekolah atau pengajaran. Pada proses pengajaran diharapkan adanya interaksi yang baik dari siswa dengan sumber belajar. Sumber belajar ini tidak hanya guru saja namun dapat diartikan sebagai media yang dapat memberikan sumbangan terhadap capaian belajar yaitu buku, alat peraga maupun media-media lain yang mendukung. Pencapaian hasil belajar sering disebut dengan hasil belajar merupakan tingkat kompetensi yang dicapai peserta didik yang mencakup tiga ranah, yaitu ranah kognitif, ranah afektif, dan ranah psikomotor (Mardapi, 2012, p. 2). Hal tersebut juga didukung oleh pendapat Krathwohl, Bloom, & Masia (1973, p. 6) menjelaskan bahwa dalam pendidikan terdapat tiga domain atau ranah yang perlu diperhatikan yaitu ranah kognitif, afektif, dan psikomotorik. Pada ketiga domain atau ranah tersebut perlu diperhatikan karakteristiknya serta tidak ada yang lebih dominan, artinya guru sebagai fasilitator dalam proses pembelajaran perlu mempersiapkan capaian pembelajaran pada ketiga ranah tersebut, yaitu kognitif, afektif, dan psikomotorik. Menurut Haryadi & Aripin (2015, p. 41) kog-



nitif merupakan suatu pokok bahasan yang berhubungan dengan kognisi, dengan tujuan akhir berupa pengetahuan yang didapat melalui percobaan, penelitian, penemuan, dan pengamatan. Kognitif berhubungan erat dengan pikiran, memori, nalar, intelektual, kemampuan berhitung, logika, eksakta, sains, numerik dan akademik. Ranah psikomotor berhubungan dengan hasil belajar yang pencapaiannya melalui keterampilan manipulasi yang melibatkan otot dan kekuatan fisik. Mata pelajaran yang berkaitan dengan psikomotor adalah mata pelajaran yang lebih berorientasi pada gerakan dan menekankan pada reaksi-reaksi fisik dan keterampilan tangan. Tujuan pengukuran ranah psikomotor adalah selain untuk memperbaiki pencapaian tujuan instruksional oleh siswa pada ranah psikomotor khususnya pada tingkat imitasi, manipulasi presisi, artikulasi, dan naturalisasi, juga dapat meningkatkan kemampuan gerak reflex, gerak dasar, keterampilan perseptual, keterampilan fisik, gerak terampil, dan komunikasi non-diskusif siswa. Ranah ketiga yaitu afektif. Haryadi & Aripin (2015, p. 41) menjelaskan bahwa afektif memiliki cakupan yang berbeda dengan kognitif, afektif lebih berhubungan dengan psikis, jiwa dan rasa. Dijelaskan lebih lanjut bahwa pada afektif meliputi sikap yang ditunjukkan. Afektif juga dapat diartikan sebagai proses yang menekankan pada perasaan, emosi, atau tingkat penerimaan atau penolakan. Tujuan afektif bervariasi dari perhatian sederhana pada fenomena terpilih hingga kualitas karakter dan hati nurani yang kompleks namun konsisten secara internal (Krathwohl et al., 1973, p. 7).

Salah satu penelitian yang dilakukan oleh Setyawan & Dimiyati (2015, p. 1) menjelaskan bahwa dalam menyusun proses pembelajaran yang berupa model permainan aktivitas luar kelas perlu memperhatikan pencapaian pembelajaran pada tiga ranah yaitu kognitif, afektif dan psikomotorik dengan hasil penelitian menunjukkan bahwa model permainan juga efektif untuk mengembangkan ranah kognitif, afektif, dan psikomotorik, yang dibuktikan dengan terjadinya pening-

katan persentase nilai rata-rata siswa sebesar 21%.

Keterkaitan pelaksanaan pembelajaran yang memiliki capaian pada tiga ranah kognitif, afektif dan psikomotorik tidak terlepas dari penilaian yang wajib dilakukan oleh guru setelah dilakukan proses belajar. Menyelenggarakan penilaian dan evaluasi proses dan hasil belajar merupakan kompetensi pedagogik yang harus dilaksanakan guru. Penilaian menurut Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 23 Tahun 2016 tentang Standar Penilaian (Menteri Pendidikan dan Kebudayaan Republik Indonesia, 2016) pada pasal 1 menjelaskan bahwa penilaian adalah proses pengumpulan dan pengolahan informasi untuk mengukur pencapaian hasil belajar peserta didik. Hasil belajar siswa merupakan informasi bagi guru untuk melihat keberhasilan belajar siswanya. Pentingnya hasil belajar tersebut, maka guru memerlukan alat untuk mengumpulkan informasi berupa data. Alat yang digunakan untuk mengumpulkan data tersebut disebut instrumen penilaian hasil belajar.

