REiD (RESEARCH AND EVALUATION IN EDUCATION)
Vol. 3, No. 1, June 2017

Indexed in:

Research and Evaluation in Education

Vol. 3, No. 1, June 2017

Vol. 3, No. 1, June 2017 ISSN 2460-6995

Research and Evaluation in Education

Publisher:
PROGRAM PASCASARJANA
UNIVERSITAS NEGERI YOGYAKARTA

REiD (Research and Evaluation in Education)
ISSN 2460-6995

Publisher
Program Pascasarjana Universitas Negeri Yogyakarta

Published biannually, in June and December
REiD disseminates articles written based on the results of research focusing on assessment, measurement, and evaluation in various educational areas

Editorial
Department of Educational Research and Evaluation, Graduate School of Yogyakarta State University
3rd Floor Pascasarjana UNY New Building, Colombo Street No. 1, Karangmalang, Yogyakarta 55281
Telephone: 0274 586168 ext. 229 or 0274 550836, Facsimile: 0274 520326
E-mail: reid.ppsuny@uny.ac.id, reid.ppsuny@gmail.com

Foreword

We are very pleased that REiD (Research and Evaluation in Education) is releasing its fourth edition. We are also very excited that the journal has been attracting papers from foreign country such as the United States. The variety of submissions from different countries will help the journal in reaching its aim in becoming a global initiative.

REiD (Research and Evaluation in Education) contains and spreads out the results of research which is not limited to the area of educational evaluation, but also comprises the results of research in education in a broader coverage, such as social science, science education, language education, educational quality, teacher competence, and academic performance, with focuses on assessment and evaluation.

The editorial board expects comments and suggestions for the betterment of the future editions of the journal. Special gratitude goes to the reviewers of the journal for their hard work, contributors for their trust, patience, and timely revisions, and all staffs of the Graduate School of Universitas Negeri Yogyakarta for their assistance in publishing this journal.

Yogyakarta, June 2017

Editor in Chief

TABLE OF CONTENT

REiD (Research and Evaluation in Education), 3(1), 2017, 1-11
Available online at: http://journal.uny.ac.id/index.php/reid
Research Article

Mapping elementary school students' creativity in science process skills of life aspects viewed from their divergent thinking patterns

***[1]Bambang Subali; [2]Paidi; [3]Siti Mariyam**
*Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta
Jl. Colombo No. 1, Depok, Sleman 55281, Yogyakarta, Indonesia
*Email: b_subali@yahoo.co.id

Submitted: 10 March 2017 | Revised: 11 July 2017 | Accepted: 11 July 2017

Abstract

The purpose of this study was to map elementary school students' creativity in science process skills (SPS) of life aspects in science subjects viewed from their divergent thinking patterns using written tests whose items were fitted with Partial Credit Model (PCM). The measurement used a test validated using the IRT approach published in JEE journal in 2015. The trials employed four sets of test, each comprising 20 items completed with anchor items which were fitted referring to PCM. The measurements were performed with larger scale on 14 regional technical implementation unit (RTIU) in Yogyakarta Special Province in five regencies/cities to students of grades IV, V, and VI. The findings show that the higher the grade level, the higher of the testees' scores would be. There were some testees who did not have divergent thinking ability and they obtained a score of 0 The divergent thinking ability of the students was not related to the regency/city where an RTIU was located.

Keywords: *creativity, divergent thinking, science process skills, partial credit models*

How to cite item:

Subali, B., Paidi, P., & Mariyam, S. (2017). Mapping elementary school students' creativity in science process skills of life aspects viewed from their divergent thinking patterns. *REiD (Research and Evaluation in Education), 3*(1), 1-11. doi:http://dx.doi.org/10.21831/reid.v3i1.13294

Introduction

The core of teaching natural sciences is to teach the students to investigate natural phenomena to look for a scientific product by experiencing a scientific process with reference to scientific attitude (Carin & Sund, 1989). A scientific process involves aspects of science process skills. A scientific process arranged in a particular order is called scientific method (Towle, 1989). The teaching that can enhance learners to master every aspect of the science process skills is badly needed in order that they can master the scientific process. The science process skills should be taught to students partially at the beginning. After mastering the aspects of science process skills, they are taught the science process skills as a unit of scientific method.

Science Process Skills

According to Rezba et al. (2007), science process skills can be divided into two aspects, namely basic skills and integrative skills. Basic skills include observing, communicating, classifying, measuring metrically, inferring, and predicting. Meanwhile, integrative skillls consist of identifying variables, constructing a table of data, constructing a graph, describing relationships between variables, acquiring and processing data, analyzing investigations, constructing hypotheses, defining variables operationally, designing experiments, and experimenting.

Unlike Rezba et.al., Bryce, McCall, MacGregor, Robertson, and Weston (1990) divide science process skills into three aspects, namely basic skills, process skills, and investigative skills. Basic skills comprise of observational skills, recording skills, measurement skills, manipulative skills, procedural skills, and following instruction skills. While process skills are skills of inference and selection of procedures. Furthermore, investigative skills include skills to make plan and carry out a practical investigation.

In reference to The American Association for the Advancement of Science in 1965 (Chiappetta, 1997), science process skills are categorized into two, namely basic skills and integrated skills. Basic skills are skills of observing, classifying space time relations, using numbers, measuring, inferring, and predicting. Furthermore, integrated skills include such skills as defining, formulating models, controlling variables, interpreting data, hypothesizing, and experimenting.

Wenning (2005) says that science process skills can be classified into rudimentary skills, basic skills, intermediate skills, integrated skills, and advanced skills. Rudimentary skills comprise of observing, collecting and recording data; drawing conclusions; communicating and classifying results; measuring metrically; estimating; decision making 1; explaining; and predicting. Basic skills are skills of identifying variables, constructing a table of data, constructing a graph, describing relationships between variables, acquiring and processing data, analyzing investigations, defining variables operationally, designing investigations, experimenting, hypothesizing, decision making 2, developing models, and also controlling variables. Integrated skills include skills of identifying problems to investigate, designing and conducting scientific investigations, using technology and mathematics during investigations, generating principles through the process of induction, and communicating and defending a scientific argument. Advanced skills may consist of solving complex real world problems, synthesizing complex hypothetical explanations, establishing empirical laws on the basis of evidence and logic, analyzing and evaluating scientific arguments, constructing logical proofs, generating predictions through the process of deduction.

In 2010, Wenning (2010) revises the formulation of science process skills by adding one new skill namely culminating skills. Also, the revision includes the elaboration of each existing skill. The formulation of science process skills according to Wenning (2010) consists of rudimentary skills, basic skills, intermediate skills, integrated skills, culminating skills, and advanced skills. Rudimentary skills are skills of observing, formulating concepts, estimating, drawing conclusions, communicating results, and classifying results. Moreover, basic skills include predicting, explaining, estimating, acquiring and processing data, formulating and revising scientific explanations using logic and evidence, recognizing and analyzing alterative explanations and models. Meanwhile, intermediate skills comprise measuring, collecting and recording data, constructing a table of data, designing and conducting scientific investigations, using technology and math during investigations, and describing relationships.

The second aspect is integrated skills which include measuring metrically, establishing empirical laws on the basis of evidence and logic, designing and conducting scientific investigations, using technology and math during investigations. Furthermore, culminating skills comprise collecting, assessing, and interpreting data from a variety of sources, constructing logical arguments based on scientific evidence, making and defending evidence-based decisions and judgments, clarifying values in relation to natural and civil rights, and practicing interpersonal skills. Advanced skills are skills of synthesizing complex hypothetical explanations, analyzing and evaluating scientific arguments, generating predictions through the process of deduction, revising hypotheses and predictions in light of new evidence, and solving complex real-world problems.

Creativity and Divergent Thinking

Solving problems to find new products through scientific method is a process of inquiry. According to Mayer (1980), all science

is inquiry. Biology is one kind of science. Biologists try to answer questions about living things.

Finding new products is a creative work. Creative thinking belongs to the high cognitive level in Bloom's taxonomy referring to Anderson, Krathwohl, and Bloom (2001) and Dettmer (2005). This tells that creativity can be taught to elementary school students. Meanwhile, Miller (2008) states that something that is not duplicated/imitated is categorized as creative.

In addition, Rule, Schneider, Tallakson, and Highnam (2012) who have quoted several sources state that elementary and middle school students who are high-achieving in science and who exhibit creativity are often not challenged or given the opportunity to fully utilize their abilities in regular classrooms. Many gifted students drop out because school is boring, repetitious, and lacks relevance to real life. They expect more exciting and challenging learning processes. Unfortunately, many classroom teachers lack sufficient background knowledge to design stimulating, advanced science projects for these students; some avoid science altogether.

In reference to Csikszentmihalyi's model of creativity (Peppler & Solomou, 2011), individuals build on culturally valued practices and design to produce new variations of the domain, which, if deemed valuable by the community (i.e. the field), becomes part of what constitutes the evolving domain. Each component of the system continues to influence one another over time.

The rethinking of design for knowledge sharing is an important part of creating new work processes and has to evolve hand in hand with space planning (Mitchell, Inouye, & Blumenthal, 2003). Hadzigeorgiou, Fokialis, and Kabouropoulou (2012) cited the opinion of Barrow (2010) that the inquiry in science will be able to develop students' creativity if there is an imaginative and divergent thinking process.

Measuring Creativity and Divergent Thinking Skills

Students' mastery of creativity should be measured. According to Kelly (2004), the existing research on creativity aims at measuring the divergent thinking as proposed by Torrance and creative personality developed by Gough. There is only little research which measures creativity as a multidimensional phenomena using self report scales which are valid and easy to administer.

The main problem in measuring creativity is ensuring what is measured is really creativity and is not affected by the measurement of intelligence (Cramond, 1994). Many studies regarding the strategies to measure creative thinking ability are compiled by Kind and Kind (2007). A detailed explanation about creativity tests which include a test to measure divergent thinking process is presented by Cropley (2000). Viewed from how to measure creativity, there are many ways and aspects that are measured. For instance, one of the strategies to measure the ability of divergent thinking can be classified based on the content and the products as reported by Meeker (1969).

Olivant (2009) says that according to Guilford (1950), creativity could and should be studied in non-eminent, 'everyday' people using psychometrics such as divergent thinking tasks (or paper and pencil tasks) to measure creative thinking. Torrance's Tests of Creative Thinking was created by Torrance (1979), and they are probably the best-known and most widely used creativity psychometric instruments. Sternberg and Lubart (1999) in Torrance (1979) state that many researchers viewed the tests as trivial and inadequate measures, while others charged that the tests, while possibly measuring aspects of creativity, failed to capture their essence.

The context dependency of creativity among students has been elaborated by Diakidoy & Constantinou in 2000-2001 (Kind & Kind, 2007) by getting as many responses as possible from three open ended assignment forms and scored based on divergent thinking skills of Guilford, namely: (a) fluency, i.e. the considerations in a given solution, (b) flexibility, i.e different types of solutions.

The science process skill measurement on different thinking aspects in siology subject of senior high school students in DIY and Central Java was performed by Subali

(2009). In this case, the standardization of instrument utilized Item Response Theory or the IRT approach. This approach creates a calibration that puts learners' ability and item difficulty on the same scale. Therefore, they can be compared. The results show that the average score of creativity ability is much lower than the item difficulty index of the item to measure creativity.

Subali (2011) also measured high school students' creativity in the science process skills in biology subjects. The results are also relatively low. Subali and Mariyam (2013) have conducted a research concerning the development of science process skill creativity related to the aspects of life on science subjects that has been done by elementary school teachers. Most of the teachers stated that creativity had been taught to students in science subjects. However, the student's mastery on creativity has not been studied. Therefore, the creativity mastery of elementary school students on life aspects viewed from divergent thinking skills need to be investigated.

This research aimed at measuring the students' creativity in science process skills of life aspects viewed from divergent thinking patterns consisting of two aspects namely basic and process skills. The basic skill aspects have been published in the Journal of Asia-Pacific Forum on Science Learning and Teaching, Volume 17, Issue 1, Article 2 (Jun., 2016) (Subali, Paidi, & Mariyam, 2016).

The research aimed to map the creativity in science process skills (SPS) of life aspects of elementary school students in sciences subjects viewed from the divergent thinking pattern using written tests skills of which test items are fitted based on Partial Credit Model (PCM).

Method

The research was conducted for three years and consisted of three stages. The first stage is divided into two phases. In the first phase, the blue print of Science Process Skills (SPS) is developed. The blue print of SPS is formulated based on SPS blue prints produced by the research conducted by Subali (2009) used for measuring divergent thinking ability of SPS in biology subjects for senior high school students. In addition, the blue print is developed referring to several sources such as Rezba et al. (2007), Bryce et al. (1990), and Cox (1958). The SPS aspects include (a) basic skills and (b) process skills. This is considered as the difficulty to teach investigative skills to students of grades IV and V.

On the second phase, based on the blue print of SPS, creativity tests for SPS consisting of 63 items are developed. All items were judged by experts consisting of three lecturers --- all holding doctoral degrees --- of Biology Education Department. Using the divergent scoring model of Diakidoy and Constantinou (Kind & Kind, 2007), the items were tested to 637 students of grades V and VI. The report of the instruments in this research were validated using the IRT approach in 2015 and was published in *Journal of Elementary Education* (JEE) Vol.25, No. 1 pp. 91-105 by Subali and Mariyam (2015). Based on the IRT approach, an item is declared to be able to measure the ability if it is fitted with the the model, in this case 1-PL (Rasch model). If all items are fitted with the model, the instruments can also be declared as valid (Wright & Masters, 1982). The testing of fitted items on Rasch Model was carried out using the Quest program (Adams & Khoo, 1993). On the third phase, the instrument was administered in large scale from elementary schools in Regional Technical Implementation Unit (RTIU) in DIY (Yogyakarta Special Province). The sample was established by using the purposive sampling technique by considering the characteristics of RTIU and school achievement through national examination. The sample testees were taken from 10 RTIUs in five regencies/cities in DIY. Two RTIUs from each regency/city were selected purposively. One of the RTIUs was located in the national capital and another was located far from the national capital, except for the RTIU in the city of Yogyakarta because both were in the city center. Moreover, two private elementary schools and four public elementary schools from each RTIU were selected. The test participants included students of grades IV, V, and VI. There were 2,563 testees of grade IV, 2,685 testees of grade V, and 2,619 from grade VI.

Findings and Discussion

Findings

After the instrument was administrated to elementary school students of grades IV, V, and VI, the findings of the research are presented as follows. Table 1 shows that there is a reasonable increase of scores performed by elementary school students of grades IV to VI. This means that the higher the grade level, the greater the creativity score in Science Process Skill on life aspects mastered by the students will be. Compared to the fact that the total score which was achieved by grade 6 was 40 with the average score of 20, it can be said that the achievement of the average score of 18.5 with the minimum score of 0 and maximum score of 38 is still relatively low, seen from the aspect of competence mastery.

Table 2 shows that the highest ranking of the creativity scores of divergent thinking model on Science Process Skills aspects of the fourth grade students in the five regencies/ cities is achieved by Sleman Regency. The score is higher than Yogyakarta City score as its capital. Meanwhile, the lowest score is achieved by Kulonprogo Regency. This indicates that the test results are not related to the characteristics of city or non-city regions.

Table 3 shows that the highest score of creativity in Science Process Skill on life aspects of the fifth grade elementary school students in five regencies/cities is achieved by the city of Yogyakarta, and the lowest score is Bantul Regency. This situation is different from that in the fourth grade.

Table 1. Creativity scores based on divergent thinking model of science process skill on life aspects in Natural Sciences subjects based on grades in DIY province.

Grade	N	Score				
		ȳ	S	Min	Max	Total
Grade IV	2563	12.8	6.7	0	37	40
Grade V	2685	15.3	6.5	0	36	40
Grade VI	2619	18.5	6.4	0	38	40

Table 2. Creativity scores based on divergent thinking model of science process skill on life aspects in Natural Sciences subjects of the fourth grade students based on types of locations in DIY province

Grade IV	N	Score				
		ȳ	S	Min	Max	Total
Yogyakarta	553	12.6	7.3	0	36	40
Bantul	593	12.6	6.7	0	37	40
Sleman	605	14.0	6.5	0	33	40
Kulonprogo	380	11.4	6.0	0	29	40
Gunungkidul	432	12.9	6.7	0	31	40

Table 3. Creativity scores based on divergent thinking model of science process skill on life aspects in Natural Sciences subjects of the fifth grade students based on types of locations in DIY province

Grade V	N	Score				
		ȳ	S	Min	Max	Total
Yogyakarta	534	16.9	6.6	0	36	40
Bantul	632	13.5	6.3	0	34	40
Sleman	688	16.4	6.3	1	35	40
Kulonprogo	361	14.2	6.0	0	32	40
Gunungkidul	470	15.2	6.5	0	36	40

Table 4. Creativity scores based on divergent thinking model of science process skill on life aspects in Natural Sciences subjects of the sixth grade students based on types of locations in DIY province

Grade VI	N	Score				
		ȳ	S	Min	Max	Total
Yogyakarta	571	18.72	5.90	0	36	40
Bantul	603	17.58	6.19	0	33	40
Sleman	620	19.49	7.02	0	38	40
Kulonprogo	335	18.15	6.04	2	34	40
Gunungkidul	490	18.10	6.22	0	34	40

Table 5. The mean scores and standard deviation of science process skills creativity on life aspects in Natural Sciences subject based on the types of RTIUs of the fourth grade students in DIY province

Grade IV	N	Score				
		ȳ	s	Min	Max	Total
Yogyakarta						
East Yogyakarta	134	23.2	17.5	0	88	120
West Yogyakarta	419	34.5	18.8	0	88	120
Bantul						
Bantul Selatan	140	37.0	20.1	2	100	120
Banguntapan	240	33.4	18.3	0	84	120
Piyungan	213	25.6	14.9	0	65	120
Sleman						
Sleman	182	37.7	18.3	5	84	120
Kalasan	256	34.8	16.7	0	86	120
Ngemplak	167	32.7	16.7	2	88	120
Kulonprogo						
Pengasih	105	32.2	16.3	4	73	120
Kalibawang	127	26.5	16.6	0	71	120
Sentolo	148	27.0	14.1	0	70	120
Gunungkidul						
Wonosari	196	37.6	17.1	3	73	120
Panggang	130	24.0	16.8	0	82	120
Purwosari	106	32.9	15.5	0	71	120

Table 4 shows that the highest score of Science Process Skills creativity of life aspects of the sixth grade students in five regencies/ cities is achieved by Sleman Regency, and the lowest is achieved by Kulon Progo Regency.

The followings are the results of the creativity measurement of Science Process Skill on life aspects in RTIUs of each regency/city ranging from grade IV to VI. Table 5 presents the results of measurements on the fourth grade.

Table 5 shows that the highest score of Science Process Skills creativity of life aspects of the fourth grade students in the five regencies/cities is achieved by the RTIU in Sleman Regency. It is followed by Wonosari RTIU in Gunungkidul Regency, and South Bantul RTIU in Bantul Regency. While the low score that ranks XII is achieved by Piyungan RTIU in Bantul Regency, rank XIII is achieved by RTIU Panggang Gunungkidul and rank XIV is achieved by RTIU of East Yogyakarta. This may imply that the mastery of Science Process Skill creativity of life aspects on the fourth grade students is not dominated by students of the elementary school located in the capital of the province.

Table 6 presents the results of measurements of the grade V. It shows that the highest score of Science Process Skills creativity of

life aspects of the fifth graders of elementary school in five regencies/cities in Yogyakarta RTIUs is achieved by the North Yogyakarta. Sleman Regency comes second and the third rank is achieved by RTIU of Ngemplak in Sleman Regency. While the low ranks, i.e rank XII is achieved by Pengasih RTIU in Kulon Progo Regency, rank XIII is achieved by Banguntapan RTIU in Bantul Regency and rank XIV is achieved by Piyungan RTIU in Bantul Regency.

Table 6. The average scores and creativity standard deviation of science process skill of life aspects in Natural Sciences subject based on the types of RTIU of gade V students in DIY

Grade V	N	Score				
		ȳ	S	Min	Max	Total
Yogyakarta						
East Yogyakarta	122	35.8	18.4	0	88	120
West Yogyakarta	412	45.5	17.4	0	98	120
Bantul						
Bantul Selatan	135	40.3	18.0	4	89	120
Banguntapan	250	33.1	17.3	0	86	120
Piyungan	247	31.3	15.4	1	72	120
Sleman						
Sleman	180	42.9	17.3	7	87	120
Kalasan	297	40.4	18.0	3	93	120
Ngemplak	211	41.6	17.0	6	94	120
Kulonprogo						
Pengasih	111	34.2	14.6	0	69	120
Kalibawang	117	35.4	19.9	0	86	120
Sentolo	133	36,0	12.9	10	81	120
Gunungkidul						
Wonosari	227	40.9	16.8	3	89	120
Panggang	131	34.5	16.3	0	78	120
Purwosari	112	37.3	20.3	0	99	120

Table 7. The average scores and creativity standard deviation of science process skill of life aspects in Natural Sciences subject based on the types of RTIU of grade VI students in DIY

Grade VI	N	Score				
		ȳ	S	Min	Max	Total
Yogyakarta						
East Yogyakarta	149	43.1	15.8	8	84	120
West Yogyakarta	422	48.7	16.2	0	98	120
Bantul						
Bantul Selatan	127	47.2	17.8	10	93	120
Banguntapan	256	43.6	18.7	0	86	120
Piyungan	220	43.7	14.4	9	78	120
Sleman						
Sleman	162	53.6	21.9	10	101	120
Kalasan	277	47.3	19.4	3	108	120
Ngemplak	181	49.4	18.2	0	103	120
Kulonprogo						
Pengasih	102	40.5	15.4	11	75	120
Kalibawang	102	52.5	17.1	16	92	120
Sentolo	131	44.2	15.4	4	84	120
Gunungkidul						
Wonosari	207	49.4	16.2	4	88	120
Panggang	138	43.6	18.2	0	90	120
Purwosari	145	41.9	15.9	3	85	120

The results of the measurements on grade VI is presented in Table 7 which shows that the highest score of Science Process Skills creativity of life aspects of the sixth grade of elementary school students in five regencies/cities in Yogyakarta RTIUs is achieved by Sleman RTIU in Sleman Regency, rank II by Kalibawang RTIU in Kulon Progo Regency, and rank III by Wonosari RTIU in Gunung Kidul Regency. While the low ranks, i.e. rank XII is achieved by East Yogyakarta, rank XIII is achieved by Purwosari RTIU in Gunungkidul Regency and rank XIV is achieved by Pengasih RTIU of Kulon Progo Regency.

Discussion

The results of the research show that the average creativity ability of SPS on life aspects of the elementary school students of grades IV, V, dan VI in 14 RTIUs is low. On the contrary, based on the research conducted by Subali and Mariyam (2013), most teachers said that they had taught creativity to the students. This is probably because the teachers do not know well how to develop student's creativity. According to Dettmer (2005, pp. 70–78), creativity learning ideally must use an applied learning and an ideational learning model. In addition, teachers could encourage the students to be creative by giving examples on how to (a) substitute/replace, (b) combine, (c) adapt, (d) modify, add, (e) put something for another use, (f) eliminate or reduce and (g) reconstruct or reverse (Michalko, 2000).

Another reason is that the target of the teaching focuses on concept understanding. Therefore, creativity is not the main teaching target. Whereas, according to Burke-Adams (2007), it is very important to consider the learning needs of talented students in integrating creativity into a standard-based system.

Teachers are not aware that the goal of creativity development in natural sciences teaching is to direct the students to perform opened-discovery or inquiry or do the relevant tasks. Meanwhile, teachers are supposed to develop student's thinking in order that they can perform logical thinking creatively (Kind & Kind, 2007, pp. 1–37). Teachers concentrate more on developing students in order that they can understand the concept and automatically develop their convergent thinking skills. Teachers will rarely give questions with divergent answers (Croom & Stair, 2005).

The teacher's worry regarding not to teach creativity to low academic potential students may not happen. The research findings of Ferrando, Prieto, Ferrandiz, and Sanchez (2005) tell that smart students are not always creative. Moreover, Cromie (2003) says that not all studies tell a correlation between students' IQ and creativity. In addition, Rawat, Qazi, and Hamid (2012) state that the development of creativity is closely linked to the development of skills to form a corresponding consideration in different situations. With regard to this, teachers should develop students' creativity as early as possible.

The findings indicate that there are two possibilities why elementary schools located in a big city are not always showing the highest scores. The first possibility is that the children are not potential. Thus, although the teachers develop creativity, the result may not be optimal. The second possibility is that children are assessed by their parents to be potentials so that they ask their children to go to elementary schools that are good based on the society assesment. For example, Ungaran, Serayu, and Muhammadiyah Sapen Elementary Schools. However, the score of RTIU in North Yogyakarta is not always the highest. Elementary schools in the area of RTIU Sleman are partially assessed by local people to be good schools and rank top. However, elementary schools in Kalibawang RTIU for grade VI rank second out of 14 RTIUs eventhough Kalibawang RTIU is located in the remote areas. Therefore, it seems that the teacher's role in developing the creativity of learners may not be optimal. Moreover, students in the sixth grade of elementary schools at cities probably more focus on the achievement of the high score national achievement in order to be received at junior high schools that are assessed good by the community based on the achievement on national examination.

Conclusion and Suggestions

Based on the findings of the research, it can be concluded that a measuring instrument for Science Process Skills creativity of life aspects produced and tested in 2015 is relatively low. Recommendations are necessary to improve the ability of teachers in teaching Science Process Skill creativity of life aspects to students. The findings indicate that elementary schools located in remote areas RTIU may achieve high score probably because elementary school teachers in the city are more focusing on developing students to reach a high score of UN. It is worth exploring further using eksposfacto retrospective approach.

Acknowledgement

The deepest gratitude is addressed to the Directorate of Research and Community Service of the Ministry of Research, Technology, and Higher Education which has sponsored this research so that it can be carried out.

References

Adams, R. J., & Khoo, S.-T. (1993). *Acer Quest version 2.1*. Camberwell, Victoria: Australian Council for Educational Research.

Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York, NY: Longman. https://doi.org/10.1207/s15430421tip4104_2

Bryce, T. G. K., McCall, J., MacGregor, J., Robertson, I. J., & Weston, R. A. J. (1990). *Techniques for assessing process skills in practical science: Teacher's guide*. Oxford: Heinemann Educational Books.

Burke-Adams, A. (2007). The benefits of equalizing standards and creativity: Discovering a balance in instruction. *Gifted Child Today*, *30*(1), 58–63.

Carin, A. A., & Sund, R. B. (1989). *Teaching science through discovery*. Columbus, OH: Merrill Publishing Company. Retrieved from https://books.google.co.id/books?id=HopLAAAAYAAJ&source=gbs_book_other_versions

Chiappetta, E. L. (1997). Inquiry-based science: Strategies and techniques for encouraging inquiry in the classroom. *Science Teacher*, *64*(10), 22–26.

Cox, D. R. (1958). *Planning of experiments*. New York, NY: John Wiley & Sons.

Cramond, B. (1994). We can trust creativity tests. *Educational Leadership*, *52*(2), 70. Retrieved from http://ezproxy.lib.ucalgary.ca:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=ehh&AN=9411032030&site=ehost-live

Cromie, W. J. (2003). Creativity tied to mental illness: Irrelevance can make you mad. *Harvard Gazette*. Retrieved from https://news.harvard.edu/gazette/story/2003/10/creativity-tied-to-mental-illness/

Croom, B., & Stair, K. (2005). Getting from Q to A: Effective questioning for effective learning. *The Agricultural Education Magazine*, *78*(1), 12–14. Retrieved from http://search.proquest.com/docview/224994858?accountid=8579

Cropley, A. J. (2000). Defining and measuring creativity: Are creativity tests worth using? *Roeper Review*, *23*(2), 72–79. https://doi.org/10.1080/02783190009554069

Dettmer, P. (2005). New blooms in established fields: Four domains of learning and doing. *Roeper Review*, *28*(2), 70–78. https://doi.org/10.1080/02783190609554341

Ferrando, M., Prieto, M. D., Ferrandiz, C., & Sanchez, C. (2005). Intelligence and creativity. *Electronic Journal of Research in Educational Psychology*, *3*(3), 21–50. Retrieved from http://www.investigacion-psicopedagogica.org/revista/new/english/ContadorArticulo.php?101

Hadzigeorgiou, Y., Fokialis, P., & Kabouropoulou, M. (2012). Thinking about creativity in science education. *Creative Education*, *3*(5), 603–611. https://doi.org/10.4236/ce.2012.35089

Kelly, K. E. (2004). A brief measure of creativity among college students. *College Student Journal*, *38*(4), 594–596.

Kind, P. M., & Kind, V. (2007). Creativity in science education: Perspectives and challenges for developing school science. *Studies in Science Education*, *43*(1), 1–37. http://dx.doi.org/10.1080/03057260708560225

Mayer, W. V (Ed.). (1980). *Biological science: An inquiry into life* (4th ed.). Denver, CO: Biological Sciences Curriculum Study.

Meeker, M. N. (1969). *The structure of intellect: Its interpretation and uses*. Columbus, OH: Merrill Publishing Company.

Michalko, M. (2000). Four steps toward creative thinking. *The Futurist*, 18–21. Retrieved from https://www.questia.com/read/1G1-62026052/four-steps-toward-creative-thinking

Miller, P. W. (2008). *Measurement and teaching*. Munster, IN: Patrick W. Miller & Associates.

Mitchell, W. J., Inouye, A. S., & Blumenthal, M. S. (Eds.). (2003). *Beyond productivity: Information, technology, innovation, and creativity* (Committee). Washington, DC: The National Academies Press. Retrieved from https://www.nap.edu/read/10671/chapter/1#ii

Olivant, K. F. (2009). *An interview study of teachers' perceptions of the role of creativity in a high-stakes testing environment*. Fresno, CA: California State University.

Peppler, K. A., & Solomou, M. (2011). Building creativity: Collaborative learning and creativity in social media environments. *On the Horizon*, *19*(1), 13–23. https://doi.org/10.1108/10748121111107672

Rawat, K. J., Qazi, W., & Hamid, S. (2012). Creativity and education. *Academic Research International*, *2*(2), 264–275.

Rezba, R. J., Sparague, C. S., Fiel, R. L., Funk, H. J., Okey, J. R., & Jaus, H. H. (2007). *Learning and assessing science process skills* (3rd ed.). Dubuque, IA: Kendall/Hunt.

Rule, A. C., Schneider, J. S., Tallakson, D. A., & Highnam, D. (2012). Creativity and thinking skills integrated into a science enrichment unit on flooding. *Creative Education*, *3*(8), 1371–1379. https://doi.org/10.4236/ce.2012.38200

Subali, B. (2009). *Pengukuran keterampilan proses sains pola divergen dalam mata pelajaran biologi SMA di Provinsi DIY dan Jawa Tengah*. Universitas Negeri Yogyakarta. Retrieved from http://eprints.uny.ac.id/4541/

Subali, B. (2011). Pengukuran kreativitas keterampilan proses sains dalam konteks assessment for learning. *Cakrawala Pendidikan*, *30*(1), 130–144. https://doi.org/10.21831/cp.v1i1.4196

Subali, B., & Mariyam, S. (2013). Pengembangan kreativitas keterampilan proses sains dalam aspek kehidupan organisme pada mata pelajaran IPA SD. *Cakrawala Pendidikan*, *32*(3), 365–381. https://doi.org/10.21831/cp.v3i3.1625

Subali, B., & Mariyam, S. (2015). Measuring the Indonesian elementary schools student's creativity in science processing skills of life aspects on natural sciences subject: In Yogyakarta Special Province (DIY). *Journal of Elementary Education*, *25*(1), 91–105. Retrieved from https://www.google.co.id/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjK0tv7qNHWAhUJpo8KHbLECmsQFggnMAA&url=http%3A%2F%2Fpu.edu.pk%2Fimages%2Fjournal%2FJEE%2FPDF-Files%2F6_v25_no1_15.pdf&usg=AOvVaw0ZFHYs1ed_mQAOEFlsRpb9

Subali, B., Paidi, P., & Mariyam, S. (2016). The divergent thinking of basic skills of

sciences process skills of life aspects on natural sciences subject in Indonesian elementary school students. *Asia-Pacific Forum on Science Learning and Teaching*, *17*(1). Retrieved from http://proxy.library.vcu.edu/login?url=http://search.proquest.com/docview/1871581264?accountid=14780%5Cnhttp://vcu-alma-primo.hosted.exlibrisgroup.com/openurl/VCU/vcu_services_page?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&genre=article

Torrance, E. P. (1979). A three-stage model for teaching for creative thinking. In A. E. Lawson (Ed.), *The psychology of teaching for thinking and creativity* (pp. 226–253). Columbus, OH: Association for Education of Teachers of Science, Ohio State University.

Towle, A. (1989). *Modern biology*. Austin, TX: Holt, Rinehart and Winston.

Wenning, C. J. (2005). Levels of inquiry: Hierarchies of pedagogical practices and inquiry processes. *Journal of Physics Teacher Education Online*, *2*(3), 3–12. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Levels+of+inquiry:+Hierarchies+of+pedagogical+practices+and+inquiry+processes#0

Wenning, C. J. (2010). Level of inquiry: Using inquiry spectrum learning sequences to teach science. *Journal of Physics Teacher Education Online*, *5*(3), 11–20. Retrieved from www2.phy.ilstu.edu/pte/publications/learning_sequences.pdf

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: Mesa Press.

REiD (Research and Evaluation in Education), 3(1), 2017, 12-27
Available online at: http://journal.uny.ac.id/index.php/reid
Research Article

Practitioner-informed improvements to early childhood intervention performance checklists and practice guides

***[1]Carl J. Dunst; [2]Deborah W. Hamby; [3]Linda L. Wilson; [4]Marilyn Espe-Sherwindt; [5]Donna E. Nelson**
*Orelena Hawks Puckett Institute
128 South Sterling Street Morganton, NC 28655 United States
*Email: cdunst@puckett.org

Submitted: 25 May 2017 | Revised: 25 July 2017 | Accepted: 27 July 2017

Abstract

Results from four early childhood practitioner field-tests of performance checklists and early intervention practice guides are reported. Findings from the first field-test were used to make changes and improvements in the checklists and practice guides were evaluated in the second and third field-tests, and findings from the latter two field-tests were used to improve the checklist and practice guide evaluated in the fourth field-test. The results indicated that changes made in response to practitioners' suggestions and feedback were associated with (1) progressive increases in the practitioners' social validity judgments of the checklists, practice guides, and checklist-practice guide correspondence; and (2) progressive decreases in the number of practitioner suggestions and feedback for improving the early intervention materials. The field-test research demonstrates the importance of practitioner input, suggestions, and feedback for improving the usefulness of early childhood intervention practices.

Keywords: *early childhood intervention, performance checklists, practice guides, social validity, practitioner appraisals*

How to cite item:

Dunst, C., Hamby, D., Wilson, L., Espe-Sherwindt, M., & Nelson, D. (2017). Practitioner-informed improvements to early childhood intervention performance checklists and practice guides. *REiD (Research and Evaluation in Education), 3*(1), 12-27. doi:http://dx.doi.org/10.21831/reid.v3i1.14158

Introduction

Early childhood intervention for infants, toddlers, and preschoolers with identified disabilities or developmental delays and their families, as well as for young children who are at-risk for poor developmental outcomes for biological or environmental reasons is now common practice throughout the world (e.g., Farrell, Kagan, & Tisdall, 2016; Groark, Eidelman, Maude, & Kaczmarek, 2011; Guralnick, 2005; Odom, Hanson, Blackman, & Kaul, 2003; Sukkar, Dunst, & Kirkby, 2017). Early childhood intervention includes the experiences and learning opportunities afforded young children to promote acquisition of functional behavior (Bailey & Wolery, 1992; Dunst & Espe-Sherwindt, 2017) and the supports and resources provided to or procured by parents and other family members to strengthen family functioning (Cowan, Powell, & Cowan, 1998; Dunst, 2017b).

The field of early childhood intervention has a relatively short but rich history (Dunst, 1996; McLean, Sandall, & Smith, 2016; Meisels & Shonkoff, 2000). In the 50+ years since Hunt (1961) first noted that experiences early in a child's life could alter developmental outcomes, and later, that respon-

sive caregiving was an important factor in shaping those outcomes (Hunt, 1987), considerable advances have been made in terms of understanding which experiences under which conditions have which kinds of outcomes and benefits (e.g., Britto, Engle, & Super, 2013; Farrell et al., 2016; Odom & Wolery, 2003; Reichow, Boyd, Barton, & Odom, 2016). Early childhood intervention practitioners now have many choices and options in terms of the intervention practices they can use in their work with young children and their families.

Many factors influence practitioners' adoption and use of different kinds of early childhood intervention practices, including, but not limited to, personal beliefs about practice-outcome relationships and one's ability to use a practice competently and confidently (Bruder, Dunst, & Mogro-Wilson, 2011; Trivette, Dunst, Hamby, & Meter, 2012). These beliefs include the social validity appraisals of early childhood intervention practices and also their intended outcomes (Kazdin, 2005).

The practical importance of social validity appraisals is that these types of judgments can help explain why a practitioner does or does not use an intervention practice. Subjective judgments of the importance and acceptability of intervention goals, practices, and outcomes likely influence practitioners' adoption and use of different kinds of intervention procedures (Foster & Mash, 1999). According to Strain et al. (2012), intervention practices are not likely to be used by practitioners (or parents) if the practices themselves are not viewed as socially valid and worth the time and also effort to use. Dunst, Raab, and Hamby (2016), for example, found parents' social validity judgments of interest-based child language learning practices were directly related to parents' fidelity of use of the practices and indirectly related to child language development mediated by fidelity of use of the practices.

The study described in this paper is part of a line of research and practice on (a) the development of evidence-informed early childhood intervention performance checklists and both practitioner and parent practice guides and (b) the influences of practitioner feedback and suggestions on the improvement of both sets of materials. The study involved four field-tests that solicited practitioner social validity judgments of selected checklists and practice guides as well as suggestions for the improvement of both products. Each field-test involved practitioner review and evaluation of a different performance checklist and a different practice guide where feedback and suggestions were used to inform improvements in both sets of products. The findings from the first field-test were used to make changes in the checklists and practice guides in the second and third field-tests, and findings from the latter two field-tests were used to inform improvements to the checklist and practice guide in the fourth field-test. Preliminary findings from this line of research and practice indicated that changes made to the checklists and practice guides in response to practitioner evaluations of the intervention materials were associated with stronger social validity appraisals of revised checklists and practice guides (Dunst, 2017a).

Evidence-Informed Checklists and Practice Guides

The performance checklists and practice guides that were the focus of field-test research were developed at the *Early Childhood Technical Assistance Center* at the University of North Carolina – Chapel Hill, United States (www.ectacenter.org). Checklists include lists of the key characteristics of a method or procedure that operationally defines the active ingredients of desired performance. Early childhood intervention checklists include the key characteristics of intervention practices that are used to produce observable changes or improvement in child or family functioning. Practice guides include descriptions of everyday intervention activities that can be used to affect changes in child or family functioning.

Performance Checklists

Early childhood intervention performance or procedural checklists provide concrete reminders for using intervention practices in a competent manner (Gawande, 2009;

Wilson, 2013). The checklists were developed using a conceptualization-operationalization-measurement framework (Babbie, 2009; Dunst, 2017c; Dunst, Trivette, & Raab, 2015) where research findings from primary research syntheses and reviews (Dunst, in preparation) informed checklist indicator selection or development. Performance checklists differ from other types of checklists by specifying a "list of tasks or steps required to complete a procedure [intervention practice] successfully" (Wilson, 2013, p. 4). According to Gawande (2009), these kinds of checklist indicators provide practitioners concrete reminders for how to implement an intervention practice consistently, reliably, and competently.

Twenty-nine performance checklists were developed by first using the Division for Early Childhood Recommended Practices (Division for Early Childhood, 2014) to identify internally consistent sets of practice indicators for different types of intervention practices where the final selection of checklist practice indicators were informed by research evidence. The checklists were all formatted in the same way because "applying organizations to new learning causes learners to focus on the meaning" [intent] of the checklist indicators (Schwartz, 2014, p. 107).

Each checklist includes (1) a brief description of a checklist practice and how the checklist can be used, (2) a list of evidence-informed practice indicators, (3) a rating scale for doing a self-evaluation or coach-facilitated evaluation of the use of the practice indicators, and (4) space for recording notes about a practitioners' experience using the checklist practices. Appendix A shows the performance checklist that was the focus of practitioner evaluation in the fourth field-test. The reader is referred to Dunst (2017) for a more detailed description of the procedures used to develop the checklists.

Practice Guides

Two sets of practice guides were developed using the checklist indicators as the sources of intervention activities: One set for parents and other primary caregivers and a second set for early childhood intervention practitioners. The practice guides were also all formatted in the same way. Each practice guide includes: (1) a description of a practice and its intended outcome, (2) examples of activities for using a practice, (3) videos of parents or practitioners using the practice, (4) a vignette of parents or practitioners implementing a practice, (5) functional outcome indicators for determining if the practice had expected benefits, and (6) a link to external resources for additional ideas (activities) for using a practice. Appendix B shows the practice guide for the *Family Capacity-Building Practices Checklist* used in the fourth field-test. The practice guides are modeled after the ones that have been extensively field-tested and evaluated by parents, practitioners, technical assistance providers in previous research and intervention studies (e.g., Dunst, Masiello, Meter, Swanson, & Gorman, 2010; Dunst, Trivette, Gorman, & Hamby, 2010; Trivette, Dunst, & Hamby, 2010).

Hypotheses

The analyses focused on two primary and two secondary hypotheses. The two primary hypotheses were:

H_1: The social validity judgments of the performance checklists and practice guides will increase linearly as a result of changes and improvements made in response to practitioners' evaluations as evidenced by the sizes of effect for the linear increases and associated improvement indices.

H_2: The number of practitioner suggestions for improving the checklists and practice guides will decrease linearly as a result of changes and improvements made in response to practitioners' evaluations as evidenced by the sizes of effect for the linear increases and associated improvement indices. The two secondary hypotheses were:

H_3: The sizes of effects and improvement indices for Field-Test 1 vs. Field-Test 4 will be larger than those for Field-Tests 2 + 3 vs. Field Test 4 as a result of the progressive changes and improvements made in response to practitioners' evaluations of the checklists and practice guides.

H_4: The sizes of effects and improvement indices for Field-Tests 1 + 2 + 3 vs.

Field-Test 4 will provide the best estimates of the cumulative changes made in response to practitioners' evaluations of the checklists and practice guides. Thus, the four hypotheses were tested by *a priori* linear and orthogonal contrasts for between-field-test comparisons in the analyses of the field-test research data.

Method

The participants consisted of 67 practitioners from an early head start program in one state and two early childhood intervention programs in other states. The three programs have a history of using innovative practices where the program practitioners are knowledgeable about contemporary evidence-informed early childhood intervention practices. There were no between-group differences in the percentage of participants in the different field-test studies, $\chi^2 = 6.68$, $df = 6$, $p = .3516$, and nor was there in the percentage of participants in the type of early childhood program in the field-tests, $\chi^2 = 2.77$, $df = 6$, $p = .8375$.

Table 1. Background characteristics of the field-test participants

Characteristics	Number	Percent
Education degree		
AA	14	20
BA/BS	24	35
MA/MS	26	40
Ph.D/Ed.D	3	5
Discipline		
Early childhood	42	63
Early childhood special education/special education	16	24
Other[a]	9	13
Years of experience		
<1	4	6
2-5	13	19
6-10	15	22
11-15	10	15
16-20	14	21
21+	11	16
Primary practitioner role		
Child-focused	25	37
Family-focused	42	63

Note: [a]Speech and language pathologists, child and family specialists, early interventionists, and social workers/family workers.

The background characteristics of the participants are shown in Table 1. The majority of practitioners (75%) had either bachelor's or master's degrees. Most of the practitioners had degrees in early childhood education or early childhood special education/ special education. The participants' median years of experience ranged between 6 and 10 with 78% having from 6 to 20+ years of experience. Nearly two-thirds of the participants worked primarily with parents and their children (family-focused) and 37% worked primarily with children (child-focused). There were no between field-test differences for any of the participant background characteristics shown in Table 1, $\chi_s = 0.17$ to 10.73, $df_s = 3$ to 15, $p_s = .1004$ to $.9817$.