Instrumen penilaian hasil belajar dikelompokkan menjadi dua yaitu tes dan non-tes. Penilaian tes dilakukan untuk menguji kemampuan kognitif siswa. Pada kemampuan afektif dan psikomotorik siswa dapat menggunakan bentuk penilaian non-tes. Guru perlu melakukan penilaian pada ketiga ranah tersebut agar dapat menggambarkan keberhasilan siswa dalam belajar secara utuh. Namun, hal ini tidak sejalan dengan proses penilaian yang terjadi di lapangan. Berdasarkan hasil penelitian yang dilakukan oleh peneliti di Kabupaten Klaten menunjukkan bahwa dari 30 responden dari guru yang diberikan kuesioner menunjukkan bahwa 83,3% guru tidak memiliki instrumen penilaian afektif. Hal tidak sejalan dengan pemahaman guru tentang pentingnya penilaian afektif yaitu sebesar 96%. Sedangkan pada pelaksanaan penilaian ditemukan sebesar 93,3% guru tidak melakukan penilaian afektif. Peneliti menindaklanjuti dengan wawancara. Hasil wawancara menunjukkan bahwa ketidakterdapatnya instrumen afektif dikarenakan guru mengalami kesulitan dalam



membuat instrumen penilaian afektif. Kesulitan tersebut disebabkan minimnya contoh penilaian afektif. Wawancara juga dilakukan terkait kendala waktu, sejauh ini kendala yang dialami guru selain kesulitan membuat instrumen yaitu waktu. Guru tidak memiliki cukup waktu untuk membuat instrumen penilaian afektif sehingga guru hanya observasi terhadap sikap yang ditunjukkan siswa pada saat proses pembelajaran maupun pada saat diluar pembelajaran namun masih dalam lingkup sekolah. Berdasarkan Kementerian Pendidikan dan Kebudayaan melalui Direktorat Pembinaan Pendidikan Dasar (2016, p. 6) pada buku Panduan Teknis Pembelajaran dan Penilaian di Sekolah Dasar menjelaskan bahwa dalam melakukan penilaian afektif yang berupa sikap perlu dilakukan konfirmasi sehingga tidak hanya dilakukan dengan observasi. Kurikulum yang diterapkan pemerintah yaitu Kurikulum 2013 mewajibkan guru untuk melakukan penilaian afektif. Penilaian afektif pada Kurikulum 2013 terdapat pada sikap sosial siswa yaitu sikap disiplin, jujur, peduli, percaya diri, santun dan tanggung jawab. Hal tersebut penting dilakukan terkait informasi yang didapatkan untuk menilai afektif siswa berdasarkan observasi terdapat unsur subjektivitas yang dilakukan guru sehingga menghasilkan informasi yang bias dalam penilaian afektif siswa.

Mardapi (2012, p. 15) menjelaskan bahwa instrumen yang dapat digunakan untuk menilai kemampuan afektif adalah bentuk kuesioner dengan disajikan kasus yang selanjutnya siswa diminta untuk memberi respon. Kuesioner yang baik tentu dalam pembuatannya telah melalui prosedur. Terkait hal itu McCoach, Gable, & Madura (2013, p. 26) menjelaskan bahwa:

*“In selecting an affective characteristic for measurement, instrumen developers should carefully consider the theory underlying the construct. Only with a clear conceptual understanding can one proceed to create valid operational definitions (i.e., statements/items) for each affective characteristic.”*

Definisi tersebut dapat diartikan bahwa dalam memilih karakteristik pada pengembangan instrumen ranah afektif perlu

berhati-hati dalam mempertimbangkan pemilihan teori secara konstruktif. Dengan pemahaman yang baik dapat menghasilkan definisi operasional yang valid untuk masing-masing karakter pada ranah afektif. Hal ini menunjukkan bahwa dalam melakukan penilaian afektif tidak boleh sembarangan dalam membuat definisi operasional yang kemudian dikembangkan menjadi indikator-indikator sikap yang akan dinilai. Guru tentu perlu memahami hal ini sebagai upaya membantu mengenali sikap yang tercermin dari diri siswa. Selain itu, pada penyusunan butir-butir perlu dilakukan pengujian validitas dan mengestimasi reliabilitas. Validitas dilakukan untuk memastikan layak digunakan untuk mengukur apa yang seharusnya diukur, dan estimasi reliabilitas dimaksudkan instrumen tersebut memiliki keajegan atau konsisten dalam mengukur kemampuan afektif siswa, dalam hal ini adalah sikap siswa. Hasil penelitian yang dilakukan peneliti menunjukkan bahwa tidak ada guru yang melakukan uji validitas dan reliabilitas pada instrumen penilaian afektif siswa. Berdasarkan hal tersebut maka dapat dikatakan bahwa instrumen yang dibuat guru tersebut tidak layak untuk mengukur afektif siswa. Instrumen yang tidak layak untuk mengukur afektif siswa menyebabkan tidak tergambar kemampuan siswa secara detail. Berdasarkan hal tersebut maka diperlukan instrumen yang layak untuk mengukur afektif siswa berupa sikap sosial siswa yaitu sikap disiplin, jujur, peduli, percaya diri, santun dan tanggung jawab secara detail dengan klasifikasi tingkatan afektif Krathwohl, Bloom & Masia. Tingkatan tersebut adalah (1) menerima yaitu Kesadaran, kemauan untuk menerima, perhatian terseleksi. Pada tahap ini pembelajar peka terhadap keberadaan fenomena atau rangsangan, (2) merespon yaitu Berpartisipasi aktif sebagai bagian dari pembelajar. Menyimak dan bereaksi terhadap suatu fenomena tertentu. Pada tahap ini pembelajar cukup termotivasi untuk berperan serta dan menghadapi rangsangan yang datang berupa gagasan, benda atau sistem nilai, (3) menilai yaitu terdiri atas menerima nilai, memilih nilai, dan komitmen. Pada tahap ini pembelajar mema-