Procedure

The performance checklists and practice guides that were the focus of the field-test evaluations are shown in Table 2. The four topic areas included child, parent-child, parent, and family-focused intervention practices. Both the checklists and practice guides included different kinds of interventions for (a) using everyday activities as sources of child learning opportunities and (b) parent sensitivity and responsiveness to child behavior in the activities as the primary caregiver practice to reinforce child competencies and sustain child engagement in the activities.

The checklist in the first field-test included practice indicators for strengthening caregiver and child relationships that focused on bidirectional, reciprocal interactions between interactive partners (Eshel, Daelmans, Cabral de Mello, & Martines, 2006). The practice guide for the checklist indicators included a number of different socially interactive games that caregivers could use to engage young children in your turn-my turn interactive episodes (e.g., Dunst, Pace, & Hamby, 2007).

Table 2. Performance checklists and practice guides that were the focus of practitioner social validity judgments and feedback

Field-test	Topic area	Performance checklists	Practice guides
1	Interaction	Adult-child interactions	Social games
2	Environment	Natural learning opportunities	It's natural
3	Instruction	Naturalistic instruction	Learning comes naturally
4	Family	Family capacity-building	Everyday child learning

The checklist in the second field-test included indicators for identifying everyday activities that provide the most opportunities for child learning (Dunst, Bruder, Trivette, & Hamby, 2006). The practice guide included ideas and strategies for engaging a child in the activities (Dunst, Raab, & Trivette, 2013b). The checklist in the third field-test included indicators for using naturalistic teaching practices for reinforcing child behavior initiations and elaborations while engaged in everyday activities (Dunst, Raab, & Trivette, 2011). The practice guide included different kinds of intervention activities and strategies for using natural reinforcing consequences for reinforcing child behavior (e.g., Dunst, Raab, & Trivette, 2013a).

The checklist in the fourth field-test included methods for strengthening family capacity to provide a child everyday learning opportunities (Swanson, Raab, & Dunst, 2011). The practice guide included a set of step-by-step instructions for practitioners to use to encourage and support parent-mediated everyday child learning (e.g., Raab, Dunst, & Trivette, 2013).

The checklist and practice guide in the first field-test had not been subjected to prior review and feedback, and practitioner evaluations of both products were used as the baseline for evaluation of subsequent revisions and improvements to the checklists and practice guides in the second and third field-tests. The changes to the checklists in response to practitioner feedback and suggestions included clarifying the purpose of the checklist instructions and intended users (practitioners), rewording the checklist indicators to improve meaning and intent, clarifying how to use the checklist indicators to plan intervention sessions with parents, and the way how to use the rating scale to do a self-evaluation of how many and how well the checklist indicators were used with a child or parent. The changes to the practice guides included adding captions to the videos of parents or practitioners using the practices, adding additional intervention activities to the practice guides, including suggestions for making adaptations to the practice guide activities (where appropriate), and clarifying how to use the outcome indicators for evaluating the benefits of the practice guide activities.

Feedback and suggestions on the second and third field-tests were used to make additional changes to the checklist and practice guide in the fourth field-test. The changes to the checklist included clarifying the difference between using the checklist indicators for *a priori* intervention planning and doing a *post hoc* self-evaluation of the use of the indicators and clarifying the instructions for how to use the checklist indicators for completing a self-evaluation. The changes to the practice guide included additional specificity in terms of the focus and intent of the practice guide as well as the practice guide activities.

Field-Test Survey

The survey included four sections: (1) practitioner social validity judgments of the checklists, practice guides, and correspondence between practice guides and checklist indicators; (2) open-ended questions asking for suggestions to improve the checklists and practice guides; (3) levels of experience needed for a practitioner to understand and use checklists and practice guides; (4) background information about the field-test participants (Table 1). Each field-test involved an emailed invitation sent to the directors of each program that included instructions for participation in the field-tests, PDFs of the checklists and practice guides, and a URL link to the survey. The program directors were asked to forward the emailed invitation to their staff. Participation in the field-tests was voluntary,

and the field-test research was considered exempt from human subjects review because practitioners were asked only to evaluate materials designed for routine early childhood intervention. The surveys were completed online using Qualtrics Survey Soft-ware.

The social validity items for the performance checklists, practice guides, and checklist-practice guide correspondence (four per each section) were developed by using Foster and Mash's (1999) framework for developing indicators for measuring the importance and acceptability of intervention practices and outcomes. In addition, the social importance of the checklists and practice guides was measured in terms of the subjective value attributed to the intervention materials (e.g., *The checklist items are easy to understand and follow; The practice guide activities would be engaging to most children*). The social acceptability of the checklists and practice guides was measured in terms of judgments about the fit of the practices to everyday life (e.g., *The checklist indicators would be easy to use with a parent or child; The practice guide would be worth my time and effort to use*). The social validity items were each rated on a 5-point scale ranging from *Do Not Agree At All (*with the survey items) to *Agree a Great Deal* (with the survey items). The items were adopted from the ones used in field-tests of other intervention practices (e.g., Dunst et al., 2007; Dunst, Trivette, et al., 2010).

The principal component factor analysis of the three sets of items in each field-test with orthogonal rotation each produced a single-factor solution indicating that summated scores were warranted as measures of social validity judgments (Spector, 1992). The average coefficient alpha for the checklist indicators was .89 (Range = .81 to .97), the average alpha for the practice guide indicators was .85 (Range = .77 to .91), and the average alpha for the correspondence between the checklists and practice guides was .92 (Range = .85 to .95). The alpha's in all 12 factor analyses reached acceptable levels of internal reliability (Nunnally & Bernstein, 1994).

The open-ended questions for improving the checklists asked for suggestions about the (1) checklist instructions, (2) checklist indicators, (3) self-evaluation scale, and (4) any other suggestions for improvement. The open-ended questions for improving the practice guides asked for suggestions about the (1) practice guide format, (2) practice guide activities, (3) videos of the practices, (4) child outcomes, and (5) any other suggestions to improve the practice guides.

Methods of Analysis

The 3 Between Field Test ANOVAs with preplanned linear and between group contrasts were used to evaluate the effects of changes to the checklists and practice guides on participants' social validity judgments. The independent variable was the different field-tests (Field-Test 1 vs. Field-Tests 2 + 3 vs. Field-Test 4). The linear contrasts and between-field-test comparisons permitted tests of the four study hypotheses. The dependent measures in three ANOVAs were the summated social validity scores for the performance checklists, practice guides, and correspondence between the checklists and practice guides.

The primary metrics for testing the study hypotheses were Cohen's *d* effect sizes and associated improvement indices (What Works Clearinghouse, 2014). Effect sizes rather than statistical significance testing is the preferred metric for substantive interpretation because effect sizes and not *p*-values are the best estimates of the magnitude of the differences between two groups or contrasts (Coe, 2002). As a general rule, effect sizes between .20 and .49 are considered small, those between .50 and .79 are considered medium, those between .80 and 1.19 are considered large, and effect sizes equal to or greater than 1.20 are considered very large. Improvement indices are measures of the practical importance of the changes made to the checklists and practice guides (Durlak, 2009). The indices convert effect sizes into a percentile change (gain) score by a target group. These indices vary from -50 to 50 where a positive difference between later and earlier field-tests provides a measure of the amount of improvement that occurred as a result of changes made to the checklists and practice guides. *ZCalc* was used to evaluate the improvement indices (Neill, 2006).

Primary analyses of the practitioners' social validity judgments were supplemented by computing the percent of indicators rated a 4 or 5 on the 5-point scale to ascertain the overall levels of agreement with the indicators. As found in consumer sciences research, the larger the percent of indicators rated a 4 or 5 on a 5-point scale, the stronger the endorsement of a product, practice, or service (Mackiewicz & Yeats, 2014; Reichheld, 2003). The Mantel-Haenszel test for linear trends was used to determine if there were progressive increases in the percent of practitioners rating the social validity items a 4 or 5 from the first to fourth field-tests (SPSS Inc., 2005).

The effects of the changes made to the checklists and practice guides in response to practitioner suggestions were tested by both 3 Between Field-Test ANOVAs for the total number of practitioner suggestions and by 3 Between Field-Test Chi-Square analyses for dichotomous responses for each open-ended section. The same linear contrasts for the social validity appraisals were made for evaluating changes in the practitioner suggestions.

Findings and Discussion

Social Validity Judgments

Figure 1 shows the mean social validity scores for four field-tests for each set of importance and acceptability judgments. The 3 between field-test ANOVAs produced between group differences for the practitioner social validity judgments of the performance checklists, $F(2, 64) = 3.49$, $p = .0364$, and checklist/practice guide correspondence, $F(2, 64) = 4.94$, $p = .0101$, but not for the practice guides, $F(2, 64) = 0.42$, $p = .6562$.

The results for linear contrasts and between-field-test comparisons are presented in Table 3. There were small linear increases to the practice guides to medium linear increases to both performance checklists and checklist/practice guide correspondence increases from the first to fourth field-tests as evidenced by the sizes of effect for the linear trends. The effect sizes were associated with improvement indices of 9, 22, and 27 percent, respectively, in response to the progressive changes in the practice guides, checklists, and correspondence comparisons. The findings are consistent with Hypothesis 1 that changes made in response to the practitioners' feedback and suggestions would be related to improvements in the social validity judgments of the checklists and practice guides.

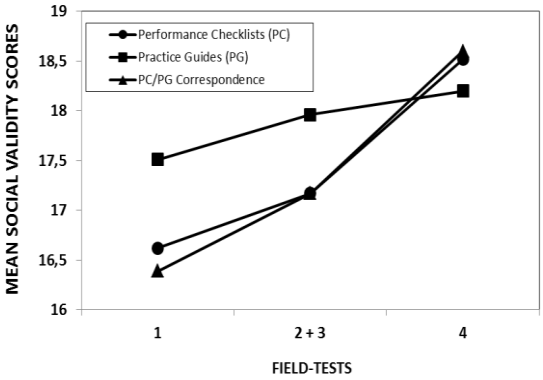

Figure 1. Mean practitioner social validity scores for the four field-tests.

Table 3. Linear contrasts and between-field-test comparisons and associated significance levels, effect sizes, and improvement indices

Product	Linear trend	Field-test comparisons		
		1 vs. 4	2 + 3 vs. 4	1+2+3 vs. 4
Statistical significance				
Performance checklists (PC)	$p = .0107$	$p = .0040$	$p = .1410$	$p = .0052$
Practice guides (PG)	$p = .1802$	$p = .1845$	$p = .2908$	$p = .2090$
PC/PG correspondence	$p = .0025$	$p = .0012$	$p = .1706$	$p = .0014$
Cohen's *d* effect sizes				
Performance checklists (PC)	0.59	0.93	0.22	0.67
Practice guides (PG)	0.23	0.29	0.18	0.21
PC/PG correspondence	0.74	1.09	0.33	0.78
Improvement indices				
Performance checklists (PC)	22	32	9	25
Practice guides (PG)	9	11	7	8
PC/PG correspondence	27	36	13	28

Note: Both the linear trends and field-test comparisons all have numerator degrees of freedom = 1.

There were small (practice guides) to large (checklists and checklist/practice guide correspondence) effect sizes for the differences between the Field-Test 1 vs. Field-Test 4 social validity judgments (Table 3). These were associated with improvement indices of 11, 32, and 36 percent, respectively, for the practice guides, checklists, and correspondence judgments. The effect sizes for the Field-Tests 2 + 3 vs. Field-Test 4 for the between-field-test comparisons were small for both the performance checklists and checklists/practice guides correspondence. The between-field-test comparisons were associated with improvement indices of 9% for the checklist differences and 13% for the checklist/practice guide correspondence differences. The comparisons of the two sets of results in Table 3 shows, as hypothesized, that the sizes of effect and associated improvement indices for Field-Tests 1 vs. 4 are considerably larger than those for Field-Tests 2 + 3 vs. 4.

The cumulative effects of the progressive changes made in response to the practitioner evaluations are evidenced from the Field-Tests 1 + 2 + 3 vs. Field-Test 4 comparisons. There were small (practice guides) to medium (checklists and checklists/practice guides correspondence) effect sizes for these between-field-test comparisons. The effect sizes were associated with improvement indices of 8, 25, and 28 percent, respectively. The results are consistent with the hypothesized relationships between changes made in response to practitioner feedback and suggestions and improvements in the social validity judgments of the intervention practices.

The percent of social validity items rated a 4 or 5 on each section of the survey for the different field-tests are shown in Figure 2. There were linear increases in the percent of indicators rated a 4 or 5 for the performance checklists, $\chi = 9.04$, $df = 1$, $p = .003$, $d = .79$, practice guides, $\chi^2 = 5.88$, $df = 1$, $p = .015$, $d = .62$, and checklist/practice guide correspondence, $\chi^2 = 10.98$, $df = 1$, $p = .001$, $d = .97$. The effect sizes for the linear trends were medium to large and associated with improvement indices of 29, 23, and 33 percent, respectively. The smaller effect size for the linear increase in the social validity ratings of the practice guides was not unexpected given the fact that practitioner judgments of the practice guides were higher than those for the checklist on the first three field-tests. As shown in Figure 2, 98% to 99% of the social validity items received the highest two ratings in the fourth field-test which are noticeably higher than that in the other three field-tests.

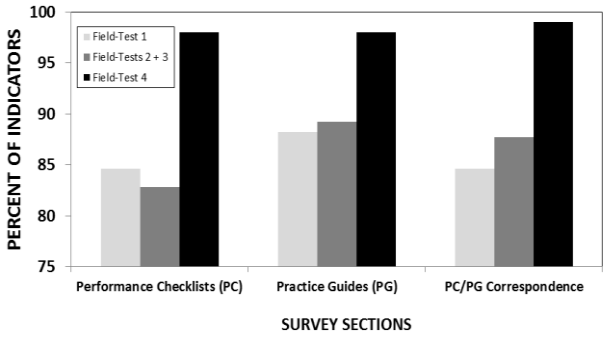

Figure 2. Percent of social validity items judged as acceptable and important by the field-test participants

Practitioner Suggestions

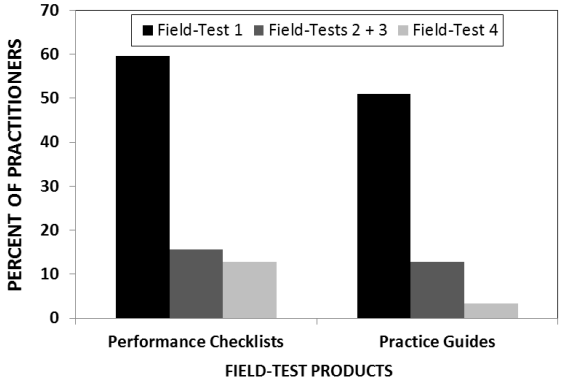

Figure 3. Percent of practitioners making suggestions for improving the performance checklists and practice guides

Figure 3 shows the percent of practitioners who made suggestions for improving the checklists and practice guides in the different field-tests. The 3 between-field-test ANOVAs for the total number of practitioner suggestions produced between field-test differences for both the performance checklists, $F(2, 64) = 7.11$, $p = .0016$, and practice guides, $F(2, 64) = 10.51$, $p = .0001$. There

were linear decreases in the number of suggestions for the checklists, $F(1, 64) = 11.41$, $p = .0012$, $d = .85$, and practice guides, $F(1, 64) = 18.31$, $p = .0001$, $d = 1.07$. Both sizes of effects were large for the linear decreases in the number of practitioner suggestions. The effect sizes were associated with the improvement indices of 30% and 36% respectively. The patterns of results are consistent with the hypothesis that the practitioners would suggest fewer changes as a function of improvements made in response to their feedback and suggestions.

Further examination of the suggestions for improving the checklists found linear decreases in the percent of practitioners who made suggestions for changes to the checklist instructions, $\chi^2 = 6.43$, $df = 1$, $p = .011$, $d = .77$, the checklist indicators, $\chi^2 = 6.96$, $df = 1$, $p = .008$, $d = .83$, and the self-evaluating rating scale, $\chi^2 = 2054$, $df = 1$, $p = .0555$, $d = .45$. The sizes of effects were medium, large, and small, respectively, and associated with improvement indices between 17% and 30%. There were also linear decreases in the percent of practitioners making suggestions to improve the practice guide format, $\chi^2 = 12.43$, $df = 1$, $p = .0000$, $d = 1.15$, practice guide activities, $\chi^2 = 9.49$, $df = 1$, $p = .001$, $d = .87$, practice guide outcome statements, $\chi^2 = 2.06$, $df = 1$, $p = .051$, $d = .38$, and videos of parents or practitioners using the practices, $\chi^2 = 17.17$, $df = 1$, $p = .0000$, $d = 1.36$. The effect sizes were small to very large and associated with improvement indices between 15% and 41%. These findings, taken together, further support hypothesized relationships between the changes made in response to the practitioners' evaluations of the checklists and practice guides and fewer suggestions for improving the intervention practices.

Discussion

The findings provide support for the two primary hypotheses that changes made to the performance checklists and practice guides in response to early childhood practitioners feedback and suggestions would be related to the study outcomes. The results showed that practitioners' social validity ratings increased as a function of the improvements to both the performance checklists and practice guides and also to the checklist/practice guide correspondence. The results also showed that there were fewer suggestions for making changes to the checklists and practice guides as a function of using practitioner feedback to improve both sets of products.

The patterns of results also provide support for the two secondary hypotheses. The sizes of effect for the first vs. fourth field-tests were larger than those for the second and third vs. fourth field-tests (Table 3). These results were expected because fewer suggestions for changes to the checklists and practice guides were made on the second and third field-tests compared to the first field-test (Figure 3). The cumulative effects for the changes made in response to practitioners' suggestions were evidenced by the sizes of effect for the first three field-tests vs. the fourth field-test. Both the effect sizes for these comparisons and improvement indices (Table 3) indicated that the progressive sets of changes made in response to practitioners' suggestions were associated with the highest social validity rating (Figure 1) and fewest suggestions for change (Figure 3) on the fourth field-test.

The fact that the effect sizes and improvement indices for the practice guides were smaller than those for the performance checklists and checklist-practice guide correspondence deserves comment in order to place the results in empirical context. The practice guides were modeled after the ones that had previously been field-tested with parents and practitioners where the results were used to improve the intervention materials (e.g., Dunst, Trivette, et al., 2010; Trivette, Dunst, Masiello, Gorman, & Hamby, 2009). It was therefore not unexpected that the majority of social validity indicators for the practice guides on the first three field-tests were higher than those for the checklists and checklist-practice guide correspondence (Figure 3). This was the case because the practice guide format and content were informed by lessons learned in previous research and practice.

The focus on the social validity of the performance checklists and practice guides was based on research indicating that subjective judgments of the importance and accept-

ability of intervention practices and outcomes are related to both adoption and fidelity of use of the practices (e.g., Dunst et al., 2016; Strain et al., 2012; Trivette, Raab, & Dunst, 2014; Wainer & Ingersoll, 2013; Wehby, Maggin, Moore Partin, & Robertson, 2011). As noted by Strain et al. (2012), these "liking-implementation with fidelity relationships" (p. 197) are important because they help explain at least the likelihood of early childhood intervention practices being used as intended.

Conclusion and Suggestions

The study described in this paper has both strengths and limitations. One strength is the fact that the procedures used to inform changes in the checklists and practice guides illustrates how the consumer level input can be used to improve social validity appraisals of the intervention materials constituting the focus of evaluation. Another strength is establishing the inverse relationship between increases in social validity ratings and concomitant decreases in practitioner suggestions for changes. In another set of analyses in this line of research and practice, practitioners' cognitive judgments of the performance checklists and practice guides were the only variable accounting for variations in the social validity ratings of the intervention materials (Dunst & Hamby, 2017).

One limitation of the study is that the field-tests were conducted in only three early childhood intervention programs. Therefore, it is not known if practitioners in other early childhood intervention programs would judge the checklists and practice guides in the same or different ways. Another limitation is the fact that only 4 out of 29 performance checklists and only 4 out of 67 practice guides were evaluated in the field-tests. Whether other checklists and practice guides would be judged similarly is therefore not known.

Advances in our understanding of the role social validity judgments play in practitioners' and parents' use of different kinds of early childhood intervention has broadened our knowledge of the antecedents for and conditions under which intervention practices are used with fidelity (Leko, 2014; Strain et al., 2012). One simple way of assessing practitioners' and parents' social validity appraisals is to ask the question "Was using XYZ practice worth your time and effort or was it more trouble than it was worth?" If a practitioner or parent responds that it was not worth the trouble to use, it is unlikely that the practice will be used with fidelity or used at all.

Although the field-test process was used to inform improvements in early childhood intervention performance checklists and practice guides, the process itself could easily be used in other fields for achieving performance excellence. This is especially the case in professions where there needs to be practitioner buy-in to ensure actual performance mirrors desired performance.

Acknowledgements

The field-test research described in this paper was supported, in part, by a subcontract to the Orelena Hawks Puckett Institute from the Early Childhood Technical Assistance (ECTA) Center, Frank Porter Graham Child Development Institute, University of North Carolina--Chapel Hill. The ECTA Center is funded by the U.S. Department of Education, Office of Special Education Programs (Grant #H326P120002). The opinions expressed, however, are those of the authors and no endorsement by the funder, University, or ECTA Center should be implied or inferred.

Special thanks to the practitioners who participated in the field-test evaluations and for providing assessments of and feedback on the performance checklists and practice guides.

References

Babbie, E. R. (2009). *The practice of social research* (12th ed.). Belmont, CA: Wadsworth.

Bailey, D. B., Jr., & Wolery, M. (1992). *Teaching infants and preschoolers with disabilities* (2nd ed.). Upper Saddle River, NJ: Merrill.

Britto, P. R., Engle, P. L., & Super, C. M. (Eds.). (2013). *Handbook of early childhood development research and its impact on global policy*. New York, NY: Oxford.

Bruder, M. B., Dunst, C. J., & Mogro-Wilson, C. (2011). Confidence and competence appraisals of early intervention and preschool special education practitioners. *International Journal of Early Childhood Special Education, 3*(1), 13-37.

Coe, R. (2002, September). *It's the effect size, stupid: What effect size is and why it is important.* Paper presented at the Annual Conference of the British Educational Research Association, University of Exeter, England. Retrieved from http://www.leeds.ac.uk/educol/documents/00002182.htm.

Cowan, P. A., Powell, D., & Cowan, C. P. (1998). Parenting interventions: A family systems perspective. In W. Damon, I. E. Sigel, & K. A. Renninger (Eds.), *Handbook of child psychology: Vol. 4. Child psychology in practice* (5th ed., pp. 3-72). New York, NY: Wiley.

Division for Early Childhood. (2014). *DEC recommended practices in early intervention/ early childhood special education*. Retrieved from http://www.dec-sped.org/recommendedpractices

Dunst, C. J. (1996). Early intervention in the USA: Programs, models, and practices. In M. Brambring, H. Rauh, & A. Beelmann (Eds.), *Early childhood intervention: Theory, evaluation, and practice* (pp. 11-52). Berlin, Germany: de Gruyter.

Dunst, C. J. (2017a). Early childhood practitioner judgments of the social validity of performance checklists and parent practice guides. *Journal of Education and Training Studies, 5*(3), 176-187. Retrieved from http://redfame.com/journal/index.php/jets/article/view/2162 doi:10.11114/jets.v5i3.2162

Dunst, C. J. (2017b). Family systems early childhood intervention. In H. Sukkar, C. J. Dunst, & J. Kirkby (Eds.), *Early childhood intervention: Working with families of young children with special needs* (pp. 38-60). Abingdon, Oxfordshire: Routledge.

Dunst, C. J. (2017c). Procedures for developing evidence-informed performance checklists for improving early childhood intervention practices. *Journal of Education and Learning, 6*(3), 1-13. Retrieved from http://www.ccsenet.org/journal/index.php/jel/article/view/66005 doi:10.5539/jel.v6n3px

Dunst, C. J. (in preparation). Research foundations for evidence-informed early childhood intervention performance checklists.

Dunst, C. J., Bruder, M. B., Trivette, C. M., & Hamby, D. W. (2006). Everyday activity settings, natural learning environments, and early intervention practices. *Journal of Policy and Practice in Intellectual Disabilities, 3*, 3-10. doi:10.1111/j.1741-1130.2006.00047.x

Dunst, C. J., & Espe-Sherwindt, M. (2017). Contemporary early intervention models, research, and practice for infants and toddlers with disabilities and delays. In J. M. Kauffman, D. P. Hallahan, & C. P. Pullen (Eds.), *Handbook of special education* (2nd ed., pp. 831-849). New York, NY: Routledge.

Dunst, C. J., Masiello, T., Meter, D., Swanson, J., & Gorman, E. (2010). Technical assistance providers' evaluation of the Center for Early Literacy Learning practice guides. *CELLpapers, 5*(3), 1-4. Retrieved from http://www.earlyliteracylearning.org/cellpapers/cellpapers_v5n3.pdf

Dunst, C. J., Pace, J., & Hamby, D. W. (2007). *Evaluation of the Games for Growing tool kit for promoting early contingency learning.* Asheville, NC: Winterberry Press.

Dunst, C. J., Raab, M., & Hamby, D. W. (2016). Interest-based everyday child language learning. *Revista de Logopedia, Foniatria y Audiologia, 36*, 153-161. doi:10.1016/j.rlfa.2016.07.003

Dunst, C. J., Raab, M., & Trivette, C. M. (2011). Characteristics of naturalistic language intervention strategies. *Journal of Speech-Language Pathology and Applied*

Behavior Analysis, 5(3-4), 8-16. Retrieved from http://www.baojournal.com/SLP-ABA%20WEBSITE/index.html

Dunst, C. J., Raab, M., & Trivette, C. M. (2013a). Getting in step with responsive teaching. *Everyday Child Language Learning Tools*, Number 5, 1-4. Retrieved from http://www.puckett.org/CECLL/ECLLTools_5.pdf

Dunst, C. J., Raab, M., & Trivette, C. M. (2013b). Methods for increasing child participation in interest-based language learning activities. *Everyday Child Language Learning Tools*, Number 4, 1-6. Retrieved from http://www.puckett.org/CECLL/ECLLReport_7_LearnOps.pdf

Dunst, C. J., Trivette, C. M., Gorman, E., & Hamby, D. W. (2010). Further evidence for the social validity of the Center for Early Literacy Learning practice guides. *CELLpapers, 5*(1), 1-3. Retrieved from http://www.earlyliteracylearning.org/cellpapers/cellpapers_v5n1.pdf

Dunst, C. J., Trivette, C. M., & Raab, M. (2015). Utility of implementation and intervention performance checklists for conducting research in early childhood education. In O. N. Saracho (Ed.), *Handbook of research methods in early childhood education: Vol. 1. Research methodologies* (pp. 247-276). Charlotte, NC: Information Age Publishing.

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology, 34*(9), 917-928. doi:10.1093/jpepsy/jsp004

Eshel, N., Daelmans, B., Cabral de Mello, M., & Martines, J. (2006). Responsive parenting: Interventions and outcomes. *Bulletin of the World Health Organization, 84*(12), 991-998. doi:10.1590/S0042-96862006001200016

Farrell, A., Kagan, S. L., & Tisdall, E. K. M. (Eds.). (2016). *The SAGE handbook of early childhood research*. Thousand Oaks, CA: SAGE Publishing.

Foster, S. L., & Mash, E. J. (1999). Assessing social validity in clinical treatment research issues and procedures. *Journal of Consulting and Clinical Psychology, 67*, 308-319. doi:10.1037/0022-006X.67.3.308

Gawande, A. (2009). *The checklist manifesto: How to get things right*. New York, NY: Metropolitan Books.

Groark, C. J., Eidelman, S. M., Maude, S., & Kaczmarek, L. (2011). *Early childhood intervention: Shaping the future for children with special needs and their families*. Santa Barbara, CA: Praeger.

Guralnick, M. J. (Ed.) (2005). *The developmental systems approach to early intervention*. Baltimore, MD: Brookes.

Hunt, J. M. (1961). *Intelligence and experience*. New York, NY: Ronald Press.

Hunt, J. M. (1987). The effects of differing kinds of experiences in early rearing conditions. In I. Uzgiris & J. M. Hunt (Eds.), *Infant performance and experience* (pp. 39-97). Urbana, IL: University of Illinois Press.

Kazdin, A. E. (2005). Social validity. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (Vol. 4, pp. 1875-1876). Chichester, England: John Wiley & Sons.

Leko, M. M. (2014). The value of qualitative methods in social validity research. *Remedial and Special Education, 35*(5), 275-286. doi:10.1177/0741932514524002

Mackiewicz, J., & Yeats, D. (2014). Product review users' perceptions of review quality: The role of credibility, informativeness, and readability. *IEEE Transactions on Professional Communication, 57*(4), 309-324. doi:10.1109/TPC.2014.2373891

McLean, M., Sandall, S. R., & Smith, B. J. (2016). A history of early childhood education. In B. Reichow, B. A. Boyd, E. E. Barton, & S. L. Odom (Eds.), *Handbook of early childhood special education*

(pp. 3-20). Switzerland: Springer International.

Meisels, S. J., & Shonkoff, J. P. (2000). Early childhood intervention: A continuing evolution. In J. P. Shonkoff & S. J. Meisels (Eds.), *Handbook of early childhood intervention* (2nd ed., pp. 3-31). Cambridge, England: Cambridge University Press.

Neill, J. (2006). *ZCalc: For converting a standardised mean effect size (ES) into a z score and in expressing various ways (Version 0.1).* Retrieved from www.wilderdom.com/research/ZCalc.xls

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Odom, S. L., Hanson, M. J., Blackman, J. A., & Kaul, S. (2003). *Early intervention practices around the world.* Baltimore, MD: Brookes.

Odom, S. L., & Wolery, M. (2003). A unified theory of practice in early intervention/early childhood special education: Evidence-based practices. *Journal of Special Education, 37*, 164-173. doi:10.1177/00224669030370030601

Raab, M., Dunst, C. J., & Trivette, C. M. (2013). Adult learning procedure for promoting caregiver use of everyday child language learning practices. *Everyday Child Language Learning Reports, Number 3*, 1-9. Retrieved from http://www.cecll.org/download/ECLL_Report_3_AdultLearning.pdf

Reichheld, F. F. (2003). The one number you need to grow. *Harvard Business Review, 81*(12), 46-54.

Reichow, B., Boyd, B. A., Barton, E. E., & Odom, S. L. (Eds.). (2016). *Handbook of early childhood special education.* Switzerland: Springer International.

Schwartz, B. L. (2014). *Memory: Foundations and applications* (2nd ed.). Los Angeles, CA: SAGE.

Spector, P. E. (1992). *Summated rating scale construction: An introduction*. Newbury Park, CA: Sage.

SPSS Inc. (2005). *SPSS 14.0. Statistical package for the social sciences*. Chicago, IL: Author.

Strain, P. S., Barton, E. E., & Dunlap, G. (2012). Lessons learned about the utility of social validity. *Education and Treatment of Children, 35*, 183-200. doi:10.1353/etc.2012.0007

Sukkar, H., Dunst, C. J., & Kirkby, J. (Eds.). (2017). *Early childhood intervention: Working with families of young children with special needs*. Abingdon, Oxfordshire: Routledge.

Swanson, J., Raab, M., & Dunst, C. J. (2011). Strengthening family capacity to provide young children everyday natural learning opportunities. *Journal of Early Childhood Research, 9*, 66-80. doi:10.1177/1476718X10368588

Trivette, C. M., Dunst, C. J., & Hamby, D. W. (2010). Acceptability and importance of adaptations to literacy learning practices for young children with disabilities. *CELLpapers, 5*(4), 1-4. Retrieved from http://www.earlyliteracylearning.org/cellpapers/cellpapers_v5n4.pdf

Trivette, C. M., Dunst, C. J., Hamby, D. W., & Meter, D. (2012). Research synthesis of studies investigating the relationships between practitioner beliefs and adoption of early childhood intervention practices. *Practical Evaluation Reports, 4*(1), 1-19. Retrieved from http://www.puckett.org/Practical%20Evaluation%20reports/CPE_Report_Vol4No1.pdf

Trivette, C. M., Dunst, C. J., Masiello, T., Gorman, E., & Hamby, D. W. (2009). Social validity of the Center for Early Literacy Learning parent practice guides. *CELLpapers, 4*(1), 1-4. Retrieved from http://www.earlyliteracylearning.org/cellpapers/cellpapers_v4n1.pdf

Trivette, C. M., Raab, M., & Dunst, C. J. (2014). Factors associated with Head Start staff participation in classroom-

based professional development. *Journal of Education and Training Studies, 2*(4), 32-45. doi:10.11114/jets.v2i4.449

Wainer, A., & Ingersoll, B. (2013). Intervention fidelity: An essential component for understanding ASD parent training research and practice. *Clinical Psychology: Science and Practice, 20*(3), 335-357. doi:10.1111/cpsp.12045

Wehby, J. H., Maggin, D. M., Moore Partin, T. C., & Robertson, R. (2011). The impact of working alliance, social validity, and teacher burnout on implementation fidelity of the Good Behavior Game. *School Mental Health, 4*(1), 22-33. doi:10.1007/s12310-011-9067-4

What Works Clearinghouse. (2014). *What Works Clearinghouse procedures and standards handbook*. Washington, DC: U. S. Department of Education, Institute of Education Sciences. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf.

Wilson, C. (2013). *Credible checklists and quality questionnaires: A user-centered design method*. Waltham, MA: Morgan Kaufman.

Appendix A

Early childhood intervention performance checklist

Family Capacity-Building Practices Checklist

This checklist includes practices for promoting parents' and other family members' use of child-level interventions to enhance child learning and development in ways that strengthen parenting confidence and competence. The capacity-building practices are used by a practitioner to promote a parent's understanding and use of everyday activities as sources of child learning opportunities.

The checklist indicators can be used by a practitioner to plan intervention sessions with parents and other family members. The checklist rating scale also can be used to do a self-evaluation to determine if practitioner capacity-building practices actively involved parents in providing their children everyday learning opportunities.

Practitioner: _____ Child: _____ Date: _____

Please indicate which practice characteristics you were able to use as part of parent and family member involvement in providing child learning opportunities:	**Seldom or Never** (0-25%)	**Some of the Time** (25-50%)	**As Often As I Can** (50-75%)	**Most of the Time** (75-100%)	**Notes**
1. Describe the use and benefits of everyday activities as sources of child learning opportunities	☐	☐	☐	☐	
2. Illustrate or demonstrate child engagement in a variety of everyday activities	☐	☐	☐	☐	
3. Describe and illustrate the importance of child interests and preferences for promoting child learning	☐	☐	☐	☐	
4. Use an everyday activity checklist to have a parent select which activities would be easiest for the parent to use	☐	☐	☐	☐	
5. Together with the parent, engage the child in a familiar everyday activity	☐	☐	☐	☐	
6. Illustrate or demonstrate how adult responsiveness to child behavior is used to promote child learning in everyday activities	☐	☐	☐	☐	
7. Provide supportive guidance, feedback, and suggestions to the parent throughout the capacity-building activities	☐	☐	☐	☐	
8. Together with the parent, identify five or six everyday activities that will be used as child learning opportunities	☐	☐	☐	☐	
9. Engage the parent in conversations about which activities will be used for child learning and which parent responses will be used to promote learning	☐	☐	☐	☐	

ECTA Center
Early Childhood Technical Assistance Center

http://www.ectacenter.org/decrp/

Appendix B

Early childhood intervention practice guide

Everyday Child Learning Opportunities

Family capacity-building practices are used to support and strengthen parents and other caregivers' abilities to provide their children everyday learning opportunities. This is accomplished using a number of different strategies for supporting and strengthening parents' use of everyday activities to promote child learning and development.

Watch a video of this Learning Guide

Learning Guide: Family-Provided Child Learning Opportunities

- Parents are more likely to use intervention practices with their children if they understand the benefits of everyday child learning opportunities. Begin by explaining both the purpose and goal of the learning opportunities and the specific benefits to young children.
- Illustrate or demonstrate how to engage a child in everyday learning activities. Video vignettes of parents providing their children everyday learning opportunities generally work best. Be sure to point out the key characteristics of the practices.
- Engage the parents in real-life (authentic) activities to provide their children everyday learning opportunities. The more familiar and routine the activities are to parents, the more likely they will feel confident using the activities as sources of child learning opportunities.
- Focus on two important child and parent behaviors as part of everyday child learning opportunities. The first is the importance of **child interests and preferences** as factors increasing child engagement in everyday activities. The second is the role **parents' responsiveness to child behavior** in everyday activities plays in supporting child learning.
- Provide parents supportive guidance and feedback during and after parents' use of everyday learning opportunities with their child. Point out which characteristics of the practice were used by the parents and describe the child benefits of the practices.
- Engage the parents in conversations, discussions, or review of their confidence using the practices. Provide specific positive comments, feedback, and suggestions in response to parents' descriptions to reinforce their sense of confidence.
- Jointly identify four or five everyday activities that the parents can use to provide learning opportunities for their child. The best activities are ones that provide the child multiple opportunities to engage in interactions with people, toys, and other materials.
- Identify things you can do to provide the parents regular, ongoing support to encourage the continued use of the practices. It is important to plan to take time together to review and evaluate the learning opportunities to decide which activities should be continued, modified, and added.

A Quick Peek

Felicity is a 3-year-old with multiple disabilities. Her mother, Emma, tells her daughter's speech therapist that Felicity is beginning to show increased interest in looking at other people and is making more sounds than usual. Mom asks the therapist about the best times to work on her daughter's speech. The therapist describes how there are many opportunities throughout the day that can be used to encourage Felicity to use sounds in interactions with mom, dad, and other family members. The therapist uses video she has stored on a tablet computer to show Felicity's mom and dad how other parents have used everyday activities to encourage their young children to make sounds and "talk more." She points out the children's high level of interest in the activities and how the parents' imitation of their children's sounds gets the children to continue making the sounds. At the therapist's next visit, Emma says that Felicity now "talks up a storm" while she plays lap games with her daughter. The therapist asks Emma to show her what this looks like and suggests trying to do the same thing in a few more activities. It isn't long before Emma identifies more than a dozen activities during which she can encourage her daughter to make more sounds.

You'll know the practice is working if ...

- Lots of everyday activities are used for child learning
- Parents are responsive to their children's behavior in the activities
- The children are interacting with people and objects in the activities

Learn more about helping families make use of everyday learning opportunities from online activities and resources such as:

Tools and Guides to Facilitate Family Engagement on the EI Excellence website

ECTA Center
Early Childhood Technical Assistance Center

http://www.ectacenter.org/decrp/

REiD (Research and Evaluation in Education), 3(1), 2017, 28-41
Available online at: http://journal.uny.ac.id/index.php/reid
Research Article

A model of *madrasa ibtidaiya* quality evaluation

***[1]Dwi Istiyani; [2]Zamroni; [3]Suharsimi Arikunto**
*Faculty of Education and Teacher Training, Universitas Islam Negeri Walisongo Semarang
Jl. Prof. Dr. Hamka Km. 2, Ngaliyan, Semarang 50181, Central Java, Indonesia
*Email: dwieistiyani@gmail.com

Submitted: 04 May 2017 | Revised: 18 September 2017 | Accepted: 26 September 2017

Abstract

The study was to develop a *madrasa ibtidaiya* quality evaluation model (MIQEM) for measuring the quality of *madrasa ibtidaiya* (MI), including the input, the process, the short-term outcome, and the long-term outcome. The method implemented in this research was research and development (R & D). The subjects of the study were the principals and teachers of state *madrasa ibtidaiya*. The study concluded that: (1) the a MIQEM was developed through the product selection, review and result analysis, and model draft; (2) the implementation of the developed MIQEM was done by the principals and the teachers by operating the MIQEM instrument; (3) the results of the developed MIQEM implementation showed that it could improve *madrasa ibtidaiya* quality; (4) the results of the model fitness test indicated that the model was fit into the data.

Keywords: *evaluation model, quality, madrasa ibtidaiya*

How to cite item:

Istiyani, D., Zamroni, Z., & Arikunto, S. (2017). A model of madrasa ibtidaiya quality evaluation. *REiD (Research and Evaluation in Education), 3*(1), 28-41. doi:http://dx.doi.org/10.21831/reid.v3i1.13902

Introduction

Improving the quality of educational institutions is one of the requirements for us to enter globalization era that is full of competition. The existence of *madrasa* as an Islamic educational institution will not be apart of the global competition. Therefore, quality improvement has been the main agenda in improving the quality of *madrasa* in order that it can survive in the global era. The development of science and technology has brought about changes in all aspects of the human life in which multiple problems might be solved by the efforts of mastering and increasing science and technology. In addition to bringing the benefits into human life, such changes have also brought human beings into the more intensed global competition. In order to be able to play a certain role in the global competition, as a nation, we should continually improve and increase the human resources quality.

The demand for quality applies in all educational institutions, including *madrasa.* A school or a *madrasa* will be considered as being well-qualified if it has met the national standards that have been stipulated by the government and that have been governed in Indonesia. The graduates of educational institutions are expected to have capacity, science, skills, and behaviors that should be in accordance with the trend. Therefore, there should be efforts for improving the *madrasa* quality.

The number of *Madrasa ibtidaiya* (MI) schools in Indonesia is 21,529 units that have been spread in 34 provinces of Indonesia. The data in 2015 showed that the number of MI schools located in the Province of Central Java was 3,903 units. Based on the data, the number of state MI is very small, namely 114 units. However, in this year the number in-

creased due to the fact that some of the private MI's turned into state MI's. Now, the total number of state MI schools is 1,662 units in the Province of Central Java. This figure is not an overall figure because some of the newly-changed MI are not yet established through the Decree provided by the Ministry of Religious Affairs regarding the newly state institutions and the letter has not been provided yet until the report of the study has been composed.

According to the existing laws and regulations, actually the existence of *madrasa* schools has the same opportunity as that of general schools. The recognition of the government and the community of the graduates of *madrasa* school is equal to that of general schools and even the graduates of *madrasa* schools have a greater opportunity to achieve success because in madrasa schools there are characteristics and culture development efforts especially in the development of Islamic Religious Education (Law No. 20 Year 2003 of Republic of Indonesia on National Education System article 17 and Government Regulation No. 19 Year 2005, on National Education Standard, article 10). The existence of *madrasa* in the middle of the wave of change also follows the development by putting the first priority on innovation in order to achieve the dreams that the quality of *madrasa* has truly become the society's idealism. Nowadays, there have already been innovative policies especially the ones stipulated by the Directorate of *Madrasa* Education. Another regulation that governs the implementation of *madrasa* schools is the Regulation of Minister Number 90 Year 2013. In the Regulation of the Minister of Religious Affairs, No. 90 Year 2013, article 7, *madrasa* education is held by the government or the society. In Article 8 of the same regulation, it is stated that the establishment of *madrasa* that has been held by the government will be regulated by the Minister and the establishment of *madrasa* that has been held by the society will be governed by the regional office under the permission of the Minister in the form of operational permit provision.

In general, the status of *madrasa* is private; *madrasa* somestimes is a private institution that has been short of many aspects. For example, the teachers of *madrasa* schools do not have sufficient income, the book availability is uncertain and the number of facilities might be low in general. Similarly, the learning loads of the students there are very heavy. In relation to the situation, the recent definition regarding *madrasa* is that a *madrasa* is a general school that has the characteristics of Islam. The lessons that have this peculiar characteristics are plenty; as a result, these lessons become a certain aspect for the students. M*adrasa* students attend two types of examination namely *Madrasa* Examination (the peculiar lessons) and the National Examination. Ironically, people only view the *madrasa* quality based on the results of National Examination and not on the other achievements, for example the students' achievements in affective and spiritual intelligence (Suprayogo, 2007, p. 90).

According to Hidayat Nur Wahid (in Supriyadi & Rachman, 2015), the Deputy Chairman of the People's Consultative Council of Republic of Indonesia (MPR RI), there is an unfairness in the budget provision for general schools and for religion-based schools especially in the welfare fund intended for the teachers of *madrasa*; these teachers have not even received welfare fund. In his opinion, the budget disparity between general schools and religion-based schools, in this case that madrasa schools, has been uncommon. The welfare fund for *madrasa* teachers is only Rp 1,000,000.00 per month. The amount is very different when it comes to the welfare fund of the teachers who work for general schools, in this case the state schools. The welfare fund for the state schools in Jakarta is around IDR 4,000,000.00 – IDR 5,000,000.00 per month. Another aspiration that he conveys is related to the school operational aid (*Bantuan Operasional Sekolah*, BOS) for madrasa schools. The BOS for madrasa schools should be bigger than that of the state schools because there are five religion-based lessons that lead to the increasing cost of book provision. In addition to the two fund-related problems, he also conveys another aspiration in relation to the facility and equal opportunities for *madrasa* students to participate in the National Science

Olympic (*Olimpiade Sains Nasional*, OSN). "So far, the students of *madrasa* schools have not been invited to join the OSN" (Wakhid, via Supriyadi & Rachman, 2015). Therefore, there should be policies that involve MI in local, national, and even international competitions.

According to the headmaster of MI Al Bidayah, Kholid Mawardi (in Aji, 2015), based on the circulation letter by the Director of Elementary School Coaching, the Ministry of Education and Culture, at least there are five events that the students of MI might attend. In his opinion, the five events of students' competitions are the International Mathematics and Science Olympic (INSO), the National Students' Sport Olympic (*Olimpiade Olahraga Siswa Nasional*, O2SN), the National Students' Art Competition and Festival (*Festival dan Lomba Seni Siswa Nasional*, FLS2N), the Elementary School Students' Literature Appreciation (*Apresiasi Sastra Siswa Sekolah Dasar*, AS3D) and the Students' Art Creation Competition (*Lomba Cipta Seni Pelajar*, LCSP). MI schools actually expect that there will be equality in terms of achievements between the elementary schools and the *madrasa ibtidaiya* schools. However, in practice, there are still many views that doubt the achievements of MI school students. One of the evidence is the failure of three students from an MI school in the County of Semarang to attend the National Science Olympic (*Olimpiade Sains Nasional*, OSN) at the provincial level whereas the three students won the OSN at the county level. Therefore, MI students who have achievements should be involved in any competitions.

The entry of MI into the part of National Education System does not decrease the act of discrimination toward *madrasa*. Madrasa schools, or in this case MI schools, still suffer from unfair treatment in displaying the quality of their students in the competitions. The reason was that *madrasa* schools have been considered less qualified for such a long time; as a result, the students of general elementary schools have uncertainty and inconfidence when they have to compete with those of *madrasa* schools which always suffer from drawbacks, disparity, and marginalization. The disparity perhaps is one of the reasons why *madrasa* schools have been labeled as second-class institutions since the operationalization of *madrasa* schools has been in very low condition. However, *madrasa* schools are educational institutions that have been closer to the lower class people because most of these schools are located in the villages.

The observers of *madrasa* are frequently unfair in comparing *madrasa* schools and general schools. The achievements of *madrasa* schools in delivering their students to study highly religious values and nobility so that the students will not be trapped in the underworld that has been widely spread nowadays have gained less attention and even have been forgotten. The phenomena found in the field show that the students of *madrasa* schools, starting from the *madrasa ibtidaiya*, *madrasa* tsanawiya to *madrasa* aliya, do not commit any worse juvenile misbehavior. This aspect might be one of the *madrasa* strengths in comparison to general schools; in other words, the students of *madrasa* have the characteristics called as *al-akhlaq-al karimah* (Suprayogo, 2007, p. 7).

On the other hand, there is a different reality, namely, the appearance of *madrasa* schools that frequently gain achievements in the city, county, provincial, or even in the national level. According to Kuswanto (2015) in the official blog of MIN Purwokerto, in the Decree of Decision by the Directorate General of Islam Education Number 4472 Year 2015 regarding the *Madrasa* Science Competition, the *Madrasa* Art and Sport Competition, the 2015 National Expo and Science Paper Competition, MIN Purwokerto Central Java won the silver medal in the national mathematics and science competition. MIN Purwokerto is one of the MI schools in the Province of Central Java that frequently gains achievements in several competitions.

The achievements that have been gained by the students of MIN 1 Purwokerto in Central Java province, as having been stated by the Supervisor of Islam Education for the Sub-County of East Purwokerto, Rustanto, in his speech, make the people in the Sub-County of East Purwokerto proud of the *madrasa* school students. In addition, MIN Purwokerto won the 1st Place of 2013 Male English Speech Competition in the Province of Cen-

tral Java and the 2nd Place of 2015 *Musabaqah Tilawatil Quran* (MTQ, an Indonesian Islamic religious festival held at national level, aimed at glorification of the Quran, in which the participants compete at reciting al-Quran employing *qira'at*–a method of recitation) competition in the Province of Central Java. These achievements are the evidence that the students of MI schools might gain achievements. The improvement in MIN Purwokerto becomes a solid proof that *madrasa* schools will have better performance than general schools if madrasa schools are well-managed.

According to Jamun Effendi, the Head of *Madrasa* Education, the Regional Office of Ministry of Religious Affairs for the Province of Central Java, in an article of *Sejahtera* Magazine published by the Sub-Department of Information and the Human Relationship for the Regional Office of Ministry of Religious Affairs for the Province of Central Java, since the issuance of Law No. 20 Year 2003 of Republic of Indonesia on National Education System, the position of *madrasa* schools become equal to that of other formal schools. The position of *madrasa* schools becomes stronger and more robust after the issue of Law No. 14 Year 2005 of Republic of Indonesia about Teachers and Lecturers, the Government Regulation No. 19 Year 2005, on National Education Standard, and the Government Regulation No. 55 Year 2007 on Religion Education and Religious Education. The development of *madrasa* schools in the Province of Central Java has been very rapid. In 2005 the number of *Madrasa ibtidaiya* schools was 3,903 units. The very rapid development has been followed by the interest of the society to entrust their children's formal education to madrasa schools. In 2013/ 2014, the number of *madrasa ibtidaiya* schools was 1,296,315 units, while in 2014/2015 the number was 1,347,206 units. This situation shows that the society has increased their trust toward madrasa schools in terms of their children's education.

The government gives positive responses by planning to turn the private status of *madrasa* schools into state status through the Ministry of Religious Affairs, especially for madrasa schools located in the far remote area, as an effort to increase and equalize the *madrasa* quality. However, the government should pay attention to the exclusiveness of these *madrasa* schools. According to the Director of *Madrasa* Education, Kholis (2015), the Ministry of Religious Affairs has issued Regulation of the Minister of Religious Affairs, No. 14 Year 2014, on Establishment of Government-Held Madrasa Schools and the Changing of Private Status into State Status for Society-Held Madrasa Schools, the status change is conducted in order to ensure the quality improvement of institution-based *madrasa* schools. In the Ministery of Religious Affairs Regulation Number 14 Year 2014 Verse 1 Article 3, it is stated that a *madrasa* school is a formal educational unit under the supervision of Ministry of Religious Affairs that holds general education with Islam's uniqueness and one of such schools is *madrasa Ibtidaiya* or MI.

The explanation describes that *madrasa* school has actually has an equal opportunity with general school and thus to be a quality Islamic education institution will be easier for *madrasa* schools. The opportunity might be found in the recognition that takes the form of laws and government regulations. However, improving the quality of *madrasa* schools is not easy because it demands revision in multiple aspects and support from all stakeholders. The societal image of the madrasa schools has been better; madrasa schools that used to be labeled as the second-class institution now might not be underestimated. Some madrasa schools have shown their achievements in the local, national, and even international level competitions. With their uniqueness, they have started to prove themselves as an educational institution that might be equal to general schools and even be greater than general schools.

According to Supaat (2011, pp. 166–167), the materials of Islam Education in accordance with the national curriculum only provide elementary or very limited religious knowledge and, therefore, the materials are considered not suitable for the identity of *madrasa* schools. Based on this fact, all madrasa schools add the Islamic education subject with several reviews in order to enrich the

Islamic content. This action impacts the curriculum of MI schools in the form of additional lesson (overload) that leads to the increasing burden of the students (overburden). Meanwhile, the peculiarity of Islam education is as follows: (1) an education that has been encouraged by the teachings of Islam; and (2) an education that has been encouraged by the religious situations. The uniqueness of *madrasa* schools is that in addition to developing the cognitive, affective and psychomotoric aspects, madrasa schools also develop the spiritual aspect. The indicator of spiritual aspect development in madrasa schools is the existence of Islamic Education development. In other words, MI schools have their different characteristics, namely having spiritual aspect development or Islamic Education development.

Looking at the different characteristics of *madrasa* schools, MI schools have an added value. According to Jubaedi (2013), the most prominent peculiarity that MI schools have is that the students of MI schools are not involved in the drug abuse and the massive brawl. Up to now, the students of MI schools have not been involved in such negative behaviors. This peculiarity has been the evidence of spiritual aspect development. The schools that have spiritual aspect development will be different from general schools. Compared to general schools, MI schools demand a special evaluation instrument for measuring the quality of MI schools. Actually, in evaluating the quality of *madrasa* school achievements, there should be a different evaluation model because implementing the evaluation model similar to that of the general school will cause unfairness to the *madrasa* school. So far there has not been an effective MI evaluation model that evaluates the MI quality. Therefore, the recent study developed an evaluation model for measuring the MI quality.

Method

The study is a research and development study adopting the model suggested by Borg and Gall (1983, p. 772) who state that "Educational research and development (R & D) is a process used to develop and validate educational product."

This study used several methods namely descriptive, evaluative and survey methods. The descriptive method was used in the preliminary study in order to gather the data regarding the MI school condition. Then, the evaluative method was used in order to evaluate the process of a product development. According to Tyler in Mertens (2010, p. 56), an evaluation method attempts to develop certain objectives based on certain evaluation models that have been implemented. In this study, the researchers implemented the logic model that had been adapted to the evaluation model under study. Then, the survey method was applied in order to describe quantitatively the tendencies, the behaviors or the opinions of a population by gathering the sample of the population. Based on the sample, generalization was made (Creswell, 2010, p. 216). According to Neuman (2013, p. 343), through the survey method the researchers might gather accurate, reliable and valid information or data but the survey should be conducted under serious efforts and paradigm. In practice, the survey method was conducted by asking the respondents several questions regarding their opinions, characteristics or attitudes toward several matters that had been studied.

The nine steps of research and development efforts which have been proposed by Borg and Gall would be summarized into four stages as follows. In the first stage, the preliminary study was conducted, including gathering the information by means of both literature and field review of the description of well-qualified MI schools. In the second stage, the MI quality evaluation model (EQ-MI Model) was drafted, including the model, the guidelines, and the instrument. As having been explained, up to now there has not been any appropriate evaluation instrument that might be implemented for evaluating the MI school quality. Therefore, the draft of the instrument would be based on the results of observations by the researchers. After drafting the instrument, the researchers performed the feasibility test in terms of readability and the expert judgement. Next, in the third stage, the

model was tested, including the limited scale test in the Kedungwuni State *Madrasa Ibtidaiya* school located in the County of Pekalongan and the Sumurrejo State *Madrasa Ibtidaiya* School located in the Capitol of Semarang, and the expanded scale test in 10 state *Madrasa Ibtidaiya* schools in the Province of Central Java. Last but not least, in the the fourth stage, the final product was revised.

The study had three components that would be evaluated namely the components of Islamic boarding school input, the process quality, and the output of alumni. The component of input was related to the evaluation of the concept of Islamic boarding school (vision, mission and objectives), the competencies of *kiai*, the competencies of *ustaz,* and the facility. Then, the component of process quality included the learning evaluation and the assessment evaluation. Next, the component of output discussed the competencies of Islamic boarding school alumni.

The subjects in the study were the principals and the teachers of *madrasa ibtidaiya* schools. In gathering the sample, the researchers had certain considerations regarding the category of MI schools. The MI schools that had been selected as the population and the sample of the study were the state MI schools, not private MI schools, located in the Province of Central Java since the characteristics of both MI schools were different; state MI schools were developed by the government, whereas private MI schools were developed by the society. The sample of state MI schools was selected from ex-residence in the Province of Central Java. The sample was established using the stratified purposive sampling technique, a technique that had been implemented for gathering the sample based on several requirements (Borg & Gall, 1983, p. 248).

The overall number of respondents was 183 teachers and 9 principals of state MI schools. The data were gathered through a questionnaire, documentation, observation, and interviews. The questionnaire was used to attain the data regarding the respondents' attitudes towards the measurement of the MI school quality. The data gathered through observation, documentation and interviews were the data regarding the *madrasa* school characteristics such as the physical condition relation of state MI schools, the data regarding the MI school students' attitude in the short-term and the middle-term outcome and the data regarding the internal environmental condition of state MI schools.

The developmental study was conducted in two main stages. The first stage was the preliminary stage that included the activity of gathering information based on the theory and preliminary product development based on the results of field observation regarding the well-qualified MI schools. These activities were intended to design the draft of MI quality evaluation model. The second stage was expert judgement and field testing that consisted of the stage number five to number ten in the procedures proposed by Borg and Gall.

In the first stage or the preliminary activity, the researchers conducted a preliminary study and the developmental study. The preliminary activity was conducted in order to analyze the academic matters regarding madrasa schools and the needs for evaluation of the MI schools. In the second stage, the researchers performed the preliminary product development starting from the draft of model design and the instrument until the implementation experiment and analysis. The development of MIQEM began with the concept and the field condition by involving *Madrasa* education experts and practitioners. In order to produce a well-qualified instrument, the researchers implemented the expert validation in order to improve the model under development. The validation was conducted in order to get suggestions and criticisms to improve the MIQEM.

In the following step, the researchers performed the instrument testing in order to measure the validity, reliability and respondents' assessment regarding the MIQEM. The testing was performed in two stages, namely the limited scale testing and the expanded scale testing.

The instrument validity was measured in two aspects. The first aspect was the content validity, conducted through the consideration of *madrasa* teachers and principals, through the agreement of experts, and also

through the process of expert judgement. The second aspect was the construct validity conducted by performing instrument testing toward 30 teachers through the implementation of SPSS 16 software. The analysis technique used was the SPSS software.

In the limited scale testing, the reference used in measuring the validity of instrument quality was that the validity score should be > 0.30. According to Azwar (2007, pp. 179–181), there is a general agreement that the coefficient of validity might be considered satisfying if the validity score is above $T_{xy} = 0.30$. On the other hand, the reliability score of the instrument is in the range between 0.60 and 0.80 (Stevens, 1996, p. 372).

The data analysis technique used in the expanded scale testing was the descriptive statistical analysis through SPSS Version 17.00 Program and the Second Order Confirmatory Factory Analysis (CFA) through the Lisrel program. The Confirmatory Factor Analysis is a model designed under an assumption that the model would describe, explain or measure the empirical data in several relative parameters. The model was based on the information priority regarding the data structure in the form of theories, specifically or hypothetically (Joreskog & Sorbom, 1993, p. 22).

The *madrasa ibtidaiya* quality evaluation (MIQE) hypothetical model was tested empirically in the study and the testing included the input quality, the process quality, and the outcome quality. For the model theoretical fitness into the empirical data of the evaluation model in the study, the researchers referred to several criteria of Goodness of Fit (Sitinjak & Sugiarto, 2006, p. 65).

Findings and Discussion

Based on the stages of the study and the development that had been conducted, the results of the study are explained as follows. The preliminary stage was the preliminary study that had been conducted by means of literature review and field study and the preliminary stage resulted in the selection stage through the observation and reading activity. The stage finally narrowed down into the condition of *Madrasa* schools in the Province of Central Java, specifically into the condition of *madrasa ibtidaiya*. From the results of observation of the *madrasa* condition, most *madrasa ibtidaiya* schools have been marginalized in comparison with general schools because there is disparity in the educational quality between madrasa schools and general schools. On one side, madrasa schools have been more prominent in terms of academic and non-academic achievements in comparison to general schools. Madrasa schools that have national achievements are *Madrasa* Kudus, State *Madrasa* Purwokerto, State *Madrasa* Kedungwuni in the County of Pekalongan, and Muhammadiyah *Madrasa Ibtidaiya* Purbalingga. These *madrasa ibtidaiya* schools have been able to prove themselves as the well-qualified educational institutions.

The portrait of *madrasa* as an educational institution with different characteristics in comparison with general schools should also be evaluated differently. Madrasa schools not only teach general subjects but also develop the subjects of Islam Education in a wider scale namely: Hadits and Quran, Arabic Language, *Fiqih,* Islamic Theology and Character (*akidah akhlak*), and the History of Islam Culture. Therefore, in determining the well-qualified *madrasa* schools there should be different evaluation instrument.

In the next stage, the researchers performed the literature review in order to investigate and to review the theories that supported the model under development. The researchers conducted theoretical review from several theories related to the development of evaluation concept, logic model, quality and *madrasa* schools. The researchers adopted the logic model as the tool for evaluating the MI school quality. The sequence of logic model that the inspired the researchers was the basic logic model that had been adapted by reducing one of its components. The results of investigation toward the logic model showed that there were many variants of logic model. Therefore, the researchers referred to and adopted the basic logic model. In order to attain the valid data regarding the MI schools located in the Province of Central Java, the researchers also conducted another investigation by means of Education Management Information System (EMIS) toward the

Ministry of Religious Affairs in the Province of Central Java regarding *madrasa ibtidaiya* schools in 2012.

The field study was conducted specifically to the State *Madrasa Ibtidaiya* Kedungwuni, the County of Pekalongan, in August 2012. State *Madrasa Ibtidaiya* Kedungwuni was an elementary-*madrasa* school that became parents' choice. The researchers conducted observation and interviews in the state *madrasa* school in order to explore the data regarding the input, the process, and the output of *madrasa* schools. The preliminary study resulted in the data regarding several variables and indicators of well-qualified MI and these variables and indicators might be a matter of consideration for the parents who entrusted the education of their children to the state *madrasa ibtidaiya* schools.

According to Siti Fatimah, the Vice Principal for curriculum in State *Madrasa Ibtidaiya* Kedungwuni, the student enrollment was conducted selectively because the number of the participants that were always over the limit. The selection process involved the administrative aspect and the test, namely the selection of age and the test of basic capacity; as a result, every year State *Madrasa Ibtidaiya* Kedungwuni always refused the student candidates due to the limited capacity and the abundant number of entrance test participants. The society's high interest, responses and enthusiasm might be caused by the fact that the school had shown success in competition with other schools or other *madrasa* schools. The school had shown its superiority over other schools in its academic and non-academic achievements in the county level, the ex-residence level and the provincial level.

The academic and also non-academic achievements of the State *Madrasa Ibtidaiya* Kedung-wuni became an evidence that the MI school might gain several achievements; as a result, the vision of such *madrasa* school would be specific superiority. An example of the academic achievement that the state *madrasa* school had attained was that several students of State *Madrasa Ibtidaiya* Kedungwuni used to win the first place of optic olympic that was held by the Walisongo State Islamic Institution in February 2012. Other competitions that the students of State *Madrasa Ibtidaiya* Kedungwuni won were the mathematics olympic, the science olympic, the provincial level *Musabaqoh Tilawatil Quran* (MTQ), the first place of Nationally Standardized *Madrasa* Final Examination (*Ujian Akhir Madrasah Berstandar Nasional*, UAMBN), the first place of healthy school competition, the first place of boyscout competition, the badminton competition and the karate competition.

The next stage in the study was the planning stage. The objective of the planning stage was to elaborate the evaluation concept, the quality, the *madrasa* school, the logic model, the input, the process and the outcome. Then, the researchers defined all of the logic model components including the input, the process, and the outcome. In order to evaluate the MI quality, the next preparation was elaborating the objectives of performing the evaluation of the MI quality. Next, the researchers constructed m*adrasa ibtidaiya* quality evaluation model (MIQEM). The design of the evaluation model was developed from the logic model that referred to the theoretical foundation of quality, evaluation and *madrasa*h so that the components developed in the model included the input, the process, the short-term outcome and the mid-term outcome of MI schools. After that, the researchers elaborated the indicators of the input, the process and the outcome variables.

The preliminary product development of MIQEM started from the example of well-qualified MI as the initial concept that had been built on the theory and the field study. After viewing the indicators of effective and well-qualified school and State *Madrasa Ibtidaiya* Kedungwuni, the researchers designed a model that was adjusted to several theoretical concepts that had been designed previously. From these indicators, the researchers drafted the guidelines of *MIQEM*. In this stage, the researchers also designed the profile of the model, the manual of model implementation, the instrument, the assessment rubric, the manual of interview and the observation. The results of preliminary MIQEM are described in Figure 1.

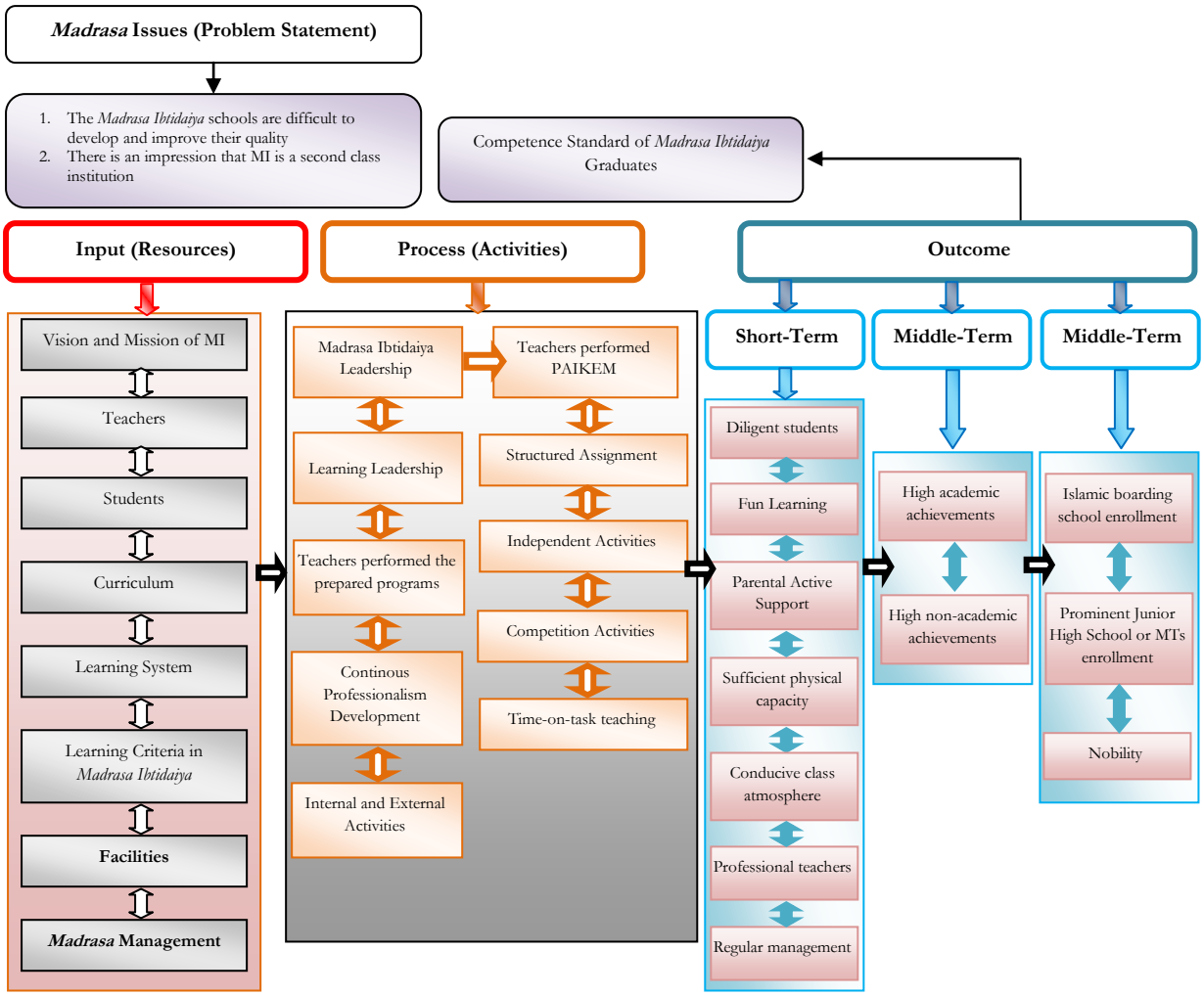

Figure 1. The result of preliminary *madrasa ibtidaiya* quality evaluation model (MIQEM)

Afterwards, the researchers performed the expert validation after the model had been drafted in the form of: model, manual of model implementation, guidelines, and instrument. The expert judgement involved several experts in the domain of evaluation, model development, methodology and basic educational concept. The feedback of the model aspects in the expert judgement according to the expert of model charts was as follows: the variables and the sub-variables should be put appropriately, had not been in a good order and should be grouped. Therefore, the researchers did a sub-variable reposition in the more appropriate places and rearranged the indicators in the input, process, and outcome variable.

The study did not need to include the long-term outcome. Therefore, the researchers revised the outcome variable by limiting the study into the middle-term outcome. In this stage, the researchers rearranged the MIQEM especially in the parts that belonged to the input, process, and output components.

After the instrument had been validated by the experts, the researchers performed preliminary testing. The preliminary testing that was implemented was the readibility test. The objective of implementing the readability test was to re-check whether or not the instrument that had been designed could be understood in terms of content, language, writing mechanics and measurement manner so that in the actual testing in the field the respondents would not be confused or would directly understand the content of the MIQEM instrument. In relation to the preliminary testing, the researchers performed the limited

scale testing with State *Madrasa Ibtidaiya* Kedungwuni located in the County of Pekalongan and State *Madrasa Ibtidaiya* Sumurrejo located in Gunung Pati, the Capital of Semarang, as the population. The sample was 30 teachers from the two *madrasa* schools. After conducting the limited scale testing toward the sample teachers, the researchers attained the empirical data and found that there were several items whose validity was below 0.30. As a result, instead of eliminating these items, the researchers revised these items for the next stage. The less-functioned or the less-valid items were improved in terms of language structure, substance and writing mechanics. The items whose validity was below standard of validity measures were improved maximally in this stage.

After the items had been improved, the researchers conducted the main testing in 10 state *madrasa ibitidaiya* schools in the Province of Central Java involving 183 teachers and 9 principals. Based on the standards had been determined by the researchers, the results of the modified development by Sudijono (2004, pp. 329–339), in the quantitative data analysis, the researchers found the mean score from the quantitative data that had been attained by means of evaluation instrument and convered the mean score into the five-point scale assessment; after that, the researchers described the data. Based on the results of the description, the researchers defined the basis for assessing the developed evaluation model. The conversion of the quantitative data into the qualitative data by means of five-point scale technique made use of the guidelines designed by Sudijono (2004, pp. 329–339). For the conversion guidelines, the researchers referred to the aspects summarized in Table 1 for defining the MI quality.

Based on the mean score and the classification of the results of the development, the evaluation model, the instrument and the evaluation manual as the results of development with the following assessment standards were assessed. The percentage 183 *madrasa* teachers' evaluation of the 10 state *madrasa ibtidaiya* schools by means of special instrument designed for the teachers are shown in Table 3, Table 4, and Table 5.

Table 1. The guidelines of qualitative data into quantitative data conversion

Formula	Mean Score	Classification
$X > \bar{x}_i + 1{,}8\, x\, sb_i$	> 4.2	Very Good
$\bar{X}_i + 0{,}6\, x\, sb_i < X \leq \bar{x}_i + 1{,}8\, x\, sb_i$	>3.4 – 4.2	Good
$\bar{X}_i - 0{,}6\, x\, sb_i < X \leq \bar{x}_i + 0{,}6\, x\, sb_i$	> 2.6 – 3.4	Moderate
$\bar{X}_i - 1{,}8\, x\, sb_i < X \leq \bar{x}_i - 0{,}6\, x\, sb_i$	> 1.8 – 2.6	Poor
$X > \bar{x}_i - 1{,}8\, x\, sb_i$	≤ 1.8	Very Poor

Table 2. Standards of assessment

Mean Score	Classification	Conclusions
> 4.2	Very Good	Very Well-Qualified
> 3.4 – 4.2	Good	Well-Qualified
> 2.6 – 3.4	Moderate	Moderate
> 1.8 – 2.6	Poor	Ill-Qualified
≤ 1.8	Very Poor	Very Ill-Qualified

Table 3. Percentage of teachers' evaluation of the MI input

Mean Score	Input	Conclusions
> 4.2	11,1 %	Very Well-Qualified
> 3.4 – 4.2	84.2 %	Well-Qualified
> 2.6 – 3.4	4.7%	Moderate
> 1.8 – 2.6	0.0%	Ill-Qualified
≤ 1.8	0.0%	Very Ill-Qualified

Table 4. Results of assessment percentage from the teachers toward the MI process

Mean Score	Process	Conclusions
> 4.2	10.70%	Very Well-Qualified
> 3.4 – 4.2	87.90%	Well-Qualified
> 2.6 – 3.4	1.50%	Moderate
> 1.8 – 2.6	0.0%	Ill-Qualified
≤ 1.8	0.0%	Very Ill-Qualified

Table 5. Results of teachers' assessment percentage toward the MI outcome

Mean Score	Outcome	Conclusions
> 4.2	8%	Very Well-Qualified
> 3.4 – 4.2	85%	Well-Qualified
> 2.6 – 3.4	7%	Moderate
> 1.8 – 2.6	0%	Ill-Qualified
≤ 1.8	0%	Very Ill-Qualified

The results of the measurement of the model fitness based on the results of printed out data analysis by means of LISREl 8.50 showed the following data of EQ-MI model fitness results. The results of Measurement Model Fit Test in the input construct were that the Root Mean Square of Approximation (RMSEA) were equal to 0.048 and the Goodness of Fit (GFI) score was equal to 0.828 or $0.82 \leq GFI \geq 0.90$; in other words, the model was marginally fit. Then, the Comparative Fit Index (CFI) score was equal to 0.912 and the Incremental Fit Index (IFI) score was equal to 0.914; in other words, the model was fit into the data. Therefore, the researchers concluded that the input variable construct model that was proposed was fit into the data. The results of measurement model fitness test in the process construct showed that the Root Mean Square Error of Approximation (RMSEA) was equal to 0.051, the Non Normed Fit Index (NNFI) score was equal to 0.822, the CFI score was equal to 0.84 and the IFI score was equal to 0.845; in other words, the model was fit into the data. In the outcome construct, the Root Mean Square Error of Approximation (RMSEA) was smaller than 0.080, namely $0.050 < 0.080$, the Comparative Fit Index (CFI) score was bigger than 0.900, namely $0.931 > 0.900$, the Non Normed Fit Index (NNFI) was also bigger than 0.900, namely $0.916 > 0.900$. In other words, the developed model was fit for correcting the NNFI measure; as a result, the model was fit into the data.

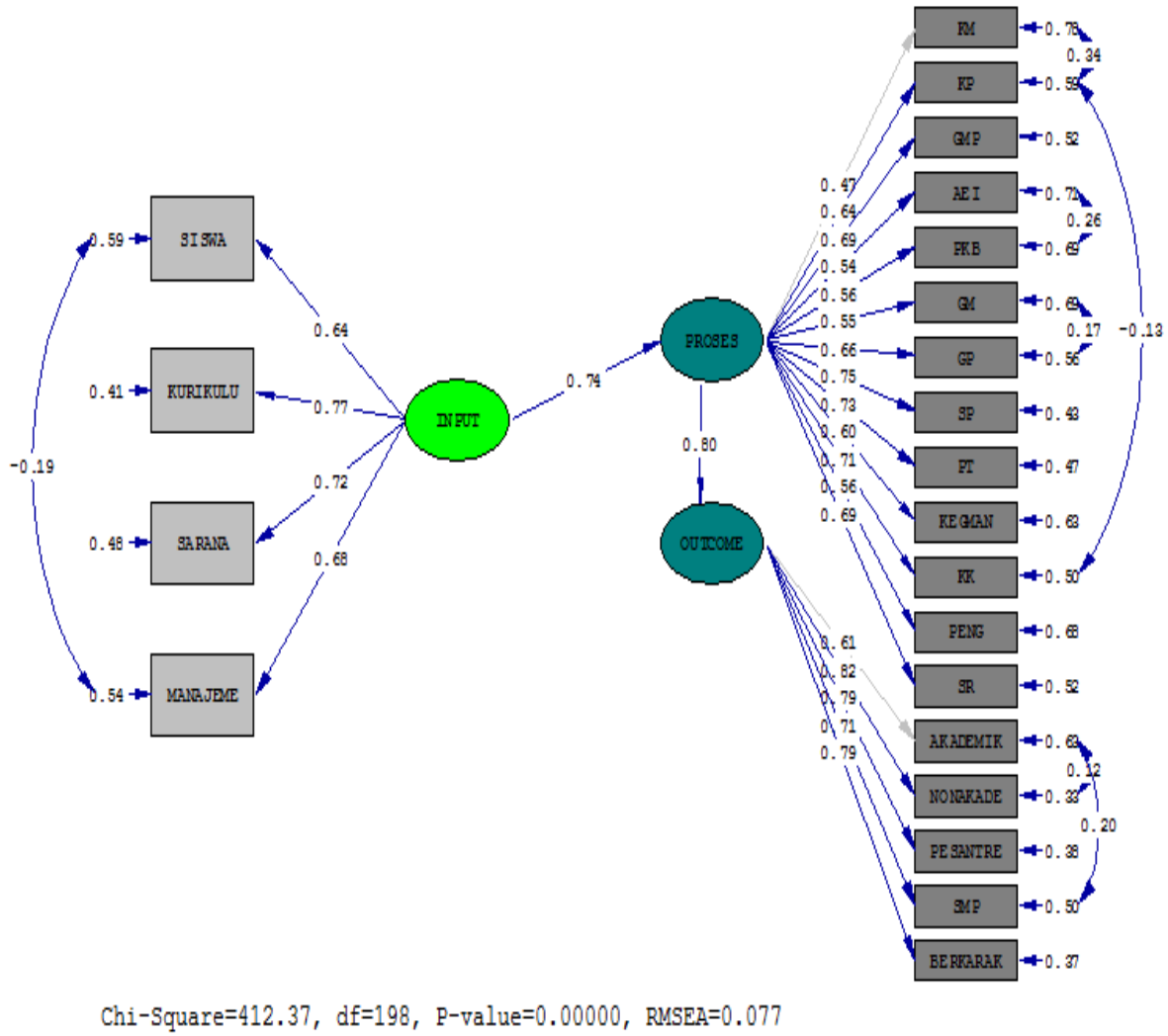

Figure 2. Results of model fitness

Based on the results of the analysis of all of the MIQEM components including the input, process and output, MIQEM feasibility or fitness testing was performed to identify the inter-variable causality. The results of fitness testing indicated that the model was fit into the data. The results were explained by the Comparative Fit Index (CFI) score of 0.907≥0.900, the Root Mean Square Error of Approximation (RMSEA) score of 0.077 ≤ 0.800 and the Root Mean Square Residual (RMSR) of 0.018 < 1.000. There were three criteria that had been met; in other words, the model was fit into the data. The results of the significance testing of the estimates of overall loading factor coeffcient showed that the significance and standardized coefficient of the loading factor of each indicator had provided the above cut-off value score (the above cut-off value should be at least 0.500). Based on the results of measurement model fitness, the final figure of MIQEM can be seen in Figure 2. The results of hypothesis testing in the study were as follows.

The Input Has a Significant Positive Effect on the Process

The structure in Figure 2 shows that from the relationship between the input and the process, the researchers found the correlation value equal to 0.740 and the t-count equal to 5.780. That the t-count was bigger than the t-table (5.780 > 1.977) showed that the variables were significant and the coefficient score was positive, thus the researchers concluded that the hypothesis which stated that the input had a significant positive effect on the process was "accepted". The effect of the input on the process was 54.7%, and the other 55.3% effect on the process came from other variables.

The Process Has a Significant Positive Effect on the Outcome

The structure in Figure 2 shows that from the relationship between the process and the outcome. The researchers found the correlation value was equal to 0.803 and the t-count was equal to 5.262. That the t-count was bigger than the t-table (5.262>1.977) showed that the variables was significant and the coefficient score was positive, and thus the researchers concluded that the hypothesis which stated that the process had a significant positive effect on the outcome was "accepted".

Based on the results of model fitness test, the second stage revision was performed. This stage was the final stage of MIQEM development and in this stage the researchers viewed the results of the tryout of a big amount of data in each variable. The instrument items whose score was still below standards were eliminated and omitted. The reason was that the model was not fit into the data if the researchers kept these items. The items which had low validity were considered to be malfunctioned or not providing contribution toward the model; as a result, the model was not fit into the data. In this stage the researchers also rearranged the model instrument so that it would be a model that might be implemented by the *madrasa ibtidaiya* schools. The researchers rearranged the instrument in each variable. Then, in several dimensions there were some eliminated items and, as a result, the number of the items decreased. After the instrument had been rearranged, the MIQEM was rearranged as well because there was one dimension that should be omitted whereas the dimension used to be the part of the model.

After several revisions and tests, the MIQEM could be implemented for evaluating the quality of the *madrasa ibtidaiya* schools. By implementing the MIQEM, the users might define whether *madrasa ibtidaiya* schools were qualified or not. The final MIQEM is presented in Figure 3.

Conclusion and Suggestions

Based on the results of the study, several conclusions are drawn. First, MIQEM was developed by performing the product selection, reviewing and drafting the model. Then, the researchers have also performed the preliminary product development, expert validation, limited scale experiment, first stage revision, expanded scale experiment, and second stage revision as part of the model development. As a result, the researchers have attained the MIQEM.

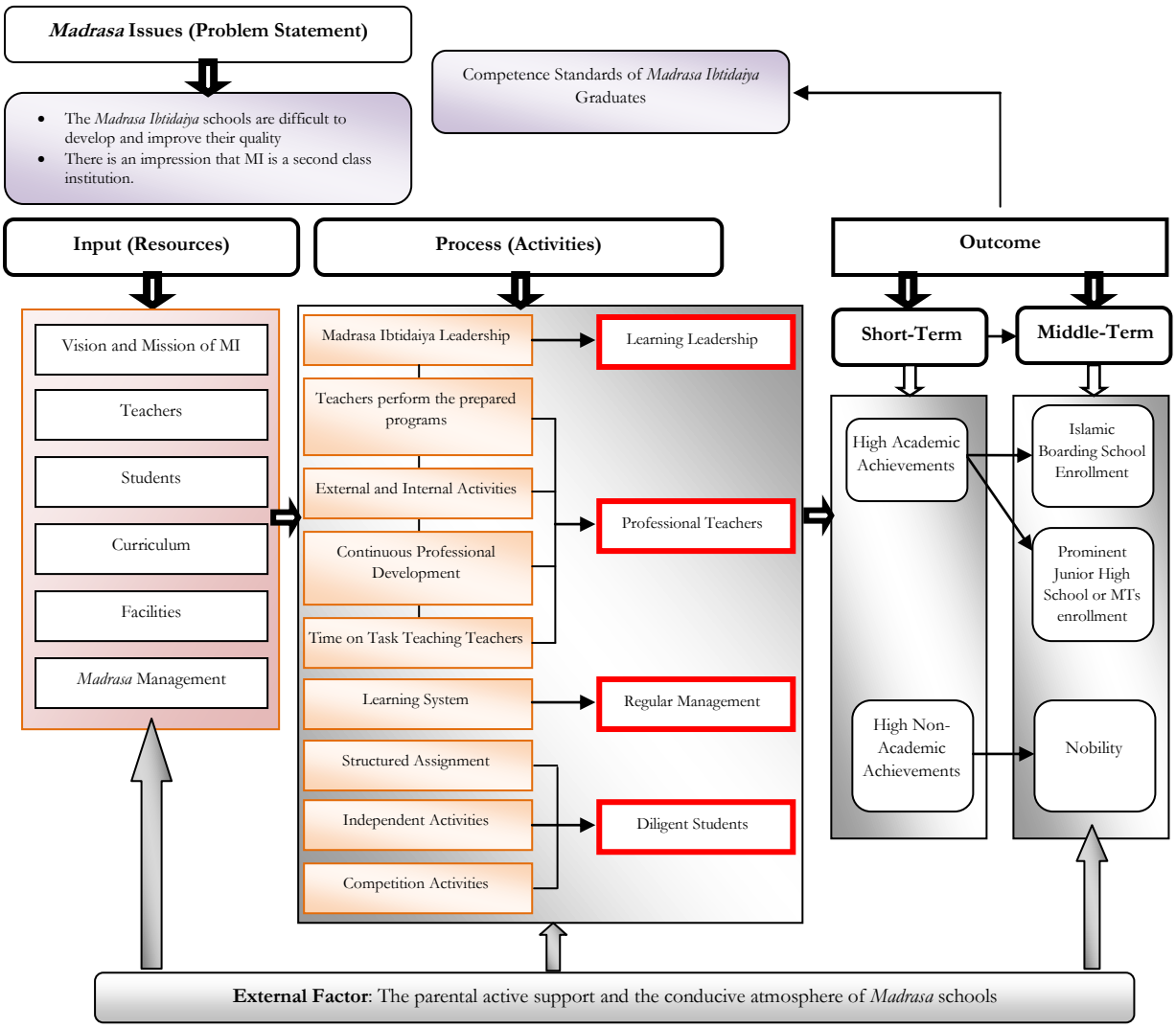

Figure 3. The final *madrasa ibtidaiya* quality evaluation model (MIQEM)

The implementation of the MIQEM was conducted by the teachers and the principals of *madrasa* schools by making use of the MIQEM instrument. The results of the implementation might be used for improving *madrasa* school quality. The results of model fitness test indicated that the model was fit into the data. These results are explained by the Comparative Fit Index (CFI) score that is equal to 0.907, the Root Mean Square Error of Approximation (RMSEA) score that is equal to 0.018 and the Root Mean Square Residual (RMSR) score that is equal to 0.018. The results of meaningfulness test toward the overall estimates of loading factor coefficient in is very significant and the score of each indicator is above the cut-off value which should be at least 0.50. Based on the results of readability test, the expert judgement and the limited-scale experiment and the expanded scale experiment in the field, the MIQEM could be considered effective in measuring MI quality.

The implication of the study is that the MIQEM will provide information regarding MI schools so that it could be used for improving the quality of MI schools, including the input, the process and the outcome. The researchers suggest that the government may implement the developed MIQEM to measure and improve MI school quality.

References

Aji, R. (2015). Juara olimpiade sains dari MI merasa dibegal maju ke provinsi. *Tribun Jateng*. Retrieved from http://jateng.

tribunnews.com/2015/03/10/juara-olimpiade-sains-dari-mi-merasa-dibegal-maju-ke-provinsi

Azwar, S. (2007). *Metodologi penelitian*. Jakarta: Pustaka Pelajar.

Borg, W. R., & Gall, M. D. (1983). *Educational research: An introduction* (4th ed.). New York, NY: Longman.

Creswell, J. W. (2010). *Research design qualitative, quantitative, and mixed methods approaches* (A. Fawaid,). Yogyakarta: Pustaka Pelajar.

Government Regulation No. 19 Year 2005, on National Education Standard, article 10 (2005). Republic of Indonesia.

Government Regulation No. 55 Year 2007 on Religion Education and Religious Education (2007). Republic of Indonesia.

Joreskog, K. G., & Sorbom, D. (1993). *Lisrel 8: Structural equation modeling with the simplis command language*. Chicago, IL: Scientific Software International.

Jubaedi, D. (2013). Siswa madrasah tidak terlibat narkoba. Retrieved September 11, 2017, from https://www2.kemenag.go.id/berita/118692/siswa-madrasah-tidak-terlibat-narkoba?lang=id

Kholis, N. (2015). Kemenag terbitkan PMA penegerian madrasah. Retrieved from https://www2.kemenag.go.id/berita/200640/kemenag-terbitkan-pma-penegerian-madrasah

Kuswanto, A. (2015). Meriahnya lomba kompetensi siswa MI, MIN Purwokerto juara umum. Retrieved October 30, 2017, from http://minpurwokerto.blogspot.co.id/2015/11/meriahnya-lomba-kompetensi-siswa-mi-min.html

Law No. 14 Year 2005 of Republic of Indonesia about Teachers and Lecturers (2005).