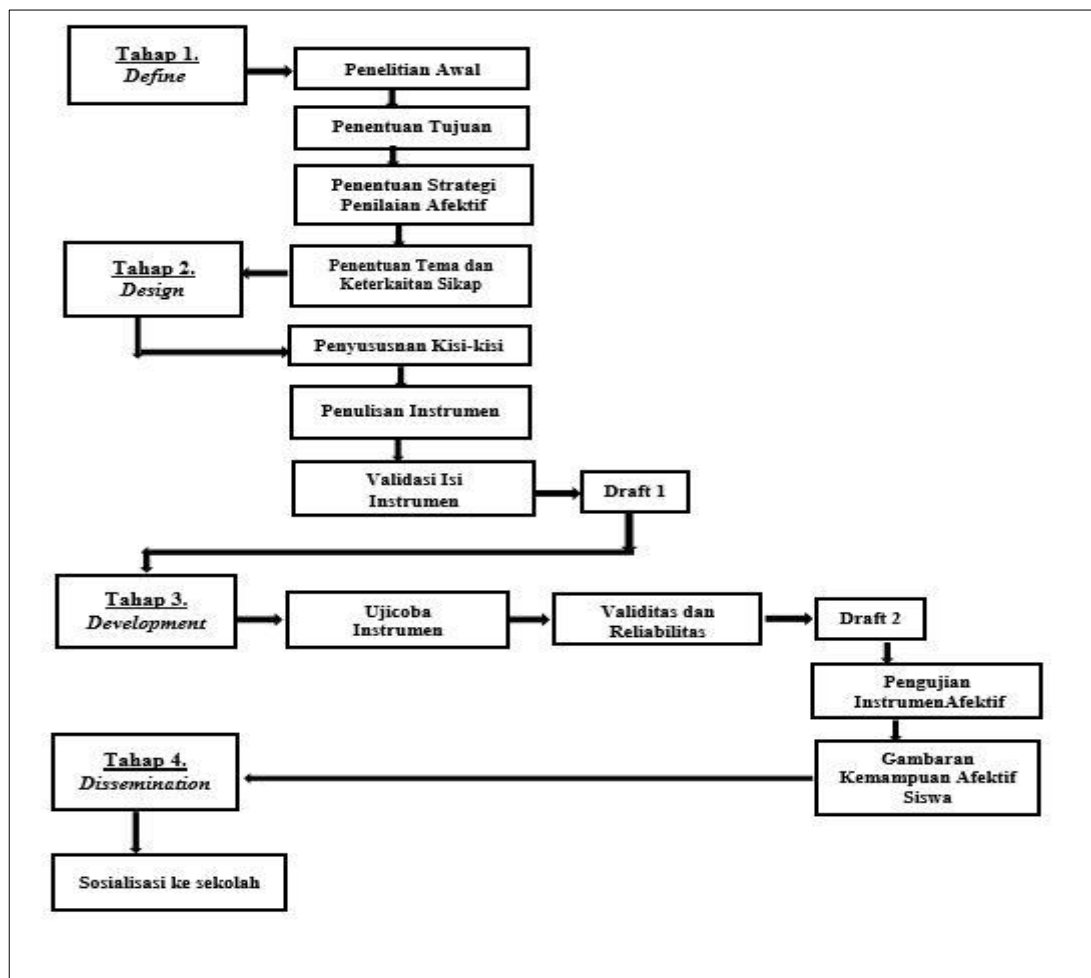
hami bahwa benda-benda, gejala atau suatu perilaku mempunyai nilai. (4) mengorganisasi yaitu adanya prioritas untuk membandingkan perbedaan nilai, dan menciptakan suatu sistem nilai yang unik dan (5) karakterisasi yaitu memiliki sistem nilai yang mengontrol perilakunya, ajeg dapat diramalkan, dan yang paling penting merupakan karakteristik dari pembelajar sehingga tataran ini disebut pula karakterisasi nilai serta gambaran afektif siswa berdasarkan acuan Kurikulum 2013.

### Metode Penelitian

Penelitian ini merupakan jenis penelitian pengembangan. Model yang digunakan pada penelitian ini mengacu pada model dari Thiagarajan, et al. Model pengembangan Thiagarajan, Semmel, & Semmel (1974) atau

sering disebut dengan Four-D Model ini terdiri dari 4 tahapan pengembangan yaitu *define* (pendefinisian), *design* (perancangan), *develop* (pengembangan) dan *dissemination* (penyebaran). Model pengembangan tersebut disajikan pada Gambar 1.

Instrumen yang dikembangkan menggunakan model Thiagarajan adalah instrumen penilaian afektif untuk siswa sekolah dasar khususnya kelas IV. Jenis instrumen yang dikembangkan adalah penilaian diri. Siswa melakukan penilaian diri terhadap sikap disiplin, jujur, peduli, percaya diri, santun, dan tanggung jawab. Pengembangan instrumen pada tahap awal dilakukan dengan kontrak teori dari keenam sikap tersebut. Konstrak teori yang didapatkan pada setiap sikap dikembangkan menjadi indikator yang dapat disusun oleh peneliti bersama guru dalam FGD.



Gambar 1. Desain Pengembangan Penilaian Afektif

Hasil FGD tersebut lalu diuji validitasnya dan diestimasi reliabilitasnya. Validitas diuji secara isi dan kontrak. Validitas isi diuji oleh ahli dan dianalisis menggunakan rumus Aiken's. Retnawari (2016, p.19) mengatakan bahwa besarnya koefisien validitas yang dapat diterima dan dianggap sudah layak adalah  $\geq 0,40$ . Validitas kontrak dilakukan dengan uji analisis faktor eksploratori dan konfirmatori. Wagiran (2015, p. 300) mengatakan bahwa analisis faktor eksploratori dilakukan untuk memastikan apakah butir-butir tertentu mendukung faktornya dan faktor-faktor tersebut mendukung variabelnya.

Kriteria analisis faktor eksploratori menurut Wagiran (2015, p. 303) adalah sebagai berikut:

Tabel 1. Kriteria analisis faktor eksploratori

Kriteria	Nilai
Keyser Mayer Oikin (KMO)	> 0,5
Barlett's Test of Sphericity	> 0,05
Anti image correlation	MSA > 0,5
Nilai Eigen	> 1
Rotasi faktor	> 0,4

Berdasarkan hasil analisis faktor eksploratori maka di dapatkan instrumen yang valid. Instrumen yang valid kemudian dilakukan uji lanjut yaitu analisis faktor konfirmatori. Hendryadi & Suryani (2014, p. 63) menyatakan bahwa analisis tersebut memiliki tujuan untuk melihat apakah model yang diusulkan memiliki kecocokan yang baik (*good fit*) atau belum. Parameter kecocokan analisis faktor konfirmatori menurut Hendryadi & Suryani (2014, pp. 18–19) disajikan pada Tabel 2.

Tabel 2. Parameter Kecocokan Model Analisis Faktor Konfirmatori

Kriteria	Nilai
Chi Kuadrat	P value > 0,05
Root Means Square Error of Approximation (RSMEA)	RSMEA $\leq$ 0,08
Goodness Of Fit Indeks (GFI)	GFI $\geq$ 0,09
Standardized Root Mean Square Residual (SRMR)	SRMR $\leq$ 0,05
Adjusted Goodness Fit Of Indeks (AGFI)	AGFI $\geq$ 0,90
Comparative Fit Indeks (CFI)	CFI $\geq$ 0,90
Incremental For Index (RFI)	RFI $\geq$ 0,90

Pada tahap selanjutnya, instrumen yang telah dinyatakan valid diestimasi reliabilitasnya. Kriteria untuk menentukan reliabilitas instrumen didasarkan atas koefisien Alpha > 0,5 (Wagiran, 2015). Reliabilitas kontrak instrumen dapat didapatkan dari hasil perhitungan analisis faktor konfirmatori menggunakan formula *Construct Reliability* (CR) dan *Variance Extract* (VE). Hendryadi & Suryani (2014, p. 17) menyebutkan bahwa nilai reliabilitas CR disarankan lebih dari 0,7 sedangkan nilai VE disarankan lebih besar dari 0,5. Formula untuk menghitung reliabilitas kontrak menurut Hendryadi & Suryani (2014, pp. 17–18) adalah sebagai berikut:

*Construct Relialit*

$$= \frac{(\sum \text{Standarized loading})^2}{(\sum \text{Standarized loading})^2 + (\sum \text{SMeasurement Error})}$$

*Variance Extract*

$$= \frac{\sum \text{Standarized loading}^2}{\sum \text{Standarized loading}^2 + (\sum \text{SMeasurement Error})}$$

Uji coba dilakukan pada siswa Sekolah Dasar Kelas IV di Kabupaten Klaten pada semester 2 tahun ajaran 2016/2017. Uji coba dilakukan pada sekolah dasar yang menggunakan Kurikulum 2013 di Kabupaten Klaten. Daftar SD tersebut disajikan pada Tabel 3.

Subjek penelitian dalam penelitian ini adalah siswa kelas IV sekolah dasar di Kabupaten Klaten pada semester 2 yang disajikan pada Tabel 4.

Penentuan subjek coba pada penelitian ini menggunakan *non-probability sampling* dengan teknik *purposive sampling*. Wagiran (2015, p. 210) menjelaskan bahwa pada teknik *purposive sampling* dipilih berdasarkan tujuan dan maksud tertentu. Tujuan dipilihnya subjek coba pada penelitian ini merupakan sekolah yang terpilih yaitu sekolah yang telah menerapkan Kurikulum 2013 secara keseluruhan. Karakteristik sekolah yang sesuai dengan penelitian ini terdapat pada 13 SD dan 6 SD sebagai subjek ujicoba. Sekolah-sekolah tersebut oleh Dinas Pendidikan Kabupaten Klaten ditunjuk sebagai Sekolah Dasar percontohan penerapan Kurikulum 2013 dalam implementasi kurikulum guru yang akan diajak pada proses FGD tidak mengalami

kesulitan. Berdasarkan hal tersebut semua siswa kelas IV sekolah tersebut dilibatkan sebagai subjek coba.