Law No. 20 Year 2003 of Republic of Indonesia on National Education System article 17 (2003).

Mertens, D. M. (2010). *Research and evaluation in education and psychology*. California, CA: Sage.

Neuman, W. L. (2013). *Social research methods: Qualitative and quantitative approaches* (E. T. Sofi). Jakarta: Indeks.

Regulation of the Minister of Religious Affairs, No. 14 Year 2014, on Establishment of Government-Held Madrasa Schools and the Changing of Private Status into State Status for Society-Held Madrasa Schools (2014). Republic of Indonesia.

Regulation of the Minister of Religious Affairs, No. 90 Year 2013, on the Curriculum of Madrasa 2013 of the Subject of Islamic Religion Education (2013). Republic of Indonesia.

Sitinjak, J. R. T., & Sugiarto. (2006). *Lisrel*. Yogyakarta: Graha Ilmu.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Sudijono, A. (2004). *Pengantar statistik pendidikan*. Jakarta: Raja Grafindo.

Supaat, S. (2011). Transformasi madrasah dalam sistem pendidikan nasional. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *15*(1), 155–186. https://doi.org/10.21831/pep.v15i1.1092

Suprayogo, I. (2007). *Quo vadis madrasah gagasan, aksi dan solusi pembangunan madrasah*. Yogyakarta: Hikayat.

Supriyadi, E., & Rachman, T. (2015). Hidayat Nur Wahid: Perbedaan anggaran sekolah umum dan madrasah tak wajar. *Republika*. Retrieved from http://www.republika.co.id/berita/mpr-ri/berita-mpr/15/08/14/nt1psy219-hidayat-nur-wahid-perbedaan-anggaran-sekolah-umum-dan-madrasah-tak-wajar

REiD (Research and Evaluation in Education), 3(1), 2017, 42-49
Available online at: http://journal.uny.ac.id/index.php/reid
Research Article

Exploring the construct of school readiness based on child development for kindergarten children

***[1]Farida Agus Setiawati; [2]Rita Eka Izzaty; [3]Agus Triyanto**
*Faculty of Education, Universitas Negeri Yogyakarta
Jl. Colombo No. 1, Depok, Sleman 55281, Yogyakarta, Indonesia
*Email: farida_as@uny.ac.id

Submitted: 13 April 2017 | Revised: 05 June 2017 | Accepted: 06 June 2017

Abstract

Indonesian government has regulated that the basic age of readiness of a child to attend elementary schools is 7 years old. In fact, some children are not exactly 7 years old when they first go to school because they develop more rapidly. This study is aimed at investigating some aspects of child development which affect their readiness to attend elementary school. The subjects were 101 grade 1, 2, and 3 teachers of elementary schools in Yogyakarta, a special Region in Indonesia. The data were collected through interviews. The results of the data collection were analyzed using both descriptive quantitative and qualitative techniques. The results of the study show some aspects of child development affecting their readiness to attend elementary schools, including: cognitive and language ability, social emotional skills, fine motor skills, gross motor skills, arts, religion and moral values, and some others. Beside these aspects, some problems in grades 1, 2, and 3 are also found. This study is expected to give significant indicators to create the construct of school readiness.

Keywords: *school readiness, elementary schools, child development*

How to cite item:

Setiawati, F., Izzaty, R., & Triyanto, A. (2017). Exploring the construct of school readiness based on child development for kindergarten children. *REiD (Research and Evaluation in Education), 3*(1), 42-49. doi:http://dx.doi.org/10.21831/reid.v3i1.13663

Introduction

Elementary school admission is conducted every year. When kindergarten age children begin to attend elementary school, parents usually curious and frequently raise a question of what skills and capability their children must have to be considered as having school readiness.

The government has already regulated new student admission through Government Regulation No. 17 Year 2010 on educational management and implementation (2010) and Ministry of National Education (2011) regulation number 04/VI/PB/2011, consisting of national education regulations of new student admission in kindergarten or primary schools. This regulation says that the main indicator of new student admission is being 7-12 years old. However, some problems which arise are that many children who were born on certain months cannot be exactly 7 years old when the admission is conducted. Moreover, there are also some children who develop more rapidly in their age. Therefore, even though they are not exactly 7 years old, they have more rapid development at this age.

School readiness is a factor which plays the most important role for chidren to suceed in the learning process. School readiness consists of children's ability in various aspects that a human has, including emotion, cognition, language, social, and motor skills. The concept of school readiness has been widely

defined and redefined by experts through various points of view. Some theories of child development and learning are employed to explain the meaning of school readiness. From some existing definitons, there are two types of meaning for readiness. The first meaning is studying readiness which consists of the description of children's readiness to be involved in learning physical material. The second is school readiness which consists of the description of capabilities from various aspects that human has, such as cognitive, language, social, and motor skills which are related to the curriculum which is used (Lewit & Baker, 1995). Moreover, Dockett & Perry (2002) add that chronological age must also be considered, because the finding of research indicates that chronological age has a positive correlation with the mental readiness and the development of each individual.

The National Association for the Education of Young Children in Dockett & Perry (2002) states that in determining children's school readiness, the policy makers must consider these three aspects: (1) considering the children's existing experience, including skills or abilities possessed, in order to be able to predict whether children can be involved in learning activities in higher educational level or not. (2) Realizing individual differences in children, including the language and culture differences used. (3) Employing appropriate wishes and reasons of chidlren's ability which must be fullfilled as one of school readiness requirements.

From some cumpulsiary considered aspects, *Technical Planning Group* (in Dockett & Perry, 2002), has identified some dimensions which become indicators in conducting an assessment of children's school readiness, including: (1) Motor development and physical condition; (2) Social and emotional development; (3) Learning approaches; (4) Language use; (5) Developmet of cognition and general knowledge.

Some experts have created instruments to detect learning readiness in elementary schools. Chew created an instrument named Lollipop test which measures school readiness (Lemelin & Boivin, 2007). This instrument was created to detect children's school readiness in France. Some aspects which are measured by this test highly emphasize children's cognitive skills. Various indicators explained by this instrument include: (1) Color and shape identification; (2) Spatial recognition; (3) Number identification and calculation; (4) Letter identification and writing system. These four development indicators are given to children in the form of an interview or an oral test.

Another instrument developed to detect children's school readiness is Early Development Instrument (EDI). This instrument was developed by Janus et al. (2006). This test reveals some areas of child development, including: (1) physical health and well-being, (2) social competence anf emotional maturity, (3) language and cognitive development, (4) communication and general skills.

In Indonesia, the Minister of National Education has created some standards to be applied for early childhood education. The standards are accommodated by the Ministry of National Education (2014) in the minister regulation number 137 year 2014. One of the standards included in this regulation is standard 1. This standard contains basic task of development, which must be taught to children in each development stage. At the end of early childhood, or age 5-6, some of the developmental characteristics which have been possessed are: (1) Religious and moral values; (2) Physical-motor skills, including gross motor skill, fine motor skill, health and safety behavior; (3) Cognitive, including learning and problem solving, logical and symbolic thinking; (4) Language skill, including language comprehension, language expression skills, and literacy; (5) Social-emotional, consisting of self-awareness, sense of responsibility to self and others, and also pro-social behaviors; (6) Arts, consisting of children's ability to enjoy various songs, melodies, or voices, and interest in art activities.

The regulation shows that there are various dimensions or aspects that have to be taught to early children. Those various aspects have become comptencies which are expected to appear in 5-6 year-old children. However, among those afore-mentioned aspects, the extent aspect which has a significant role in

preparing school readiness has not been revealed yet. Therefore, it is crucial to identify the aspects which contribute to preparing children's school readiness. If certain aspects do not appear or are not in children's characters, then there will be a lot of problems which have to be encountered by children while studying in elementary schools.

The afore-mentioned explanation also portrays the existence of differences which underlie school readiness. For instance, the opinion of Dockett & Perry (2002) is different to Lemelin & Boivin (2007). This difference is affected by teacher and expert perceptions in determining the developmental aspects which underlie children's school readiness. In addition, the situation, condition, and culture of learning processes also influence general readiness. Therefore, examining the aspects which affect school readiness based on distinctive characteristics of a region is significant. Thus, two research questions for the first year of this research are proposed: (1) What are the developmental aspects which affect children's school readiness in elementary schools in Indonesia? (2) What are the problems related to children's lack of readiness to attend elementary schools?

Method

This research is qualitative and quantitative research. The type of the research in this first year is exploratory and that in the second year is developmental.

The subject of this research was 101 elementary school teachers of early grades from and four districts in Yogyakarta Special Region, including Bantul, Sleman, Kulonprogo, and also Yogyakarta municipality. The data collecting technique which was used in this research was snowball sampling technique. This technique was based on the previous sample investigation. The investigations began with the key person of elementary teacher group, or *Kelompok Kerja Guru (KKG)* in Indonesian language, board management. They were asked to select some teachers in each elementary school. After identifying the teachers, the researchers interviewed them by phone.

This research was a survey which was aimed at revealing the developmental aspects which affect children's readiness to attend elementary schools. In collecting the data, the researchers employed an interview technique with the assistance of interview guidelines, and the questions about the aspects of school readiness and the problems of the students who were not ready yet to attend elementary school. The interview results were identified, and the collected information was coded and extracted. Therefore, the data revealed some developmental aspects which played important roles in the children's school readiness. In analyzing the data, descriptive analysis technique was employed. The revealed data were analyzed by percentace of subjects' responses.

Findings and Discussion

Aspects of School Readiness

The findings of extracted data reveal some factors or dimensions which underlie children's learning readiness to attend elementary schools. The findings are presented in Table 1.

Based on Table 1, there are 29 child performances which influence school readiness. Of the 29 aspects, the five most influential aspects are concentration (15%), imitating movements (running, jumping, standing on one leg) and dancing (9.82%), team work (9.22%), recognizing letters (8.62%), and also reading/comprehending reading texts (7.21). On the other hand, the five lowest aspects are obeying the rules/discipline (0.4%), drawing curves, straight lines, circles, as well as squares (0.4%), religious activities (praying, charity) (0.2%), sharing/helping others (0.2%) and creativity (0.2%).

The result shown in Table 1 is child performances which influence readiness. The performances are classified into six developmental aspects in the activity programs of kindergartens: cognitive, social emotional, and fine motor skills; gross motor skills; fine motor skills; art; religion and moral; and other aspects.

Table 1. Child performance which influences school readiness

Child Performances	Data Finding (%)	Aspects of Development
Concentration	15	Other Factors
Imitating movements (running, jumping)	9.82	Gross Motor
Team work	9.22	Social Emotional
Recognizing letters	8.62	Cognitive
Reading/comprehending reading texts	7.21	Cognitive
Recognizing numbers	6.81	Cognitive
Parents (caring, breakfast, children's readiness)	5.61	Other Factors
Socialization with peers	4.61	Social Emotional
Classroom adaptation	4.21	Social Emotional
Fine motor (snipping and sticking)	4.21	Fine Motor
Vocabulary	3.81	Language
Writing	3.41	Fine Motor skills
Following rhythms, sounds, and tones	2.81	Art
Story telling	2.2	Language
Independence	2	Social Emotional
Apprehending the rules	1.6	Social Emotional
Age	1.4	Other Factors
Apprehending instruction and information	1.2	Cognitive
Counting	1	Cognitive
Fluent speaking	1	Language
Problem solving	0.8	Cognitive
Drawing	0.8	Art
Holding stationery	0.6	Fine Motor skills
Recognizing colors and their uses	0.6	Cognitive
Obeying rules/ discipline	0.4	Social Emotional
Drawing curves, straight lines, circles, and squares	0.4	Fine Motor skills
Religious activities (praying, reading Al Quran, charity)	0.2	Religious and Moral Values
Sharing/helping others	0.2	Social Emotional
Creativity	0.2	Art

Table 2. Child performances on cognitive which influence school readiness

Child Performances	%
Reading/comprehending reading texts	7.2
Recognizing numbers	6.8
Recognizing letters	8.6
Counting	1.0
Vocabulary building	3.8
Apprehending instruction and information	1.2
Story telling	2.2
Fluent speaking	1.0
Problem solving	0.8
Recognizing colors and their usage	0.6
Sum	33.5

Based on Table 2, there are 10 cognitive child performances. The three highest cognitive performances of school readiness are recognizing letters (8.6%), reading/comprehending texts (7.2%), and recognizing numbers (6.8%). The following cognitive aspects are vocabulary building (3.8%), story-telling (2.2%), comprehending instructors, and information (1.2%), and fluent speaking, and also counting. The two lowest ranks are problem solving (0.8%) and recognizing colors (0.6%).

Table 3. Child performances of children's readiness based on social emotional aspects

No.	Social Emotional Performances	%
1.	Team work	9.2
2.	Socialization with peers	4.6
3.	Classroom adaptation	4.2
4.	Independence	2.0
5.	Obeying rules / discipline	2.0
6.	Sharing/helping others	0.2
	Sum	22.3

Table 3 shows that there are six child performances of school readiness based on social emotional aspects. The highest performance is team work (9.2%), the second rank is socialization with peers (4.6%) followed by classroom adaptation (4.2%), independence (2.0%), and obeying rules/discipline (2.0%). Meanwhile, the lowest rank is sharing/helping others (0.2%).

Table 4. Child performances of fine motor skills children's which influence school readiness

Performance of Fine Motor Skills	%
Cutting and pasting	4.2
Writing	3.4
Holding stationery	0.6
Drawing curves, straight lines, circles, and squares	0.4
Sum	8.6

Table 4 shows that there are four child performances of school readiness based on fine motor skills, including cutting and pasting, writing, holding stationery, drawing some curves, straight lines, and also squares. The other motoric skills are gross motor skill (see Table 5), which has only one aspect, namely: imitating movements (including running, jumping, standing on one leg) or dancing (9.8%), and this aspect becomes the second highest aspect.

Table 5. The performance of school readiness based on gross motor skills

No.	Performances of Gross Motor Skills	%
1.	Imitating movements (running, jumping, standing on one leg) or dancing	9.8

Table 6. Child performances of school readiness based on art

No.	Child Performances of Art	%
1.	Following rhythms, sounds, and tones	2.8
2.	Drawing	0.8
3.	Creativity	0.2
	Sum	4.8

The art aspect is divided into three points (as presented in Table 6). The highest is following rhythm, sounds, and tones (2.8%), followed by drawing (0.8%) and then creativity (0.2%).

Table 7. Child performances of school readiness based on religion and moral

Religion and Moral Performance	%
Religious activities (praying, reading Al Quran, charity)	0.2

As presented in Table 7, the next aspect is that of religion and morality which consists of religious activities (0.2%). Such phenomenon needs further identification because religion and moral aspect, which is one of the aspects of school readiness, has the lowest percentage among the other influential aspects.

Table 8. Child performances of school readiness based on other aspects

Other Performance	%
Concentration	15.0
Parents' Caring	5.6
Age	1.4
Sum	22.0

The last aspect of school readiness as presented in Table 8 is dealing with other aspects which are not appropriately categorized into one of the existing aspects of development. This category is divided into three parts, namely concentration (15%), parents' caring (5.6%), and age (1.4%). The three aspects give great contribution, especially the aspects of concentration and parents' caring. In fact, the most important requirement of student admission in Indonesia is age. Therefore, the role of the other aspects, in particular concentration and parents' caring, must also be considered.

The findings show a number of developmental aspects which contribute to children's learning readiness to attend elementary school. In terms of roles, ordered from the most to the least dominant, the roles are cognitive, social-emotional, other (concentration, parents' caring, and age), gross motor skills, fine motor skills, religion, and moral aspects. Those aspects will be used as a draft for developing the construct instruments of school readiness in elementary school.

Finding Aspects of School Readiness Learning Problems of Early Grade Students

There are many problems extracted by the data related to children's lack of readiness to attend elementary schools which emerge in Yogyakarta. Table 9 shows that there are 26 aspects of learning problems faced by early grade students.

Table 9. Learning problems of lower grade students

No.	General Description	%
1.	Children's focusing on playing	15.4
2.	Hard to follow the rule	10.5
3.	Daydreaming	9.37
4.	Stop studying	7.44
5.	Low learning result	6.34
6.	Slow task working	5.23
7.	Disturbing friends	4.41
8.	Slow instruction comprehending	4.41
9.	Being timid/afraid of asking	4.41
10.	Exiting class	4.41
11.	Wandering/running around the classroom	4.13
12.	Asking to be waited by parents	3.58
13.	Being unconfident	3.58
14.	Story telling/cheating with friends	3.31
15.	Lackluster	2.48
16.	Crying	2.2
17.	Dependence on teacher	2.2
18.	Keeping silent	1.65
19.	Feeling bored	1.38
20.	Easily exhausted	1.1
21.	Not completing the task	0.83
22.	Frequently asking for going home	0.83
23.	Not doing homework	0.28
24.	Moving around/Not sitting still	0.28
25.	Being alone/isolated	0.28

Table 9 shows that learning problems are always found in each educational level, and thus, some efforts to solve these problems are highly needed. Through the interview which had been conducted, this research identifies many problems faced by the first up to third graders of elementary school.

Table 10 explains the various learning problems encountered by grade 1, 2, and 3 students of elementary school. It also shows that the higher the grade, the fewer the problems will be. However, it is also clearly shown that the problems faced by students of grade 1, 2, and 3 are identical. Their problems are in terms of motivation, learning, independence, concentration, interaction, liveliness, motor skills, and comprehension. Meanwhile, the problems shared by each grade related to learning problems are stopping learning, not completing the task, not doing assignment/ homework, daydreaming, keeping silent, and slowness in comprehending in-struction. The similar problems possessed by grade 1 and 2 students are frequently asking to go home. In addition, other common problems encountered by elementary school students are feeling bored, lackluster performance, slowness in comprehending instruction, and not sitting still. Through in-depth identification, it is clearly shown that the problems related to social-emotional aspects more commonly occur than cognitive aspects. This condition is in line with the previous research findings which reveal that one of the aspects influencing children's learning readiness is social and emotional aspects. Therefore, social and emotional aspects determine children's learning readiness.

Table 10. Learning problems of students of grades 1, 2 and 3 of elementary school

Learning Problem	Grade 1	Grade 2	Grade 3
Stop studying	✓	✓	✓
Not completing the task	✓	✓	✓
Slow task working	✓	✓	✓
Not doing assignment/homework	✓	✓	✓
Low learning result	✓	✓	✓
Asking to be waited by parents	✓	✓	✓
Being not dependent	✓	✓	✓
Hard to follow the rules	✓	✓	✓
Crying	✓	✓	✓
Children focus on play	✓	✓	✓
Story telling/cheating with friends	✓	✓	✓
Wandering/running around the classroom	✓	✓	✓
Daydreaming	✓	✓	✓
Disturbing friends	✓	✓	✓
Keeping silent	✓	✓	✓
Easily exhausted	✓	✓	✓
Moving around/ Not sitting still	✓	✓	✓
Frequently asking for going home	✓		
Dependence on teacher	✓		
Exiting class	✓		
Feeling bored	✓		
Lackluster behavior	✓		
Slow instruction comprehension		✓	✓

Conclusion and Suggestions

The findings of this research show that there are some performing developmental aspects influencing children's learning readiness to attend elementary school. Ordered from the most to the least dominant, the effects are

explained as follows: (1) cognitive and language aspects, consisting of recognizing letters, reading, recognizing numbers, counting, vocabulary building, comprehending instruction and information, fluent speaking, story telling, and problem solving; (2) social emotional aspects, consisting of team work, socialization with peers, classroom adaptation, independence, apprehending rules, discipline, helping others; (3) other aspects, consisting of concentration, parents' caring, and age; (4) fine motor skills aspects, consisting of writing, snipping and sticking, holding stationeries, and drawing lines; (5) gross motor skills aspects consisting of imitating movements (running, jumping, and standing on one leg) or dancing; (6) arts, consisting of memorizing poem, drawing, and creativity; and (7) religion and other moral aspects, which consist of religious activities such as praying, reading Holy Quran, and charity.

Parents' caring played great role in preparing their children to go to school. Parental style and involvement at school and at home give big influences on the academic success of their children (Magdalena, 2014). Also, the parental characteristics and practices affect children's success. Furthermore, Tunçeli and Akman (2013) found that parents' age and education affect the school readiness of their child.

In addition to the afore-mentioned aspects, this research also finds some problems faced by elementary school students of grade 1, 2, and 3. The most dominant problems are related to socio-emotional, such as children focus on playing, hard to follow the rule, and stopping studying. Therefore, social-emotional aspects highly determine the readiness of children to attend elementary school compared to other aspects.

Based on the research findings, the following suggestions are proposed: (1) Children's learning readiness to attend elementary school is affected not only by age, but also by other developmental aspects which play significant roles in determining readiness. Therefore, identifying some aspects to construct the component of school readiness is needed. (2) Moreover, cognitive aspects, social-emotional aspects, concentration, as well as parental caring also play a big role to determe children's learning readiness to attend elementary school. Therefore, teachers, educators, or psychologists need to give special considerations to these aspects. Such action contributes to easily decide children's learning readiness to attend elementary school, especially those who do not meet the age requirement. (3) For instrument developers, the findings of this research need to be followed up by composing new instruments which can detect some appropriate indicators of school readiness and are not only oriented to cognitive aspects and fine motor skills, but also social-emotional aspects, concentration, parents' caring, age, gross motor skills, and arts.

This research was focused only on many aspects of child development. Meanwhile the roles of environment factors are not considered yet though many researchers have studied the role of environment on students' school readiness. Shaari & Ahmad (2016) found that physical and social environment have effect on school readiness. The environment has indirect effect on stimulating many aspects of child development.

References

Dockett, S., & Perry, B. (2002). Who's ready for what? Young children starting school. *Contemporary Issues in Early Childhood*, *3*(1), 67–89. https://doi.org/10.2304/ciec.2002.3.1.9

Government Regulation No. 17 Year 2010 on educational management and implementation (2010). Indonesia.

Janus, M., Brinkman, S., Duku, E., Hertzman, C., Santos, R., Sayers, M., … Walsh, C. (2006). *The early development instrument: A population-based measure for communities. A handbook on development, properties, and use*. Ontario: Offord Centre for Child Studies.

Lemelin, J.-P., & Boivin, M. (2007). Success starts in grade 1: The importance of school readiness. *Québec Longitudinal Study of Child Development*, *4*(2), 1–12.

Lewit, E. M., & Baker, L. S. (1995). School readiness. *The Future of Children*, *5*(2),

128–139. https://doi.org/10.2307/1602361

Magdalena, S. M. (2014). The effects of parental influences and school readiness of the child. *Procedia - Social and Behavioral Sciences*, *127*, 733–737. https://doi.org/10.1016/j.sbspro.2014.03.345

Ministry of National Education. Minister regulation no. 04/VI/PB/2011 on student admission at kindergarten/raudhatul athfal/bustanul athfal and school/madrasah (2011).

Ministry of National Education. Minister regulation no. 137 year 2014 on the national standard of early childhood education (2014).

Shaari, M. F., & Ahmad, S. S. (2016). Physical learning environment: Impact on children school readiness in Malaysian preschools. *Procedia - Social and Behavioral Sciences*, *222*, 9–18. https://doi.org/10.1016/j.sbspro.2016.05.164

Tunçeli, H. I., & Akman, B. (2013). The investigation of school readiness level of six years old preschool children in terms of different variables. *Procedia - Social and Behavioral Sciences*, *106*, 2899–2905. https://doi.org/10.1016/j.sbspro.2013.12.3 35

REiD (Research and Evaluation in Education), 3(1), 2017, 50-63
Available online at: http://journal.uny.ac.id/index.php/reid
Research Article

The structural equation modeling of reading interest psycho-behavioural constructs: How are they related across different modes of reading?

***[1]Nur Hidayanto Pancoro Setyo Putro; [2]Jihyun Lee**
*Faculty of Languages and Arts, Universitas Negeri Yogyakarta
Jl. Colombo No. 1, Depok, Sleman 55281, Yogyakarta, Indonesia
*Email: noersabar@yahoo.com

Submitted: 03 April 2017 | Revised: 13 May 2017 | Accepted: 13 May 2017

Abstract

The present study examines the relationships between the psycho-behavioral constructs underlying undergraduate students' reading interest. The *a priori* framework in conceptualizing the sub-components of reading interest is based on two modes of reading (printed-text-based and also Internet-based), and three types of psycho-behavioral motives/intentions of reading (affective, cognitive, and behavioral). Participants in this study were students (M = 20.14 years old) from an Indonesian university (n = 993). Exploratory and confirmatory factor analyses show the salience of 10 factors across reading modes and psycho-behavioral domains of reading. The most acceptable SEM models that explore the relationships among the sub-components of reading interest have the student reading interest in the print mode preceded interest in reading online materials. Implications of these findings are discussed for theory development and practice.

Keywords: *reading interest, affect, cognition, behavior, factor analysis*

How to cite item:

Putro, N. & Lee, J. (2017). The structural equation modeling of reading interest psycho-behavioural constructs: How are they related across different modes of reading?. *REiD (Research and Evaluation in Education), 3*(1), 50-63. doi:http://dx.doi.org/10.21831/reid.v3i1.13530

Introduction

In the past twenty years, there is a growing body of literature suggesting that current young generation has adopted multiple modalities in reading. Much of this work highlighted the increasing practice of reading online materials among school-aged and university students (Coiro & Dobler, 2007; Karim & Hasan, 2007; Liu & Huang, 2008; McKenna, Conradi, Lawrence, Jang, & Meyer, 2012). Other research revealed the emerging practice of reading from social media platforms, e.g., *Facebook* and *Twitter* (Junco, 2012; Kirschner & Karpinski, 2010). Although these studies have indicated an increasing trend towards reading online materials and social media reading, there is also a large volume of published studies showing reading printed materials has not been completely eclipsed (e.g., Buzzetto-More, Guy, & Elobaid, 2007; Liu, 2005).

As multimodal literacy becomes more widespread (Walsh, 2010), a great deal of previous research into reading motivation has focused on how frequent access to the Internet is associated with low interest in reading from printed materials and a decline in academic achievement (Alterman, 2007; DeWaal, Schönbach, & Lauf, 2005; Kirchhoff, 2010; Lee & Leung, 2008; Mokhtari, Reichard, & Gardner, 2009). Similarly, there is a growing body of literature suggesting how social media disempower today's youth book reading motivation and lead to low academic achievement (Junco, 2012; Kirschner & Karpinski, 2010).

While much research has explored the relationships between how the amount of

time spent on digital reading and the amount of time spent on reading printed books, little is known about how the psycho-behavioural constructs of reading interest within different modes (i.e., printed, online, and social media) are related to one another. Consequently, there is a need to further our understanding of interrelationships among the constructs within reading interest. This study is designed to address this lacuna.

The following sections briefly review some evidence suggesting the relationships among interest in reading in print settings, interest in reading online materials, and interest in social media reading. The first section outlines previous studies indicating the relationships among reading printed materials, reading online materials, and reading social media. The next section of the literature reviews and explores the evidence for interrelationships among the constructs within reading interest. Clearly, this is largely an unexplored area. This brief literature review is then followed by the results from an exploration of the relationships and interrelationships between the dimensions of reading interest, based on the data from 993 undergraduate students from one Indonesian university. The discussion then builds on the findings from both survey analyses and brief review of the literature. It also provides some steps to explore the issues raised further in current thinking and practice.

Relationship between Different Modes of Reading

Previous research that examined the relationship among reading in printed materials, reading online materials, and reading social media has shown mixed results. Many studies that claimed a close and positive relationship between reading in print settings and reading online materials have made its argument based on the amount of time people spent on reading across different types of settings (OECD, 2011; Veenhof, 2006). For instance, Veenhof (2006) investigated the social impact of the Internet use based on the Statistics Canada 2005 General Social Survey data, and found that readers who spent more time reading from the Internet also spent more time reading printed books. Similarly, the PISA 2009 project (OECD, 2011) also showed that students who read more frequently from online sources also read printed materials more frequently. A study by Tenopir, Volentine, and King (2013) also reported that readers who used social media more frequently also read scholarly materials more frequently.

While much research has shown the close positive relationships between reading patterns involving different modes of reading, other studies demonstrated the contrasting results. When it comes to the time spent on reading, the negative relationship was documented as well between the amount of time spent on reading online materials and that spent on reading printed materials (DeWaal, Schönbach, & Lauf, 2005; Lee & Leung, 2008). For instance, the study by Lee & Leung (2008) investigated the replacement effects of the Internet and found that use of the Internet for reading is negatively related to reading printed newspapers (r = -.23) and magazines (r = -.39), indicating that those who frequently read online are less likely to read printed materials. Another line of studies emphasizing the differences has brought about the widespread trend of reading newspapers online. Online newspapers have now become the preferred news source for young people over the printed newspapers (Alterman, 2007; Kirchhoff, 2010). Together, these studies suggest that there has been some partial shift in the mode of reading from all printed settings to online/digital/social media reading. Young generations, in particular, have adopted reading online materials as an alternative mode to conventional reading in print settings. These studies also indicate that reading can happen in three different formats: print, online, and social media. As such, reading interest in the present study is also investigated from these three modes of reading.

Relationships among the Psycho-Behavioural Constructs of Reading Interest

Following Putro's work (2017), reading interest in the present study is conceptualized to incorporate both mode and psycho-behavioral dimensions (i.e., affective, cognitive, and behavioral). Specifically, Putro (2017) claimed that each of these psycho-behavioral

dimensions is situated in a particular mode of reading. Interest in reading printed materials involves three psycho-behavioral constructs (i.e., elaboration, enjoyment, and competence experience); interest in reading online materials involves five psycho-behavioral constructs (i.e., value, confidence, enjoyment, competence experience, and flow); and interest in reading social media comprises two psycho-behavioral constructs (i.e., sense of belonging and enjoyment). While there have not been systematic reviews or empirical studies that examined the links between all these psycho-behavioral constructs of reading interest, the relationships between certain pairs or groups of the constructs (e.g., enjoyment and flow) have been explored and their close – either conceptual or empirical – links have been demonstrated. The following segments provide some evidence on the empirical relationships between the ten constructs of reading interest.

Enjoyment, Flow, and Competence

There have been few studies conducted (e.g. Shernoff, Csikszentmihalyi, Shneider, & Shernoff, 2003; Sherry, 2004; Weber, Tamborini, Westcott-Baker, & Kantor, 2009) which have considered the role of enjoyment and competence/knowledge/cognitive abilities in generating flow experiences. People would experience being completely absorbed in an activity that they find intensely enjoyable (e.g., Shernoff et al., 2003; Sherry, 2004). Without enjoyment, an intense experience of flow is unlikely to occur (Csikszentmihalyi, 1997; Sherry, 2004). Further, in a study by Shernoff, Csikszentmihalyi, Shneider, and Shernoff (2003), it is reported that when students found classroom activities interesting, easy to concentrate, and enjoyable, their flow condition was also high. Some scholars also argue that competence is important to sustain enjoyment and to transfer it to the flow condition (Carroll & Loumidis, 2001; Csikszentmihalyi, 1997; Sherry, 2004).

Enjoyment, Competence, and Achievement

While the causal relationships among these constructs are proven to be hard to demonstrate, they are at least interrelated with each other. For example, confidence in performing a task can lead to higher competence level but the reverse relationship, i.e., competence leading to feeling confidence, is also highly likely (Clanton et al., 2014; Dunst & Dempsey, 2007; Pajares & Johnson, 1994). Not surprisingly, students' self-evaluation of competence is significantly correlated with their achievement and confidence (Pajares & Johnson, 1994). Examples of domains demonstrating the close relationships of competence, confidence, and achievement are abundant: writing (Pajares & Johnson, 1994), reading (McGeown et al., 2015), and general cognitive abilities (Stankov & Lee, 2008).

Enjoyment, Competence, Value, and Achievement

Many empirical studies have been conducted within the framework of expectancy-value theory (Eccles, 1983; Wigfield, 1994; Wigfield & Eccles, 2002; Wigfield & Tonks, 2002) to investigate how students' enjoyment, values, and perceived competence beliefs are related to academic outcomes attainment. It appears that competence beliefs, enjoyment (intrinsic value), and utility value would positively reinforce each other (Chouinard, Karsenti, & Roy, 2007; Cocks & Watt, 2004; Wilson et al., 2008). While the achievement-related outcomes are employed as the final destination of this achievement-motivational theory, a more realistic picture would include reciprocal relationships (Marsh & Martin, 2011) among these constructs especially when developmental perspectives (Wigfield & Eccles, 2002) are taken into account.

Elaboration, Enjoyment, Value, Competency, and Sense of Belonging.

Empirical studies were able to demonstrate the links between students' use of elaboration strategy in reading and enjoyment in reading (Lau & Ho, 2016). Even when people read for a targeted purpose (e.g., doing homework, conducting a project), reading with elaboration can be a useful strategy in attaining the goals. The Program for International Student Assessment (PISA) data also showed that students' use of elaboration strategies are positively linked to competency beliefs, anxiety, and interest (Schleicher, 2016); students who use elaboration strategies more frequent-

ly in their reading reported higher self-competence beliefs in their ability, less anxiety and more interest in reading. Students who are confident in their abilities in learning tended to report using more elaboration strategy as well (Perry & Smart, 2007). In recent studies people's tendency to use the elaboration strategy were also found to have greater intrinsic motivation, sense of belonging, competence, and autonomy (Sundar, 2015).

A caveat should be registered; it appears that different studies use slightly different labels for the same constructs. In this study, enjoyment and intrinsic value are considered interchangeable, so are relatedness and sense of belonging. Value in this study is referred to the perception of usefulness, i.e., utility value, perceived value, and perceived utility value. Competence means self-beliefs in one's own capability in completing a task, which is also interchangeably used with confidence, competence beliefs, and perceived competence. Experiences of competence are referred to as memories of prior experiences about achievement or mastery of skills or tasks.

Method

The participants were undergraduate students in an Indonesian university, a medium-sized university with about 25,000 students enrolling in 2014. A total of 993 undergraduate students volunteered to participate in the study. The survey data were collected between the 17th of August and the 16th of November in 2014. Seventy one percent of the participants were female students. The majority of these students were in their second year (45%) and third year (35%).

Students' reading interest across the three modes (print, online, social media) was measured with 36 items from reading interest scale developed by Putro (2017), in which the 36 items were converged into 10 factors:

> *elaboration in print settings, enjoyment in print settings, competence experience in print settings, utility value in online reading, confidence in online reading, enjoyment in online reading, competence experience in online reading, flow in online reading, sense of belonging in social media reading, and enjoyment in social media reading.*

In the present study, these 10 factors were referred to as dimensions of reading interest.

The survey items were written in a way that includes a particular reading mode. The survey respondents were asked to rate their interest in reading in three different formats, i.e., reading in print settings, reading online materials, and reading through social media. All items were measured on a 5-point response category, ranging from *Strongly Disagree* (1) to *Strongly Agree* (5) with the middle point of *Neither Disagree nor Agree* (3).

Statistical Analysis

The main analyses of the present study were confirmatory factor analysis (CFA) and structural equation modeling (SEM). CFA was used to confirm the structure of the reading interest dimensions. SEM was used to test the relationships between the psycho-behavioral constructs of reading interest within and across modes of reading. *Mplus* version 7.2 (Muthén & Muthén, 1998-2012) was used for both the CFA and SEM results reported in this study. The maximum likelihood estimation with robust standard errors (MLR) was used to adjust for non-normality of the survey responses of the data, as suggested in Bentler (2005). As the model fit indices, the Comparative Fit Index (CFI > .90), Tucker-Lewis index (TLI > .90), Root Mean Square Error of Approximation (RMSEA < .05), and Standardized Root Mean Square Residual (SRMR < .05) were used to indicate a good model fit (criteria cut-off scores indicated, also see Byrne (2006). In addition, a ratio of 1/3 or less between the degrees of freedom (*df*) and chi-square statistics (x^2) was used as an acceptable model fit criterion (see Wang & Wang, 2012) instead of the significance of x^2. The Cronbach's α scale reliability for each factor was calculated with *SPSS* version 21.

Findings and Discussion

Nature of Reading Interest

The result of the confirmatory factor analysis (CFA) showed that the 36 items converged into 10 factors was a very good fit (x^2= 984.12, df = 549, x^2/df= 1.8, RMSEA = .03, SRMR= .04, CFI = .97, and TLI = .97).

Table 1. Confirmatory factor analysis on reading interest

Item	Factor									
	1	2	3	4	5	6	7	8	9	10
1. I always connect what I read in printed materials to my background knowledge.	.81									
2. When I read in printed settings, I always try to understand the materials better by relating to my personal experiences.	.80									
3. When I read in printed settings, I always figure out how the information fits in with what happens in my real life.	.78									
4. I enjoy reading printed materials		.84								
5. Reading printed materials makes me feel good.		.77								
6. I feel happy if I receive a printed book as a present.		.70								
7. I had good marks because I liked reading printed materials.			.88							
8. My reading in print settings skill continues to help me get good grades.			.85							
9. I did well in school due to my ability in reading printed materials.			.83							
10. I did well in my courses because I read printed materials fast.			.69							
11. Reading online materials helps me think about new concepts and ideas.				.72						
12. Reading online materials advances my general knowledge.				.69						
13. New ideas come to my mind when I read online.				.69						
14. I learn about what is going on in the world from reading online materials.				.65						
15. Reading online materials makes me feel linked to the world.				.65						
16. I obtain a great deal of information whenever do reading online materials.				.61						
17. Reading online materials is very easy for me.					.88					
18. I never have problems in reading online materials.					.76					
19. When I read from screen (e.g., computer screen, cell-phone, etc.), I am a good reader.					.70					
20. Reading online materials is one of my favourite activities.						.84				
21. Reading online materials makes me feel relaxed.						.81				
22. I always try to read online for my own enjoyment.						.69				
23. I did well in my studies at university because of reading online materials.							.91			
24. I did well in school because of my reading online materials ability.							.91			
25. I had good grades because I liked reading online materials.							.85			
26. My academic achievement has been influenced by my ability in reading online materials.							.83			
27. I feel fascinated when I read online.								.93		
28. When I read online, I forget about other things.								.82		
29. Time goes faster when I read online.								.66		
30. I feel linked to others who read the same things from social media sites (e.g. Facebook, WhatsApp).									.86	
31. Reading from Social media (e.g. Facebook, WhatsApp) makes me feel connected to the world.									.81	
32. Reading from Social media (e.g. Facebook, WhatsApp) makes me feel belonged to a certain group.									.76	
33. Reading from Social media (e.g. Facebook, WhatsApp) makes me communicate better with others.									.71	
34. Social media reading is one of my favourite activities (e.g. Facebook, WhatsApp).										.82
35. Most of the knowledge I obtained is from my social media reading.										.68
36. Once I read social media sites (e.g. Facebook, WhatsApp), I always reading for hours.										.65
Cronbach's α	**.84**	**.80**	**.89**	**.82**	**.82**	**.82**	**.95**	**.84**	**.87**	**.76**

The standardised factor loadings were all significant and substantial, ranging from β = .61 to β = .91 across all 36 items. Alpha coefficients for scores on the 10 reading interest dimensions ranged from .76 (enjoyment in social media reading) to .95 (competence experience in online reading), indicating reasonably good internal consistency for each of the scales. The standardized factor loadings of the CFA results are presented in Table 1, together with the Cronbach's α of each factor.

Relationships between the Psycho-behavioral Constructs of Reading Interest

Subsequent to the CFA, the model building strategy was to first construct the model in a way to test the relationships among the psycho-behavioural dimensions situated in a particular mode (e.g., reading online materials). The dimensions representing other modes of reading (i.e., print, mode, and social media) were then added to build a more comprehensive model of reading interest that represented all three models of reading.

Reading Online Materials

Model A (see Figure 1) was constructed to examine the relationships among the reading interest dimensions within the context of online reading. Because there are more online reading variables (five) than print modes of reading (three) and social media reading (two), the model was built with the dimensions related to online reading first. This model reflects four propositions: (a) enjoyment in reading online materials facilitates flow in online reading (Shernoff, Csikszentmihalyi, Shneider, & Shernoff, 2003; Sherry, 2004; Csikszentmihalyi, 1997); (b) competence experience in reading online materials is positively linked to enjoyment in online reading (e.g., Carroll & Loumidis, 2001; Sherry, 2004); (c) confidence in reading online materials is moderately related to enjoyment in online reading (e.g., Clark & De Zoysa, 2011); and (d) the perceived value in reading online materials is positively related to enjoyment and competence in reading online materials (e.g., Wilson et al., 2008).

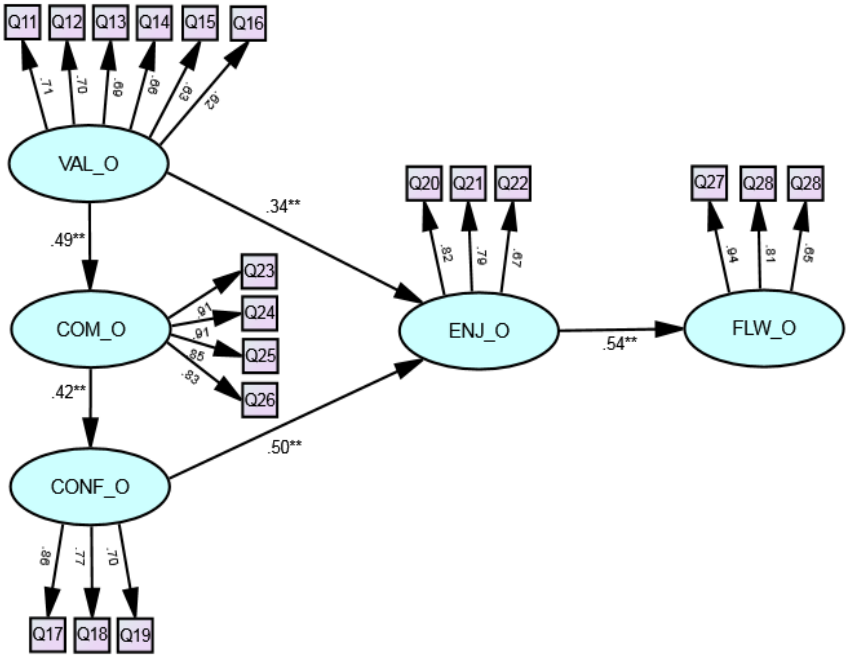

Figure 1. Model A: The relationships among the psycho-behavioural constructs within the context of online reading

Notes: VAL_O: Utility value in online reading; CONF_O: Confidence in online reading; ENJ_O: Enjoyment in online reading; COM_O: Competence experience in online reading; FLW_O: Flow in online reading.

Table 2. Standardised coefficients, standard errors, estimated standard errors, and p-Value for Model A

Path	Estimate	S.E.	Est./S.E.	Sig.
Enjoyment in reading online materials to flow in online reading	.54	.03	16.17	.00
Utility value in reading online materials to enjoyment in online reading	.34	.04	7.68	.00
Confidence in reading online materials to enjoyment in online reading	.50	.05	10.79	.00
Utility value in reading online materials to competence experience in online reading	.49	.03	14.22	.00
Competence experience in reading online materials to confidence in online reading	.42	.04	11.89	.00

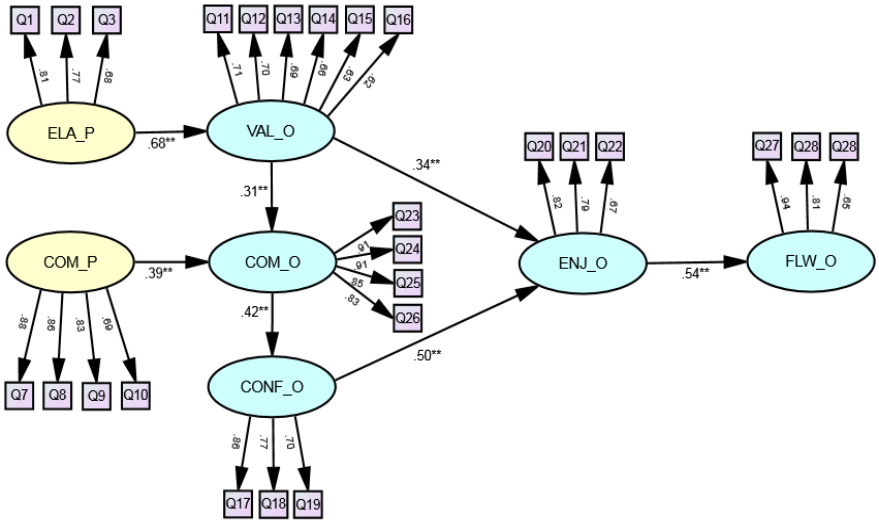

Figure 2. Model B: The relationships among the psycho-behavioural constructs within the context of online reading and reading in print settings

Notes. ELA_P: Elaboration in print settings; ENJ_P: Enjoyment in print settings; COM_P: Competence experience in print settings; VAL_O: Utility value in online reading; CONF_O: Confidence in online reading; ENJ_O: Enjoyment in online reading; COM_O: Competence experience in online reading; FLW_O: Flow in online reading.

Table 3. Standardised path coefficients, etandard errors, estimated standard errors, and p-Value for Model B

Path	Estimate	S.E.	Est./S.E.	Sig.
Enjoyment in reading online materials to flow in reading online materials	.54	.03	16.14	.00
Utility value in reading online materials to enjoyment in online reading	.34	.04	7.84	.00
Confidence in reading online materials to enjoyment in online reading	.50	.05	11.06	.00
Utility value in reading online materials to competence experience in online reading	.31	.04	8.61	.00
Competence experience in reading online materials to confidence in online reading	.42	.04	11.94	.00
Competence experience in reading in print to competence experience in online reading	.39	.04	9.95	.00
Elaboration in reading in print to utility value in online reading	.68	.03	23.16	.00

The model that showed the best fit to the data is presented in Figure 2. Model B presented in this figure, i.e., Figure 2 yielded good fit (x^2 = 1295.34, df = 577, x^2/df= 2.24, RMSEA = .04, SRMR = .08, CFI = .95, and TLI = .95), which are better or higher fit indices compared to those of other models tested. The standardized path coefficients among the seven latent variables were all significant and substantial, ranging from β = .31 to β = .68. In fact, it is an extension of Model A with additional pathways from 'competence in reading in print settings' to 'competence in reading online materials', and from 'elaboration in reading in print settings' to 'utility values in online reading'.

Among the 10 factors, there was one more variable related to reading in print settings, which is 'enjoyment in reading in print settings'. Various attempts were made to include this variable, but the addition of this variable resulted in the worsening of the overall model fit and the potential pathways, such as from 'enjoyment in reading in print settings' to 'enjoyment in online reading' (β = -.16, p < .01) and from 'enjoyment in reading in print settings' to 'flow in online reading' (β = .03, p > .05), showed weak and non-significant links. Therefore, the variable was dropped in the final model, and it was concluded that Model B is the best representation of the variables related to two reading settings (i.e., print and online reading). It also shows that reading interest in print settings precedes interest in reading online materials

Reading online materials, social media reading, and reading in print settings

Model C (see Figure 3) was constructed to examine the relationships among the reading interest dimensions from the three different types of reading modes (i.e., reading online materials, reading in print settings, and social media reading). Out the 10 factors, two variables are related to reading in social media, enjoyment in social media reading and sense of belonging through social media reading. Models were built to reflect the literature suggesting that: (a) confidence in reading is moderately related to enjoyment in reading (e.g., Clark & De Zoysa, 2011; McGeown et al., 2015) and (b) elaboration is related to sense of belonging (e.g., Sundar, 2015).

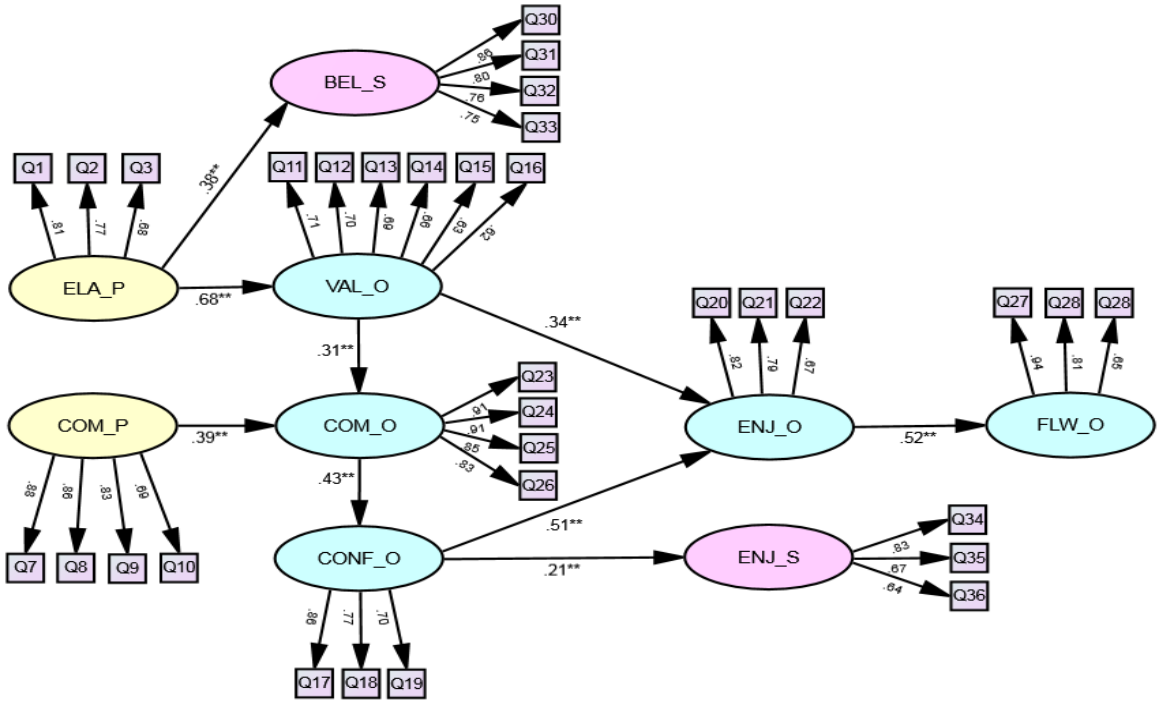

Figure 3. Model C: The relationships among the psycho-behavioural constructs of interest in online reading, social media reading, and reading in print settings

Notes: ELA_P: Elaboration in print settings; ENJ_P: Enjoyment in print settings; COM_P: Competence experience in print settings; VAL_O: Utility value in online reading; CONF_O: Confidence in online reading; ENJ_O: Enjoyment in online reading; COM_O: Competence experience in online reading; FLW_O: Flow in online reading; BEL_S: Sense of belonging in social media reading; ENJ_S: Enjoyment in social media reading.

Table 4. Standardised path coefficients, standard errors, estimated standard errors, and p-Value of Model C

Path	Estimate	S.E.	Est./S.E.	Sig.
Enjoyment in reading online materials to flow in reading online materials	.52	.03	15.16	.00
Utility value in reading online materials to enjoyment in online reading	.33	.04	7.65	.00
Confidence in reading online materials to enjoyment in online reading	.51	.05	11.18	.00
Utility value in reading online materials to competence experience in online reading	.31	.04	8.64	.00
Competence experience in reading online materials to confidence in online reading	.43	.04	12.04	.00
Competence experience in reading in print to competence experience in online reading	.39	.04	10.12	.00
Elaboration in reading in print to utility value in online reading	.72	.03	26.00	.00
Elaboration in reading in print to sense of belonging in social media reading	.41	.04	10.25	.00
Confidence in online reading to enjoyment in social media	.22	.50	4.53	.00

After several options were tested, a final model was chosen in which 'elaboration in print settings' is significantly related to 'sense of belonging in social media reading' (β = .41, p < .01) and 'confidence in online reading' is significantly linked to 'enjoyment in social media' (β = .22, p < .01). Model C yielded good fit (x^2 = 1292.38, df = 579, x^2/df = 2.23, RMSEA = .04, SRMR = .08, CFI = .95, and TLI = .95). The standardised path coefficients in Model C were all significant and substantial, ranging from β = .21 to β = .68 across all nine latent variables. This model reveals weak to moderate relationships between interest in social media reading and interest in reading online materials and printed materials. Table 4 shows the standardised parameter estimates and standard errors of all pathways included in Model C.

Discussion

Despite extensive research on reading interest, the relationships among the psychobehavioral dimensions of that construct remain unclear. The aim of this study is to examine how the dimensions of reading interest within and across modes of reading are related to one another. Noteworthy findings from the final models of the relationship among the dimensions of reading interest are considered in this section.

The first important finding is that the dimensions of interest in reading in print settings preceded those of interest in reading online materials, suggesting the importance of interest in reading in print settings for the development of interest in reading online materials. This finding supports the idea that reading in print settings is positively linked to, or may even facilitate, reading online materials (e.g., Coiro, 2011a, 2011b; Coiro & Dobler, 2007; Schmar-Dobler, 2003). It may partly be explained by the fact that, to be able to get the most from reading online materials, readers need to be proficient in reading in print settings and to be able to use their reading-in-print strategies in order to read in online settings. Fluent in-print readers need to learn additional practices and strategies, such as how to use web-based search engines and how to locate information efficiently and

effectively by adopting strategies they used when they read in print settings. To better understand what they read online, undergraduate students also need to connect what they already know from reading in print settings to what they read online.

In contrast to earlier findings (e.g., De Waal et al., 2005; Mokhtari et al., 2009), however, this study found no evidence of negative relationships between interest in reading online materials and interest in reading in print settings. A possible explanation for this is that previous studies relied on the frequency of either reading online materials or reading in print settings as the measure of reading interest. These previous studies drew this conclusion (i.e., that there is a negative relationship between reading online materials and reading in print) based on the fact that the time individuals spent reading online materials reduced the time they spent reading printed materials because they could not use the time spent on one activity for time spent on another activity (see Mokhtari et al., 2009; Valkenburg & Peter, 2007). This study, however, did not rely on frequency of reading as the measure of reading interest, which might account for the different results.

A moderate relationship between elaboration in reading in print settings and sense of belonging in social media reading was also documented in this study. This relationship may partly be explained by the nature of social media reading itself; that is, an activity performed to establish interactions among readers who share a common interest. The source of this interest may be what they read in print settings (e.g., interest in reading printed novels or comics). This result supports previous research findings that reading in print settings is related to social media reading (e.g., Cheung, Chiu, & Lee, 2011; Tenopir et al., 2013). The particular relationship between elaboration and sense of belonging has also been documented by Sundar (2015).

Third, considering the relationships between particular reading interest dimensions, the first important finding is that enjoyment in reading seems to be the only variable directly and consistently connected to flow in reading. In the three models (A, B, and C), it is evident that enjoyment is the sole predictor of flow across modes of reading. This result appears to be consistent with other research showing that flow occurred only when individuals continued to follow their sense of enjoyment in a particular object of interest (Csikszentmihalyi, 1997; Shernoff et al., 2003). A possible explanation for this is that the flow experience in reading occurs only when an individual finds the reading activity to be intrinsically enjoyable. Thus, people who know the value of the reading material and are confident in their reading skills will not experience flow if they do not find the reading activity enjoyable. This finding has important implications for the use of enjoyment in reading as one of the key predictors of flow in reading in future measurement of reading interest.

This study also found that utility value in reading was significantly connected to enjoyment in reading online materials. This result is in line with those in previous studies (e.g., Nakamura & Csikszentmihalyi, 2009; Wilson et al., 2008) and may help us to understand why some students are reluctant to read when they cannot perceive the value of what they need to read. This result may be explained by the fact that people find a reading activity enjoyable when they believe the reading activity is valuable or worth doing. In other words, the value or benefits expected from reading a text may help the reader to find the reading activity pleasurable.

Another interesting finding is that enjoyment in reading online materials is significantly linked to confidence in reading online materials and that confidence in reading online materials is significantly predicted by competence experience in reading online materials. This result supports the idea that enjoyment in reading is strongly influenced by both competence and confidence in reading (e.g., Clark & De Zoysa, 2011). It may explain the relatively significant correlation among confidence, competence, and enjoyment in the way that improvement in individuals' competence in reading leads to improvement in their confidence in reading. Improvement in their confidence may in turn lead to an increase in the pleasure or enjoyment derived from reading, as individuals will only find the

activity enjoyable when they are confident that their skills meet the associated reading challenges (Shernoff et al., 2003).

The results of this study also show that elaboration in reading is strongly connected with both utility value in reading online materials and enjoyment in recreational reading. This result is consistent with findings from other studies that elaboration is moderately related to enjoyment (e.g., Frenzel, Goetz, Stephens, & Jacob, 2009) and utility value (e.g., Brockman, 2006). One possible reason for this is that linking individuals' prior knowledge with what they are reading facilitates the creation of a balance between what they already know and the challenge from the reading process. In turn, this leads them to perceive the reading activity as enjoyable and the activity as valuable or worth doing. These findings suggest that, to help learners get the most from what they read, teachers need to involve the students' prior knowledge before gradually changing the level of reading challenge to help them enjoy the reading process and to recognize the value of what they read.

This study also found a significantly weak relationship between confidence in reading online materials and enjoyment in social media reading. This result supports the ideas of Dunst and Dempsey (2007) who found that confidence in parenting led to enjoyment in parenting. Further, research has shown that confidence in reading is moderately related to enjoyment in reading (McGeown et al., 2015). This relationship may be partly explained by the fact that when people believe they are good at a particular activity (i.e. confident in their abilities), they are more likely to enjoy performing the activity (Boyd & Yin, 1996; Carroll & Loumidis, 2001; Durik, Vida, & Eccles, 2006). Thus, undergraduate students' belief in their abilities in reading online materials appears to lead them to enjoy reading through social media platforms such as *Facebook* and *Twitter*.

Conclusion and Suggestions

This study provides evidence of how the psycho-behavioral constructs of reading interest are related to one another. Given that the constructs within interest in reading in print settings are connected to those in reading online materials, educators need to encourage students utilizing both reading modes to help them get the best from their reading. Online materials help students search for information efficiently and effectively, whereas printed materials facilitate deep understanding. The existence of moderate to strong relationships among elaboration in reading in print settings, utility value in reading online materials, confidence in reading online materials, and enjoyment in reading online materials suggests that educators can enhance students' reading interest (particularly their enjoyment in reading) by connecting reading tasks to real life experience, assigning value to the reading activity, and developing students' confidence in reading.

Although testing of the final model of reading interest dimensions yielded an acceptable fit with the data, other models might also yield an acceptable fit. Evidence from other types of investigations is required to confirm these models and to test their application.

References

Alterman, E. (2007). Out of print: The death and life of the American newspaper. *Caligrama (São Paulo. Online), 3*(3). doi: http://dx.doi.org/10.11606/issn.1808-0820.cali.2007.67395.

Bentler, P. M. (2005). EQS 6 structural equations program manual. Encino, CA: Multivariate Software, Inc.

Boyd, M. P., & Yin, Z. (1996). Cognitive-affective sources of sport enjoyment in adolescent sport participants. *Adolescence, 31*(122), 383-395.

Brockman, G. (2006). *What factors influence achievement in remedial mathematics classes?* (Doctoral dissertation). University of Southern California, Los Angeles, CA.

Buzzetto-More, N., Guy, R., & Elobaid, M. (2007). Reading in a digital age: E-books are students ready for this learning object? *Interdisciplinary Journal of E-Learning and Learning Objects, 3*, 239-250.

Byrne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and*

programming: Mahwah, NJ: Lawrence Erlbaum Associates.

Carroll, B., & Loumidis, J. (2001). Children's perceived competence and enjoyment in physical education and physical activity outside school. *European Physical Education Review, 7*(1), 24-43. doi: https://doi.org/10.1177/1356336X010071005

Cheung, C. M., Chiu, P.-Y., & Lee, M. K. (2011). Online social networks: Why do students use Facebook?. *Computers in Human Behavior, 27*(4), 1337-1343. doi: 10.1016/j.chb.2010.07.028

Chouinard, R., Karsenti, T., & Roy, N. (2007). Relations among competence beliefs, utility value, achievement goals, and effort in mathematics. *British Journal of Educational Psychology, 77*(3), 501-517. doi: 10.1348/000709906X133589

Clanton, J., Gardner, A., Cheung, M., Mellert, L., Evancho-Chapman, M., & George, R. L. (2014). The relationship between confidence and competence in the development of surgical skills. *Journal of Surgical Education, 71*(3), 405-412. doi: 10.1016/j.jsurg.2013.08.009

Clark, C., & De Zoysa, S. (2011). *Mapping the interrelationships of reading enjoyment, attitudes, behaviour and attainment: An exploratory investigation*. London: National Literacy Trust.

Cocks, R. J., & Watt, H. M. (2004). Relationships among perceived competence, intrinsic value and mastery goal orientation in English and maths. *The Australian Educational Researcher, 31*(2), 81-111. doi: 10.1007/BF03249521.

Coiro, J. (2011a). Predicting reading comprehension on the Internet: Contributions of offline reading skills, online reading skills, and prior knowledge. *Journal of Literacy Research, 43*(4), 352-392. doi: 10.1177/1086296x11421979

Coiro, J. (2011b). Talking about reading as thinking: Modeling the hidden complexities of online reading comprehension. *Theory Into Practice, 50*(2), 107-115. doi: 10.1080/00405841.2011.558435

Coiro, J., & Dobler, E. (2007). Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information Internet. *Reading Research Quarterly, 42*(2), 214-257. doi: 10.1598/RRQ.42.2.2

Csikszentmihalyi, M. (1997). *Flow and the psychology of discovery and invention*. New York, NY: Harper Collins.

DeWaal, E., Schönbach, K., & Lauf, E. (2005). Online newspapers: A substitute or complement for print newspapers and other information channels? *Communications, 30*(1), 55-72. doi: 10.1515/comm.2005.30.1.55

Dunst, C. J., & Dempsey, I. (2007). Family–professional partnerships and parenting competence, confidence, and enjoyment. *International Journal of Disability, Development and Education, 54*(3), 305-318. doi: 0.1080/10349120701488772

Durik, A. M., Vida, M., & Eccles, J. S. (2006). Task values and ability beliefs as predictors of high school literacy choices: A developmental analysis. *Journal of Educational Psychology, 98*(2), 382-393. doi: 10.1037/0022-0663.98.2.382.

Eccles, J. S. (1983). Expectancies, values, and academic behaviors. In J. Spence (Ed.), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 75-146). San Francisco, CA: Freeman.

Frenzel, A. C., Goetz, T., Stephens, E. J., & Jacob, B. (2009). Antecedents and effects of teachers' emotional experiences: An integrated perspective and empirical test. In P. Schutz (Ed.), *Advances in teacher emotion research: The impact on teachers' lives* (pp. 129-151). Dordrecht: Springer.

Junco, R. (2012). Too much face and not enough books: The relationship between multiple indices of Facebook use and academic performance. *Computers in Human Behavior, 28*(1), 187-198. doi: 10.1016/j.chb.2011.08.026

Karim, N. S. A., & Hasan, A. (2007). Reading habits and attitude in the digital age. *The Electronic Library, 25*(3), 285-298. doi: 10.1108/02640470710754805

Kirchhoff, S. M. (2010, September). *US newspaper industry in transition*. Paper presented in Congressional Research Service Report for Congress.

Kirschner, P. A., & Karpinski, A. C. (2010). Facebook® and academic performance. *Computers in Human Behavior, 26*(6), 1237-1245. doi: https://doi.org/10.1016/j.chb.2010.03.024

Lau, K.-l., & Ho, E. S.-c. (2016). Reading performance and self-regulated learning of Hong Kong students: What we learnt from PISA 2009. *The Asia-Pacific Education Researcher*, 25(1), 159-171. doi: https://doi.org/10.1007/s40299-015-0246-1.

Lee, P. S., & Leung, L. (2008). Assessing the displacement effects of the Internet. *Telematics and Informatics, 25*(3), 145-155. doi: 10.1016/j.tele.2006.08.002

Liu, Z. (2005). Reading behavior in the digital environment: Changes in reading behavior over the past ten years. *Journal of Documentation, 61*(6), 700-712. doi: 10.1108/00220410510632040

Liu, Z., & Huang, X. (2008). Gender differences in the online reading environment. *Journal of Documentation, 64*(4), 616-626. doi: http://dx.doi.org/10.1108/00220410810884101

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *British Journal of Educational Psychology, 81*(1), 59-77. doi: 10.1348/000709910X503501

McGeown, S. P., Johnston, R. S., Walker, J., Howatson, K., Stockburn, A., & Dufton, P. (2015). The relationship between young children's enjoyment of learning to read, reading attitudes, confidence and attainment. *Educational Research, 57*(4), 389-402. doi: 10.1080/00131881.2015.1091234

McKenna, M. C., Conradi, K., Lawrence, C., Jang, B. G., & Meyer, J. P. (2012). Reading attitudes of middle school students: Results of a U.S. survey. *Reading Research Quarterly, 47*(3), 283-306. doi: 10.1002/rrq.021

Mokhtari, K., Reichard, C. A., & Gardner, A. (2009). The impact of internet and television use on the reading habits and practices of college students. *Journal of Adolescent & Adult Literacy, 52*(7), 609-619. doi: 10.1598/JAAL.52.7.6

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus User's Guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Nakamura, J., & Csikszentmihalyi, M. (2009). The concept of flow. In C. R. Snyder & S.J. Lopez, *Oxford Handbook of Positive Psychology* (pp. 89-105). New York, NY: Oxford University Press.

OECD. (2011). *PISA 2009 results: Students on line: Digital technologies and performance* (Vol. 6). Paris: OECD Publishing.

Pajares, F., & Johnson, M. J. (1994). Confidence and competence in writing: The role of self-efficacy, outcome expectancy, and apprehension. *Research in the Teaching of English*, *28*(3), 313-331. doi: 10.2307/40171341.

Perry, R. P., & Smart, J. C. (2007). *The scholarship of teaching and learning in higher education: An evidence-based perspective*. Dordrecht: Springer Science & Business Media.

Putro, N. (2017). *Reading interest in a digital age* (Doctoral dissertation). University of New South Wales, Sydney.

Schleicher, A. (2016). Teaching excellence through professional learning and policy reform: Lessons from around the world. *International summit on the teaching profession*. OECD Publishing. doi: 10.1789/9789264252059.en

Schmar-Dobler, E. (2003). Reading on the Internet: The link between literacy and technology. *Journal of Adolescent & Adult*

Literacy, 47(1), 80-85. doi: 10.2307/40026906

Shernoff, D. J., Csikszentmihalyi, M., Shneider, B., & Shernoff, E. S. (2003). Student engagement in high school classrooms from the perspective of flow theory. *School Psychology Quarterly, 18*(2), 158-176. doi: 10.1521/scpq.18.2.158.21860

Sherry, J. L. (2004). Flow and media enjoyment. *Communication Theory, 14*(4), 328-347.

Stankov, L., & Lee, J. (2008). Confidence and cognitive test performance. *Journal of Educational Psychology, 100*(4), 961-976. doi: 10.1037/a0012546

Sundar, S. S. (2015). *The handbook of the psychology of communication technology*. Chicester: John Wiley & Sons. http://unsw.eblib.com/patron/FullRecord.aspx?p=1895448

Tenopir, C., Volentine, R., & King, D. W. (2013). Social media and scholarly reading. *Online Information Review, 37*(2), 193-216. doi: 10.1108/OIR-04-2012-0062

Valkenburg, P. M., & Peter, J. (2007). Online communication and adolescent well-being: Testing the stimulation versus the displacement hypothesis. *Journal of Computer-Mediated Communication, 12*(4), 1169-1182. doi: 10.1111/j.1083-6101.2007.00368.x

Veenhof, B. (2006). *The Internet: Is it changing the way Canadians spend their time?*. Statistics Canada. Available at http://www.statcan.gc.ca/pub/56f0004m/56f0004m2006013-eng.htm.

Walsh, M. (2010). Multimodal literacy: What does it mean for classroom practice? *Australian Journal of Language and Literacy, 33*(3), 211–239.

Wang, J., & Wang, X. (2012). *Structural equation modeling: Applications using Mplus*. Chicester: John Wiley & Sons.

Weber, R., Tamborini, R., Westcott-Baker, A., & Kantor, B. (2009). Theorizing flow and media enjoyment as cognitive synchronization of attentional and reward networks. *Communication Theory, 19*(4), 397-422. doi: 10.1111/j.1468-2885.2009.01352.x

Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review, 6*(1), 49-78. doi: 10.2307/23359359

Wigfield, A., & Eccles, J. S. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In A. Wigfield, & J. Eccles (Eds.), *The development of achievement motivation* (pp. 91-120). San Diego, CA: Academic Press. doi: https://doi.org/10.1016/B978-012750053-9/50006-1

Wigfield, A., & Tonks, S. (2002). Adolescents' expectancies for success and achievement task values during the middle and high school years. In T. Urdan & F. Pajares (Eds.), *Academic motivation of adolescents* (pp. 53-82). Greenwich, CT: Information Age.

Wilson, N., Bouhuijs, P., Conradie, H., Reuter, H., Van Heerden, B., & Marais, B. (2008). Perceived educational value and enjoyment of a rural clinical rotation for medical students. *Rural Remote Health, 8*(3), 999. Available at: http://www.rrh.org.au/articles/subviewnew.asp?ArticleID=999.

REiD (Research and Evaluation in Education), 3(1), 2017, 64-76
Available online at: http://journal.uny.ac.id/index.php/reid
Research Article

A construct of the instrument for measuring junior high school mathematics teacher's self-efficacy

***[1]Rachmadi Widdiharto; [2]Badrun Kartowagiran; [3]Sugiman**
*Centre for the Development and Empowerment of Mathematics Teachers and Educational Personnel (PPPPTK Matematika) of Yogyakarta
Jl. Kaliurang Km. 6, Sambisari, Condong Catur, Depok, Sleman 55281, Yogyakarta, Indonesia
*Email: rachmadiw@yahoo.com

Submitted: 06 April 2017 | Revised: 15 July 2017 | Accepted: 15 July 2017

Abstract

The aim of this study was to develop a construct of the instrument for junior high school mathematics teacher self-efficacy and its mapping in the Special Region of Yogyakarta. The population was 816 junior high mathematics teacher, and a sample of 274 teachers was selected through proportionate random sampling technique. The data were analyzed using Confirmatory Factor Analysis (CFA), using Lisrel 8.80 software through the first order and the second order stages. The result of data analysis toward four dimensions obtained: 11 fit items for dimension of personal efficacy (PE), 12 fit items for dimension of general teaching efficacy (GTE), 13 fit items for dimension of subject matter teaching efficacy (STE), and 8 fit items for dimension of outcome efficacy (OE). Afterward, the result of selecting 54 items in the first order stage was examined for the second order CFA, which shows the model is fit to the data and obtains 25 fit items. The loading factors for each dimension PE, GTE, STE, and OE consecutively were: 0.46; 0.84; 0.89, and 0.92, and the mapping of mathematics teacher self-efficacy level shows: 43.07% in low category, 55.47% in medium category, and 1.46% in the high category.

Keywords: *self-efficacy, construct, mathematics, junior high school*

How to cite item:

Widdiharto, R., Kartowagiran, B., & Sugiman, S. (2017). A construct of the instrument for measuring junior high school mathematics teacher's self-efficacy. *REiD (Research and Evaluation in Education), 3*(1), 64-76. doi:http://dx.doi.org/10.21831/reid.v3i1.13559

Introduction

Increasing and improving teacher quality is continually implemented by the government through the fulfillment of academic qualification of S-1/D-IV, teacher certification, block grant for the continuation of the study, the revitalization of Teachers Working Group (*Kelompok Kerja Guru*/KKG) for elementary school teachers, Subject-matter Teacher Forum (*Musyawarah Guru Mata Pelajaran /MGMP*) for junior and senior high school teachers, and program *BERMUTU* (Better Education through Reformed Management and Universal Teacher Upgrading) (Jalal et al., 2009, p. 124). However, the government's efforts still failed to give satisfactory results when the condition of teacher teaching practices do not support the ability of students in mathematics achievement.

The results of TIMSS (Trends in International Mathematics and Science Study) Video Study 2007 (Leung & Ragatz, 2010, p. 33) states that most junior high school mathematics teachers in Indonesia use 76% of their time to the problem activity and 24% for non-problem, while in Japan, 82% for problem activiy and 18% for non-problem, and

Hong Kong 85% of problem activity and 15% for non-problem activity.

The Report of Training Need Assessment and Recruitment (PPPPTK Matematika, 2007, p. 46) of a sample of 268 teachers in 15 provinces showed approximately 61.78% of the teachers had difficulty learning mathematics associated with problem solving. In a further interview with the respondents, there was a tendency for them to avoid delivering learning materials considered difficult. In terms of having no choice and a must to deliver the stated material, they lacked confidence in their performance or in their teaching practice.

Sumardyono (2011, p. 244) in a study of math anxiety to 89 participants of mathematics teachers training in *PPPPTK Matematika* in 2010, from the District of Banjarmasin, South Kalimantan, showed that adapting Mathematics Anxiety Rating Scale (MARS), increased the level of anxiety gradually from the higher education in which teachers teach to the lower education. It means that the high school teacher had a low level of anxiety compared to the junior high school mathematics teachers or primary school teachers.

Meanwhile, preliminary research conducted by the researchers to 38 junior high school mathematics teachers in Java, who attended *PPPPTK Matematika* training activities in 2012, by adopting Teacher Efficacy Scale (TES) from Tschannen-Moran, Hoy, and Hoy (1998), reported that 13.16% of those with high-efficacy scale categories, and approximately 86.84% efficacy scale were in a medium category. The teachers were also tested with Mathematics Teacher's Efficacy Belief Instrument (MTEBI) developed by Enoch & Smith (1997) which states that most (almost all) their mathematics belief is in the medium category, and no participant was in the low or high category.

Hastuti et al. (2009) mention that teacher certification improves the welfare and likely improves the teachers quality because they can concentrate more and become more motivated. However, he was not convinced by it because increasing the quality and performance is a matter of personal commitment. Sadtyadi and Kartowagiran, (2014, p. 291) mention that through the assessment done by the time the teacher is teaching, it is difficult to describe the actual performance of teachers, because they have a tendency to be better prepared, than when monitoring is not done, in the assessment of its performance.

Based on that condition in which the teacher competence is still not optimal, and they themselves lack of confidence in carrying out the tasks they are responsible for, it is a bit difficult to expect they will be able to teach the material to their students well. Pajares (1996, p. 544) refers to Bandura's opinion that defines self-efficacy as a belief about their ability to successfully perform certain tasks in certain situations. Self-efficacy is also defined as an assessment of the person's ability to organize and execute courses of actions required to complete a type of work that has been determined, primarily for mathematics teachers' duties associated with fostering students' mathematical power (Kastberg, D'Ambrosio, McDermot, & Saada, 2005, p. 10).

Self-Concept

Burn (1984) states that self-concept is a composite image of what we think we are, what we think we can achieve, what we think others think of us, and what would we like to be. Most social psychologists, one of whom, Rokeach, (Burn, 1984, p. 52) agree that self-concept as a set of self- attitudes that consist of four components appear to be embodied: (a) a belief, or knowledge or cognitive component, (b) an affective or emotional component, (c) an evaluation, and (d) a predisposition to respond. Attitude organizes a relatively enduring belief in the object or situation around as a person's tendency to respond in various ways that he or she likes. Thus the self-concept is more of a hypothetical construct. In other words, it is a concept or a useful way to predict the attitude or behavior of a person, but must be careful not to 'filter' or judge the constructs as a thought that exists in the real world.

Self-efficacy

Another concept related to the belief of an individual or representation of one's self is the self-efficacy. Bandura in Keller (2010, p. 146) mentions another concept related to the

belief in personal agency, i.e self-efficacy, the belief which is typically referred to as a person's belief that he or she can succeed in performing a given task. In line with this definition, Bandura also mentions self-efficacy as people's judgments on reviews of their abilities to organize and execute courses of action required to attain designated types of performances. Based on those opinions, it might be concluded that self-efficacy is defined as a belief or judgment about a person's ability to organize and execute courses of action required to complete a type of work that has been determined.

Keller (2010, p. 146) states that a person's self-efficacy is comprised of combination of belief related to three questions: Am I capable of doing the things that are necessary for success, developing a plan that will lead to success, and persisting in my effort long enough to achieve success? Thus the results of the strength or toughness of self-efficacy can be expected of a person: whether repetition or modification of planned behavior, how much effort will be made, and how long one will survive in the face of obstacles and challenging experience.

Mathematics Teacher Competency

Teacher competence, on Law No. 14 Year 2005 of Republic of Indonesia about Teachers and Lecturers, article 32, states that the promotion and development of the teaching profession as referred to in paragraph (1) includes the pedagogical competence, personal competence, social competence, and spiritual competence. In relation to the competence of mathematics teachers, there are some opinions that highlight the mastery of substance, achievement in performing or teaching in the classroom, peers assessment, or preparation of the portfolio. Fennema and Franke (Turnuklu & Yesildere, 2007, p. 2) mention that some of the components of mathematical knowledge to be possessed by a mathematics teacher: knowledge of mathematics, knowledge representation/math symbols, knowledge of the students, and knowledge about teaching and decision making.

Another opinion from Kulm and Wu (Turnuklu & Yesildere, 2007, p. 3) mentions the beliefs on a reciprocal basis underlying the substance of pedagogical content knowledge. Pedagogical content knowledge comprises of three components: content knowledge, teaching practice, and knowledge of the curriculum, each of which interacts reciprocally. In the practice of teaching, a teacher must understand the thought of his/her students (knowing students' thinking). The understanding of students' thoughts is translated into five components namely: addressing students' misconception, engaging student learning in math, student learning, promoting student thinking in mathematics, and building on student math idea. Thus, this shows that the conviction of a mathematics teacher will be the basis for the substance of pedagogical knowledge, which will ultimately lead to active student activities, anticipate misconceptions, and build mathematical ideas.

Mathematics Teacher Self-efficacy

Related to teacher efficacy, several studies support the theory that the belief in one's ability is the best predictor for the behavior of the completion of a task (Bandura, 1996, 1997; Pajares, 1996, in Leder, Pehkonen, & Torner, 2002, p. 216). Referring to Bandura's concept of self-efficacy as confidence in one's ability to organize and carry out a number of actions needed to generate the expected result, with the same understanding, Philippou and Christou (Leder et al., 2002, p. 217), mention that the teaching efficacy can be understood as a belief in the ability of teachers to organize and create effective learning environments. The activities and actions of teachers are more dependent on what they believe than on what they know, or the competence they rarely achieve. The same idea is said by Hoy and Spero (2005, p. 29) that teachers 'sense of efficacy as teachers' judgments about reviews their capabilities to promote student learning.

Gibson and Debo (Leder et al., 2002, p. 218) classify the teacher's self-efficacy into two factors: general teaching efficacy (GTE) and personal teaching efficacy (PTE). GTE refers to teachers' general feeling that their teaching and education system will be able to grow and develop students' academic achievement despite the negative influence of outside

teachers. Meanwhile, personal teaching efficacy (PTE) is intended as a reflection of teachers conviction on their own to continue the significant learning and student achievement. Furthermore, Philippou and Christou (Leder et al., 2002, p. 217) state that the efficacy of belief about the teaching of mathematics is mostly, but not entirely, shaped by one's experience and knowledge of mathematics and its pedagogy. The process skills of mathematics teachers should also be developed, i.e skills in reasoning, understanding of the concept, the relationship between concepts, representation, communication, and problem solving.

In relation to teachers' efficacy and competence, Tschannen-Moran et al. (1998) proposed an Integrated Model of Teacher Efficacy Cyclical. Furthermore, with reference to Lee (2009, p. 15) about the teacher's self-concept, as well as the opinion of Gibson and Dembo (1984) about the outcome efficacy, the approach develops a system model as presented in Figure 1.

Method

Research on the construct of the instrument for measuring junior high school mathematics teacher self-efficacy is a kind of developmental research (Borg & Gall, 1983, p. 775), to obtain a construct of dimensions or factors in relation to the self-efficacy of mathematics teachers, especially junior high school mathematics teachers in the Special Region of Yogyakarta.

This study was conducted over four months from September to December 2013) in four districts, namely Sleman, Bantul, Kulonprogro, and Gunungkidul regencies, as well as in the municipality of Yogyakarta in the of Special Region of Yogyakarta.

Research Design

The development of a modified construct in this study adapted Borg & Gall's model, which was simplified from 10 steps into 6 steps. The development of the modified construct is presented in Figure 2.

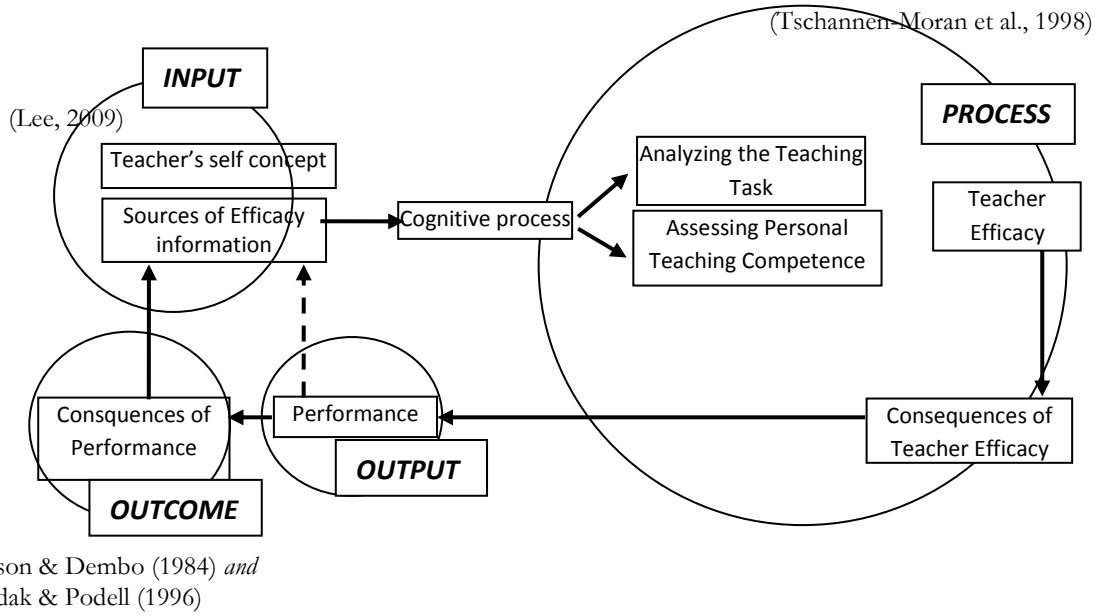

Figure 1. Development of "*The cyclical nature of teacher eficacy*" (Tschannen-Moran et al., 1998); "*Outcome efficacy*" (Gibson & Dembo, 1984; Soodak & Podell, 1996), through the system approach.

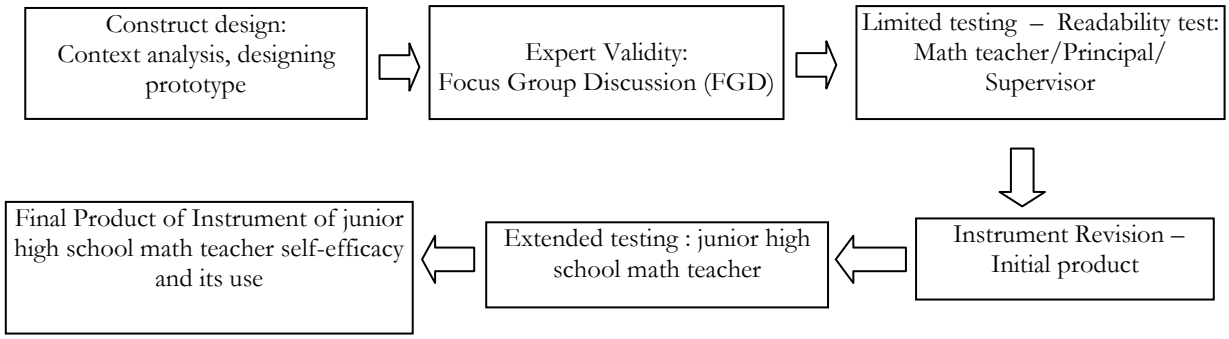

Figure 2. Research design on the development of the instrument of junior high school mathematics teacher self-efficacy

There are six stages in this research design. First, a construct was designed. This activity consisted of context analysis, relevant literature review, and prototype designing. At this stage, a preliminary instrument of mathematics teacher efficacy was developed. It consisted of four dimensions, namely: Personal Efficacy (PE), General Teaching Efficacy (GTE), Subject-Matter Teaching Efficacy (STE), and Outcome Efficacy (OE) based on the relevant theory and literature. The total number of items in this prototype instrument was 94 items consiting of: PE (25 items), GTE (26 items), STE (33 items), and OE (10 items). Likert scale was used with the rating scale of 1–4.

The second stage was validation by experts. This activity was a focus group discussion (FGD) involving eight experts or specialists from universities, consisting of: two mathematics education experts, three psychometric experts, one educational psychologist, and two experts on teacher training. The aspects assessed included: blue-print and indicators, clarity of the instruments, and the model development. The FGD results obtained content validity (content validity coefficient) through Aiken validity (Aiken, 1985, p. 132; Azwar, 2013, p. 134) was 0.71, meaning that the instrument could be used for collecting data.

The third stage was the limited testing. This activity involved 32 people, consisting of 22 mathematics teachers (of three districts in the province) and seven principals and three supervisors. The results of this readability test obtained a score of 4.13 which means that the instrument could be used.

The fourth stage was the revision or improvement. Based on the expert judgement in FGD and the limited testing, revision was done to improve the instrument in accordance with the input and advice from the experts. The fifth stage was the extended testing. In this case the test subjects were as many as 274 mathematics teachers in the Special Region of Yogyakarta.

The sixth stage was the final product and its use. The data from the extended testing were analyzed by using Lisrel 8.80 through the the first order and the second order analysis of CFA in order to obtain a suitable construct between the model and data. Based on this instrument, the researchers employed it for mapping the level of mathematics teacher self-efficacy.

Population and Sample

The population of this research was junior high school mathematics teachers, by referring to the data of the Provincial Education Department of Yogyakarta in 2012. It consisted of 816 junior high school mathematics teachers. Using the proportionate random sampling technique (Cohran, 2010, p. 85) the researcher established a sample of 274 teachers, consisting of 38 teachers from Yogyakarta City, 85 from Sleman, 70 from Bantul, 38 from Kulonprogro, and 43 from Gunungkidul. The sample size in the CFA analysis was determined by the number of the observed variables or items. According to Hair, Black, Babin, and Anderson (2006), for the sample size, it is recommended to use the estimates of the Maximum Likelihood (ML) at 100-200.

Data Analysis Technique

In data analysis, Confirmatory Factor Analysis (CFA) consisting of first-order and second order with the software of Lisrel 8.80 was used. The evaluation criteria for the model fit were by p-value on Chi-square ($\chi 2$) and the Root Mean Square Error of Approximation (RMSEA). The model was declared fit if the p-value was greater than Chi-square ($\chi 2$); was not significant if p-value > 0.05, meaning there is no significant difference between the model with the data (Joreskog & Sorbom, 2003, p. 128). The evaluation model with the RMSEA was expected to show the RMSEA value of ≤ 0.05 for the model considered as close to or the RMSEA value of ≤ 0.08 for a model declared as a good fit model. Furthermore, the fit instrument construct was used to map the level of mathematics teacher self-efficacy, referring to the score of Mathematics Teacher Self-efficacy (X): $X < \mu - 1\sigma$ (low category), $\mu - 1\sigma \leq < \mu\ X + 1\sigma$ (medium category), and $\mu + 1\sigma \leq X$ (high category).

Findings and Discussion

The First Order Analyisis

Dimension of Personal Efficacy (PE)

This dimension consists of three indicators: mathematics self-concept, math anxiety and internalizing the source of efficacy. PE consisted of 25 items, the number of the items might decrease gradually in the the first order analysis for obtaining a fit model. Items V2, V5, V6, V7, V9, and V10 were eliminated because the t-value of the loading factor < 1.96. Items V24 and V25 were also eliminated because of a negative loading factor value. Items V3, V13, V18, V19, V20 and V23 errors were eliminated because they shared a variance among items as the cause of the goodness of fit value was not a significant dimensional construct.