Tabel 3. Subjek Uji Coba

Nama Sekolah	Jumlah
SD N 1 Brajan	30
SD N 2 Brajan	30
SD N 3 Kemudo	28
SD N 1 Kebondalem Lor	52
SD N 2 Cucukan	25
SD N 1 Joho	31
Jumlah	196

Tabel 4. Subjek Penelitian

Nama Sekolah	Jumlah
SD Muh. Tonggalan	89
SD N 1 Tegalyoso	28
SD N 1 Prambanan	39
SD N2 Prambanan	39
SD N 2 Taskombang	22
SD N Kendalsari	49
SD N 1 Bonyokan	54
SD N Gatak	51
SD N Mlese	13
SD N 1 Tegalondo	28
SD N Kradenan	54
SD N 2 Cawas	38
SD N 1 Paseban	41
Jumlah	548

## Hasil Penelitian dan Pembahasan

### Validitas dan Reliabilitas Instrumen

Validitas terdiri dari validitas isi dan konstruk sedangkan reliabilitas diestimasi dengan Alpha Cronbach dan secara konstruk. Hasil dari validitas isi menunjukkan bahwa indeks Aiken sebesar 0,85 sehingga dapat dikatakan bahwa Instrumen tersebut valid secara isi. Validitas Konstrak dibuktikan dengan analisis faktor eksploratori dan konfirmatori. Hasil dari analisis faktor eksploratori disajikan pada Tabel 5.

Berdasarkan hasil yang didapatkan berdasarkan analisis faktor eksploratori dan analisis faktor konfirmatori maka dapat disimpulkan bahwa 18 butir tersebut valid. Butir yang telah didapatkan tersebut dapat

digambarkan pada hasil analisis faktor konfirmatori. Model yang sudah memiliki kesesuaian yang baik tersebut dapat dilihat pada *standardized solution* pada Gambar 2.

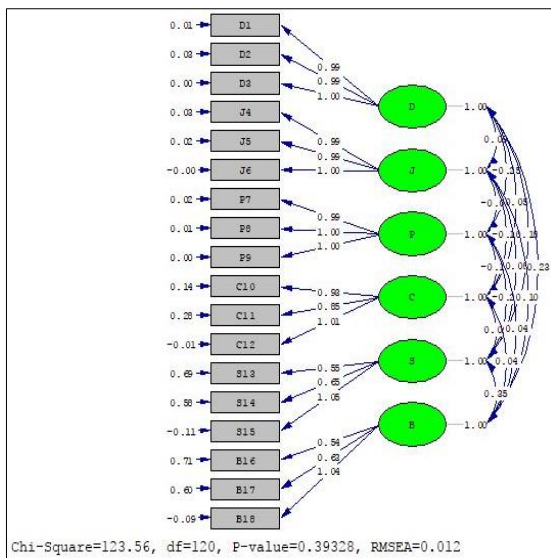
Berdasarkan gambar model yang dilihat dari *standardization solution* menunjukkan bahwa p-value 0,39 dan RSMEA yang mendekati 0 yaitu 0,012 maka dapat dikatakan bahwa model faktor yang digunakan baik atau terdapat kecocokan model. Selain pada *standardized solution*, dapat juga dilihat melalui *t value* dengan ditunjukkan oleh Gambar 3.

Tabel 5. Hasil Analisis Faktor Eksploratori

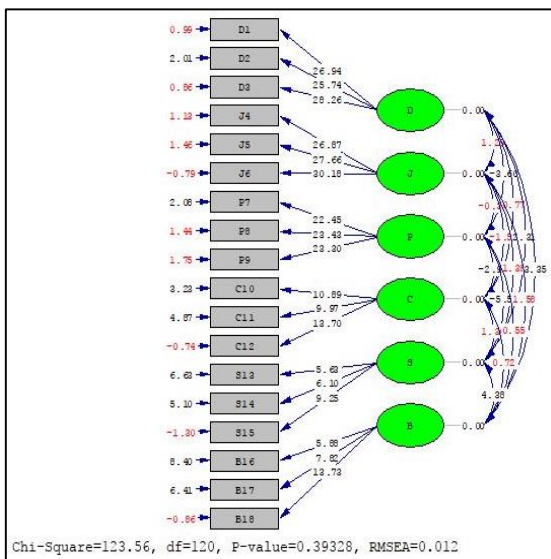
Analisis Faktor Eksploratori	Hasil
<i>Kaiser-Meyer-Olkin (KMO) sampling adequacy</i>	0,726
Bartlett's Test of Sphericity	0,000
Nilai MSA	0,779; 0,869; 0,727; 0,817; 0,734; 0,656; 0,869; 0,754; 0,688; 0,783; 0,608; 0,694; 0,655; 0,668; 0,712; 0,612; 0,663.
Nilai Eigen	4,232; 3,321; 2,494; 2,467; 2,088; 1,177
<i>Rotated Componen Matrik</i>	Disiplin: 0,98; 0,97; 0,98 Jujur: 0,98; 0,99; 0,99 Peduli: 0,97; 0,97; 0,97 Percaya Diri: 0,94; 0,91; 0,97 Santun: 0,72; 0,82; 0,86 Tanggung Jawab: 0,67; 0,85; 0,70

Tabel 6. Hasil Analisis Faktor Konfirmatori

Analisis Faktor Konfirmatori	Hasil
<i>Chi-Kuadrat</i>	0.039
RSMEA	0.012
GFI	0,93
SRMR	0,03
AGFI	0,90
CFI	1,00
RFI	1,00



Gambar 2. Basic Model Standardization Solution



Gambar 3. Basic Model T-Value

Tabel 7. Hasil Reliabilitas Berdasarkan Alpha Cronbach

Reliabilitas Alpha Cronbach	Ket.
0,694	Reliabel

Tabel 8. Hasil Reliabilitas Konstrak Berdasarkan Alpha Cronbach

Reliabilitas Konstrak		
CR	VE	Ket
0,995	0,995	Reliabel
0,994	0,995	Reliabel
0,996	0,997	Reliabel
0,949	0,951	Reliabel
0,994	0,811	Reliabel
0,993	0,798	Reliabel

Berdasarkan hasil validitas dan reliabilitas diperoleh diperoleh 18 butir dengan indikator penilaian atau kisi-kisi sebagai berikut.