The result of the first order analysis of CFA showed that the model was fit to the data by obtaining the Chi-Square = 53.61 df = 44 p-value = 0.15201 and RMSEA = 0.020. The number of items decreased from 25 items into 11 items, with the loading factor of 0.19 to 0.64. So the items on the dimensions of PE were 11 items, namely items V1, V4, V8, V11, V12, V14, V15, V16, V17, V21, and V22.

Dimension of General Teaching Efficacy

This dimension consited of four indicators: pedagogy content knowledge, classroom management, student engagement, and parental involvement. GTE consisted of 26 items, the number of item might decrease gradually in the first order analysis for obtaining a fit model. Items V29 and V30 were eliminated because the t-value of the loading factor <1.96. Items V26 and V43 were also eliminated because it had an error value variance greater than the value of the loading factor which caused the goodness of fit value was not significant.

The result of the first order analysis of CFA showed that the model was fit to the data by obtaining Chi-Square = 66.59; df = 54, p-value = 0.11670 and RMSEA = 0.029. The number of the items decreased from 26 items into 12 items with the loading factor of 0.22 to 0.71. Thus, the number of the fit items in dimensions of GTE was 12 items, namely items V28, V31, V35, V38, V41, V44, V45, V47, V48, V49, V50, dan V51.

Dimension of Subject-matter Teaching Efficacy (STE)

This dimension consisted of three indicators: knowledge of junior high school mathematics content, teaching strategies, and fostering student mathematical power. STE consisted of 33 items, the number of items might decrease gradually in the first order analysis for obtaining a fit model. Items V52, V56, V57, V58, V66, and V67 were eliminated because the value of t loading factor < 1.96. The items that were removed were Items V53, V54, V59, V60, V61, V62, V65, V68, V69, V70, V71, V74, and V78 because the error of variance was much greater than the value of the factor loading, causing the value of goodness of fit not significant.

The result of the first order analysis of CFA showed that the model was fit to the data by obtaining Chi-Square = 24.49; df = 20; p-value = 0.2216, RMSEA = 0.029. The number of the items decreased from 33 items

into 13 items with the loading factor from 0.22 to 0.40. Thus, there were 13 items left in STE's dimensional. They were items V55, V63, V72, V73, V75, V76, V77, V79, V80, V81, V82, V83, and V84.

Dimension of Outcome Efficacy (OE)

The dimension of Outcome Efficacy *(OE)* consisted of three indicators, namely: student achievement, building mathematics attitude, and the continouing study. OE consisted of 10 items, and the number might decrease gradually in the first order analysis for obtaining a fit model. All items had a t-value of loading factor > 1.96 and no negative loading factor value. However, items V86 and V91 were eliminated because they had a value of error variance much greater than the value of the loading factor, causing the value of goodness of fit not significant.

The result of the first order analysis of CFA showed that the model was fit to the data by obtainng Chi-Square = 14.60; df = 9; p-value = 0.10256; RMSEA = 0.048. The number of the items decreased from 10 items to eight items, with the loading factor from 0.32 to 0.42. Therefore, the number of items was reduced to 8 items, namely V85, V87, V88, V89, V90, V92, V93 and V94.

The Second Order Analysis

Based on the items obtained in each dimension in the first order analysis, the second order analysis of CFA was done. Several simulations and iterations among these dimensions were done for obataining a fit model, such as: PE and GTE; STE and OE; dimensions of PE, GTE, and dimension STE. Finally, iterations of the next dimensions of PE, GTE, STE, and OE, derived a construct model that was fit to the data, as presented in Figure 3.

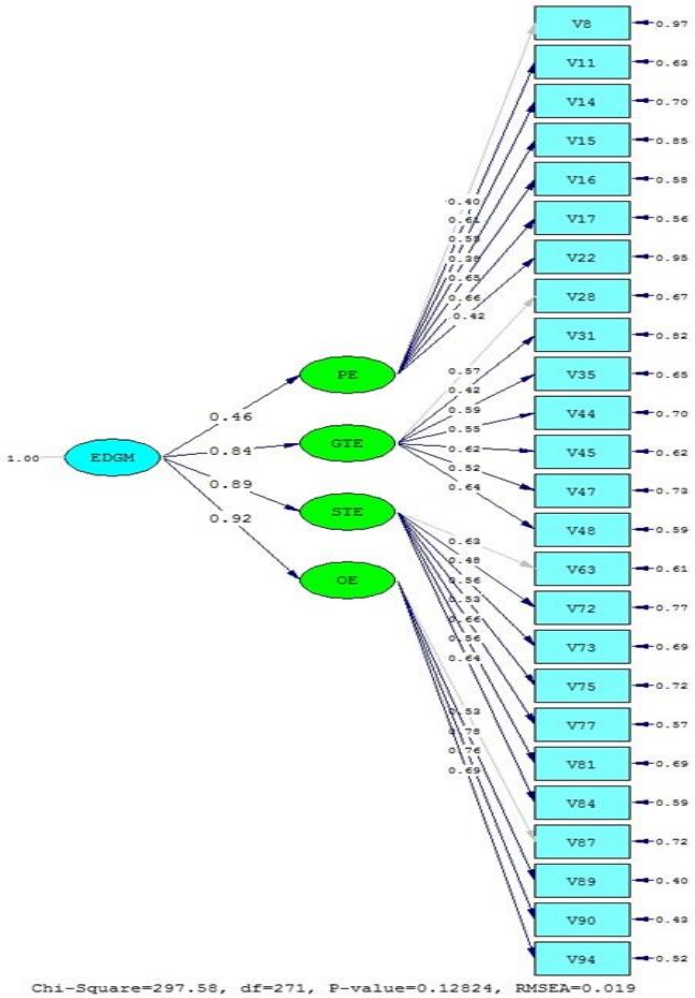

Figure 3. Path diagram of the second order analysis output

Table 1. Results of the second order CFA of the instrument of junior mathematics teachers' self-efficacy with 25 items

Item	*Loading Factor*	*t-value*	R^2	Result
Dimension: PE				
Item 8	0.40	---	0.19	*Reference Item*
Item 11	0.61	2.68	0.37	*Item Fit*
Item 14	0.58	2.62	0.30	*Item Fit*
Item 15	0.48	2.47	0.25	*Item Fit*
Item 16	0.65	2.63	0.42	*Item Fit*
Item 17	0.66	2.67	0.44	*Item Fit*
Item 22	0.42	2.53	0.21	*Item Fit*
Dimension: GTE				
Item 28	0.57	---	0.33	*Reference Item*
Item 31	0.42	6.41	0.18	*Item Fit*
Item 35	0.59	7.20	0.35	*Item Fit*
Item 44	0.55	6.05	0.30	*Item Fit*
Item 45	0.62	6.70	0.38	*Item Fit*
Item 47	0.52	5.48	0.27	*Item Fit*
Item 48	0.64	7.65	0.41	*Item Fit*
Dimension: STE				
Item 63	0.62	--	0.39	*Reference Item*
Item 72	0.48	6.04	0.23	*Item Fit*
Item 73	0.56	8.11	0.31	*Item Fit*
Item 75	0.52	7.68	0.28	*Item Fit*
Item 77	0.66	8.71	0.43	*Item Fit*
Item 81	0.56	8.35	0.31	*Item Fit*
Item 84	0.64	8.42	0.41	*Item Fit*
Dimension: OE				
Item 87	0.52	--	0.28	*Reference Item*
Item 89	0.78	8.17	0.60	*Item Fit*
Item 90	0.76	8.00	0.57	*Item Fit*
Item 94	0.69	7.31	0.48	*Item Fit*

The results of the tests performed on the measurement model of the second order analysis of CFA on 54 items resulted in p value = 0.12824 ($p > 0.05$) and RMSEA = 0.019 (RMSEA <0.05). Based on the data, p-value and RMSEA were successfully met so that it could be concluded this model was really fit with the data. The RMSEA value of 0.019 indicates that the model is very fit. In other words, all 25 items are valid indicators for measuring the instrument construct of self-efficacy of junior high school mathematics teachers. These results also showed that 25 items measured a latent variable, which was the self-efficacy of mathematics teachers. It was concluded that self-efficacy measurement instrument for mathematics teachers met unidimensionality assumptions. Table 1 is the table of all fit items of the results of second order CFA for measuring of junior high school mathematics teacher self-efficacy.

Based on the t-value of the second order CFA testing, it was known that all of the items were fit to measure junior high school mathematics teacher self-efficacy because the whole t-value was greater than 1.96. From Table 1, it is also noted that Item 89 has the highest contribution to the measuring instrument with the loading factor of 0.78, while Item 15 gives the smallest contribution to the loading fator of 0.38.

Mapping of Mathematics Teacher Self-Efficacy

The degree or level of mathematics teacher self-efficacy was obtained from the interpretation of the scores of an individual mathematics teacher as many as 274 teachers within 25 fit items. The scores obtained in the questionaire are raw scores, which need to be converted first into z-standard score, with $\mu = 0$, and $\sigma = 1$. However, because the standard z-scores allow their negative score, then for the ease of readability and interpretation, they need to be converted into t-score,with $\mu =$

50, and σ = 10. The result of the conversion of the scoring through a simple program MS Excel, and which referred to the categorization, as presented in the data analysis, shows the obtained mapping of mathematics teacher self-efficacy for each diemnsion as in Table 2. Meanwhile, the percentage of respondents' MTSE level is shown in Figure 4.

Discussion

The results of the analysis of the second order with Chi-Square = 297.58; df = 271; p-value = 0.12824; RMSEA = 0.019; 25 out of 54 items are with the factor loading (λ) of each dimension PE, GTE, STE, and OE consecutively being 0.46; 0.84; 0.89 and 0.92. Thus it can be said that the model was fit to the data.

Dimension of Personal Efficacy (PE)

The dimension of Personal Efficacy contains three indicators, from the initial 11 items (the first order) it decreases to 7 items (the second order). Those three indicators include (a) mathematics self-concept, with item descriptors: efficacy of the ability to provide necessary information known to the students in learning mathematics (V8); (b) mathematics anxiety, item descriptors; efficacy on the readiness of the teachers when they would teach mathematics (V11), tranquility or comfort during mathematics learning (V14), the level of concern toward the material that was not be acquired (V15), having difficulty in concentrating while teaching mathematics (V16), and concerns if there were other people observing their teaching (V17); (c) the internalization of the source of efficacy, with efficacy item descriptor against social persuasion such as: invitations, suggestions, and verbal advice from a colleague which can push them to perform task (V22).

Table 2. Frequency recapitulation of mathematics teacher self-efficacy (MTSE) level for each dimension

Category	Dimensions/Factors				MTSE
	PE	GTE	STE	OE	
Low(L)	82	94	94	93	93
Medium (M)	192	179	181	170	178
High (H)	0	2	4	1	4
Total Number					274

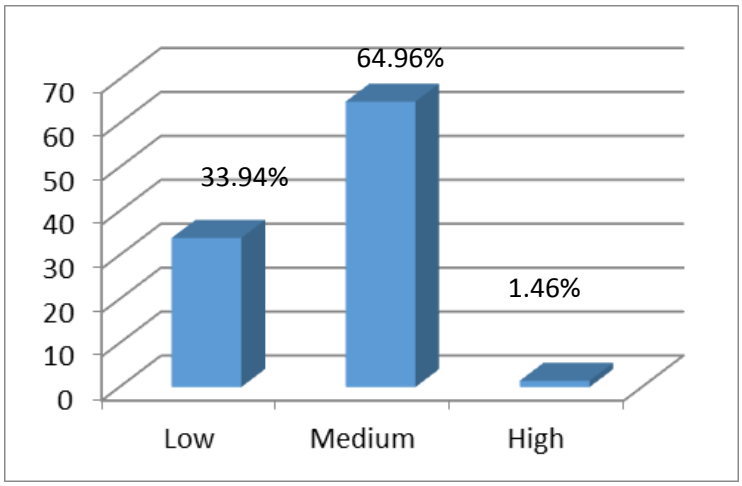

Figure 4. Graphic of Percentage MTSE level

Dimension of General Teaching Efficacy (GTE)

General Teaching Efficacy dimension contains four indicators (from the initial 12 items (the first order) decreasing to 7 items (the second order)). Those four indicators are: (a) pedagogy content knowledge of mathematics, with item descriptors including efficacy of ability to apply appropriate learning strategies in classroom practice (V28); (b) classroom management, with item descriptor efficacy toward abilities including students' drive to obey the rules in class (V31), explanation of the steps that must be performed by students in learning inside and outside the classroom (V35); (c) students' engagement with descriptor item efficacy toward abilities which help students to be actively involved in fun and meaningful learning (V44), maintain or restore the students' attention to stay focused on the material presented (V45), enhance students' understanding (V47) and assure the students that they can perform for completing the lesson task at school well (V48).

In GTE dimensions, there actually exist one more indicator, that is promoting the parental involvement in helping their children learning mathematics. At the time of the first-order analysis of GTE, all of the three items representing this indicator exist, but unfortunately on the second order CFA three items are forced to be eliminated to fit its models. In this study, the respondents might argue that the indicators were not required to measure mathematics teacher self-efficacy. Thus, in other words, the respondents believed that parental involvement might not help their children learning. The existence of out-of-school learning guidance, additional lesson, or private lesson would replace this role.

Dimensions of Subject-matter Teaching Efficacy (STE)

The dimensions of STE (Subject-matter Teaching Efficacy) consist of three indicators. The initial 13 items (the first order) decreased to 7 items (the second order). Those three indicators are: (a) the strategy of mathematics teaching, with item descriptors being the efficacy toward the ability to guide students in using a representation of an image, and symbol for mathematics learning (V63); (b) fostering students' mathematical power, with item descriptors being the efficacy in abilities to capture gaps between students' capability and competencies expected (V72), guiding students in examining the true relationship between one statement and others (V73), guiding students in developing a conjecture from available premises (V75), designing learning that encourages students to appreciate the benefits of mathematics (V77), managing the provision of questions to students (V81), and efficacy toward the ability to provide questions to students relating the idea of mathematics and its applications (V84).

In fact, there is one more indicator in STE dimension, that is acquiring mathematics content knowledge. At the time of first-order analysis of GTE, there is one out of five items as a representation of this indicator, but in the second order CFA, one item was forced to be eliminated to fit its models. It means that in this study, these indicators were not required to measure mathematics teachers' self-efficacy. In other words, acquiring the mathematics content knowledge only by asking thorugh questionaire is not enough. Using a test to measure this domain is more reasonable. The existence of Teacher Competency Testing would support the absence of this indicator.

Dimension of Outcome Efficacy (OE)

The dimension of Outcome Efficacy (OE) consists of three indicators, the initial eight items (the first order) decrease to four items (the second order). Those three indicators include: (a) student achievement, with item descriptor efficacy against the ability to: guide students to succeed in the mathematics contest or mathematics olympiad in district region (V87), (b) bulid a mathematics attitudes, with items descriptor belief that mathematics learning done by the teacher is to promote students' critical logical thinking and to be consistent (V89), efficacy toward the belief that learning is done to guide the students to be honest, disciplined and responsible (V90), (c) continouing study, with item descriptor efficacy toward the belief that learning is done to be able to equip students to practice problem solving in their future life (V94).

Based on the discussion, a framework of a construct for junior high school mathematics teacher self-efficacy in the Special Region of Yogyakarta could be made, consisting of four dimensions. The first dimension is Personal Efficacy (PE) with a loading factor (λ=0.41), consisting of indicators including mathematics self-concept, mathematics anxiety, and also internalization source of efficacy. The second dimension is General Teaching Efficacy (GTE) with a loading factor (λ = 0.84), consisting of indicators including pedagogy content knowledge, classroom management, and student engagement. The third dimension is Subject-matter Teaching Efficacy (STE) with a loading factor (λ=0.89), consisting of indicators including teaching strategies and fostering students' mathematical power. The fourth dimension is Outcome Efficacy (OE) with a loading factor (λ=0.92), consisting of indicators including student achievement, building mathematics attitude, and continuing study.

Mapping of Mathematics Teacher Self-Efficacy

Figure 4 shows that the percentage of mathematics teacher self-efficacy is 1.46% in a high category, 64.96% in a medium category, and 33.94% in a low category. With the hope of an ideal efficacy of mathematics teachers reaching high categories, as many as 98.54% of mathematics teachers should be enhanced for high category efficacy.

In order to give an idea of the profile of mathematics teacher self-efficacy in these categories and to make it easier to follow up the results of measurements of efficacy in the process of continuing professional development of mathematics teachers, and refer to the indicators and items of MTSE which have been fit and significant, Table 3 is a general description of MTSE (Mathematics Teacher Self-Efficacy) profile.

Table 3. General description of mathematics teacher self-efficacy

MTSE Category	Description of MTSE
Low **(Score: 24.96 – 49.92)**	**a. Not sure**: the importance of understanding and the role of mathematics self-concept, overcoming math anxiety, and internalization of the sources of self-efficacy, in carrying out the task of teaching responsibility. b. Not sure: able to master knowledge of the pedagogical substance, class-room management, and students engagement c. Not sure: able to perform mathematical learning strategies and fostering students' mathematical power. d. Not sure: able to improve student achievement, students' mathematical attitudes, as well as provisions for the continuation of the study at the next level.
Medium **(Score: 49.93 – 74.88)**	**a. Sure**: the importance of understanding and the role of mathematics self-concept, overcoming math anxiety, and internalization of the sources of self-efficacy, in carrying out the task of teaching responsibility. b. Sure: able to master knowledge of the pedagogical substance, classroom management, and students engagement c. Sure: able to perform mathematical learning strategies and fostering stu-dents' mathematical power. d. Sure: able to improve student achievement, students' mathematical atti-tudes, as well as provisions for the continuation of the study at the next level.
High **(Score: 74.89 – 100.00)**	**a. Very Sure**: the importance of understanding and the role of mathema-tics self-concept, overcoming math anxiety, and internalization of the sources of self-efficacy, in carrying out the task of teaching responsibility. b.Very Sure: able to master knowledge of the pedagogial substance, class-room management, and students engagement c. Very Sure: able to perform mathematical learning strategies and foster-ing students' mathematical power. d. Very Sure: able to improve student achievement, students' mathematical attitudes, as well as provisions for the continuation of the study at the next level.

Conclusion and Suggestions

Based on the findings, some conclusions are drawn. First, the results in the second order analysis of the construct of the instrument for measuring the self-efficacy of junior high school mathematics teachers in the Special Region of Yogyakarta shows the model is fit to the data, indicated by Chi-Square = 297.58; df = 271; p-value = 0.12824; RMSEA = 0.019, from 54 items, 25 items obtained with factor loading (λ) each dimension PE, GTE, STE, and OE consecutively are: 0.46; 0.84; 0.89 and 0.92.

The construct of the instrument for measuring self-efficacy of junior high school mathematics teachers in the Special Region of Yogyakarta consists of four dimensions. First, the dimensions of Personal Efficacy (PE) with a loading factor (λ=0.41) with indicators: mathematics self concept, mathematics anxiety, and the internalization source of efficacy. Second, the dimensions of General Teaching Efficacy (GTE) with a loading factor (λ = 0.84) with indicators: pedagogy content knowledge, classroom management and students engagement. Third, the dimensions of Subject-matter Teaching Efficacy (STE) with a loading factor (λ=0.89) consisting of indicators: teaching strategies and fostering students' mathematical power. Fourth, Outcome Efficacy (OE) dimension with a loading factor (λ=0.92), with indicators: student achievement, building mathematics attitude, and the continouing study.

The results of the mapping of the self-efficacy of mathematics teachers in the Special Region of Yogyakarta show that 43.07% of the teachers are categorized as low, 55.47% are categorized as moderate, and 1.46% are in high category.

Besides, some suggestions are proposed. First, further research or advanced analysis needs to determine the relationship between mathematics teacher self-efficacy and teacher performance. Second, research or further analysis is required to find out how instruments are constructed for senior or vocational high school mathematics teachers.

References

Aiken, L. R. (1985). *Psychological testing and assessment* (5th ed.). Boston, MA: Allyn and Bacon.

Azwar, S. (2013). *Penyusunan skala psikologi* (2nd ed.). Yogyakarta: Pustaka Pelajar.

Bandura, A. (1996). *Self-efficacy in changing societies*. New York, NY: Cambridge University Press.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W. H. Freeman and Company. Retrieved from https://books.google.co.id/books/about/Self_Efficacy.html?id=eJ-PN9g_o-EC&redir_esc=y

Borg, W. R., & Gall, M. D. (1983). *Educational research: An introduction* (4th ed.). New York, NY: Longman.

Burn, R. B. (1984). *The self concept, theory, measurement, development, and behavior*. New York, NY: Longman.

Cohran, W. G. (2010). *Teknik penarikan sampel* (Rusdiansya). New York, NY: John Wiley & Sons.

Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology*, *76*(4), 569–582. https://doi.org/10.1037/0022-0663.76.4.569

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2006). *Multivariate data analysis*. Upper Saddle River, NJ: Prentice-Hall.

Hastuti, H., Sulaksono, B., Akhmadi, A., Syukri, M., Sabainingrum, U., Ruhamaniyati, R., … Suryahadi, A. (2009). *Pelaksanaan sertifikasi guru dalam jabatan 2007: Studi kasus di Provinsi Jambi, Jawa Barat, dan Kalimantan Barat*. Jakarta.

Hoy, A. W., & Spero, R. B. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education*, *21*(4), 343–356.

Jalal, F., Samani, M., Chang, M. C., Stevenson, R., Ragatz, A. B., & Negara, S. D. (2009). *Teacher certification in Indonesia: A strategy for teacher quality improvement*. Jakarta: Ministry of National Education.

Joreskog, K. G., & Sorbom, D. (2003). *Lisrel 8.54 help file*. Chicago, IL: Scientific Software International.

Kastberg, S. E., D'Ambrosio, B., McDermot, G., & Saada, N. (2005). Context matters in assessing students' mathematical power. *For the Learning of Mathematics*, *25*(2), 10–15.

Keller, J. M. (2010). *Motivational design for learning and performance: The ARCS model approach*. Florida: Springer.

Law No. 14 Year 2005 of Republic of Indonesia about Teachers and Lecturers (2005).

Leder, G. C., Pehkonen, E., & Torner, G. (2002). *Beliefs: A hidden variable in mathematics education*. New York, NY: Kluwer Academic Publishers.

Lee, J. (2009). *Self construct and anxiety across cultures*. Princeton, NJ.

Leung, F., & Ragatz, A. B. (2010). *Inside Indonesia's classrooms: A TIMSS video study of teaching practices and student achievement*. Jakarta: Human Development Department of East Asia and Pacific Region, World Bank Office.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, *66*(4), 543–578.

PPPPTK Matematika. (2007). *Laporan kegiatan Training need assessment (TNA) dan rekruitmen SMP tahun 2007*. Yogyakarta.

Sadtyadi, H., & Kartowagiran, B. (2014). Pengembangan instrumen penilaian kinerja guru sekolah dasar berbasis tugas pokok dan fungsi. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *18*(2), 290–304. https://doi.org/10.21831/pep.v18i2.2867

Soodak, L. C., & Podell, D. M. (1996). Teacher efficacy: Toward the understanding of a multi-faceted construct. *Teaching and Teacher Education*, *12*(4), 401–411. https://doi.org/10.1016/0742-051X(95)00047-N

Sumardyono. (2011). Kecemasan matematika guru matematika dinas pendidikan kota Banjarmasin. *Edumat: Jurnal Edukasi Matematika*, *2*(4).

Tschannen-Moran, M., Hoy, A. W., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, *68*(2), 202–248. https://doi.org/10.2307/1170754

Turnuklu, E. B., & Yesildere, S. (2007). The pedagogical content knowledge in mathematics: Pre-service primary mathematics teachers' perspectives in Turkey. *IUMPST: The Journal*, *1*(October). Retrieved from www.k-12prep.math.ttu.edu

REiD (Research and Evaluation in Education), 3(1), 2017, 77-91
Available online at: http://journal.uny.ac.id/index.php/reid
Research Article

An assessment model of Islamic religion education teacher personality competence

***[1]Sovia Mas Ayu; [2]Marzuki**
*Faculty of Education and Teacher Training, Institut Agama Islam Negeri Raden Intan Lampung
Jl. Letkol H. Endro Suratmin Sukarame 35131, Kota Bandar Lampung, Indonesia
*Email: ayurijaya@yahoo.com

Submitted: 18 May 2017 | Revised: 17 June 2017 | Accepted: 19 June 2017

Abstract

The study aimed at developing an assessment model of Islamic religion teacher personality competence consisting of: (1) instrument constructs; (2) instrument characteristics; and (3) assessment characteristics. The method implemented in the study was research and development, developed by Borg & Gall. The instrument validation was conducted by experts through expert judgement continued by means of V-Aiken formula. The first experiment of the instrument involved 50 Islam religion education teachers, 50 non-Islamic religion education teachers and 50 students. The results from the first experiment were analyzed by means of EFA in order to prove the validity of the instrument construct. The second experiment of the instrument involved 200 Islamic religion education teachers, 200 non-Islamic religion education teachers and 200 students. The results of the second experiment were analyzed by means of CFA in order to prove the validity of the instrument constructs. The results of the study showed that: (1) the instrument construct of teacher personality consist of the *al-'Iffah*, *al-Syaja'ah*, *al-Hikmah* and *al-'Adalah* dimensions; (2) the results of the CFA analysis show that there is a compatibility between the model and the empirical data; and (3) the assessment guidelines are effective and good.

Keywords: *assessment model, personality competence, teacher*

How to cite item:

Ayu, S., & Marzuki, M. (2017). An assessment model of Islamic religion education teacher personality competence. *REiD (Research and Evaluation in Education), 3*(1), 77-91. doi:http://dx.doi.org/10.21831/reid.v3i1.14029

Introduction

Teachers have an important role in shaping students' personality and internalizing moral values to be implemented by the students. Like parents, teachers not only educate and teach students, but also are responsible to shape them to be better persons.

Teaching refers to a series of activities that include teachers' activities, students' activities, and learning processes. All of the three aspects have dynamic interaction from one to another. Interaction in the teaching process is very complex and, therefore, interaction is defined as the teachers' procedures or behaviors (Mouly, 1973, p. 1). Such dynamic interaction should be able to encompass the ethical values and the behavioral esthetics within the students' behaviors in encountering the life challenges of their community.

Based on the report of the Commission of Indonesian Children Protection (*Komisi Perlindungan Anak Indonesia*, KPAI) released in their official website on June 14th, 2015, from 2007 to 2014, there was a significant increase within the data of violence against children. In addition, based on the results of the monitoring and evaluation conducted by the same commission in 2012, the percentage of violence against children was 87.60% from nine

provinces in the Republic of Indonesia and 29% of the violence case was the violence committed by teachers to students.

The significant role and responsibilities that teachers have within the educational process encourage the government to improve and develop the teachers' professional competencies. The concrete form of the government's effort is the issuance of the Law No. 14 Year 2005 of Republic of Indonesia about Teachers and Lecturers (2005) and the Government Regulation Number 13 Year 2015 regarding National Education Standards (the second amendment was under the Regulation of the Government of Republic of Indonesia, No. 32 Year 2013, on National Education Standard (2013). The teacher competencies include: pedagogic, personality, professional, and social competencies.

Teachers then will be referred to as educators and should have academic qualifications and competencies as teaching agents, should be physically and mentally healthy and should have the capacity to manifest the objectives of national education. Teachers' competencies are formally proven by the certificates of educator that will be held after the completion of certification examination.

The implementation of teacher certification should be attended by all of the teachers who have not held the certificate of educator and who still actively teach in the schools under the direction of Ministry of Education and Culture except for the teachers of Religion education. The competencies of Islamic religion education teachers in the elementary and secondary education levels as well as in the early childhood education according to the Regulation of Minister of Religion Affairs of Republic of Indonesia, No. 16 Year 2010, on the Management of Religion Education at Schools (2010) include: pedagogic, personality, social, professional, and leadership competencies.

The assessment process for the teachers' competence fulfillment has not touched the personality competencies. The materials of the competence test for religion education teachers that has been conducted in online manner take the form of multiple choice items and objective test items with four options and the materials cover 30% of the pedagogic competence and 70% of the professional competence. The teachers who have passed the competence examination should attend the Teacher's Professional Training and Education Program (*Pendidikan dan Latihan Profesi Guru*, PLPG) or the Teacher's Professional Education Program (*Pendidikan Profesi Guru*, PPG) conducted by the accredited and certified LPTK institutions.

The term competence refers to capacity or capability. It is defined as the fundamental characteristics that one possesses in adapting himself or herself to the required criteria for effective and superior performance in accordance with the job and community demands (Spencer & Spencer, 1993, p. 9). The two statements imply that competence is an individual's characteristics in accomplishing his or her duty.

The characteristics of an individual are associated with the performance and are indicated by motives, traits, self-concept, knowledge and skills. Specifically, Spencer and Spencer (1993, pp. 9–11) mention five characteristics within the competence: (1) motives, which refer to the objects that an individual thinks, that an individual desires and that encourage an individual to perform an action; (2) traits, which refer to the derivative characteristics in the form of an individual's physical characteristics and responses toward situations and information; (3) self-concept, which consists of individual's attitudes, values and self-impression; (4) knowledge, which refers to all information and knowledge that an individual has in certain domains; and (5) skills, which refer to the capacity in performing a task both mentally and physically.

Personality is one of the five competencies that must be mastered by an Islamic religion teacher. The personality of a teacher is apparent from: the actions that are in accordance with the Indonesian's national religion; the legislative, social, and the cultural norms; his or her appearance as an honest person, a nobel person, and a good role model for the students and the community; his or her self-appearance as a settled, stable, mature, wise, and dignified person; the possesion of working ethos, high responsibility, pride in

being a teacher, and self-confidence; and the respect toward teacher's code of conduct.

The term personality is derived from a Latin word, *persona*, meaning a mask used by the actors in a game or a stage. Cattel (1965, p. 27) states that personality is the basis of an individual's behaviors in overcoming a situation and has been performed under full consciousness. Further, according to Allport in Aiken (1999, p. 22), personality is a dynamic organization of psycho-physical individual processes that define the peculiar manner of an individual in adjusting himself or herself toward his or her neighborhood. From both statements, the researchers may conclude that personality is a peculiar behavior of an individual in adjusting himself or herself to the neighborhood.

In Islam, as stated by Wahab (2010, p. 1), personality is defined as nobility. In Arabic, personality has been shown by the word *sulukiyyah* (behavior), *khuluqiyyah* (nobility), *infi'aliyyah* (emotion), *jasadiyyah* (physics), *qadarah* (competence) and *muyyul* (interest). It is defined as *khuluqiyyah* (nobility) which refers to the behavior or the will. Behavior is an action that has been performed repetitively (Amin, 1975, p. 74). Therefore, personality is a will-based action performed repetitively so that the action will turn into a habit.

Rajab (1961, p. 21) defines personality as *al-thab'u* (characters) and as *al-sajiyah* (habit/ *al'adah*). *Al-thab'u* itself is defined as the internal description that resides within the human beings. The internal description refers to all of the matters that have been created by Allah SWT, God The Almighty, since they were born. On the contrary, *al-sajiyah* or the habit/ *al'adah* refers to the results of integrating the humane characters and the performed activities. Such opinion an has asserted that personality refers to the actions performed by an individual that may reveal his or her characters.

The Islamic personality refers to the Islamic normative teachings-based attitude or the personality that has been conducted by a Moslem both individually and also socially (Muhaimin, 2005, pp. 273–274). From the statement, it can be concluded that the intended attitude or behavior refers to the action which is based on the religious norms and regulations.

According to Prihatini, Mardapi, and Sutrisno (2013, p. 349), personality is frequently associated with the moral or habit that has been perceived and performed and, specifically, personality is also associated to knowledge regarding moral. A person with good personality is frequently referred to as a person with good moral behaviors and, similarly, a person with bad personality is frequently referred to as a person with bad moral behaviors. Therefore, there should be a moral reference.

The word teacher in Arabic is addressed by several words that mean profession namely *mudarris*, *muallim*, *murabbi* and *muaddib*. These words have similar meanings but different characteristics. In the dictionary of *Mu'jam al-Wasith* (Anis, 1972, p. 526), *mudarris* is derived from the root form *darasa* that means to study and, therefore, the term *mudarris* is defined as a person who teaches lessons to students. Then, *mu'allim* is derived from the root form *'allama* that means to tell or to inform and, therefore, the term *mu'allim* is defined as a person who delivers knowledge.

According to Muhaimin and Mujib (1993, p. 164), the term *muaddib* is derived from the root form *addaba* that refers to behaviors and, therefore, the term *muaddib* is defined as a person who teaches behaviors namely all kinds of movement, attitudes and deeds as well as good values based on the Islamic teachings. From these definitions of a teacher, the researchers might conclude that a teacher is a person who delivers knowledge, skills or experiences to other persons so that the knowledge, skills, or experiences will be the basis for the students to actualize their attitude and behaviors.

Al-Nahlawi (1995, p. 170) states that the main duty of a teacher is: (1) to purify, namely to serve as a purifier, a preserver, and a developer of human purity; and (2) to teach, namely to internalize the religious values and to transform the knowledge into human beings. It might be inferred that the function of a teacher in the Islam religion education is to serve as the executor of transformation and internalization process of the values of Islam-

ic teachings and knowledge in order to develop the purity and the basic capacity that the students have in order to achieve the balance and the equality within all life aspects.

The form of a teacher's personality, according to Daradjat (1982, p. 3), might be found from his or her appearance, words, socialization, dressing manner and ways of dealing with other people. The attitudes and behaviors of a religion teacher will describe the self-identity, the characters, and the personality that he or she has.

According to Asyari (1238, pp. 32–34), teacher's personality is divided into three aspects namely: (1) the teacher's ethics toward himself, namely, adhering himself or herself to Allah, being afraid of the Allah's wrath, being careful to each word and action and paying attention to *tawadhu*; (2) the teacher's ethics during the teaching process, namely, greeting the students as they enter the classrooms, having conversation with nice words, explaining the lessons that become prominence and not raising the tone of voice; and (3) the teacher's ethics to the students, namely, delivering the knowledge only because of the order of Allah, avoiding the insincere attitude, loving the students as loving his or her own self, and explaining the learning materials clearly.

In addition, according to Miskawaih (2010, p. 24), the concept of personality is established through the four dimension of a Moslem's personality, namely: (1) the dimension of *al-'Iffah* (modesty); (2) the dimension of *Syaja'ah* (courage); (3) the dimension of *al-Hikmah*; and (4) the dimension of *al-'Adalah*.

Al-'Iffah (modesty) refers to the educatedness of lust through the education of idea and *syariat*. The dimension of *Al-'Iffah* might be internalized within human beings through the habituation since early childhood. The matters that might grow the dimension of *Al-'Iffah* within an individual are: faith and piety, marriage, and embarrasment.

Miskawaih (1398, p. 32) explains that there are 12 branches of prominence in the dimension of *Al-'Iffah*, namely: (1) *Al-haya*, the self control for being afraid of committing bad action; (2) *Al-da'ah*, self-calmness when the desire appears; (3) *Al-shabr*, the desire of not being lured by the negative impact of deliciousness; (4) *Al-sakha*, a balanced attitude in terms of gift; (5) *Al-hurriyyat*, self-prominence in attaining, providing, and refusing wealth under the appropriate manner; (6) *Al-qana'at*, attitude of being moderate in terms of food, drink and jewelry; (7) *Al-damasah*, the self-inclination toward the good aspects and the immediate realization of the good aspects; (8) *Al-intizham*, the self-condition that values and arranges many aspects in a very good manner; (9) *Husn Al-hady*, the fond of decorating one's self with the good aspects; (10) *Al-musalamat*, self-capacity to leave something that has bad consequence; (11) *Al-waqar*, the self-calmness when the demand of desire starts being demanded; and (12) *Al-wara*, the continuity in performing good deeds.

Al-Syaja'ah, also known as courage, is one of the characteristics that an *istiqamah* has in fighting within the way of Allah. The intended courage is not a baseless courage; instead, the intended courage refers to the courage in performing and encouraging the truth based on the religious values. A teacher is the leader in the classroom; as a leader, a teacher has authority in the implementation of learning activities.

Miskawaih (2010, p. 30) states that there are nine characteristics of *Al-Syaja'ah* that a leader has, namely: (1) dignity (كبارالنفس/*kibar al-nafs*); (2) unyielding (النجدة/*al-najdah*); (3) calmness (**لهمةامظع**/*'idzmu al-himmat*); (4) perseverance (الثبات/*al-tsabat*); (5) patience (الصبر/*al-shabr*); (6) generosity (الحلم/*al-hilm*); (7) self restraint (**دمعالطيش**/*adam al-thaisy*); (8) powerfulness (الشحمة/*al-syahamat*); and (9) enormous durability/fond of hardworking (**احتمال الكاد**/*ihtimal al-kadd*).

Al-Hikmah is a self-condition that understands the right and wrong choice in all behaviors that have *ikhtiar* (choices). *Hikmah* as the main nobility, according to Al-Ghazali (1957, p. 20), refers to the capacity of an individual in controlling the desire. Desire within oneself includes the rage and the lust. The capacity to control the desire will lead an individual to the life happiness and success.

The prominence that belongs to the *Al-Hikmah*, according to Miskawaih (2010, p. 27) is as follows: (1) intelligence acuity (الذكاء/*ad-*

zakaa); (2) memory durability (الذكر/*ad-zikru*); (3) rationality (التعقل /*at-ta'aqqul*); (4) clear mind and agility (سرعةالفهم/*sur'ah al-fahmi*); (5) clear memory/understanding (صفاءالذهن/*shofa al-zihni*); and also (6) easy studying (سهولة التعلم/ *suhuulah at-ta'allum*). An individual might achieve the prominence through the knowledge and the learning perseverance.

Al-'Adalah, according to Al-Ghazali's view (1957, p. 31), refers to the success of human beings upon their action and they will be responsible for their actions. Justice, in the meantime, refers to the self-condition and self-strength in overcoming and mastering the emotion and lust in accordance with their needs.

According to Miskawaih (2010, p. 28), the prominence that belongs to the *Al-'Adalah* are as follows: (1) honesty (الصدقة/*ash-shidqah*); (2) fondness of having association (الألفة/*al-ulfah*); (3) fraternity establishment (صلةالرحم/ *shilatu ar-rahim*); (4) reward provision (المكافاة/ *al-mukaafaah*); (5) good working relationship (حسن الشركة/*husnu asy-syirkah*); and (6) good decision making (حسن القضاء/*husnu al-qhada*). An individual might achieve the prominence through the knowledge and the experience.

The concept of *Al-'Iffah* in the study will be developed into five indicators namely: reluctance, politeness, piety, discipline and humbleness. Then, the concept of *Al-Syaja'ah* in the study will be developed into five indicators namely: dignity, responsibility, generousity, patience, and self-restraint. Next, the concept of *Al-Hikmah* in the study will be developed into five indicators, namely: wide knowledge, rationality, creativity, criticality and optimism. Last but not least, the concept of *Al-'Adalah* in the study will be developed into five indicators, namely: strictness, equality, not ruthlessness, honesty, and positive thinking.

Method

The study is categorized as research and development referring to the procedures of research and development proposed by Borg and Gall (1983, p. 772). The study aims to generate an assessment model for measuring the personality competence of an Islamic religion education teacher.

The research and development model proposed by Borg & Gall was modified into several stages, namely: (1) information gathering; (2) planning; (3) preliminary product development; (4) preliminary testing; (5) preliminary product revision; (6) limited testing; (7) main product testing and revision; (8) expanded testing; and (9) final revision.

The study was conducted at several senior and vocational high schools in Bandar Lampung, Central Lampung and Southern Lampung, Indonesia. Preliminary testing was implemented in 14 senior/vocational high schools located in Bandar Lampung from January to February 2015. The second testing was conducted in 75 senior/vocational high schools located in the Province of Lampung from April to June 2015.

The subjects in the study include Islamic religion education teachers, non-Islamic religion education teachers and the students. In order to establish the sample, the researchers implemented purposive sampling technique based on several criteria (Borg & Gall, 1983, p. 248). The researchers' consideration in establishing the sample was that the number of Islamic religion education teachers in each school was very limited. Then, the number of respondents in the final product testing was 200 Islamic religion education teachers, 200 non-Islamic religion education teachers and 200 senior/vocational high school students.

Procedures

The research and development (R & D) was conducted in two stages, namely development stage and experiment stage. The development stage consisted of: (1) the preliminary study, including the information gathering and theoretical review; (2) planning, including the objective formulation and period planning, the fund and field officers; (3) initial design development, including the development of preliminary measurement draft and the preliminary product revision; and (4) the expert validation, including the readability test and the instrument clarity.

The second stage consisted of the first and second product testing that aimed at measuring the model validity, reliability, compatibility and the assessment of the model guide-

line. The phases in the product testing consisted of: (1) the first product testing; (2) the analysis and the first product revision; (3) the second product testing; (4) the analysis and the second product revision; (5) the instrument printing; and (6) the guideline composition.

Data, Instruments and Data Gathering Techniques

For the data gathering method, the researchers utilized an observation sheet and questionnaire. The observation sheet was submitted through the observation on the non-Islamic religion education teachers and the students regarding the personality of Islamic religion education teacher. The questionnaire was in the form of self-report, regarding the personality of the Islamic religion education teacher in relation to self-concept and performance within the school. The Islamic religion education teachers were asked to fill in the questionnaire.

The assessment instrument for the personality of Islamic religion education teacher was coded A, B and C. Instrument A and Instrument B were in the form of an observation sheet for the students and the non-Islamic religion education teachers with the four-scale assessment. On the other hand, Instrument C was in the form of self-report for the Islamic religion education teachers with four-scale assessment. The results of the assessment provided by the students, non-Islamic religion education teachers, and Islamic religion education teachers were calculated in order to find the mean values that served as the basis for the final assessment of the personality of Islamic religion education teachers.

Data Analysis Techniques

The content validity test from the experts was based on the mean values and the V-Aiken analysis. The reliability of the instrument for the assessment of the personality of Islamic religion education teachers used the following formula: Alpha Cronbach > 0.70 (Nunally, 1981, p. 230). An item would be considered good if its factor loading is bigger than 0.30. Further, the cummulative presentation shows the compatibility or the incompatibility to serve as an indicator if the number of presentation is > 50% (Solimun, 2002, p. 81). The factors that met the Eigenvalue > 1.00 were the factors that might be implemented as the indicators for a characteristic or a trait.

The data analysis technique implemented in the second product testing was the descriptive statistical analysis by means of SPSS 17.0 software and the Second-Order Confirmatory Factor Analysis (Second-Order CFA) by means of 8.51 LISREL software. In addition to testing the validity and the reliability, the confirmatory factor analysis was also implemented in order to test the compatibility between the theoretical model and the empirical data resulted from the field. According to Hair, Anderson, Tatham, and Black (2006, p. 579), the confirmatory factor analysis might be implemented to analyze the construct of an amendment.

The evaluation of the instrument of the teacher personality employed the analysis of the mean score with the ideal standard deviation for each assessed component and the ideal standard deviation referring to the modification model by Sudijono (2003, pp. 329–339) as presented in Table 1.

From the results of the mean score and classification of development results based on Table 1, the researchers evaluated the model, instrument, and assessment guideline by implementing an evaluation standard as presented in Table 2.