Tabel 9. Kisi-Kisi Penilaian Afektif Siswa

Sikap Siswa	Indikator	No. Butir
Disiplin	Tepat waktu dalam melakukan atau mengerjakan sesuatu	1, 2
	Tertib dalam menjalankan peraturan	3
	Jujur	Menyampaikan sesuatu dengan keadaan yang sebenarnya
Peduli	Mengaku kekurangan ataupun keterbatasan	6
	Berpartisipasi dalam kegiatan	7, 8
Percaya Diri	Menunjukkan perhatian kepada orang lain ataupun lingkungan sekitar	9
	Bergantung pada kemampuan sendiri	10
Santun	Berani mengungkapkan sesuatu	11, 12
	Menghormati orang lain dengan cara berbicara atau bertutur kata halus	13, 14
Tanggung jawab	Dapat mengendalikan perilaku atau perbuatan walaupun mendapatkan kesulitan atau masalah	15
	Melaksanakan tugas dengan baik	16
	Menerima resiko dari tindakan yang dilakukan	17, 18

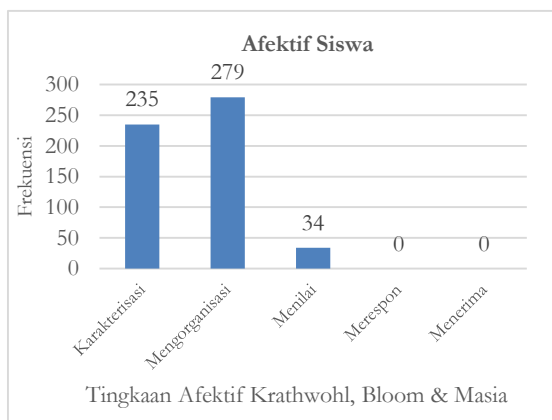
Gambaran Sikap Afektif Siswa

Gambaran afektif dalam penelitian ini disajikan dalam: (1) tingkatan afektif Krathwohl, Bloom & Masia; (2) acuan Kurikulum 2013. Gambaran afektif siswa secara keseluruhan berdasarkan tingkatan afektif Krathwohl, Bloom dan Masia disajikan dalam Tabel 10.

Tabel 10. Afektif Siswa secara Keseluruhan

Skor	Kategori	f	%
> 75,6	Karakterisasi	235	42,9
>61,2-75,6	Mengorganisasi	279	50,9
> 56,8-61,2	Menilai	34	6,2
>32,4-46,8	Merespon	0	0
8-32,4	Menerima	0	0
Jumlah		548	100

Tabel 10 menunjukkan bahwa kemampuan siswa kelas IV Kabupaten Klaten dilihat dari kategori berdasarkan tingkatan afektif Krathwohl, Bloom & Masia dari 548 siswa terdapat 235 siswa yang mencapai karakterisasi dengan persentase sebesar 42,9%, 279 siswa mencapai tingkatan mengorganisasi dengan persentase 50,9%, 34 siswa mencapai tingkatan menilai dengan persentase 6,2 % sedangkan pada tingkatan merespon dan menerima persentase 0% atau dapat dikatakan tidak ada siswa yang mencapai pada kedua tingkatan tersebut. Kemampuan afektif siswa berdasarkan tingkatan Krathwohl, Bloom & Masia ditunjukkan pada Gambar 4.



Gambar 4. Afektif Siswa secara Keseluruhan Berdasarkan Tingkatan Afektif Krathwohl, Bloom & Masia

Gambaran pada setiap sikap berdasarkan tingkatan afektif Krathwohl, Bloom dan Masia adalah sebagai berikut:

#### Disiplin

Gambaran sikap disiplin siswa berdasarkan tingkatan afektif Krathwohl, Bloom dan Masia digambarkan pada Tabel 11.

Tabel 11. Sikap Disiplin

Skor	Kategori	f	%
> 12,6	Karakterisasi	155	28
>10,2- 12,6	Mengorganisasi	274	50
> 7,8 - 10,2	Menilai	110	20
>5,4 - 7,8	Merespon	9	2
3- 5,4	Menerima	0	0
Jumlah		548	100

#### Jujur

Gambaran sikap jujur siswa berdasarkan tingkatan afektif Krathwohl, Bloom & Masia digambarkan pada Tabel 12.

Tabel 12. Sikap Jujur

Skor	Kategori	f	%
> 12,6	Karakterisasi	171	31
>10,2 - 12,6	Mengorganisasi	171	31
> 7,8 - 10,2	Menilai	173	32
>5,4 - 7,8	Merespon	28	5
3- 5,4	Menerima	5	1
Jumlah		548	100

#### Peduli

Gambaran sikap peduli siswa berdasarkan tingkatan afektif Krathwohl, Bloom dan Masia digambarkan pada Tabel 13.

Tabel 13. Sikap Peduli

Skor	Kategori	f	%
> 12,6	Karakterisasi	326	59,5
>10,2- 12,6	Mengorganisasi	121	22,1
> 7,8 - 10,2	Menilai	61	14,8
>5,4 - 7,8	Merespon	18	3,3
3- 5,4	Menerima	2	0,4
Jumlah		548	100

#### Percaya Diri

Gambaran sikap percaya diri siswa berdasarkan tingkatan afektif Krathwohl, Bloom dan Masia digambarkan pada Tabel 14.

Tabel 14. Sikap Percaya Diri

Skor	Kategori	f	%
> 12,6	Karakterisasi	289	52,7
>10,2- 12,6	Mengorganisasi	199	36,3
> 7,8 - 10,2	Menilai	57	10,4
>5,4 - 7,8	Merespon	1	0,2
3- 5,4	Menerima	2	0,4
Jumlah		548	100

*Santun*

Gambaran sikap santun siswa berdasarkan tingkatan afektif Krathwohl, Bloom dan Masia digambarkan pada Tabel 15.

Tabel 15. Sikap Peduli

Skor	Kategori	f	%
> 12,6	Karakterisasi	358	65,3
>10,2- 12,6	Mengorganisasi	119	21,7
> 7,8 - 10,2	Menilai	64	11,7
>5,4 - 7,8	Merespon	7	1,3
3- 5,4	Menerima	0	0
Jumlah		548	100

*Tanggung Jawab*

Gambaran sikap tanggung jawab siswa berdasarkan tingkatan afektif Krathwohl, Bloom dan Masia digambarkan pada Tabel 16.

Tabel 16. Sikap Peduli

Skor	Kategori	f	%
> 12,6	Karakterisasi	322	58,8
>10,2- 12,6	Mengorganisasi	129	23,5
> 7,8 - 10,2	Menilai	75	13,7
>5,4 - 7,8	Merespon	22	4,0
3- 5,4	Menerima	0	0
Jumlah		548	100

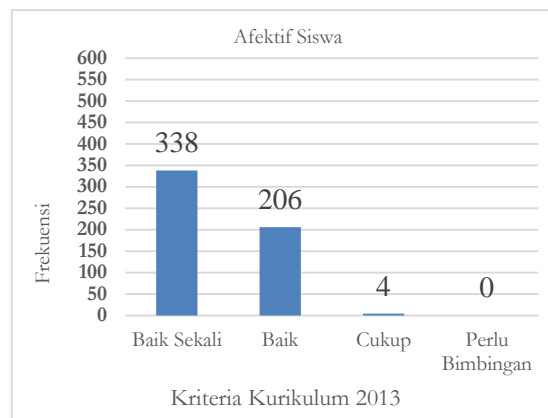
Gambaran afektif siswa secara keseluruhan berdasarkan acuan Kurikulum 2013 disajikan dalam Tabel 17.

Tabel 17. Afektif Siswa Secara Keseluruhan

Skor	Kategori	f	%
> 72	Baik Sekali	338	62
> 54-72	Baik	206	38
> 36 – 54	Cukup	4	1
18-36	Perlu Bimbingan	0	0
Jumlah		548	100

Berdasarkan Tabel 17 menunjukkan bahwa kemampuan afektif siswa kelas IV Kabupaten Klaten dilihat dari kategori Kurikulum 2013 terdapat 338 siswa yang masuk dalam kategori Baik Sekali dengan persentasenya 62%. Pada kategori Baik terdapat 206 siswa dengan persentase 38% siswa sedangkan kategori cukup menunjukkan terdapat 4 siswa yang masuk dalam kategori ter-

sebut dengan persentase 1%, pada kategori perlu bimbingan sebesar 0% yang dapat digambarkan pada Gambar 5.



Gambar 5. Afektif Siswa secara Keseluruhan Berdasarkan Acuan Kurikulum 2013

Gambaran pada setiap sikap berdasarkan acuan Kurikulum 2013 adalah sebagai berikut:

*Disiplin*

Gambaran afektif siswa pada sikap disiplin secara keseluruhan berdasarkan acuan Kurikulum 2013 disajikan dalam Tabel 18.

Tabel 18. Sikap Disiplin

Skor	Kategori	f	%
> 12	Baik Sekali	155	28,3
> 9 – 12	Baik	345	63
> 6 – 9	Cukup	46	8,3
18-36	Perlu Bimbingan	2	0,4
Jumlah		548	100

*Jujur*

Gambaran afektif siswa pada sikap jujur secara keseluruhan berdasarkan acuan Kurikulum 2013 disajikan dalam Tabel 19.

Tabel 19. Sikap Jujur

Skor	Kategori	f	%
> 12	Baik Sekali	171	31,2
> 9 – 12	Baik	229	41,8
> 6 – 9	Cukup	134	24,5
18-36	Perlu Bimbingan	14	2,6
Jumlah		548	100

*Peduli*

Gambaran afektif siswa pada sikap peduli secara keseluruhan berdasarkan acuan Kurikulum 2013 disajikan dalam Tabel 20.

Tabel 20. Sikap Peduli

Skor	Kategori	f	%
> 12	Baik Sekali	326	59,5
> 9 – 12	Baik	153	27,9
> 6 – 9	Cukup	59	10,8
18-36	Perlu Bimbingan	10	1,8
Jumlah		548	100

*Percaya Diri*

Gambaran afektif siswa pada sikap percaya diri secara keseluruhan berdasarkan acuan Kurikulum 2013 disajikan dalam Tabel 21.