Table 1. Assessment category

No.	Formula	Score Mean	Category
1.	$X > X i + 1{,}8 \times sb_i$	>4.2	Very Good
2.	$x_i + 0{,}6 \times sb_i < X \leq x_i + 1{,}8 \times sb_i$	>3.4 – 4.2	Good
3.	$x_i - 0{,}6 \times sb_i < X \leq x_i + 0{,}6 \times sb_i$	>2.6 – 3.4	Moderate
4.	$x_i - 1{,}8 \times sx_i < X \leq \ddot{X}_i - 0{,}6 \times sb_i$	>1.8 – 2.6	Poor
5.	$X \leq X i - 1{,}8 X sb_i$	≤ 1.8	Very Poor

Note:
$\ddot{X}_i$ (ideal mean) = ½ (ideal maximum score + ideal minimum score)
sb_i (ideal standard deviation) = $1/6$ (ideal maximum score – ideal minimum score)
X = empirical score

Table 2. The standards of model, instrument, and guideline evaluation

No.	Score Mean	Category	Conclusions
1.	>4.2	Very Good	Might serve as an example
2.	>3.4 – 4.2	Good	Might be implemented without any revision
3.	>2.6 – 3.4	Moderate	Might be implemented with minor revision
4.	>1.8 – 2.6	Poor	Might be implemented with major revision
5.	≤ 1.8	Very Poor	Might not be implemented

Table 3. The final assessment for the personality competence of Islam religion education teacher

No.	Score Mean	Category	Personality Criterion
1.	≥ 2.6 – 4.0	Relevant	Good Personality/ *Al-Akhlak al-Mahmudah*
2.	≥ 1.6 – 2.5	Irrelevant	Less Good Personality/*Al-Akhlak al-Madzmumah*

The assessment of the personality competence of Islamic religion education teachers referred to *al-akhlaq al-mahmudah* (good personality) and *al-akhlaq al-madzmumah* (worse personality). The formulation proposed by Lawshe (1975, p. 5) that had usually been implemented to decide the relevant and irrelevant items on the process of quantifying consensus in the research and development analogized the criteria that had been relevant to those of *al-akhlaq al-mahmudah* (good personality) and of *al-akhlaq al-madzmumah* (worse personality). The system of final assessment of the personality competence of Islamic religion education teachers is presented in Table 3.

In the study, the criteria of model fitness determination referred to the criteria of fitness index in the form of Chi Square and Root Mean Square Error of Approximation (RMSEA). The probability of Chi Square in the testing was expected not to be significant ($p > 0.05$) in order that the model would be fit into the data. The second fitness index was the Root Mean Square Error of Approximation (RMSEA) < 0.08 which indicated that the model was fit. The criteria for determining the model fitness is presented in Table 4.

Table 4. Goodness of Fit Index

No.	Index	Cut of Value	Note
1.	Chi-Square (χ^2)	Small	n=100-200. χ^2
2.	Probability (ϱ)	≥ 0.05	and df are not significantly different
3.	Root Mean Square of Error Approximation (RMSEA)	≤ 0.08	For huge n
4.	Goodness of Fit Index (GFI)	≥ 0.90	Maximum 1

Source: Solimun (2008, p. 80).

Findings and Discussion

The development of the assessment model was conducted by performing several activities that had been arranged systematically, starting from the theoretical review, the empirical review to the field survey. As a result, these activities generated the concept of assessment instrument for the Islamic religion education teachers' personality competence.

The results of the theoretical review in the preliminary study are categorized into four dimensions of teacher personality based on the four main points of nobility in Islam, namely: *al-'Iffah* (modesty), *al-Syaja'ah* (courage), *al-Hikmah* (wisdom) and *al-'Adalah* (justice).

The dimension of *al-'Iffah* includes embarrasment, politeness, piety, discipline and humbleness. The dimension of *al-Syaja'ah* consists of dignity, responsibility, generosity, patience, and self-restraining. Next, the dimension of *al-Hikmah* covers wide knowledge, rational thinking, creativity, critical thinking and optimism. Last but not least, the dimension of *al-'Adalah* covers assertiveness, non-favoritism, non-arbitrariness, objectivity, and positive thinking.

The initial draft of the instrument within the preliminary study was discussed with the educational practitioners namely: Islamic religion education lecturers, Islamic religion education teachers, non-Islamic religion education teachers and students (for the reability test). From the results of the discussion, the researchers made the initial draft for the assessment of the personality of Islamic religion education teachers.

Then, the results of practitioner review in the preliminary stage were validated by experts consisting of *Tasawuf*/spiritual experts, Islamic education science experts, and measurement and evaluation experts. From the results of the expert judgement or the expert agreement (within the expert validation), the researchers attained a feedback on the instrument content, instrument readability, language and punctuation use, composition method and arrangement of each dimension. These activities were also conducted in order to validate the products that are in the form of procedures, instruments, and also guidelines on the model implementation in the second stage.

The results of the expert judgement were analyzed descriptively in order to view the mean score and the analysis was continued by the content validity testing. For the content validity testing, the researchers implemented the V-Aiken formula in order to calculate the content validity of the personality instrument. The results of the expert judgement are presented in Table 5.

Based on the mean and the results of content validity test with the V-Aiken presented in Table 5, the assessment instrument is accepted and has already possessed the criteria of goodness, sufficiency and implementability. After the researchers revised the instrument based on the experts' suggestions and opinions, they defined the fixed hypothetical constructs of personality into four dimensions and each dimension has five indicators.

The next stage in the study was field testing, which was performed twice; the first field testing is the limited scale testing, while the second field testing is the expanded scale testing. The first field testing was performed in 14 senior/vocational high schools located in Bandar Lampung. The subjects in the first field testing were 50 students, 50 non-Islamic religion education teachers, and 50 Islamic religion education teachers. The data were analyzed by means of Exploratory Factor Analysis (EFA) in order to prove the validity of the instrument construct and the results of the analysis is presented in Table 6.

Table 5. The results of V-Aiken analysis for the dimension of personality of Islam religion education teacher

No.	Personality	Indicator	V-Aiken	Criteria	Mean	Criteria
1.	*Al-Iffah* dimension	Embarrasment	0.969	Accepted	5.00	Very good
		Politeness	0.891	Accepted	4.83	Very good
		Piety	0.844	Accepted	4.67	Very good
		Discipline	0.953	Accepted	5.00	Very good
		Humbleness	0.875	Accepted	4.83	Very good
2.	*Al-Syaja'ah* dimension	Dignity	0.969	Accepted	5.00	Very good
		Responsibility	0.922	Accepted	5.00	Very good
		Generosity	0.844	Accepted	4.33	Very good
		Patience	0.859	Accepted	4.83	Very good
		Self-restraint	0.812	Accepted	4.33	Very good
3.	*Al-Hikmah* dimension	Wide knowledge	0.969	Accepted	5.00	Very good
		Rationality	0.922	Accepted	4.67	Very good
		Creativity	0.937	Accepted	4.83	Very good
		Critical thinking	0.906	Accepted	4.83	Very good
		Optimism	0.859	Accepted	4.83	Very good
4.	*Al-'Adalah* Dimension	Assertiveness	0.891	Accepted	4.83	Very good
		Non-favoritism	0.937	Accepted	4.83	Very good
		Non-arbitrariness	0.891	Accepted	4.67	Very good
		Objectivity	0.937	Accepted	4.83	Very good
		Positive thinking	0.937	Accepted	5.00	Very good

Table 6. The summary of descriptive analysis results for the first field testing

No.	Dimension	N Resp.	N Item	Students' Mean (A)	Non-Islam religion education Teachers' Mean (B)	Islam Religion Teachers' Mean (C)	Total Mean	Category	Total Personality Mean
1.	*Al-Iffah*	50	15	3.5720	3.6080	3.5373	3.5724	Good	
2	*Al-Syajaah*	50	15	3.6000	3.5626	3.6573	3.6066	Good	**3.6083**
3.	*Al-Hikmah*	50	15	3.4706	3.6226	3.7333	3.6088	Good	**(Good)**
4.	*Al-Adalah*	50	15	3.5906	3.6866	3.6586	3.6453	Good	

Table 7. The results of variant analysis test

	Sum of Square	df	Mean Square	F	Sig.
Between Groups	46.188	2	23.094	2.820	.063
Within Groups	1203.656	147	8.188		
Total	1249.844	149			

Table 8. The results of first-stage instrument validity test and instrument reliability test

No.	Dimension	N	Coefficient Alpha	KMO	Bartlett's Test			Criteria
					Chi-Square	df	Sig.	
1.	*Al-'Iffah*	50	0.870	0.715	594.361	105	0.000	Good
2.	*Al-Syaja'ah*	50	0.879	0.690	513.750	105	0.000	Good
3.	*Al-Hikmah*	50	0.900	0.801	485.473	105	0.000	Good
4.	*Al-'Adalah*	50	0.873	0.729	505.504	105	0.000	Good

The results of the assessment conducted by students, non-Islamic religion education teachers, and Islamic religion education teachers were analyzed by using Anava technique to determine whether or not the mean score of the subjects was significantly different. The objective of implementing Anava technique within the study is to find the independent variables within the study and to define the intervariable interaction and its influence on the treatment. The results of the variant analysis show a the significance value of 0.063 > 0.050; in other words, there is no significant difference in the results of the assessment of the personality competence conducted by students, non-Islamic religion education teachers, and Islamic religion education teachers. The results of the analysis are presented in Table 7.

Table 7 show that there is no significant difference in the assessment of the personality competence conducted by the students, non-Islamic religion education teachers, and Islamic religion education teachers. Therefore, the researchers calculate the mean score of the personality as-sessment provided by the three groups of re-spondents to determine the final score of the teacher's personality assessment.

The results of teacher's personality assessment conducted by the students in the first stage testing were tested by means of exploratory analysis in order to prove the construct validity. The summary of the results of the instrument validity testing and of instrument reliability testing is presented in Table 8.

The results of the analysis in Table 8 show that the instrument had met the criteria of a good instrument. Then, from the results of Eigen value and components of personality variance, the researchers attained 5 as the score of Eigen value, bigger than 1. Therefore, it is concluded that there are five factors in each dimension of Islamic religion education teacher's personality that are explained. The summary of the Eigen values for each personality dimension is presented in Table 9.

The second stage testing was conducted at 75 senior/vocational high schools located in Bandar Lampung, Southern Lampung, and Central Lampung. The subjects were 200 students, 200 non-Islamic religion education teachers, and 200 Islamic religion education teachers. The results of the descriptive analysis of the second-stage testing are presented in Table 10.

Table 9. The summary of the results of eigenvalue in the first stage testing

No.	% Variance Explained							
	Al-'Iffah	% Var	*Al-Syaja'ah*	% Var	*Al-Hikmah*	% Var	*Al-'Adalah*	% Var
1.	6.587	43.91	5.907	39.38	6.523	43.49	6.406	42.710
2.	2.166	14.44	2.371	15.81	2.151	14.34	1.736	11.580
3.	1.278	8.52	1.664	11.09	1.174	7.83	1.527	10.180
4.	1.103	7.35	1.203	8.02	1.122	7.48	1.180	7.8670
5.	1.039	6.93	1.008	6.72	1.011	6.74	1.006	6.7100
Total	**81.15**		**81.02**		**79.88**		**79.05**	

Table 10. The results of descriptive analysis of the second-stage testing

No.	Dimension	N Resp.	N Item	Students' Mean (A)	Non-Islamic religion education Teachers' Mean (B)	Islamic religion education Teachers' Mean (C)	Total Mean	Category	Total Personality Mean
1.	*Al-'Iffah*	200	15	3.746	3.588	3.605	3.646	Good	
2	*Al-Syaja'ah*	200	15	3.778	3.652	3.612	3.681	Good	**3.711**
3.	*Al-Hikmah*	200	15	3.770	3.734	3.706	3.737	Good	**(Good)**
4.	*Al-'Adalah*	200	15	3.815	3.828	3.694	3.779	Good	

Table 11. The results of instrument significance testing

Al-'Iffah			*Al-Syaja'ah*			*Al-Hikmah*			*Al-'Adalah*		
KMO		0.648	**KMO**		0.699	**KMO**		0.643	**KMO**		0.627
Barlets Test of Spher-icity	X^2	645.928	**Barlets Test of Spher-icity**	X^2	806.186	**Barlets Test of Spheric-ity**	X^2	1573.243	**Barlets Test of Spheric-ity**	X^2	1143.904
	Df	105		Df	105		Df	105		Df	105
	sig	0.00		sig	0.00		sig	0.00		Sig	0.00

The objective of second stage testing is to test the instrument that has been resulted from the first stage revision toward the respondents that might represent several characteristics that has already been defined. The results of data gathering in the second stage were analyzed by means of confirmatory technique to test the appropriateness of the theoretical construct of the assessment model for the personality competence of Islamic religion education teachers toward the empirical data.

The construct testing for the instrument within the study refers to the second order confirmatory technique performed for testing the construct model in the multiple dimension variables (Hadi, 2008, p. 25). Before performing the confirmatory analysis, the researchers should test whether the instrument was feasible or not for testing by means of exploratory technique and the confirmatory analysis. The results of instrument significance testing are presented in Table 11.

The results of instrument significance testing by means of exploratory analysis show that *al-'Iffah* instrument has the Kaiser Meyer Olkin Measure of Sampling Adequacy (KMO) score that has been equal to 0.648, the Barletts Test of Sphericity and the Chi-Square score that has been equal to 645.928, the df that has been equal to 105 and the significance that has been equal to 0.00. Then, *al-Syaja'ah* instrument has the Kaiser Meyer Olkin Measure of Sampling Adequacy (KMO) score that is equal to 0.699, the Barletts Test of Sphericity and the Chi-Square score that is equal to 860.186, the df that is equal to 105 and the significance that is equal to 0.00. Next, *al-Hikmah* instrument has the Kaiser Meyer Olkin Measure of Sampling Adequacy (KMO) score that is equal to 0.643, the Barletts Test of Sphericity and the Chi-Square score that are equal to 1573.243, the df that is equal to 105 and the significance that is equal to 0.00. Last but not least, *al-'Adalah* instrument has the Kaiser Meyer Olkin Measure of Sampling Adequacy (KMO) score that is equal to 0.627, the Barletts Test of Sphericity and the Chi-Square score that are equal to 1143.904, the df that is equal to 105 and the significance that is equal to 0.00.

The results of the instrument significance testing show that the model construct test requirements are met. Thus, the analysis of the construct testing by means of Confirmatory Factor Analysis could be proceeded.

The results of confirmatory factor analysis for the *al-Iffah* dimension show that the Chi-Square (x^2) value = 97.10, the *p*-value = 0.13813 ($p > 0.05$), the Root Mean Square Error of Approximation (RMSEA) = 0.029 (RMSEA ≤ 0.08), the Goodness of Fit Index (GFI) = 0.94 (GFI > 0.90) and the Adjusted Goodness of Fit Index (AGFI) = 0.91 (AGFI > 0.90). The results are presented in Figure 1.

Figure 1 presents the correlation of error between X1 and X7, between X2 and X4 and between X1 and X5. The correlation shows that the respondents' perception of the statements in the observation sheet is equal.

The results of the CFA for *al-Syaja'ah* dimension show that the Chi-Square (x^2) value = 98.05 with df = 83, *p*-value = 0.12396 ($p > 0.05$), RMSEA = 0.030 (RMSEA < 0.08), GFI = 0.93 (GFI > 0.90) and AGFI = 0.90 (AGFI > 0.90). The results of the construct model analysis for *al-Syaja'ah* dimension meet the require-ments for Goodness of Fit Index and the re-sults are presented in Figure 2.

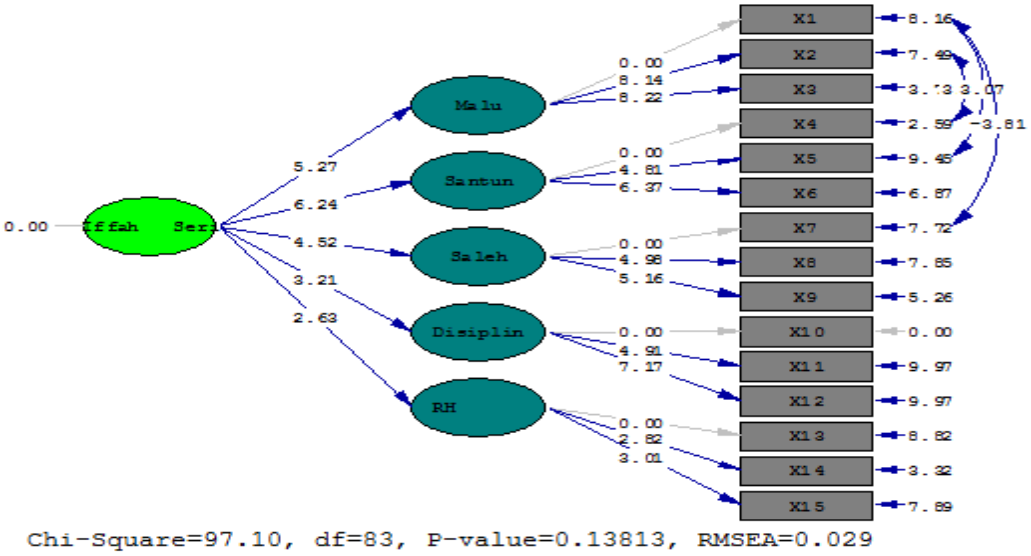

Figure 1. The T-Value scores for *al-Iffah* dimension

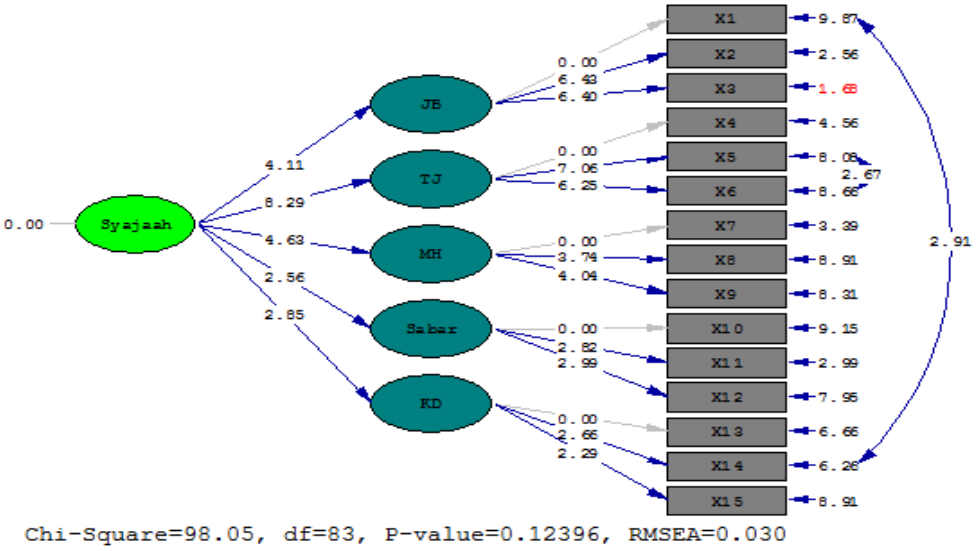

Figure 2. The T-Value scores for *al-Iffah* dimension

Figure 2 presents the correlation of errors between X1 and X14 and between X5 and X6. The correlation shows that the perception of the respondents toward the statements in the observation sheet is equal. The results of the correlation in the model make the model become fit because these results meet the requirements that had been defined.

The results of the confirmatory factor analysis for *al-Hikmah* dimension show that the Chi-Square (x^2) value = 101.78 with the df = 85, the *p*-value = 0.10361 ($p > 0.05$), the RMSEA = 0.031 (RMSEA < 0.08), the GFI = 0.93 (GFI > 0.90) and the AGFI = 0.90 (AGFI > 0.90). These results are presented in Figure 3.

Figure 3 presents the correlation of error between X1 and X3 and between X4 and X14; in addition, Figure 3 also adds the item X6 to the indicators of Optimist. The correlation of error shows that the perception of the respondents toward the statements in the observation sheet has been equal.

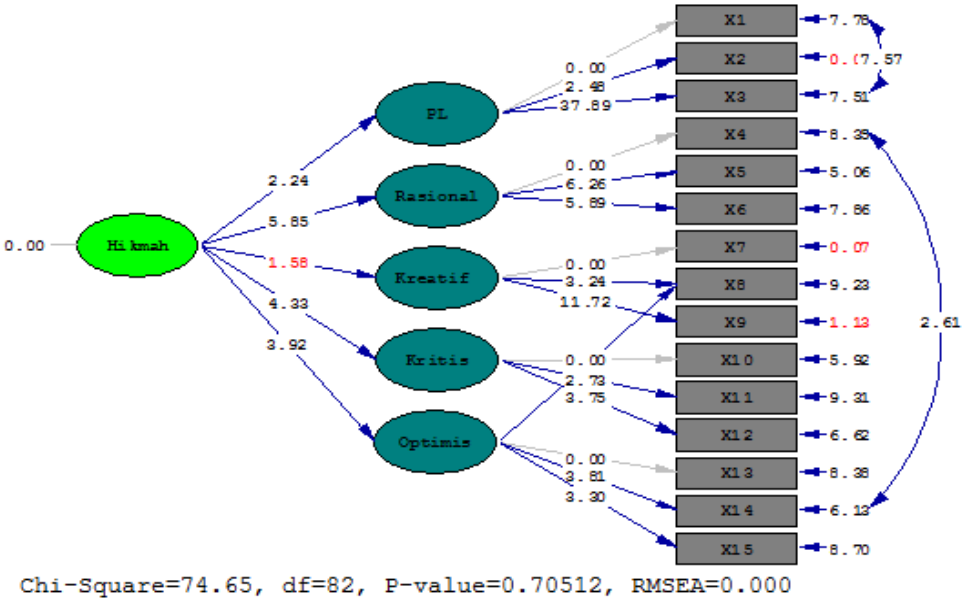

Figure 3. The T-Value scores for *al-Hikmah* dimension

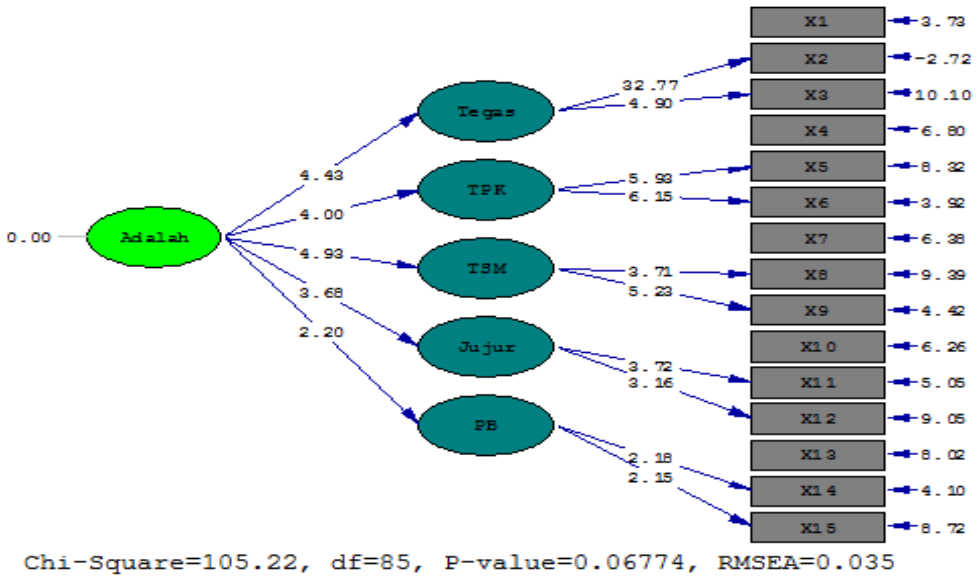

Figure 4. The T-Value scores for *al-'Adalah* dimension

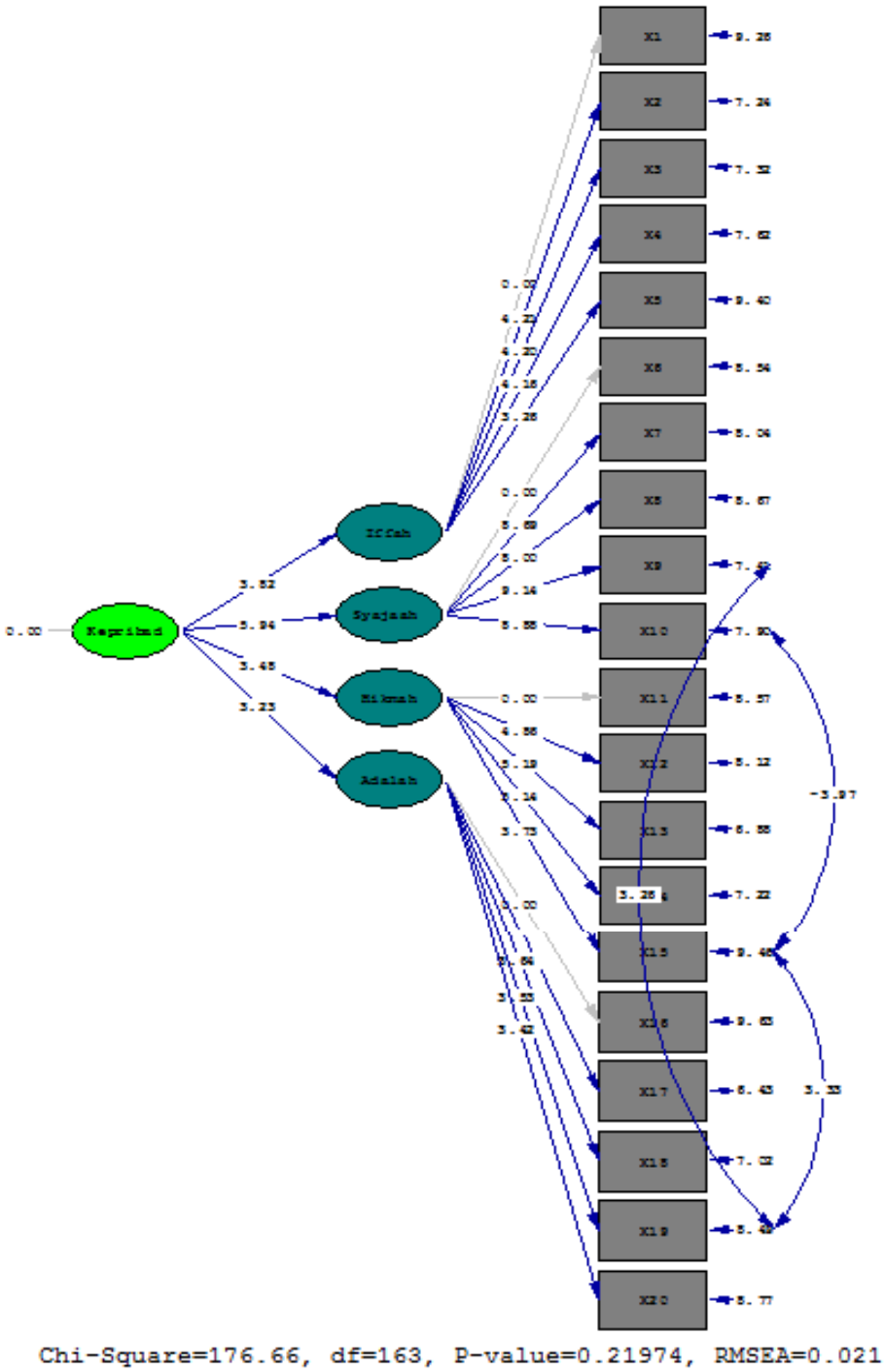

Figure 5. The overall model of teacher personality

Table 12. The summary on results of analysis toward the overall model fitness test

No.	Goodness of Fit Index	Results of Analysis	CF	Criteria
1.	Chi-Square	176.66	Small	-
2.	Probability (p-value)	0.21974	> 0.05	Good
3.	GFI	0.92	≥ 0.90	Good
4.	AGFI	0.89	≥ 0.90	Marginal
5.	RMSEA	0.021	≤ 0.08	Good
6.	RMR	0.049	< 1	Good

The results of confirmatory factor analysis for *al-'Adalah* dimension show that the Chi-Square (x^2) value = 86.14 with df = 83, the *p*-value = 0.38502 (*p*> 0.05), the RMSEA = 0.014 (RMSEA < 0.08), the GFI = 0.96 (GFI > 0.90) and the AGFI = 0.93 (AGFI > 0.90). These results are presented in Figure 4.

The overall personality model resulted from the second stage testing or the expanded scale testing show the model compatibility and the empirical data in the field. The overall model fitness test was calculated from the Goodness of Fit Index (GFI) score resulted from the second stage testing and that of the conceptual model. From the results of the confirmatory analysis toward the overall model upon the teacher personality, the researchers found that Chi-Square (x^2) = 176.66 with df = 163, *p*-value = 0.21974 (*p*-value > 0.05), RMSEA = 0.021 (RMSEA < 0.080), GFI = 0.92 (GFI > 0.90) and AGFI = 0.89 (AGFI > 0.90).

Then, from the results of confirmatory analysis toward the overall model upon the teachers' personality, the researchers gained suggestions to perform the correlation of error between X9 and X14, between X10 and X15 and between X15 and X19. The correlation of error displayed that the perception among the respondents had been equal. The results of analysis toward the overall constructs of the model upon the teacher personality are presented in Figure 5 and summarized in Table 12.

The results of model fitness test presented in Table 12 show that the data from the testing and those from the conceptual model have been fitted. The fitness is explained by the fact that the Root Mean Square Error of Approximation (RMSEA) = 0.021 (RMSEA < 0.080) and the Root Mean Square Residual (RMR) = 0.0499 (RMR < 0.100). The criteria for the *p*-value of > 0.05 are met (*p*-value = 0.21974), which prove that there is no difference between the empirical data and the theoretical model. Therefore, the results of the testing analysis of the model can be accepted and the model can be implemented for assessing the personality competence of Islamic religion education teachers.

Conclusion and Suggestions

Based on the findings and the results of analyses, it is concluded that the construct model of assessment instrument for the personality competence of Islamic Religion Analysis (MPKKG-PAI) that has been developed includes *al-'Iffah* (modesty) dimension, *al-Syaja'ah* (courage) dimension, *al-Hikmah* dimension (wisdom) and *al-'Adalah* (justice) dimension. Each dimension consists of five indicators with three observed variables. The assessment model for the personality competence of Islamic religion education teachers has been tested empirically by means of Confirmatory Factory with the LISREL 8.51 program. The results of the analysis show the Chi-Square (x^2) of 176.66, *p*-value of 0.21974, RMSEA of 0.021 and GFI of 0.92.

The characteristics of the assessment instrument on the personality competence of Islamic religion education teachers are in the good category and the evidence of possessing the good category is that all of the loading factor scores in all dimensions and all indicators are bigger than 0.3 and that the instrument reliability of all personality dimensions have met the minimum requirement, namely bigger than 0.7.

Based on the results of the readability testing, expert validation, first stage testing, second stage field testing, the implementation guidelines on the assessment of the personality competence of Islamic religion education teachers is effective and it is in a good category. As a result, the model might be implemented for measuring the personality competence of Islam religion education teachers.

The assessment model of the personality competence of Islamic religion education teachers might be developed further and is not limited to assessing the personality competence of Islamic religion education teachers of senior/vocational high school level. The same assessment model might also be implemented for assessing the personality competence of Islamic religion education teachers of Madrasa Aliyah and Islamic boarding schools. In addition, the same assessment model might be implemented for assessing the personality competence of non-Islam religion education teachers.

References

Aiken, L. R. (1999). *Personality assessment: Methods and practices* (3rd ed.). Kirkland, WA: Hogrefe & Huber. Retrieved from https://www.scribd.com/document/350348192/Aiken-Personality-Assessment-Methods-Practices-3rd-Edition

Al-Ghazali. (1957). *Ihya ulum al-diin*. Al-Qahirah (Cairo): Haramain.

Al-Nahlawi, A. R. (1995). *PendidikanIslam di rumah, sekolah, dan masyarakat*. Jakarta: Gema Insani Press.

Amin, A. (1975). *Etika (Ilmu akhlak)*. Jakarta: PT. Bulan Bintang.

Anis, I. (1972). *Al-mu'jam al-wasith*. Cairo, Mesir: Dar al-Ma'arif.

Asyari, H. (1238). *Adabu al-'alim wa al-muta'allim*. Jombang: Tebu Ireng.

Borg, W. R., & Gall, M. D. (1983). *Educational research: An introduction* (4th ed.). New York, NY: Longman.

Cattel, R. B. (1965). *The scientific analysis of personality*. New York, NY: Penguin.

Daradjat, Z. (1982). *Psikologi kepribadian guru*. Jakarta: PT. Rineka Cipta.

Hadi, S. (2008). *Analisis butir untuk instrumen angket, tes, dan skala nilai dengan BASICA*. Yogyakarta: Andi Offset.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (2006). *Multivariate data analysis* (6th ed.). Jersey City, NJ: Pearson Education.

Law No. 14 Year 2005 of Republic of Indonesia about Teachers and Lecturers (2005).

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

Miskawaih, I. (1398). *Tahdzib al-akhlaq*. Beirut: Masyurat Dar al-Maktabat al-Hayat.

Miskawaih, I. (2010). *Tahdzib al-akhlaq*. Islamic Book. Retrieved from http://shamela.ws/books/127/12720.rar

Mouly, G. J. (1973). *Psychology for effective teaching* (3rd ed.). New York, NY: Holt, Rinehart and Winston.

Muhaimin. (2005). *Kawasan dan wawasan studi Islam*. Jakarta: Prenada Media.

Muhaimin, & Mujib, A. (1993). *Pemikiran pendidikan Islam, kajian filosofis dan kerangka dasar operasionalnya*. Bandung: Trigenda Karya.

Nunally, J. C. (1981). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Prihatini, S., Mardapi, D., & Sutrisno, S. (2013). Pengembangan model penilaian akhlak peserta didik madrasah aliyah. *Jurnal Penelitian Dan Evaluasi Pendidikan*, *17*(2), 347–368. https://doi.org/10.21831/pep.v17i2.1705

Rajab, M. A. (1961). *Taammulat fi falsafat al-akhlaq* (3rd ed.). Cairo, Mesir: Maktabat al-Anhalu al-Mishriyyah.

Regulation of Minister of Religion Affairs of Republic of Indonesia, No. 16 Year 2010, on the Management of Religion Education at Schools (2010).

Regulation of the Government of Republic of Indonesia, No. 32 Year 2013, on National Education Standard (2013).

Solimun. (2002). *Structural equation modeling, lisrel, dan amos*. Malang: Universitas Negeri Malang.

Solimun. (2008). *Structural equation modeling (SEM), lisrel, dan amos*. Malang: Fakultas MIPA Universitas Brawijaya.

Spencer, L. M., & Spencer, S. M. (1993). *Competence at work: Models for superior performance*. New York, NY: John Wiley & Sons.

Sudijono, A. (2003). *Pengantar statistik pendidikan*. Jakarta: Raja Grafindo.

Wahab, T. F. (2010). *Pendalaman materi kompetensi profesional*. Bandung: Fakultas Tarbiyah dan Keguruan, Universitas Islam Negeri Sunan Gunung Jati.

REiD (Research and Evaluation in Education), 3(1), 2017, 92-105
Available online at: http://journal.uny.ac.id/index.php/reid
Research Article

Developing assesment instruments for the sensory acceptability of food products

***[1]Sugiyono; [2]Djemari Mardapi; [3]I Gusti Putu Suryadarma**
*Faculty of Engineering, Universitas Negeri Yogyakarta
Jl. Colombo No. 1, Depok, Sleman 55281, Yogyakarta, Indonesia
*Email: giyono@uny.ac.id

Submitted: 05 May 2017 | Revised: 24 May 2017 | Accepted: 24 May 2017

Abstract

The study investigated the constructs and characteristics of physiological, psychological, and sensory instruments. The data were collected by means of a questionnaire, physiological instrument, psychological instrument, and sensory instrument. The expert judgment conclusions were calculated by means of Aiken formula; the instrument construct validity was analyzed with confirmatory factor analysis by the goodness-of-fit test at the significance level of 0.05; the reliability estimate by generalizability with a G-study coefficient ≥ 0.7 and an ICC coefficient ≥ 0.7; and the instrument characteristics were analyzed by D-study. The results of the study are as follows: (1) the physiological instrument consisted of nine constructs: four constructs of psychological instrument and five constructs of sensory instrument; (2) three instruments had good face validity, content validity and construct validity, supported by the empirical evidence at $p > 0.05$; (3) the reliability estimate of the three instruments was good and the reliability estimate was supported by empirical evidence with G coefficients of > 0.7 and ICC coefficients of > 0.7, (4) the three instruments had characteristics that might be appropriate to be used by the university students of culinary program and the vocational high school students of culinary program.

Keywords: *physiological instrument, psychological instrument, sensory instrument*

How to cite item:

Sugiyono, S., Mardapi, D., & Suryadarma, I. (2017). Developing assesment instruments for the sensory acceptability of food products. *REiD (Research and Evaluation in Education), 3*(1), 92-105. doi:http://dx.doi.org/10.21831/reid.v3i1.13940

Introduction

The instruments used in assessing food product from the results of food processing have not been standardized. The assessment is still based on the experience of culinary lecturers and teachers as the instructors. The use of standardized instruments will eliminate rater's subjectivity in assessing the results of the practice. The instruments, as a measurement tool, should have the evidence of validity and reliability (Mardapi, 2008, p. 2).

Food products are used for fulfilling human's basic needs both quantitatively and qualitatively. Food product quantity measurement is conducted by implementing physical measurement tool; on the other hand, the measurement of food product quality is conducted by implementing sensory acceptability assessment instrument related to human sense responses. The acceptability involves feelings so that the instrument belongs to affective domain.

Food processing is a process of turning the raw material into the product that might be consumed. Food processing for generating food product is conducted by means of certain manners or methods. Food product is used for meeting human's basic needs both quantitatively and qualitatively. The quantity measurement of food product employs the physical measurement tools, while the quality measurement of food product makes use of a

sensory assessment instrument, which is related to human senses.

The measurement of human sense response toward food product involves the principles of sensory sensing, physiological process, and psychological processes (Stone & Sidel, 2004, p. 16). The physiological process should be controlled in order to decrease the physiological error and the psychological process should be controlled in order to decrease the psychological error that might occur. The physiological and psychological error might be controlled by means of physiological and psychological instruments; as a result, the two instruments should be developed.

The physiological aspects consist of attributes of eyes, nose, ears, tongue and skin on the fingers and mouth. The physiological aspects are latent; therefore, the indicators of the attribute measurement should be identified. The attributes of eyes have the measurement indicators of vision and visible light. The attributes of nose have the measurement indicators of odor intensity in the form of olfactory. Then, the attributes of ears have the measurement indicators of ear drum, which is influenced by noise. The sensitivity of tongue is influenced by the temperature of the tested product and the hunger. Last but not least, the skin on the fingers and mouth have the indicators of finger feeling for solid materials and the indicators of mouth feeling for solid or liquid materials.

On the other hand, the psychological aspects that will be involved in sensory measurement are attitude, motivation, adaptation, and situational aspects. The psychological aspects are latent, therefore, the measurement indicators of these attributes should be identified. Attitude is measured through affection or feeling indicators, cognition or knowledge indicators, and conation or tendency indicators. Then, motivation has the measurement indicators of boredom, fatigue and environmental condition. Next, adaptation will influence the sensory benchmark of the measured product and there will be certain changes of response due to the adaptation. Last but not least, situational aspect that might influence the response error of the panelists will be measured by means of errors due to the pre-testing expectation, errors due to the position of sample presentation, errors due to the stimulus, and errors due to the sample difference, which might be very contrast in the testing sequence.

Psychometric measurement, in addition, should have the stimulus that will be delivered to an individual and then will be processed psychologically in order to generate response. Food product has the attributes that will release the stimulus and these attributes will be sent to an individual. Next, the stimulus will be retrieved by the receptors in human senses and will generate a sensation. The sensation will be sent to the brain and the brain will process the sensation in order to generate perception (Walgito, 1989, p. 54). Based on the perception, there will be responses toward food products under measurement in the form of sensory acceptability. The response measurement toward food products involves feeling that belongs to affective domain.

The measurement scale of affective study that has been widely implemented is Thurstone Scale, Likert Scale and Semantic Differential Scale (Mardapi, 2008, p. 117). Stone and Sidel (2004, p. 8) argue that the interval data scale has been attained from the line-scale measurement scale. Thurstone measurement scale in the form of category data scale is implemented in order to measure interest. Then, Likert scale in the form of ordinal data scale is implemented in order to measure attitude. Semantic Differential Scale is a scale that is limited by the highest and the lowest extreme values. The line scale links (anchors) the lowest and the highest continua. In the line scale, the rater puts a sign on the horizontal line as a reflection of certain sensory characteristic intensity. The measurement scale should be adjusted to the data scale that a researcher demands.

Stone and Sidel (2004, p. 84) state that in order to measure the preference for a product, a researcher might implement hedonic measurement scale technique, for example the nine-point hedonic scale. Watts *et al.* (1989, p. 49) argue that the hedonic scale belongs to the interval data scale. The hedonic scale is very similar to semantic difference scale. The data from the results of sensory testing obser-

vation by means of hedonic scale might be analyzed by implementing parametric statistical test. The hedonic scale should not use the nine-point scale; instead, the hedonic scale might be adjusted to necessity in the sensory testing. A study by Shumate *et al.* (2007, p. 357) has concluded that the number of point scale basically will influence the estimation of generalization. The estimation of generalization increase is in accordance with scale point increase up to a certain level. Up to date, the sensory test instrument refers to Likert Scale for most of the time. However, for conclusion drawing, many researchers refer to the parametric statistical analysis.

The sensory perception might be generated from the process of sensory testing if there are stimuli and receptors. The stimuli come from the attributes of food product and the receptors are possessed by human senses. Perception will occur if the stimulus that has been released by food products is retrieved by the receptors of human senses. The receptors of human senses are unique for they are able to accept certain appropriate stimuli. In addition to having been determined by the stimuli from the attributes of food products, the perception that has been generated is also influenced by the physiological and psychological aspects of the raters.

The objective of the study is to develop a valid, reliable and applicable sensory acceptability assessment instrument of food products, which is expected to be applied as the means for assessing the sensory acceptability of food products resulted from the food processing practice in the culinary study program under the culinary department in vocational high schools and universities.

Method

The study applied research & development (R&D) approach. The stages in the study had been started with field preliminary survey to attain the existing problem of discrepancy as the basis for the stages of development. The developmental stage is conducted by using Borg and Gall (1989) step modification, namely: (1) review of literature and product design planning, (2) preliminary product development, (3) preliminary field testing, (4) product revision, (5) main field testing, and (6) final product revision.

The research was conducted from December 2014 to March 2015. The study was conducted in two locations, namely: Faculty of Engineering of Universitas Negeri Yogyakarta, and State Vocational High School 4 Yogyakarta.

The population of the study was divided into two: the population from the university and vocational high school. The population from the university was the lecturers and students practicing food processing. On the other hand, the population from the vocational high school consisted of the teachers and students practicing food processing.

The sample of the study was established randomly. Eight lecturers and 36 students of food processing practice course at the Department of Culinary, Faculty of Engineering, Universitas Negeri Yogyakarta and three teachers and 60 students of food processing practice subject at State Vocational High School 4 Yogyakarta were chosen as the sample of the study.

Procedures

Acceptability involves feeling, therefore, it belongs to affective domain. For the development of the affective instrument, the researchers adopted overall steps developed by Mardapi (2008, p. 108): (1) deciding the instrument specification; (2) composing the instrument; (3) deciding the instrument scale; (4) deciding the scoring system; (5) reviewing the instrumet; (6) performing experiment; (7) analyzing the instrument; (8) designing the instrument; (9) performing measurement; and (10) interpreting the measurement results. The ten steps were compulsory to perform in developing the affective instrument both for the psychological and physiological attribute measurement and for the learning process. The first step in determining the instrument specification was the substance of the instrument that would be developed. The first step was very important because it was the core of the affective instrument development. The following steps were the elaboration or the follow-up of the previous step.

For the instrument design, the researchers referred to the pattern of instrument specification design developed by Mardapi (2008, p. 109) namely: deciding the measurement objective, deciding the instrument guidelines, selecting the instrument design and format, and deciding the length of the instrument. The instrument generated from the developmental study was to measure the sensory acceptability of food products. The sensory acceptability was measured based on physiological, psychological, and food products aspects. The three attributes were latent and the measurement of the three attributes were based on certain indicators. Then, the researchers designed the operational definition of the three attributes. Based on the operational definition, the instrument guidelines were developed in order to decide the indicator measurement items. The instrument of sensory acceptability was based on the instrument guidelines.

The developmental procedures were performed by modifying the ten steps of R & D approach by Borg and Gall (1989, pp. 784–785) into the following six steps: (1) literature review and product design planning; (2) preliminary product development; (3) preliminary field testing; (4) product revision; (5) main field testing; and (6) final product revision.