Tabel 21. Percaya Diri

Skor	Kategori	f	%
> 12	Baik Sekali	289	52,7
> 9 – 12	Baik	220	40,1
> 6 – 9	Cukup	37	6,8
18-36	Perlu Bimbingan	2	0,4
Jumlah		548	100

*Santun*

Gambaran afektif siswa pada sikap santun secara keseluruhan berdasarkan acuan Kurikulum 2013 disajikan dalam Tabel 22.

Tabel 22. Sikap Santun

Skor	Kategori	f	%
> 12	Baik Sekali	358	65,3
> 9 – 12	Baik	145	26,5
> 6 – 9	Cukup	42	7,7
18-36	Perlu Bimbingan	3	0,5
Jumlah		548	100

*Tanggung Jawab*

Tabel 23. Sikap Disiplin

Skor	Kategori	f	%
> 12	Baik Sekali	323	58,9
> 9 – 12	Baik	170	31
> 6 – 9	Cukup	48	8,8
18-36	Perlu Bimbingan	7	1,3
Jumlah		548	100

Gambaran afektif siswa pada sikap tanggung jawab secara keseluruhan berdasarkan acuan Kurikulum 2013 disajikan dalam Tabel 23.

**Simpulan**

Berdasarkan hasil penelitian maka dapat disimpulkan bahwa (1) instrumen penilaian untuk mengukur afektif siswa kelas IV berjumlah 18 butir telah memenuhi syarat yaitu validitas dan reliabilitas; (2) Gambaran afektif siswa adalah (a) berdasarkan tingkatan afektif Krathwohl, Bloom & Masia pada tingkatan karakterisasi sebesar 42,9%, mengorganisasi sebesar 50,9 %, menilai 6,2% sedangkan tahapan merepon dan menerima sebesar 0%, (b) berdasarkan acuan Kurikulum 2013 pada kategori baik sekali sebesar 62%, kategori baik 38% dan kategori cukup 1%, kategori butuh bimbingan sebesar 0%.

Saran yang dapat diberikan dalam pemanfaatan produk adalah sebagai berikut: (1) pada proses pembelajaran sebaiknya guru merancang pembelajaran yang mengasah afektif siswa; (2) instrumen ini dapat digunakan untuk mengeeahui sejauhmana tingkatan yang telah dicapai siswa berdasarkan tingkatan afektif Krathwohl, Bloom & Masia yaitu karakterisasi mencapai, mengorganisasi, menilai, merespon dan menerima; (3) bagi guru-guru yang akan mengetahui kemampuan afektif siswa berdasarkan tigtkatan Krathwohl, Bloom & Masia dapat menggunakan instrumen ini dikarenakan telah layak, (4) bagi guru-guru yang ingin mengembangkan instrumen penilaian afektif siswa pada jenjang yang lain diharapkan dapat mengikuti langkah-langkah yang tersedia dalam penelitian ini.

**Daftar Pustaka**

Adibatin, A. (2016). Pendidikan karakter bangsa berbasis strategi pembelajaran pakem melalui permainan cincin di jempol tangan (karya inovasi pembelajaran sekolah dasar). *Scholaria*, 6(1). Retrieved from <http://ejournal.uksw.edu/scholaria/article/download/180/168>



- Haryadi, T., & Aripin, A. (2015). Melatih kecerdasan kognitif, afektif, dan psikomotorik anak sekolah dasar melalui perancangan game simulasi “warungku.” *Andharupa; Jurnal Desain Komunikasi Visual & Multimedia*, 1(2). Retrieved from <https://publikasi.dinus.ac.id/index.php/andharupa/article/view/963>
- Hendryadi, & Suryani. (2014). *Structural equation modeling dengan lisrel 8.80 pedoman untuk pemula*. Yogyakarta: Kaukaba Dipantara.
- Kementerian Pendidikan dan Kebudayaan melalui Direktorat Pembinaan Pendidikan Dasar. (2016). *Buku panduan penilaian sekolah dasar*. Jakarta: Kementrian Pendidikan dan Kebudayaan melalui Direktorat Pembinaan Pendidikan Dasar.
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1973). *Taxonomy of educational objectives*. London: David McKay Company.
- Mardapi, D. (2012). *Teknik penyusunan instrumen tes dan non tes*. Yogyakarta: Mitra Cendekia.
- McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). *Instrument development in the affective domain*. New York: Springer New York. <https://doi.org/10.1007/978-1-4614-7135-6>
- Menteri Pendidikan dan Kebudayaan Republik Indonesia. Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 23 Tahun 2013 tentang Standar Penilaian (2016).
- Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 23 Tahun 2015 tentang Penumbuhan Budi Pekerti (2015).
- Samani, H., & Hariyanto. (2012). *Pendidikan karakter*. Bandung: PT Remaja Rosdakarya.
- Setyawan, H., & Dimiyati, D. (2015). Model permainan aktivitas luar kelas untuk mengembangkan ranah kognitif, afektif, dan psikomotorik siswa SMA. *Jurnal Keolahragaan*, 3(2). Retrieved from <https://journal.uny.ac.id/index.php/jolahraga/article/view/6230>
- Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). *Instructional development for training teachers of exceptional children*. Bloomington: Indiana University.
- Wagiran. (2015). *Metodologi penelitian pendidikan*. Yogyakarta: Budi Utama.
- Wening, S. (2012). Pendidikan karakter bangsa melalui pendidikan nilai. *Jurnal Pendidikan Karakter*, 1(12). Retrieved from <https://journal.uny.ac.id/index.php/jpka/article/view/1452>