The modification performed toward the R & D approach by Borg & Gall is explained as follows. The first and the second step were combined into literature review and product design planning (the first step). Then, the third step was modified into preliminary product development (the second step). Next, the fourth step was modified into preliminary field testing (the third step). Afterward the fifth step was modified into product revision (the fourth step). Further, the sixth step was modified into main field testing (the fifth step). The seventh and eighth steps were omitted by considering the objective of the study. Next, the ninth step was modified into final product (the sixth step). Last but not least, the tenth step was omitted by considering the objective of the study.

The first step, namely literature review and product design planning, included the activities of gathering information and creating the product design plan. Then, the second step, namely preliminary product development, included the activities of creating a prototype based on the literature review and expert feedbacks. The result of the second step was Prototype-I. Next, the third step, namely preliminary field testing, included the activities of getting evaluation and feedbacks from the experts and practitioners. Afterwards, the fourth and the fifth steps, namely product revision and main field testing, included the activities of evaluating the product and getting the evaluation from the panels of experts. Last but not least, the sixth step, namely field test, included the activities of performing the second step based on the results of the main field testing. The results of the sixth step was a standardized instrument as the final product.

Data, Instrument, Data Gathering and Data Analysis Technique

The data were quantitative with the ordinal data scale. The data were gathered by means of three instruments implemented in the study, namely physiological instrument, psychological instrument, and sensory instrument. In the data gathering activities using the physiological instrument, the respondents as the candidates of sensory acceptability raters should assess themselves regarding their physiological preparedness. Then, the psychological instrument was applied in order to assess their psychological condition by themselves as the rater candidates. If the value that had been attained from the two instruments fulfilled the requirements of the cut value, then the related rater candidates might continue assessing the sensory acceptability of food products by means of the sensory instrument.

Both the test and the non-test instruments should have the evidence of validity and reliability (Mardapi, 2008, p. 15). The face validity was determined through the expert opinion by performing expert judgement, while the content validity was determined by experts through rating activities. The value was decided by calculating the data through the use of Aiken formula:

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

$$S = \sum s \quad (2)$$

$$s = r - lo \quad (3)$$

Legend:
V : coefficient of content validity
S : number of scores from all raters
n : number of raters
s : score from each rater
r : rater's validity rating of items
c : number of rating category
lo : lowest validity category
hi : highest validity category

Note:
If $lo < hi$ then the formula (3) applies: $s = r\text{-}lo$
If $hi<lo$ then the formula (3) becomes: $s = lo\text{-}r$

The content validity of the instrument was calculated by means of Aiken formula and the calculation generated the coefficient of V-count validity. The index of instrument validity was determined by comparing the coefficient of V-count validity and the coefficient of V-table validity. The instrument was regarded meeting the criteria of content validity if the V-count was equal to or bigger than the V-table.

The V-table value was determined by operating the V-Aiken Table on the intersection between the appropriate c column and n line. The 1% level of significance was on the upper line of the same n value and the 5% level of significance was on the lower line of the same n value.

The construct validity was determined by factor analysis with maximum likelihood and referred to the Eigen values ≥ 1. The data attained from the limited scale experiment were used for attaining the estimation of reliability by means of generalizability theory and Genova analysis with the criterion of coefficient G ≥ 0.7. Then, the data attained from the expanded scale experiment were used for attaining the estimation of reliability by means of intraclass correlation coefficient (ICC) with the criterion of coefficient ICC ≥ 0.7. According to (Murti, 2011, p. 13) and Bartko (1976, p. 763), the formula of ICC correlation coefficient is as follows:

$$r_{ICC} = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_o^2 + \sigma_e^2}$$

Note:
r_{ICC} : ICC correlation coefficient
σ_s^2 : subjects' variants (between people)
σ_o^2 : observer's variant (between people within people)
σ_e^2 : error variant (residual)

The ICC correlation coefficient might be converted into the Cronbach Alpha correlation coefficient as follows:

$$\alpha = \frac{k \times r_{ICC}}{1 + (k-1) \times r_{ICC}}$$

Note:
α : Alpha Cronbach correlation coefficient
k : number of raters
r_{ICC} : ICC correlation coefficient

The sensory acceptability assessment instrument was used for assessing food products resulted from the food processing practice. The reference referred by the researchers in the sensory acceptability instrument was criterion reference. The criterion of passing value was determined by means of Standard Setting Method with Extended Angoff (Cizek & Bunch, 2007, p. 88). The cut value calculation was as follows:

$$Cut\ score = Means \times number\ of\ items$$

The self-assessment data using physiological instruments calculated the total score obtained. The result was compared to the cut score of the physiological instruments in the rubric. If the total score ≥ cut score, then it can continue the assessment using sensory instruments.

From the self-assessment data gathered by means of the psychological instrument, the total value attained was calculated. The result of the calculation was compared to the cut value of the psychological instrument in the rubric. If the total value ≥ the cut value, then the assessment by means of sensory instrument might be continued.

From the data of measurement results by sensory instrument, the mean of the sensory instrument was calculated. The calculation result was compared to the scoring guidelines in the rubric to draw conclusion on the sensory acceptability of food products tested.

Findings and Discussion

The results of the preliminary study show that the implementation of sensory assessment on the results of food processing practice had not made use of a standardized instrument. The assessment was still based on the rater's experience. Then, the literature review shows that the sensory assessment involved physiological and psychological processes (Stone & Sidel, 2004, p. 16).

Physiological process is related to physiological organs that support the activities of sense testing. The sensory organs related to the sensory testing are as follows: sight, olfactory, taste, texture and auditory. The senses retrieve sensory information in the form of stimuli and then the stimuli is converted into electric-loaded nerve impulses that are read by the brain; thus, as a result, sensation occurs (Brennan, 2006, p. 273; Rakhmat, 2007, p. 49). The sensation is retrieved by the receptors in the sensory organs and then sent to the brain. After that, the sensation is returned to the sensory organs as perception. Mason and Nottingham (2002, p. 8) state that stimuli is the chemical or the physical activators that cause responses toward the receptors. For example, human beings have the eye receptor stimulus and sound is the ear receptor stimulus. The receptors are the cells of sensory organs that detect the stimuli, for example the taste buds receptor in the tongue and the light receptor in the retina. The stimuli that has been exposed to the receptors causes sensation which is interpreted psychologically by comparing the sensation to the past experiences; as a result, the sensation is converted into perception. Therefore, the sensory testing might be performed if there is stimuli that has been released by the attributes of food products and the receptors possessed by the senses.

Azwar (2000, p. 5) elaborates that attitude is the constellation of cognitive (mind), affective (feeling) and connative (behavior) components that have interaction from one to another in understanding, feeling and behaving toward an object. An attitude is determined not only by the situation of the object that has been encountered, but also by the past experience, present situation and also future expectation (Azwar, 2000, p. 3). Attitude itself is an evaluative response. Thus, it is determined by the situational aspect of an individual. Responses would occur only if an individual is encountered by a stimulus that urges a reaction from the individual. Evaluative responses imply that the appearance of the form of reaction that has been stated as an attitude is based on the internal evaluation process of an individual in drawing conclusions from the stimulus in the form of good-bad, positive-negative and also pleasant-unpleasant relationships, then, the individual will potentially react to the object of attitude. The psychological aspects related to the study are the cognitive (mind), affective (acceptance), psychomotoric (adaptation, motivation) and also psychological condition (situational) aspects.

The process of sensory assessment of food products is started from the stimulus that has been exposed to the sense receptors and the stimulus is delivered to the center of nerve constellation in the brain, and as a result, perception occurs (Walgito, 1989, p. 53). The sensing mechanism of food products acceptability assessment includes three stages namely physical process, physiological process and psychological process. The physical process refers to the process of accepting the stimulus that has been released by the object to the sense receptor. The object is in the form of food product and the object releases a stimulus exposed to the sense receptor, and as a result, sensation occurs. The stimulus might come from the external object that has been directly exposed to the sense or from the internal object that has been directly exposed to the sensory accepting nerve. Then, the physiological process refers to the continuation of physical processes. The stimulus that has been accepted by the sense receptor generates sensation and is delivered by the sensory nerve to the center of nerve constellation in the brain, and as a result, response occurs. In other words, the physiological process is the process of delivering the sensation to the center of nerve constellation that generates response. Next, the psychological process refers to the continuation of physiological process. The response that has

been generated in the physiological process is processed in the center of brain nerve constalleation, and as a result, perception occurs. Perception is a negotiation between the pattern that people have attained from the environment and the pattern that people have attained from the accumulation of experience (Brennan, 2006, p. 273). Rakhmat (2007, p. 51) states that perception is an experience regarding objects, events and relationships attained by interpreting and also concluding the information from the sensory stimuli. The perception that is attained from the environmental information is combined with the past experience.

In the sensory testing, during the tasting of the food sample, there is an inter-human senses interaction that is resulted in perception. The perception that has been resulted from the stimuli is not purely from one receptor; instead, the perception is the results of an interaction between two or more receptors. The occurence of the interaction itself has not been surely identified whether it is from the senses as the receptors or from the brain, however, the biggest possibility is that the process of interaction occurs in the brain. Another possibility is that the inter-stimuli interaction comes from the food sample. The perception that has been resulted is not purely from the food, instead, the perception is the result of an interaction between two or more stimuli. The final result of the process is the sensory acceptability of food products.

Based on the literature study and the feedbacks from the experts and practitioners, the instrument guidelines were designed. For the first instrument, the experts who are involved in the study are food science experts, food processing experts and practitioners consisting of food processing lecturers and food processing teachers. The guidelines of the first instrument are clearly presented in Table 1.

Then, for the second instrument, the experts consist of food science experts and psychological experts, while the practitioners consist of culinary lecturers and culinary teachers. The guidelines of the second instrument are clearly presented in Table 2.

Table 1. The physiological instrument guidelines

No.	Attributes	Indicators
1.	Sight	1. Eye health 2. Color blind 3. Lighting condition
2.	Olfactory	1. Nose health 2. Strong odor allergic 3. Smoke allergic 4. Dust allergic 5. Adaptation
3	Auditory	1. Auditory disorder 2. Auditory sensitivity
4.	Taste	1. Tongue health 2. Taste sensitivity
5.	Texture	1. Teeth health 2. Fingertips skin sensitivity 3. Artificial teeth health

Table 2. The psychological instrument guidelines

No.	Attributes	Indicators
1.	Cognitive	1. Product identification 2. Taste sensitivity 3. Olfactory sensitivity 4. Texture sensitivity
2.	Affective	1. Interestedness in assessing the product 2. Enjoyment of the product 3. willingness to test the product
3.	Psychomotor	1. Adaptation, frequently performing test 2. Adaptation, frequently tasting
4.	Psychological condition	1. Having problems 2. Having physical fatigue (stress)

Table 3. The sensory instrument guidelines

No.	Attributes	Indicators
1.	Appearance	1. Color 2. Presentation 3. Size 4. Form
2.	Smell	1. Easy detection 2. Scent 3. Scent strength 4. Odor, unpleasant scent
3.	Touch	1. Kinestetic, solid product 2. Consistency, paste product 3. Viscosity, liquid product
4.	Hearing	1. Sweet 2. Salty
5.	Taste	1. Sour 2. Bitter 3. Savory

In the third instrument, the experts consist of food science experts and sensory testing experts while the practitioners consist of culinary lecturers and culinary teachers. The guidelines of the third instrument are presented in Table 3.

The design of sensory acceptability assessment instrument that has been generated from the preliminary study consists of three instruments namely physiological instrument, psychological instrument and sensory instrument. In the preliminary product development stage, the researchers generated the prototype of the instrument, namely the first prototype of the physiological instrument, the first prototype of the psychological instrument, and the first prototype of the sensory instrument. The first prototype instruments were revised in the product revision stage based on the feedbacks and suggestions from the preliminary field test stage. The revision resulted in the second prototype of the physiological instrument, the second prototype of the psychological instrument, and the second prototype of the sensory instrument.

The second prototype of the physiological instrument, the second prototype of the psychological instrument, and the second prototype of the sensory instrument were tested in the main field test stage. These instruments were reviewed based on the feedbacks and suggestions from the state of the main field testing. The result of the revision was the final products. The reliability index of the three instruments was estimated by means of G-Study and of intraclass correlation coefficient (ICC). The validity and reliability of the three instruments were determined before the three instruments were applied as the measurement tools of food product sensory acceptability.

The face validity was determined by means of expert judgement. The experts argued that the instrument had met the face validity.

The content validity was determined by the experts. The values that the experts had provided toward the instruments were calculated by means of Aiken formula. The criterion of content validity was that the V-count coefficient of validity should be equal or bigger than the V-table coefficient of validity. The V-table coefficient of validity was read on the number of rating category (c = 4) and the number of the rater (n = 8) and the level of significance of 5% in V-table = 0.75. The results of the calculation showed that the instrument had met the content validity requirement; in other words, the developed instrument was in accordance with the theory. The results of validity calculation showed that 14 out 15 items in the physiological instrument were valid (one item was dropped); 11 items in the psychological instrument were valid; and 17 out of 18 items in the sensory instrument were (one item was dropped).

The physiological instrument was analyzed by means of factor analysis in order to determine the number of factors that should be formed in the instrument. The results of the analysis are presented in Table 4.

Table 4. KMO and Bartlett's test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.653
Bartlett's Test of Sphericity	Approx. Chi-Square	222.321
	df	91
	Sig.	.000

Table 4 shows that the KMO-MSA value is equal to 0.653 and the Chi-Square value is equal to 222.321 with the degree of freedom equal to 91 at the level of significance of 0.000. The KMO value of 0.653 > 0.500 implies that the developed instrument is in the 'Good' category. The Chi-Square value is equal to 222.321 and the level of significance is Sig. 0.000; in other words, the instrument can be applied for the factor analysis. The results of the factor analysis is that 14 items in the physiological instrument are categorized into five factors (constructs): sight, olfactory, auditory, taste and texture. The factor analysis for determining the construct validity by means of goodness of fit with the maximum likeli-hood method is presented in Table 5.

Table 5. The goodness-of-fit test

Chi-Square	df	Sig.
19.314	31	.949

The Chi-Square value from the results of calculation is equal to 19.314 on the degree of freedom of 31 at the significance level of 0.949. From the results of analysis, it is found that the p value of 0.949 is bigger than α 0.05; in other words, the constructs developed from the theory are not different from those developed from the analysis on the empirical data. Based on the explanation, it is concluded that the categorization of the items into the factors or the constructs is valid based on the construct validity. Therefore, based on the construct validity with x^2 = 19.314 and p = 0.949 in the Goodness-of-fit test analysis, the physiological instrument is valid.

The psychological instrument was analyzed by means of the factor analysis in order to determine how many factors should be formed in the instrument. The results of the analysis are presented in Table 6.

Table 6. KMO and Bartlett's test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.569
Bartlett's Test of Sphericity	Approx. Chi-Square	149.290
	df	55
	Sig.	.000

Table 6 shows that the KMO-MSA value is equal to 0.569 and the Chi-Square value is equal to 149.290 with the degree of freedom equal to 55 at the significance level of 0.000. The KMO value of 0.569 > 0.500 implies that the developed instrument is in the 'Good' category. The Chi-Square value of 149.290 at the significance level 0.000 shows that the instrument can be applied for the factor analysis.

The results of the factor analysis show that the 11 items in the psychological instrument are categorized into four factors (constructs): cognitive, affective, psychomotor, and psychological condition factors. The factor analysis to determine the construct validity by means of goodness-of-fit with the maximum likelihood method is shown in Table 7.

Table 7. The goodness-of-fit test

Chi-Square	df	Sig.
18.185	17	.377

The Chi-Square value from the results of the calculation is equal to 18.185 on the degree of freedom of 17 at the significance level of 0.377. The results of the analysis show that p 0.377 is bigger than α 0.050; in other words, the constructs that had been developed from the theory are not different from those developed from the empirical data analysis. Based on the explanation, it can be concluded that the categorization of the items into the factors or the constructs is valid based on the construct validity. Therefore, the psychological instrument is valid based on the construct validity with x^2 = 18.185 and p = 0.377 in the Goodness-of-fit Test analysis.

The sensory instrument was analyzed by means of the factor analysis to determine the number of the factors that should be formed in the instrument. The results of the analysis are presented in Table 8.

Table 8. KMO and Bartlett's test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.652
Bartlett's Test of Sphericity	Approx. Chi-Square	375.062
	df	136
	Sig.	.000

Table 8 shows that the KMO-MSA value is equal to 0.652 and the Chi-Square value is equal to 375.062 with the degree of freedom equal to 136 at the significance level of 0.000. The KMO value of 0.652 > 0.500 implies that the developed instrument is in the 'Good' category. The Chi-Square value of 375.062 at the significance level of 0.000 showed that the instrument can be applied for the factor analysis.

The results of the factor analysis, namely the 17 items in the sensory instrument, were grouped into five factors (constructs) as follows: sight, olfactory, texture, auditory and taste. The factor analysis for determining the construct validity by means of goodness-of-fit with the maximum likelihood method is presented in Table 9.

Table 9. The goodness-of-fit test

Chi-Square	df	Sig.
56.415	61	.643

The Chi-Square value from the results of calculation is equal to 56.415 with the degree of freedom of 61 at the significance level of 0.643. The results of analysis show that the p value of 0.643 is bigger than the α value of 0.050; in other words, the constructs that had been formed from the theory are not different from those formed from the empirical data analysis. Based on the explanation, it can be concluded that the categorization of the items into the factors or the constructs is valid based on the construct validity. Therefore, the sensory instrument is valid based on the construct validity with $x^2 = 56.415$ and p = 0.643 in the Goodness-of-fit Test.

For the estimation of instrument reliability in the limited scale experiment, Genova analysis was implemented. The three instruments were tested toward 36 university students majoring in culinary and 60 vocational high school students majoring in culinary. The summary of the results of the reliability estimation by means of Genova analysis is presented in Table 10.

Table 10. Results of G coefficient analysis

Instrument	Facet	Number of Item	G Coeff.	Criteria ≥ 0.70
1. Physiological	Univ. Students	14	0.9288	> 0.7
	VHS Students	14	0.9058	> 0.7
	Mean	14	0.9173	> 0.7
2. Psychological	Univ. Students	11	0.8705	> 0.7
	VHS Students	11	0.8837	> 0.7
	Mean	11	0.8771	> 0.7
3. Sensory	Univ. Students	17	0.9467	> 0.7
	VHS Students	17	0.9602	> 0.7
	Mean	17	0.9535	> 0.7

Table 10 shows that the results of physiological instrument experiment for university students majoring culinary results in the G coefficient value of 0.9288 > 0.7, which implies that the physiological instrument has met the reliability requirement. The table also presents that the results of physiological instrument experiment for vocational high school students majoring culinary results in the G coefficient value of 0.9058 > 0.7, which implies that the physiological instrument has also met the reliability requirement. The mean of coefficient G of 0.9058 > 0.7 shows that the physiological instrument will meet the reliability requirement if the physiological instrument is applied altogether to the university and vocational high school students majoring in culinary.

Table 10 shows that the results of psychological instrument experiment for the university students majoring culinary resulted in the G coefficient value of 0.8705 > 0.7, which implies that the psychological instrument has met the reliability requirement. On the other hand, the table shows that the results of psychological instrument experiment for vocational high school students majoring culinary resulted in the G coefficient value of 0.8837 > 0.7, which implies that the psychological instrument has met the reliability requirement as well. The mean of G coefficient value of 0.8771 > 0.7 implies that the psychological instrument will meet the reliability requirement if the psychological instrument is applied altogether in the university and vocational high school students majoring culinary.

Table 10 shows that the results of sensory instrument experiment for the university students majoring culinary resulted in the G coefficient value of 0.9467 > 0.7, which implies that the sensory instrument has met the reliability requirement. On the other hand, the table shows that the results of sensory experiment for vocational high school students majoring in culinary resulted in the G coefficient value of 0.9602 > 0.7, which implies that the sensory instrument has met the reliability requirement. The mean of G coefficient value of 0.9535 > 0.7 implies that the sensory instrument will meet the reliability requirement if the sensory instrument is applied altogether in the university and vocational high school students majoring culinary.

For the estimation of instrument reliability in the expanded scale experiment, the researchers implemented intraclass correlation coefficient (ICC) analysis. The three instruments were tested toward 16 university students majoring in culinary and 16 vocational high school students majoring in culinary. The summary of the results of reliability instru-

ment by means of ICC analysis is presented in Table 11.

Table 11 shows that from the testing toward university students majoring culinary, the physiological instrument resulted in the coefficient of reliability of 0.747) > 0.7, which implies that the physiological instrument is reliable and stable for the university students. On the other hand, the table shows that from the testing toward vocational high school students majoring culinary, the physiological instrument resulted in the reliability coefficient of 0.765 > 0.7, which implies that the instrument is reliable and stable for the vocational high school students majoring culinary. The mean of reliability coefficient value of 0.756 > 0.7 implies that the instrument is reliable and stable and it can be applied altogether in the university and vocational high school students majoring in culinary.

Table 11. Results of ICC coefficient analysis

Instrument	Facet	Number of Item	ICC Coeff.	Criteria ≥ 0.70
1. Physiological	Univ. students	14	0.747	> 0.7
	VHS students	14	0.765	> 0.7
	Mean	14	0.756	> 0.7
2. Psychological	Univ. students	11	0.78	> 0.7
	VHS students	11	0.715	> 0.7
	Mean	11	0.712	> 0.7
3. Sensory	Univ. students	17	0.730	> 0.7
	VHS students	17	0.733	> 0.7
	Mean	17	0.732	> 0.7

Table 11 shows that from the testing toward university students majoring culinary, the psychological instrument resulted in the coefficient of reliability of 0.78 > 0.7, which implies that the instrument is reliable and stable for the university students. On the other hand, the table shows that from the testing toward vocational high school students majoring culinary, the psychological instrument resulted in the reliability coefficient 0.715 > 0.7, which implies that the instrument is reliable and stable for the vocational high school students majoring culinary. The mean of the reliability coefficient value of 0.756 > 0.7 implies that the instrument is reliable and stable and can be applied altogether in the university and vocational high school students majoring in culinary.

Table 11 shows that the sensory instruments tested to the university students majoring culinary resulting in an ICC reliability coefficient of 0.73 exceeding the required minimum 0.7 ICC coefficient criterion, meaning reliable and stable sensory instruments for university students majoring in culinary facets. Similarly, the experiments to vocational high school students majoring culinary produce an ICC reliability coefficient of 0.733 exceeding the criteria of the coefficient of ICC at least 0.7, which means that it is reliable and stable sensory instruments for students of vocational high school majoring culinary facets. The average coefficient of sensory instrument reliability of 0.732 exceeds the criteria of the coefficient of ICC of at least 0.7, which means that the sensory instrument is reliable and stable and can be used for either university students or vocational high school students majoring culinary together.

The characteristics of the food product sensory acceptability assessment instrument were related to the number of minimum items that should be involved in the assessment of the criteria of minimum score that should be gained. For deciding the number of minimum item, D-study analysis was implemented, while for deciding the criteria of minimum score, standard setting with the Extended Angoff method was implemented. The summary of the results of analysis is presented in Table 12.

Table 12. Instrument characteristics

Instrument	Facet	Number of items	Min. score	Min. items	Min. score
1. Physiological	Univ. student	14	77.14	≥ 3	16.33
	VHS sudent	14	77.14	≥ 4	22.04
2. Psychological	Univ. student	11	59.99	≥ 4	20.36
	VHS student	11	59.99	≥ 4	20.36
3. Sensory	Univ. student	17	85.85	≥ 3	15.15
	VHS student	17	85.85	≥ 2	10.10

Table 12 shows that the number of physiological instrument items is 14 items and the criteria of minimum score that should be gained is 77.14. If the physiological instrument is applied toward the university students majoring in culinary, then the number of minimum items will be 3 items and the criteria of minimum score that should be gained would be 16.33. On the other hand, if the physiological instrument is applied toward the vocational high school students majoring in culinary, then the number of minimum items is 4 items and the minimum score that should be gained will be 22.04.

Table 12 shows that the number of psychological instrument items is 11 items and the minimum score that should be gained is 59.99. If the psychological instrument is applied toward the university students majoring in culinary, then the number of minimum items is 4 items and the minimum score that should be gained will be 20.36. On the other hand, if the psychological instrument is applied toward the vocational high school students majoring in culinary, then the number of minimum items will be 4 items and the minimum score that should be gained will be 22.04.

Table 12 shows that the number of sensory instrument items is 17 items and the minimum score that should be gained is 85.85. If the sensory instrument is applied toward the university students majoring in culinary, then the number of minimum items will be 3 items and the minimum score that should be gained will be 15.15. On the other hand, if the sensory instrument is applied toward the vocational high school students majoring in culinary, then the number of minimum items will be 2 items and the minimum score that should be gained will be 10.10.

For deciding the benchmark on the criteria of minimum score, the researchers implemented the standard setting with the Extended Angoff method. The results of the calculation on the criteria of minimum score (cut score) for the physiological instrument is presented in Table 13.

Table 13. Criteria of the physiological instrument score

Number of Item	Cut Score	Minimum Score	Maximum Score
1	5.51	1	7
2	11.02	2	14
3	16.53	3	21
4	22.04	4	28
5	27.55	5	35
6	33.06	6	42
7	38.57	7	49
8	44.08	8	56
9	49.59	9	63
10	55.10	10	70
11	60.61	11	77
12	66.12	12	84
13	71.63	13	91
14	77.14	14	98

The results of the calculation on the criteria of minimum score (cut score) for the psychological instrument is presented in Table 14.

Table 14. Criteria of the psychological instrument score

Number of Item	Cut Score	Minimum Score	Maximum Score
1	5.09	1	7
2	10.18	2	14
3	15.27	3	21
4	20.36	4	28
5	25.45	5	35
6	30.54	6	42
7	35.63	7	49
8	40.72	8	56
9	45.81	9	63
10	50.90	10	70
11	55.99	11	77

The results of the calculation on the criteria of minimum score (cut score) for the sensory instrument is presented in Table 15.

The physiological instrument was aplied for measuring the rater's self-assessment toward the physiological preparedness in assessing food product acceptability. For the criteria of minimum score that should be gained in order that the related rater might continue the sensory assessment, sensory instrument was implemented. If the criteria of the minimum

score was not met, then the rater might not continue the process of sensory assessment toward food products resulting from the food processing practice or, under certain conditions, the rater might continue the sensory assessment with the other team of raters.

Table 15. Criteria of the sensory instrument score

Number of Item	Cut Score	Minimum Score	Maximum Score
1	5.05	1	7
2	10.10	2	14
3	15.15	3	21
4	20.20	4	28
5	25.25	5	35
6	30.30	6	42
7	35.35	7	49
8	40.40	8	56
9	45.45	9	63
10	50.50	10	70
11	55.55	11	77
12	60.60	12	84
13	65.65	13	91
14	70.7	14	98
15	75.75	15	105
16	80.80	16	112
17	85.85	17	119

The psychological instrument was applied for measuring the rater's self-assessment toward the psychological condition in assessing food products acceptability. For the criteria of minimum score that should be gained in order that the related rater might continue the sensory assessment, sensory instrument was implemented. If the criteria of minimum score was met, then the rater might not continue the process of sensory assessment toward food products resulting from the food processing practice or, under certain conditions, the rater might continue the sensory assessment with the other team of raters.

The sensory instrument was implemented in order to assess the sensory acceptability of food products resulting from the food processing practice. The results of the assessment were in the form of the scores that ranged from 17 to 119 if all of the 14 items had been implemented. The score that had been attained was interpreted by means of qualitative score range guideliness from the low extreme to the high extreme. The guidelines regarding the qualitative score were presented separately in the rubric of food product sensory acceptability assessment instrument.

Conclusion and Suggestions

Based on the calculation and analysis of the research data, three conclusions can be drawn as follows: (1) the three instruments meet the face validity, content validity, construct validity requirements; (2) the three instruments meet the reliability requirement based on Genova analysis and ICC analysis; (3) the determination of the assessor to be able to assess the acceptability of sensory food products meets the cut score requirements.

Since the food product sensory acceptability assessment instruments as the result of this research and development have not been disseminated, the instruments should be disseminated in order that they might be applied in the wider scale for meeting the ever-proven validity and reliability.

References

Azwar, S. (2000). *Sikap manusia: Teori dan pengukurannya* (2nd ed.). Yogyakarta: Pustaka Pelajar.

Bartko, J. J. (1976). On various intraclass correlation reliability coefficients. *Psychological Bulletin*, *83*(5), 762–765. https://doi.org/10.1037/0033-2909.83.5.762

Borg, W. R., & Gall, M. D. (1989). *Educational research: An introduction* (5th ed.). New York, NY: Longman.

Brennan, R. L. (2006). *Educational measurement* (4th ed.). Westport, CT: Praeger.

Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage Publications.

Mardapi, D. (2008). *Teknik penyusunan instrumen tes dan nontes*. Yogyakarta: Mitra Cendekia.

Mason, R. L., & Nottingham, S. M. (2002). *Sensory evaluation manual*. Phitsanulok, Thailand: Naresuan University.

Murti, B. (2011). Validitas dan reliabilitas pengukuran. In *Matrikulasi Program Studi Doktoral*. Surakarta: Fakultas Kedokteran, Universitas Sebelas Maret.

Rakhmat, J. (2007). *Psikologi komunikasi*. Bandung: Remaja Rosdakarya.

Shumate, S. R., Surles, J., Johnson, R. L., & Penny, J. (2007). The effects of the number of scale points and non-normality on the generalizability coefficient: A Monte Carlo study. *Applied Measurement in Education*, *20*(4), 357–376. https://doi.org/10.1080/08957340701429645

Stone, H., & Sidel, J. L. (2004). *Sensory evaluation practices* (3rd ed.). Waltham, MA: Academic Press. https://doi.org/10.1016/B978-012672690-9/50013-5

Walgito, B. (1989). *Pengantar psikologi umum* (Rev. ed.). Yogyakarta: Andi Offset.

Watts, B. M., Ylimaki, G. L., Jeffery, L. E., & Elias, L. G. (1989). *Basic sensory methods for food evaluation*. Ottawa, Ontario: International Development Research Centre.

SUBJECT INDEXES

AUTHOR INDEXES

AUTHORS' BIOGRAPHY

Agus Triyanto. Currently works as a lecturer in the Department of Guidance and Counseling, Faculty of Education, Universitas Negeri Yogyakarta. He attained bachelor of guidance and counseling in 2003 from Universitas Negeri Yogyakarta. His master degree was attained in 2011 from Universitas Negeri Malang at the same major.

Badrun Kartowagiran. Was born on 25 July 1953. He works as a lecturer in the Faculty of Engineering and Graduate School of Yogyakarta State University, Indonesia. He attained his bachelor degree in 1977 from Yogyakarta Institute of Teacher Education and Educational Sciences (recently known as Yogyakarta State University) majoring in Machine Engineering Education. In 1992, he achieved his master degree on Educational Research and Evaluation from Jakarta Institute of Teacher Education and Educational Sciences (recently known as Jakarta State University). In 2005, he graduated from his doctoral study in Psychology/Psychometrics in Gadjah Mada University, Indonesia.

Bambang Subali. Was born on 12 January 1952. He works as a lecturer in the Faculty of Mathematics and Science, and Graduate School of Yogyakarta State University, Indonesia. He attained his bachelor degree from Yogyakarta Institute of Teacher Education and Educational Sciences (recently known as Yogyakarta State University) majoring Biology Education. He achieved his master degree on Forestry Science from Gadjah Mada University, Indonesia. He graduated from his doctoral study in Educational Research and Evaluation, Yogyakarta State University, Indonesia.

Carl J. Dunst. Carl J. Dunst is a Senior Research Scientist at the Orelena Hawks Puckett Institute in Asheville and Morganton, North Carolina, USA. His research and practice spans more than 40 years and includes early intervention, early childhood special education, family and social support, and family-centered practices. His primary areas of interest include evidence-based child, parent-child, and family intervention practices, meta-analyses and research syntheses of child, parent, parent-child, and family practices, and research on inservice professional development and adult learning methods and strategies. He serves as a Knowledge Generation consultant in all our data report and literature synthesis product development.
(source: http://ecpcta.org/person/carl-dunst-ph-d/)

Deborah W. Hamby. Deborah W. Hamby is a Research Analyst at the Orelena Hawks Puckett Institute.

Djemari Mardapi. Was born on January 1, 1947. He is a professor at Universitas Negeri Yogyakarta, Indonesia. He currently works as a lecturer at the Faculty of Engineering and the Graduate School of Universitas Negeri Yogyakarta. He obtained a bachelor degree in Electrical Engineering Education in 1973, from Yogyakarta Institute of Teacher Education and Educational Sciences (recently known as Universitas Negeri Yogyakarta). In 1984, he achieved his master degree on Educational Research and Evaluation from Universitas Negeri Yogyakarta. He is an alumnus of IOWA university in 1988, majoring in Educational, Measurement, and Statistics and obtained a Ph.D.

Donna E. Nelson. Donna E. Nelson is a Program Director and Adjunct Faculty at Expanding Horizons: Early Intervention Program, Center for Early Childhood, East Tennessee State University, Johnson City, Tennessee, USA.

Dwi Istiyani. Dwi Istiyani is a lecturer in the Faculty of Education and Teacher Training, Universitas Islam Negeri Walisongo Semarang, Central Java, Indonesia.

Farida Agus Setiawati. Currently works as a lecturer in the Department of Psychology, Faculty of Education, Universitas Negeri Yogyakarta. Her bachelor degre in psychology was attained in 1996 from Universitas Gadjah Mada. Her master degree in psychometry was attained in 2004 from the same university. She attained her doctoral degree in Educational Research and Evaluation from Universitas Negeri Yogyakarta in 2013.

I Gusti Putu Suryadarma. Was born in Tabanan, Bali, Indonesia, on 25 December 1951. He is a professor and lecturer in Universitas Negeri Yogyakarta who is actively teaching in Environmental Sciences, Ecology, Ethnobotany, General Biology, Soil Biology, Philosophy of Science, and Natural Sciences Learning. His bachelor degree was attained in 1979 from Yogyakarta Institute of Teacher Education and Educational Sciences (recently known as Universitas Negeri Yogyakarta) majoring Biology Education. His master degree was attained from Universitas Gadjah Mada majoring Environmental Sciences in 1989. He attained his doctoral degree in 2005 from di Institut Pertanian Bogor majoring Natural Resources and Environment Management.

Jihyun Lee. Jihyun Lee is a survey methodologist, applied statistician, and educational psychologist. She is interested in national and international, large-scale assessments. Her main research area is developing methodology to increase psychometric properties and usability of survey instruments. She has developed numerous survey instruments for more than a decades of post-PhD professional services in the US, Singapore, South Korea, and Australia. Substantively, she investigates theoretical and practical utility of education-related psychological constructs. In her research, a wide range of education-related psychological constructs are reviewed in relation to student achievement in various subject domains (reading, writing and mathematics). The constructs that Dr Lee has most extensively studied are interest, motivation, confidence, attitudes, self-belief, self-efficacy, and anxiety. She comes to UNSW from the National Institute of Education in Singapore, where she worked as Assistant Professor. Prior to that, Jihyun was a key member for test/questionnaire/survey development at Educational Testing Service (ETS), Princeton New Jersey, U.S.A. for several years, working on large-scale projects such as the National Assessment of Educational Progress (NAEP) and the Programme for International Student Assessment (PISA). She was also ETS liaison to the US federal government bureau, the National Center for Educational Statistics (NCES) within the U.S. Department of Education. She received a PhD in Measurement and Applied Statistics from Columbia University and a Master in Human Development from Harvard School of Education.

Linda L. Wilson. Linda L. Wilson is an Early Head Start Director at The Enola Group Early Head Start Program, Morganton, North Carolina, USA.

Marilyn Espe-Sherwindt. Marilyn Espe-Sherwindt is the Director of the Family Child Learning Center, Akron Children's Hospital and Kent State University, Tallmadge, Ohio, USA.

Marzuki. Was born on 21 April 1966. He is currently working as a lecturer in the Department of Civics and Law Education of the Faculty of Social Sciences, Universitas Negeri Yogyakarta. His bachelor degree in Education/*Tarbiyah* was attained in 1990 from Sunan Kalijaga State Islamic Institute. His master degree was attained in 1997 and doctoral degree was attained in 2007, both are in the same major, Islamic Studies, and from the same higher educational institute, Syarif Hidayatullah Jakarta State Islamic University.

Nur Hidayanto Pancoro Setyo Putro. Currently works as a lecturer in the Department of English Education, Faculty of Languages and Arts, Universitas Negeri Yogyakarta. He attained his bachelor degree in English Education in 2005 from Universitas Negeri Yogyakarta. His master degree in Educational Research and Evaluation was attained in 2010 from the same university. His doctoral degree was attained in 2016 from School of Education, the University of New South Wales, Sydney, Australia.

Paidi. Currently works as a lecturer in the Department of Biology Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta. He attained his bachelor degree in 1991 from Yogyakarta Institute of Teacher Education and Educational Sciences (recently known as Universitas Negeri Yogyakarta) majoring Biology Education. His master degree was attained in 1999 from Universitas Gadjah Mada majoring Biology. His doctoral degree was attained in 2008 from Universitas Negeri Malang majoring Biology Education.

Rachmadi Widdiharto. Was born on May 21, 1968 in Ponorogo, East Java, Indonesia. He attained his bachelor degree in Mathematics Education, from Yogyakarta Institute of Teacher Education and Educational Sciences (recently known as Universitas Negeri Yogyakarta) in 1992. He also completed a bachelor degree of Government Science in the Faculty of Social and Political Science, Universitas Gadjah Mada, Yogyakarta in 1994. In 1995, he started working as a civil servant/mathematics teacher at Junior Secondary School of 3 Paliyan, Gunung Kidul, Yogyakarta. In 1997-1999 he obtained a grant-education of master degree students of the Asian Development Bank (ADB) through Senior Secondary Education Project, Ministry of Education, Jakarta in Curriculum & Instruction, Graduate School, New Mexico State University, Las Cruces, USA, in the concentration of Learning Technology. In 1999 to 2003, he taught mathematics in Senior Secondary School of 9 Yogyakarta. In 1999-2011, he became extraordinary Lecturer at the Department of Mathematics Tadris, Faculty of Education, Sunan Kalijaga State Islamic Institute, which is recently known as the Faculty of Science & Technology, Sunan Kalijaga State Islamic University, Yogyakarta. Starting in 2004 until now, he has been becoming a lecturer at the Centre for the Development and Empowerment of Mathematics Teachers and Educational Personnel (*PPPPTK Matematika*) of Yogyakarta, and in 2008 he persued his study on Doctoral program in the Departement of Educational Research and Evaluation, Universitas Negeri Yogyakarta.

Rita Eka Izzaty. Currently works as a lecturer in the Department of Psychology, Faculty of Education, Universitas Negeri Yogyakarta. She attained her bachelor degree in 1996, master degree in 2006, and doctoral degree in 2012, in psychology from Universitas Gadjah Mada, Indonesia.

Siti Mariyam. Worked as a lecturer in the Department of Biology Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri Yogyakarta.

Sovia Mas Ayu. Was born in Tanjung Karang, Bandar Lampung, Indonesia, on November 30th, 1976. She has been working as a lecturer of Islam Education Science in Raden Intan Lampung State Islamic University. She earned her undergraduate degree in Arabian Language Education from Raden Intan Lampung State Islamic University in 2000. Then, she earned her master degree in Islam Education Science from Syarif Hidayatullah Jakarta State Islamic University in 2003. She is a graduate of the doctoral program of Educational Research and Evaluation from Universitas Negeri Yogyakarta.

Sugiman. Was born on February 28, 1965 in Kutoarjo, Purworejo, Central Java. In 1989, he completed his Bachelor's degree in Mathematics Education at Yogyakarta Institute of Teacher Education and Educational Sciences (recently known as Universitas Negeri Yogyakarta). In 1997, he attained his master degree in Mathematics from Bandung Institute of Technology, and his doctoral degree in Mathematics Education was attained in 2010 from Indonesia University of Education, Bandung, West Java. Since 1991, he has been working as a lecturer of the Department of Mathematics Education, Universitas Negeri Yogyakarta, with interests and expertise in Mathematics Learning.

Sugiyono. Was born in Sragen on April 12th, 1953. He pursued his undergraduate study in the Department of Agricultural Product Processing, Faculty of Agricultural Technology, Universitas Gadjah Mada and earned his bachelor degree in 1984. Then, he pursued graduate study in the department of Occupational Health Science in Universitas Gadjah Mada, and earned his master degree in 2002. He has recently been working as a lecturer in the Department of Culinary Engineering Education, Faculty of Engineering, Universitas Negeri Yogyakarta.

Suharsimi Arikunto. Currently works as a lecturer in the Graduate School and Faculty of Education of Universitas Negeri Yogyakarta, and also in Universitas Ahmad Dahlan in the expertise of educational administration and management.

Zamroni. Currently works as a senior lecturer in Universitas Negeri Yogyakarta in the expertise field of educational sociology. He completed his bachelor degree in Yogyakarta Institute of Teacher Education and Educational Sciences (recently known as Universitas Negeri Yogyakarta) focusing on economic enterprise, and his master and doctoral degrees in Florida State University, United States of America achieving M.A. and Ph.D. He has been being active in publication since he started working in Universitas Negeri Yogyakarta in 1974.

SUBMISSION GUIDELINES

- The manuscript submitted is a result of an empirical research or scientific assessment of an actual issue in the area of research, evaluation, and education in a broad sense, which has not been published elsewhere and is not being sent to other journals.
- Only articles written in English will be considered. Any consistent spelling and punctuation styles may be used. Please use single quotation marks, except where 'a quotation is "within" a quotation'. Long quotations of 40 words or more should be indented without quotation marks.
- A typical manuscript is approximately 7000 words including tables, figures, references, and captions. Manuscripts that greatly exceed this will be critically reviewed with respect to length. (A4; margins: top 3, left 3, right 2, bottom 2; double columns [Except in Abstract: single column]; single-spaced; font: Garamond, 12).
- Manuscripts should be compiled in the following order: (1) title; (2) abstract; (3) keywords; (4) main text: introduction, method, findings and discussion, conclusion and/or suggestions; (5) acknowledgements for the Funding and grant-awarding bodies (if any); (6) references; and (7) appendices (as appropriate).
- (If any) The funding or grant-awarding bodies is acknowledged in a separate paragraph. *For single agency grants*:"This work was supported by the [Name of Funding Agency] under Grant [number xxxx]."
- The title of the manuscript should clearly represent the content of the article and contain the **keywords**.
- Authors' identities under the title should be omitted, and replaced by the following item:

  Anonymous
  (Author's identity is omitted due to review process)
- An abstract that does not exceed 300 words is required for any submitted manuscript. It is written narratively containing the aim(s), method, and the result(s) of the research.
- Each manuscript should have 3 to 6 keywords written under the abstract.
- All tables and figures are adjusted to the paper length and are numbered and referred to the text.
- The citation and references are referred to American Psychological Association (APA) (Sixth Edition) style.
- APA Style format for references can be checked in http://www.citationmachine.net/apa/cite-a-website
- The author is strongly preferred to use Reference Manager applications.
- The manuscript must be in *.doc or *.rtf , and sent to **REiD's Management** via online submission by creating account in the Open Journal System (OJS) [click **REGISTER** if you have not had any account yet; or click **LOG IN** if you have already had an account].
- Authors' biography is written in the form of narration, including author's full name, place and date of birth, educational qualification/information started from bachelor degree (S1) until the latest educational degree, the affiliation in which the author is currently working, phone number, and email address.
- All Author(s)' names and identity(es) must be completely embedded in the form filled in by the corresponding author: email; affiliation; and each author's short biography (in the column of 'Bio Statement'). [if the manuscript is written by two or more authors, please click 'Add Author' in the 3rd step of 'ENTER METADATA' in the submission process and then enter each author's data.]
- All correspondences, information, and decisions for the submitted manuscripts are conducted through the email/s used for the submission.
- Word template is available for this journal. Please visit the journal's homepage at https://journal.uny.ac.id/index.php/reid
- If you have submission queries, please contact *reid.ppsuny@uny.ac.id* or *reid.ppsuny@gmail.com*